NeuroImage 105 (2015) 536–551

Contents lists available at ScienceDirect

NeuroImage

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

Review

Recent progress and outstanding issues in motion correction in resting state fMRI

Jonathan D. Power [a,*], Bradley L. Schlaggar [a,b,c,d], Steven E. Petersen [a,b,d,e,f,g]

[a] Dept. of Neurology, Washington University School of Medicine in St. Louis, 660 S. Euclid Ave., St. Louis, MO 63110, USA
[b] Dept. of Radiology, Washington University School of Medicine in St. Louis, 660 S. Euclid Ave., St. Louis, MO 63110, USA
[c] Dept. of Pediatrics, Washington University School of Medicine in St. Louis, 660 S. Euclid Ave., St. Louis, MO 63110, USA
[d] Dept. of Anatomy & Neurobiology, Washington University School of Medicine in St. Louis, 660 S. Euclid Ave., St. Louis, MO 63110, USA
[e] Dept. of Psychology, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA
[f] Dept. of Neurosurgery, Washington University School of Medicine in St. Louis, 660 S. Euclid Ave., St. Louis, MO 63110, USA
[g] Dept. of Biomedical Engineering, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA

ARTICLE INFO

Article history:
Accepted 15 October 2014
Available online 24 October 2014

Keywords:
fMRI
Motion
Artifact
Denoising
Resting state
Functional connectivity

ABSTRACT

The purpose of this review is to communicate and synthesize recent findings related to motion artifact in resting state fMRI. In 2011, three groups reported that small head movements produced spurious but structured noise in brain scans, causing distance-dependent changes in signal correlations. This finding has prompted both methods development and the re-examination of prior findings with more stringent motion correction. Since 2011, over a dozen papers have been published specifically on motion artifact in resting state fMRI. We will attempt to distill these papers to their most essential content. We will point out some aspects of motion artifact that are easily or often overlooked. Throughout the review, we will highlight gaps in current knowledge and avenues for future research.

Contents

* Corresponding author at: Washington University School of Medicine, Dept. of Neurology, 660 S., Euclid Ave, Box 8111, St. Louis, MO 63110, USA. Fax: +1 314 362 2186.
E-mail addresses: jonathan.power@nih.gov (J.D. Power), schlaggarb@neuro.wustl.edu (B.L. Schlaggar), sep@npg.wustl.edu (S.E. Petersen).

http://dx.doi.org/10.1016/j.neuroimage.2014.10.044
1053-8119/

Introduction

Over the last decade, studies of correlated fMRI signal, as opposed to task-evoked activity, have come to represent a considerable portion of published fMRI studies. Such studies, called functional connectivity studies, have been used to examine the functional organization of the brain and to describe changes in this organization over development, aging, and disease states. Functional connectivity studies often involve no explicit task (other than for subjects to lie in the scanner for some period of time) and are therefore called "resting state" studies of spontaneous (or "intrinsic") fMRI signal.

One finding that spanned many of the first functional connectivity fMRI studies is that the young, the elderly, and the diseased often exhibited "underconnectivity" relative to healthy young adults. For example, between (distant) regions of the default mode network, children had weaker signal correlations than young adults (Fair et al., 2007), and these same (distant) correlations were weaker in older subjects than in young adults (Andrews-Hanna et al., 2007). These findings were part of a larger story whereby correlation values were modulated across the lifespan in a distance-dependent manner: in childhood, short-distance correlations were strong and longer-distance correlations were weaker, then as children aged into young adulthood, short-distance correlations weakened and longer-distance correlations strengthened, and then as subjects aged into later adulthood the short-distance correlations again became stronger and long-distance correlations became weaker. Interestingly, frameworks for understanding some diseases sometimes had similar aspects of distance dependence, such as "local overconnectivity but long-distance disconnection" in autism (Courchesne and Pierce, 2005).

In 2011, three groups reported a previously unrecognized aspect of motion artifact in functional connectivity MRI studies, which is that motion adds spurious variance that tends to be more similar at nearby voxels than at distant voxels, causing distance-dependent modulation of signal correlations (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). Troublingly, even relatively small movements that were traditionally not thought to be problematic could produce these effects. This meant that, all other things being equal, a higher-motion group would have relatively stronger short-distance correlations than a lower-motion group, and, depending on the processing strategy used, that long-distance correlations would often be weaker in the higher-motion group. The implications for functional connectivity studies of development, aging, and disease were clear, since children, the elderly, and patients tend to move more in the scanner than typical and/or young adult populations. Indeed, in the 3 initial studies, previously-described developmental (Power et al., 2012; Satterthwaite et al., 2012) and aging (Van Dijk et al., 2012) differences were reported to be at least partially explained by motion artifact.

The full impact of motion artifact on functional connectivity studies of development, aging, and disease is not yet clear, since methods to counteract motion artifact are still being developed and validated (Bright and Murphy, 2013; Jo et al., 2013; Kundu et al., 2013; Muschelli et al., 2014; Patel et al., 2014; Power et al., 2014; Satterthwaite et al., 2013a; Scheinost et al., 2014; Yan et al., 2013a,b; Zeng et al., 2014). Studies that carefully control for motion artifact are beginning to determine which previous findings might relate to motion artifact, and which findings withstand newer and more stringent corrections for motion artifact (Fair et al., 2012; Satterthwaite et al., 2013b; Tyszka et al., 2014). On the one hand, some findings may be attenuated or overturned. On the other hand, improved removal of a biasing artifact may reveal effects that were previously obscured.

Artifact caused by small movements is not uniquely problematic to functional connectivity studies. Within the last year, scrutiny of small movements in diffusion weighted imaging (DWI) has determined that small movements also create previously unrecognized, spurious differences in this modality (Yendiki et al., 2013). This realization has prompted the re-evaluation of prior structural connectivity findings with newer and more stringent methods of motion correction. One of the first such re-evaluative studies now indicates that many differences in white matter tracts previously attributed to autism may be attributable to motion artifact, though differences in the inferior longitudinal fasciculus withstand the newer motion correction procedures (Koldewyn et al., 2014). This report has a parallel in the re-evaluative reports in functional connectivity studies: in data from control and high-functioning autistic subjects, binning subjects by motion creates much larger group differences than binning by diagnosis (Tyszka et al., 2014). These findings underscore the importance of identifying and removing the influences of motion in MRI studies.

Motion is unique among artifactual influences in fMRI in that it can be measured from the data itself via realignment parameters (unlike, for example, cardiac- or respiratory-related artifact, which require external recordings). Thus, all fMRI datasets contain the substrates needed for many motion-targeting denoising procedures. Over one dozen papers have been published specifically on motion artifact in functional connectivity fMRI since 2011. The aim of this review is to synthesize and consolidate the knowledge gained to date, and, in doing so, to point out directions for future research. Much of the literature on motion artifact is technical; we will attempt to communicate this literature to a broad audience. The first portions of the paper concern findings and methods that are applicable to most existing datasets. In the final portions of the paper we consider emerging methods that may be useful in future datasets.

The reader should note that the field of denoising, especially with respect to motion, is rapidly evolving. There is no consensus about which methods are most successful at removing motion artifact, or even how to measure success in removing motion artifact. Different groups would likely have different interpretations of the literature. We have tried to steer a neutral course in this article. Our figures present data from 11 studies (2 of our own) and we discuss nearly every paper on motion published after 2011 up until the time of submission (and some papers that emerged during review). Our intent is to collate most of the important new data regarding motion artifact and to help readers learn how to interpret various commonly-used analyses aimed at detecting or removing motion artifact. Our interpretation of the literature is unavoidably embedded in selecting and discussing certain data. Our intent is not to steer the field toward or away from a particular processing stream; we do not make processing recommendations. At the end of the article, we do suggest some data that, if reported, would help reviewers and readers assess motion-related effects in a given study.

Features of motion artifact

For over two decades, it has been known that motion creates strong disturbances in fMRI signal (Friston et al., 1996; Hajnal et al., 1994). In the mid-1990s, nearly all fMRI studies were task studies, which modeled evoked activity from many trials or blocks of trials. In these studies, any source of noise, including motion, would be suppressed via averaging in the modeling process, unless the noise systematically related to experimental timing. Most studies of motion at that time focused on task-associated motion, where systematic relationships between experimental timing and subject motion (e.g., a spoken response) meant that motion-related variance could be mistakenly modeled as task-related activity (Barch et al., 1999; Bullmore et al., 1999).

In functional connectivity studies, where covariance among time series is the measure of interest, motion artifact cannot be suppressed by averaging. For a pair of voxels, depending on the timing, duration, and direction of motion, motion can add similar variance to signals in both voxels, it can add variance to only one of the voxel signals, or it can add different types of variance to each voxel signal. Thus, motion can cause both increases and decreases in observed correlations. Importantly, the similarity of motion-added variance is typically high between nearby voxels, and the similarity of added signal drops with distance between voxels, meaning that modulation of correlations is different in nearby versus distant pairs of voxels.

Measuring motion

Effort in recent years has been directed toward a fuller understanding of the signal disruptions caused by motion. To identify such disruptions, it is first necessary to establish measures of motion. Nearly every fMRI dataset in any institution undergoes rigid body realignment as one of the first image processing steps. The rigid body realignment estimates are derived relative to some fixed position and describe the displacement of the head from that position at every time point in the scan. An example of such realignment estimates is shown in the top panel of Fig. 1A. Note that most fMRI data are acquired using echoplanar imaging (EPI) sequences, in which data are acquired as slices. Motion may impact all slices of an image, or only some of them, and different motions can occur within a single volume, impacting slices differentially. Almost universally across neuroimaging groups, realignment estimates are derived from volumes, not from slice data. The volume-derived estimates are a simplification and summarization of movement over the acquisition of all slices. It is possible that slice-derived information could be used for denoising purposes in more specific and, perhaps, more efficacious ways than volume-derived realignment estimates (e.g., (Beall and Lowe, 2014)).

A useful distinction can be drawn between absolute displacement of the head and relative displacement of the head. In general, it is the relative movements, from one volume of data to the next volume, which cause the most difficult-to-correct disruptions in signal. The two types of displacement are illustrated in the middle panel Fig. 1A, where absolute displacement from a fixed position is shown in dotted red lines, and relative displacement is shown in the solid red line. Below, in blue, is DVARS (Smyser et al., 2010), a measure of how rapidly and ubiquitously signal changes from volume to volume (discussed later). These traces, and similar traces in hundreds of other subjects, indicate that relative displacement identifies prominent signal disruptions caused by motion.

Multiple versions of relative displacement measures, often called framewise displacement (FD), have been proposed and used by various groups. For example Van Dijk et al. (2012) calculated the root-mean-square of the derivatives of the translational realignment estimates, whereas Power et al. (2012) summed the absolute values of the derivatives of the translational and rotational realignment estimates (after converting rotational estimates to displacement at 50 mm radius). Other versions of these measures are calculated in various software packages such as AFNI (http://afni.nimh.nih.gov/afni/) or FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/). Subscripts will be used in this review to denote various versions of FD.

Several points are worth noting about the FD measures. First, though various versions of FD are calculated differently, they target similar properties of the data, and various measures often correlate highly (Fig. 1B). For example, when using exact voxel-specific FD as a reference, the FD calculated in FSL (Jenkinson et al., 2002) correlates at $r = 0.99$, the FD calculated after Power et al. (2012) correlates at $r = 0.98$, and the FD calculated after Van Dijk et al. (2012) correlates at $r = 0.85$ (Yan et al., 2013a). Thus, across FD measures, the shapes of the FD traces are similar. A second and important point to note, however, is that the magnitudes of the measures differ. Though they are highly correlated, FD_{Power} has values almost twice those of FD_{FSL}, and FD_{FSL} has values roughly twice those of $FD_{Van\ Dijk}$ (Yan et al., 2013a). A third and less obvious point to note is that FD magnitudes depend on TR, since this determines how rapidly movements will be sampled (and subdivided). A final point worth noting is that the realignment estimates themselves, because they are derived from the fMRI images and not from external measures of motion, may not be entirely accurate, may not reflect properties relevant to all slices, and can be influenced by how data is processed prior to realignment (e.g., if outlying voxel intensities are replaced by interpolation prior to realignment (Christodoulou et al., 2013; Jo et al., 2013)). Naturally, the accuracy of techniques that utilize FD (or any QC measure) is limited by the accuracy of the measure itself. For example, if slice-specific motion is not captured in the volume-derived motion estimate or volumetric DVARS, the utility of such measures for

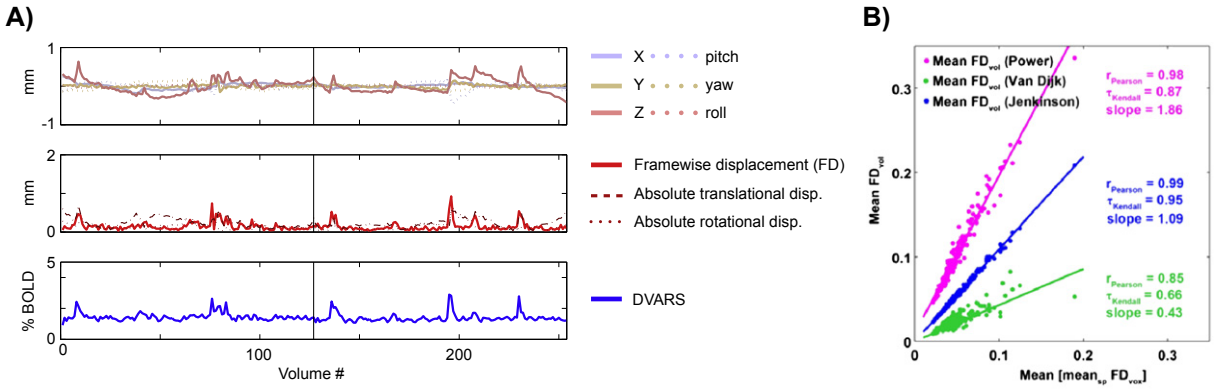

Fig. 1. Measuring motion. A) For a single subject, realignment estimates, framewise displacement, absolute rotational and translational displacements, and a DVARS trace calculated over the whole brain. Absolute displacements are the sum of the absolute values of the 3 translational and 3 rotational parameters, respectively. Framewise displacement is the sum of the absolute values of the first derivatives of the 6 realignment parameters, after converting the rotational parameters to translational displacements on a sphere of radius 50 mm (roughly the radius of the brain). DVARS is calculated as the RMS value, at each volume, of the first derivatives of all voxel time series over the whole brain. The TR is 2.5 s. Modified from Power et al. (2014). B) Comparisons of various FD measures, modified from Yan et al. (2013a).

defining data to be censored is diminished, since some data that "ought" to have been identified would not be identified. The extent to which such considerations are practically important is not well understood at present. In sum, various FD measures all target and identify similar aspects of the data, but the magnitudes of the waveforms differ across measures and according to sampling rates. *A practical corollary is that if FD measures are to be used in some dataset, FD waveforms and magnitudes should be established (and reported) for that specific dataset, since magnitudes from the literature may not be appropriately scaled to the data at hand.*

Motion-induced signal disruptions

The signal disruptions produced by motion are both stereotyped and highly variable. Modeling of both the global signal and voxel signals during and after motion (Fig. 2A) indicates that signal decreases are commonly found and that the magnitude of signal decrease scales with the magnitude of movement (Satterthwaite et al., 2013a). However, at the single subject level, when FD traces are plotted along with the signals in gray matter, it is evident that disruptions are not always stereotyped (Power et al., 2014) (Fig. 2B). Rather, signal disruptions can be increases, decreases, or complex waveforms. These changes are sometimes shared across much of the brain and are sometimes different in different parts of the brain. Such variability – both in waveforms across movements and waveforms across voxels – is likely due to the timing and duration of motion (impacting one slice, multiple slices, or all slices of a volume), the trajectory of motion, and a voxel's position relative to the skull, sinuses, and other tissues with differing signal intensity. Note that the only processing applied to the data in Fig. 2B is slice-time correction, image realignment, and removal of the mean and linear trend terms at each voxel in each run (to facilitate comparisons across runs). In other words, no frequency filtering or nuisance regression or other denoising has been performed in these data. Note also that finding stereotyped signal decreases in the TRs during and after motion (Satterthwaite et al., 2013a), but also variable and more prolonged effects for individual movements (Power et al., 2014), are not mutually incompatible findings. Importantly, unlike task fMRI, to impact resting state correlations, motion-induced signal fluctuations need only be present, they do not need to be similar across motions.

The spatial impact of motion on functional connectivity correlations

This section presents a conceptual framework for thinking about the spatial effects of motion on fMRI signal. The discussion should be understood as probabilistic, reflecting tendencies across motions and voxels. Any particular motion may produce a wide variety of signal changes that vary in duration, sign, and prevalence across voxels, and simulations that reproduce this complicated behavior are lacking. This discussion therefore summarizes the behavior of empirical data.

When a motion occurs, two voxels may experience similar disruptions in signal (increasing correlations), dissimilar disruptions (driving correlations to zero or to negative values), or only one voxel may be impacted (a dissimilar disruption driving correlations to zero). Nearby voxels often display similar disruptions. More distant voxels often display similar disruptions when the artifact is widespread, but they also frequently display dissimilar disruptions. Thus, motion acts 1) to typically increase signal correlations between nearby voxels, 2) to increase signal correlations between distant voxels if the signal disruption is widespread and similar in nature over much of the brain, and 3) to drive signal correlations to zero or possibly even cause negative correlations between voxels if the disruptions are different in character in different locations of the brain, an effect that typically occurs between distant voxels, though it can occur between nearby voxels. We will refer to these 3 types of spurious variance as types 1, 2, and 3 in the following discussion.

The balance of these different types of signal disruption determines the distance-dependent character of motion artifact. In our data and experience, type 1 effects are prominent, type 2 effects are also prominent, and type 3 effects are less prominent. Different datasets may inherently contain different balances of these types of motion artifact, but little is presently known about whether and how motion artifact is modulated by acquisition parameters (e.g., slice orientation, interleaved vs. continuous vs. spiral acquisition, etc.).

Different denoising strategies can selectively remove different types of motion artifact, yielding different post-denoising profiles of distance-dependent artifact. For example, in all datasets that have been examined, prior to denoising, correlations at all distances are elevated by motion (due to type 2 effects), but correlations at short distances are more elevated (by type 1 effects) than those at longer distances (which are somewhat decreased by type 3 effects). Many studies have examined the impact of global signal regression on this profile of distance dependence. In data that have undergone global signal regression the profile is changed: at long distances, motion acts to slightly *decrease* correlations, whereas motion still increases correlations at short distances. This effect can be explained as follows: the global fMRI signal (averaged across the entire brain) presumably reflects many type 2 disruptions, whereas focal type 1 effects must have a weaker representation, as must type 3

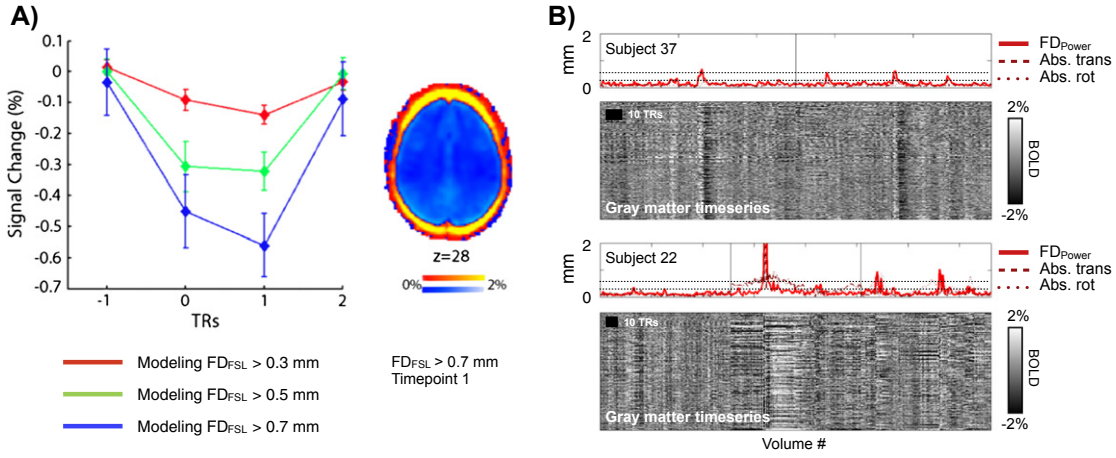

Fig. 2. Signal disruptions by motion. A) Modeling of the global signal during and after motion, for 3 magnitudes of motion, and a voxelwise map of modeled changes at the most severe level of motion. Modified from Satterthwaite et al. (2013a). B) For two subjects, motion traces and gray matter voxel time series are shown, illustrating the heterogeneity and duration of motion's impact on fMRI signal. The TR is 2.5 s, meaning the black scaling bar covers 25 s. These data have only been slice time corrected (for interleaved acquisition), realigned, registered to a target atlas, and the mean signal at each voxel has been subtracted, i.e., no frequency filtering or nuisance regression or other denoising has been performed. The dotted lines roughly denote FD = 0.5 and FD = 0.2 mm. Modified from Power et al. (2014). Note that, within a single dataset, measured magnitudes of FD_{FSL} are roughly half the magnitudes of FD_{Power}.

effects (since different signal disruptions could partially or completely cancel each other in an average signal calculation (Jo et al., 2013; Muschelli et al., 2014)). Thus, regression of the global signal is expected to remove much type 2 variance, but will fail to remove much type 1 and type 3 variance. In terms of correlations, removal of type 2 variance should reduce (spuriously increased) correlations at all distances by removing global motion-related variance, but spuriously increased short-distance and spuriously decreased long-distance correlations will be only modestly corrected since the global signal only weakly represents focal (type 1 and 3) motion-related variance. Removal of type 2 variance by global signal regression thus "unmasks" the disparity between the effects of type 1 and type 3 variance. Both modeling and empirical findings are consistent with these ideas (Jo et al., 2013; Power et al., 2014; Satterthwaite et al., 2013a). Other denoising strategies that target focal variance as opposed to global variance should produce different distance-dependent profiles, discussed below.

Techniques to reveal distant-dependent motion artifact

The impact of motion on functional connectivity correlations can be quantified in several ways, two of which are discussed here (others will be discussed later). One way to characterize motion artifact is to measure subject motion (e.g., via mean FD over a subject's scans) and to correlate, across subjects, subject motion with observed correlation values. For example, 120 subjects would yield 120 mean FD values, and these mean FD values could be correlated with a pairwise functional connectivity correlation, across subjects, to determine how that correlation is modulated by subject motion. We refer to these correlations as QC-RSFC correlations (QC referring to quality control, measured by mean FD in the present example, RSFC denoting resting state functional connectivity). Fig. 3A shows an example of QC-RSFC analyses, where FD_{FSL} was calculated in 348 subjects, and the corresponding QC vector was correlated, across subjects, with all possible observed pairwise functional connectivity correlations between 160 regions of interest (ROIs). In Fig. 3A, the 12,720 (160*159/2) QC-RSFC correlations are plotted on the y-axis, and the x-axis shows the distance between the ROIs giving rise to each pairwise correlation. At left, the data without global signal regression indicate that all correlations are increased by motion, and that the dependence of correlation values on motion is strongest at short distances. The right panel indicates that regression of the global signal decreases the dependence of correlations on subject motion at all distances (consistent with removal of type 2 disruptions), but that short-distance dependence is still positive (consistent with unremoved, prominent type 1 disruptions), while long-distance dependence is near zero or slightly negative (consistent with unremoved but less prominent type 3 disruptions). Note too, consistent with the previous discussion, how distance-dependent differences are enhanced by removing the global signal, a result that stems from near-complete elimination of long-distance motion dependence but only partial elimination of shorter-distance motion dependence.

A different way to characterize motion artifact is to observe how correlations change when motion-impacted volumes are included or excluded from correlation calculations. This technique has a long history with respect to motion, including excluding volumes contaminated by speech-related motion artifact from modeling in task fMRI (Barch et al., 1999) or excluding motion-contaminated data from resting state analyses (e.g., (Kennedy and Courchesne, 2008; Lemieux et al., 2007)). Excluding volumes from analysis is sometimes called "scrubbing", a reference to the practice of discarding unreliable data, and the procedure is also called "censoring". Fig. 3B shows the changes produced by censoring volumes with FD_{Power} >0.5 mm in a cohort of 22 children scanned at 3 T. These images complement the QC-RSFC analyses in Fig. 3A, showing that, in data prepared without any nuisance regressions, censoring high-motion volumes decreases all correlations, with the greatest effects at shorter distances. When nuisance regression (including global signal regression) is performed, short-distance correlations are decreased by censoring, and longer-distance correlations are slightly increased, but the overall magnitude of change in correlations is considerably reduced.

A point that is sometimes overlooked in discussions of distance dependence is the fact that is that the disruptions produced by motion are also orientation-dependent. The strongest correlation changes produced by the censoring analysis in Fig. 3B are shown in Fig. 3C. The blue lines, correlations that are decreased by scrubbing, are predominantly short-distance and oriented laterally, whereas correlations that are increased by scrubbing are in red and are predominantly long-distance and oriented in the anterior–posterior and vertical axes. This orientation-dependence can be appreciated when the changed

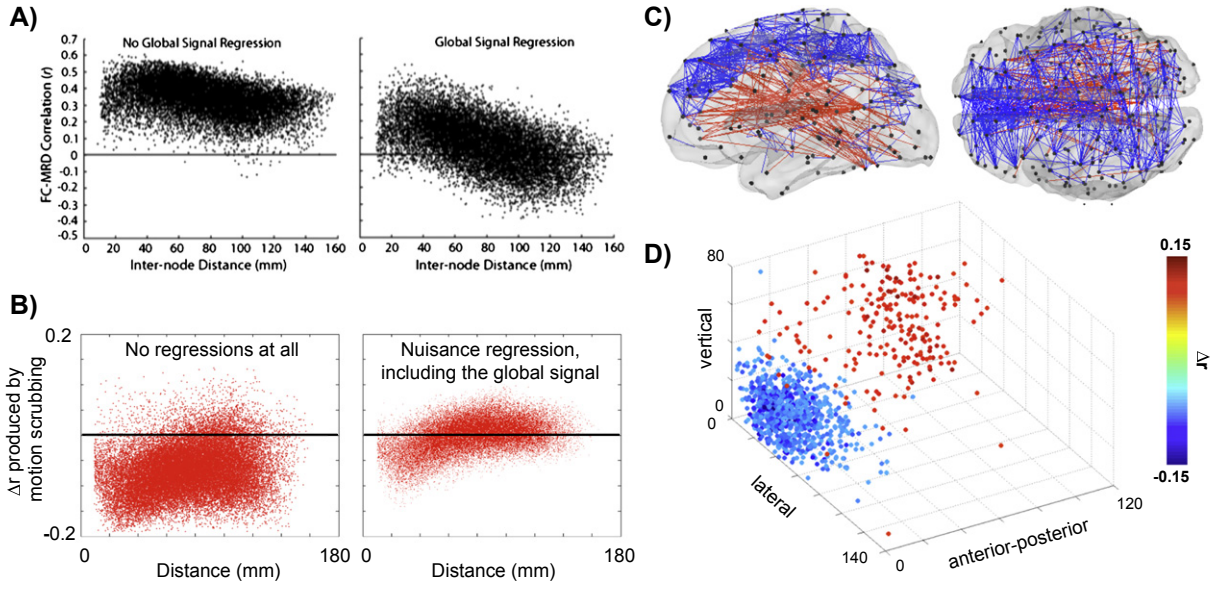

Fig. 3. Spatial aspects of motion artifact's impact on correlations. A) QC-RSFC correlations across hundreds of adult subjects, where the QC measure is FD_{FSL}, modified from (Satterthwaite et al., 2013a). B) Scrubbing analyses of FD_{Power} > 0.5 mm, modified from Power et al. (2012). C) The locations of the top 2% of changed correlations in Fig. 3B, with blue and red colors representing correlations that are decreased and increased, respectively, by scrubbing. D) Projections of the top 3% of changed correlations of Fig. 3B onto the X, Y, and Z axes.

correlations are plotted from a common origin (Fig. 3D). These data indicate that motion tends to increase correlations with predominantly lateral orientations, and decrease those with predominantly vertical or anterior–posterior orientations. It is possible that this phenomenon reflects a preponderance of motion in the form of head nodding (which is symmetric about the lateral axis), or that it is a property of specific acquisition sequences, but the cause is not presently known.

The techniques introduced in this section, QC-RSFC correlations and scrubbing techniques, have properties that are useful and/or limiting in different contexts. The QC-RSFC correlations are attractive since they can be applied to any dataset, regardless of the processing stream, to measure the success of various denoising strategies at unlinking subject motion from outcomes of interest (here, pairwise correlations were the outcome, but any outcome could be used, see (Yan et al., 2013a)). A limiting factor is that such analyses are generally not practical for single-subject or single-scan assessment, since the reliability and accuracy of the correlations scale with the number of measurements (here, scans or subjects). Another limiting factor is that if subjects (or scans) have little variability in the QC measure, it is difficult to reliably establish a correlation. For example, the 348 subjects studied in Satterthwaite et al. (2013a) span a wide range of mean FD values, allowing QC-RSFC correlations to reveal distance-dependence. But consider a scenario where all subjects have high motion. In this case, it is conceivable that QC-RSFC correlations could be near zero (if motion-related increases had a ceiling effect, for example, or if all subjects had similarly high levels of motion and the QC range was narrow). A low QC-RSFC correlation would be falsely reassuring in this situation, since those subjects would certainly be influenced by motion artifact. In relation to this point, it should be noted that when subjects are binned into low-, medium-, and high-motion groups by mean FD, the slopes of FD-RSFC relationships are highest in low-motion subjects, lower in medium-motion subjects, and still lower in high-motion subjects, consistent with a ceiling-like effect of motion on correlation values (Power et al., 2014). These caveats aside, successful removal of motion artifact should remove the distance dependent slope and the positive offset in QC-RSFC measures. It is possible that the variance of QC-RSFC values about their mean may also communicate something about a denoising strategy or dataset, and that a zero-mean distribution with a restricted range is preferable to a zero-mean distribution with a broader range of values, but this is not established.

The censoring technique is attractive because it precisely characterizes particular volumes of data, and because it can be applied to single scans. A limiting factor is that some threshold must be chosen (and justified) to define which data to censor, and a further limitation is that censoring reduces the number of datapoints from which to derive correlations. A major limitation is that *censoring analyses reveal nothing about the uncensored data*. It is easy to imagine scenarios where a denoising technique largely corrects high-motion volumes (e.g., via interpolation), but does not correct subtler influences of motion (e.g., artifact in medium-motion volumes that do not meet interpolation criteria). In such a scenario, censoring high-motion volumes might reveal little or no change in correlations, indicating that the high-motion volumes were well-corrected, but the analysis would also fail to reveal the artifact still residing in medium-motion volumes. In this way, censoring analyses too can be falsely reassuring.

For these reasons, no single technique necessarily establishes that a dataset is free of motion artifact. Instead, convergent evidence from several lines of inquiry is needed to give confidence that motion artifact has been removed. In our own work, we characterize motion artifact using scrubbing analyses and we incorporate scrubbing into the processing of data (discussed below), but the success of denoising is assessed by the elimination of QC-RSFC correlations, by the inspection of post-denoising time series and QC traces for evidence of motion-related disruption, and by the elimination of group differences between motion-binned subjects (these subjects are typical, healthy subjects who, to our knowledge, differ only in amount of motion exhibited).

The temporal impact of motion on functional connectivity correlations

The data in Fig. 2 indicate that stereotyped signal decreases are often seen in the ~5 s during and after motion, but also that individual disruptions can last longer than 5 s. Recall that to impact correlations, a signal disruption needs only to be present. An important question is: for how long, post-movement, does motion artifact alter correlations? Prior to nuisance regression (i.e., in data such as are shown in Fig. 2B), censoring volumes before, during, and after motion indicates that the alterations in correlations are not detectable before motion, are greatest during motion (defined as $FD_{Power} > 0.2$ mm, see Fig. 1A), and are detectable for roughly 8–10 s after motion (Power et al., 2014). This temporal pattern is unchanged by nuisance regression of white matter signals, CSF signals, and motion estimates (Power et al., 2014). When global signal regression is included in the nuisance regression, alterations are restricted to the period of motion and are no longer found in volumes after motion (Power et al., 2014). The duration of these effects is not easily attributed to some aspect of data processing. The asymmetric effects are present before nuisance regression (Fig. S4 of (Power et al., 2014)), and, though frequency filtering is symmetric in time, no effects are seen in volumes immediately prior to motion after frequency filtering (Fig. 8 of (Power et al., 2014)). In our data, similar duration and asymmetry of motion-induced signal disruptions are also seen when slice time correction (via sinc interpolation in our routines) is omitted from the processing stream.

The duration of motion effects in these analyses is surprising. One possibility is that motion co-occurs with other artifactual mechanisms to produce certain delayed effects. Anyone who has undergone a 10 or 20 minute resting state scan knows that such acquisitions can be boring. Yawning (even surreptitiously) causes head motion and also fluctuations in pCO_2 and heart rate. Some motions may reflect yawns, and some post-motion effects may be related to pCO_2 changes or other changes due to deep inhalation or breath-holding. Future investigations of datasets with concurrent physiological recordings may identify such instances of linked physiological and motion artifact.

Methods to reduce or remove motion-related variance

Having covered many of the recent gains in understanding motion's impact on signals and correlations, we now turn to recent developments in removing motion artifact. We first review progress in post-hoc corrections that can be applied to any existing dataset. Progress in acquisition methods is mainly relevant to future datasets and will be discussed later.

Realignment estimates

For many years, it was common practice in resting state analyses to regress the 6 realignment estimates, and sometimes their first temporal derivatives, to remove motion-related variance. The first three reports on motion artifact (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012) followed such strategies, which inadequately removed motion artifact, and subsequent studies have also confirmed the inadequacy of these regressors (Kundu et al., 2013). In 1996, Karl Friston and colleagues proposed that expansions of realignment estimates might be useful to remove spin-history related aspects of motion-related artifact (Friston et al., 1996). The expansions, for R = [X Y Z pitch yaw roll], take the form $[R\ R^2]$, $[R\ R^2\ R_{t-1}\ R^2_{t-1}]$, $[R\ R^2\ R_{t-1}\ R^2_{t-1}\ R_{t-2}\ R^2_{t-2}]$, and so forth, each expansion incorporating information further back in time. In recent papers, 24-parameter (incorporating 2 time points) (Power et al., 2014; Satterthwaite et al., 2013a; Yan et al., 2013a) and 36 parameter expansions (incorporating 3 time points) (Power et al., 2014) have been tested. Unfortunately, even expansions including terms from 3 time points leave much motion-related variance in data (see Fig. 4 for a visual illustration). Further expansions seem likely to yield diminishing returns and are unlikely to justify the degrees of freedom they would occupy in regression models. Measurable gains in explained variance are produced by increasing the number of motion regressors

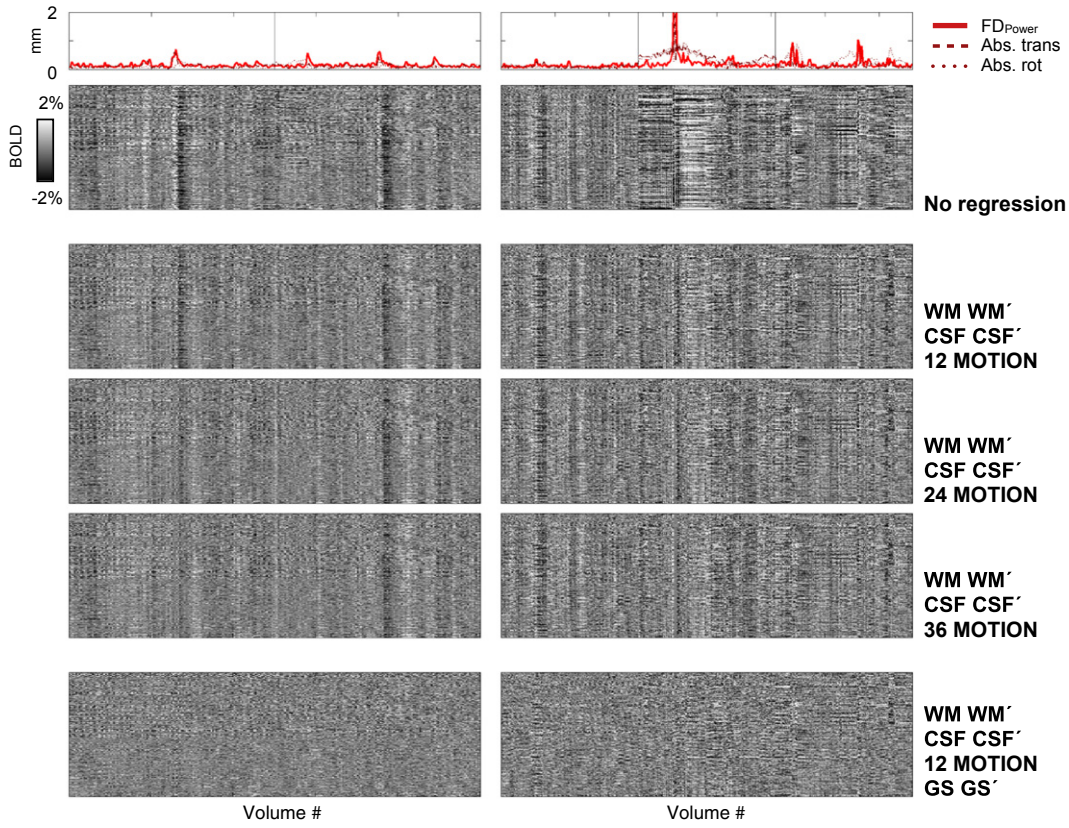

Fig. 4. Effects of different regressors on gray matter signals. At top, the data from the 2 subjects shown in Fig. 2. The middle 3 gray panels show the effect of regressing increasing numbers of motion parameters. The bottom panel shows the effect of regressing the global signal. WM: white matter signal; CSF: ventricle signal; GS: global signal; R: realignment estimates. The 12 motion parameters are [R R′], the 24 are [R R^2 R_{t-1} R^2_{t-1}], and the 36 are [R R^2 R_{t-1} R^2_{t-1} R_{t-2} R^2_{t-2}].
Modified from Power et al. (2014).

from 12 to 24 to 36, but the gains are modest and must be balanced against lost degrees of freedom. For the modest gains delivered, the cost of adding additional motion regressors may not be attractive in comparison to other methods to remove motion artifact (described below). It should also be noted that regression of voxel-specific motion measures (derived from whole-brain measures, see (Bullmore et al., 1999)) has yielded no denoising benefits over regression of whole-brain motion measures (Satterthwaite et al., 2013a; Yan et al., 2013a).

Non-gray-matter nuisance signals

The heterogeneity of motion's effects suggests that fMRI-signal-based denoising may be more successful than motion-parameter-based denoising. Regression of mean white matter and mean CSF signals have been of modest utility in reducing motion artifact (Power et al., 2014; Satterthwaite et al., 2013a; Yan et al., 2013a). Whether signals from those compartments may be used to better effect by subdividing those signals (and/or other nuisance signals) using spatial criteria (e.g. (Jo et al., 2010)) or data reduction techniques (e.g. (Behzadi et al., 2007)) is beginning to be explored.

ANATICOR (Jo et al., 2010) is a method that derives a nuisance regressor for each gray matter voxel from nearby white matter voxels within a specified distance. Because local motion effects (type 1 and type 3 disruptions) presumably cause much of the distance-dependent nature of motion artifact, ANATICOR, by deriving and removing local nuisance signals, is in principle well-positioned to address this aspect of motion artifact. In (Jo et al., 2013), ANATICOR largely suppressed distance-dependent artifact in a scrubbing analysis with FD_{AFNI} > 0.25 mm as a threshold (Fig. 5A). However, the amount of motion artifact in all other (non-scrubbed) volumes is unknown since no QC-RSFC correlations or residual time series were presented. ANATICOR thus appears to be a promising denoising strategy with respect to motion artifact, but further analyses are needed to understand its efficacy.

CompCor (Behzadi et al., 2007) is a method that defines voxels whose signal is presumably of no interest (in the white matter, CSF, or other non-gray matter locations) and uses Principal Component Analysis (PCA) to obtain a small set of nuisance regressors from the larger set of voxelwise signals of no interest. This method, like ANATICOR, could obtain more locally-specific signals that may not be well-reflected in mean CSF or mean white matter signals. One report has found that implementations of CompCor can reduce FD:DVARS correlations more than regression of mean white matter and mean CSF signal regression (Muschelli et al., 2014). However, QC-RSFC analyses indicate that after CompCor, QC-RSFC values center about r = 0.2, and increased values are observed at shorter distances, indicating incomplete removal of motion artifact (Fig. 5B).

Gray matter signals: ICA denoising

Another way to derive nuisance signals is by matrix decomposition of whole-brain data, usually via Independent Component Analysis (ICA). ICA is a blind source separation technique that reduces a set of signals to combinations of fewer signals (component signals). Ideally, in resting state fMRI, these components would separate neural signal from noise. In typical 3 T datasets, however, “neural” ICA components exhibit motion dependence, indicating that motion-related variance is not effectively partitioned from neural variance (Mowinckel et al., 2012; Satterthwaite et al., 2012; Tyszka et al., 2014). It is not yet known whether higher-dimensional ICA (e.g. (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014)) in high quality datasets (such as the Human Connectome Project data) will be able to adequately separate such variance; such investigations are underway. One point worth mentioning is that ICA, by definition, identifies maximally spatially

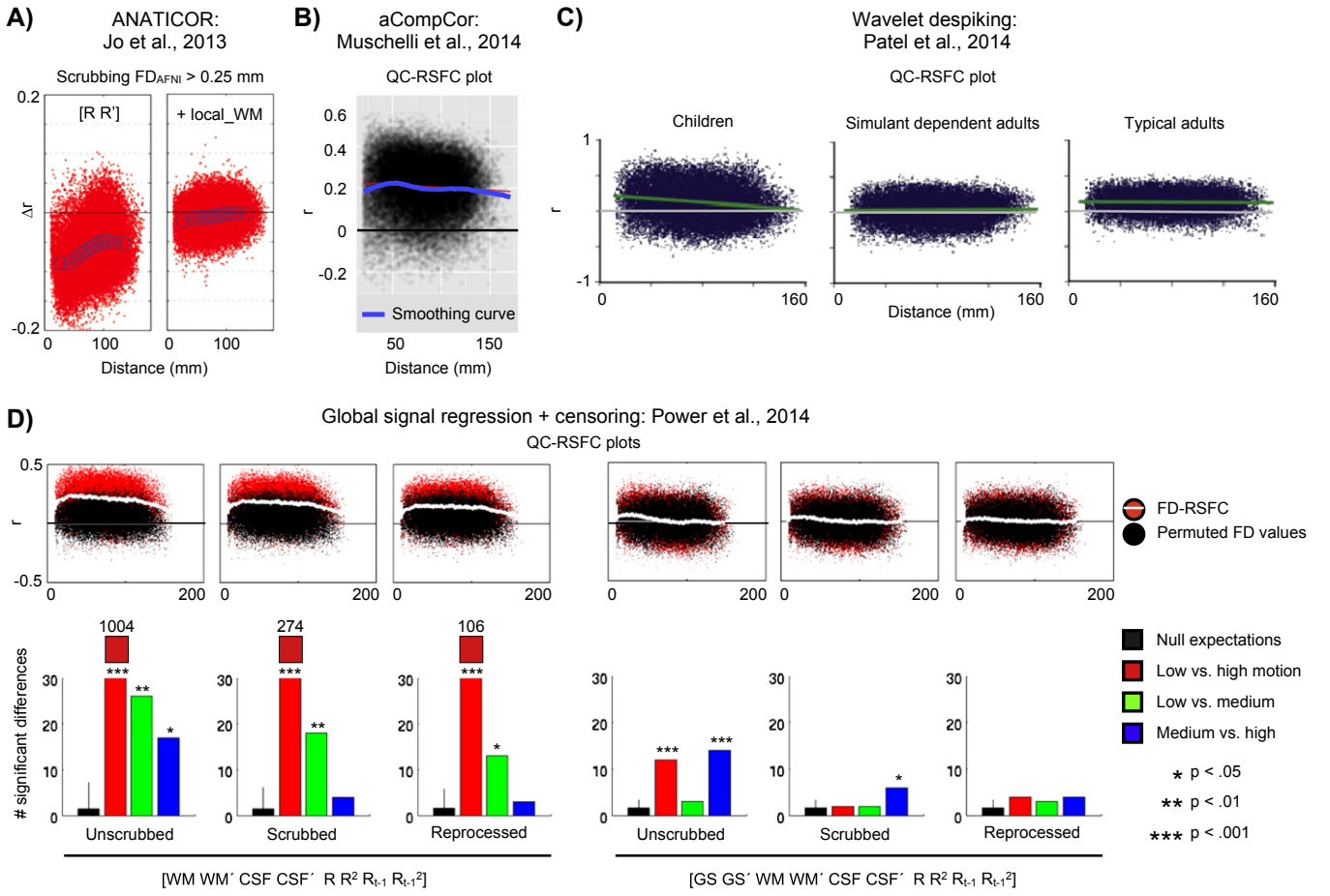

Fig. 5. Removal of motion artifact under various processing strategies. A) Scrubbing analyses of volumes with FD$_{AFNI}$ > 0.25 mm, when regressing only 12 motion parameters or the 12 motion parameters and a local white matter regressor derived by ANATICOR. B) QC-RSFC correlations after applying aCompCor to a dataset. C) QC-RSFC correlations in 3 datasets after wavelet despiking. D) QC-RSFC correlations in 120 adults under processing streams including or excluding global signal regression. "Unscrubbed" data have undergone the indicated regression, followed by temporal filtering and spatial blurring. "Scrubbed" data are obtained by censoring "unscrubbed" data with FD$_{Power}$ > 0.2 mm as a threshold. "Reprocessed" data undergo identical censoring, except the censored volumes are also withheld from the nuisance regression calculations and are interpolated over prior to temporal filtering. The bar charts indicate the number of significant differences among 40-subject cohorts of motion-binned adults, identified by two-sample, two-tailed t-tests with $p < 0.0005$. 10,000 permutations of FD values across subjects established null expectations and significance levels for these comparisons. WM: white matter signal; CSF: ventricle signal; GS: global signal; R: realignment estimates.

independent signals in the data. More local signal disruptions (type 1 and type 3) may therefore be well-separated as sources by ICA. However, globally shared artifact (type 2 disruption) may be more difficult to isolate and remove by ICA, because this variance is shared across sources.

Gray matter signals: wavelet despiking

Recently, Patel and colleagues proposed a denoising method that decomposes voxel signals into wavelet space, identifies trains of "bad" high-amplitude wavelets that likely reflect motion artifact, and reconstitutes denoised voxel signals from the "good" wavelets (Patel et al., 2014). This method removes many motion-related artifacts from the time series, but it is not clear whether motion artifact is fully removed: QC-RSFC plots indicate that, after wavelet despiking, motion dependence remains positive in 2 of 3 datasets and distance-dependence is seen in the highest motion dataset (Fig. 5C). However, the mean values of these distributions are not within the top 2.5% mean values seen in permutation testing of the same data, and the authors argue that it is therefore possible that the observed QC-RSFC values arise by chance rather than due to motion artifact.

Gray matter signals: global signal regression

Several groups have noted that regressing the global signal is very effective at removing motion artifact (Power et al., 2014; Satterthwaite et al., 2013a; Yan et al., 2013a,b). The panels of Fig. 4 visually shows the failure of several sets of motion regressors and mean white matter and mean CSF signals to remove motion artifact, and the visible efficacy of global signal regression relative to those regressors. Other successful techniques should show a similar visual elimination of motion-related variance. The utility of global signal regression in removing motion artifact (especially in combination with censoring motion-contaminated volumes at all stages of processing) can also be seen in the QC-RSFC plots in Fig. 5D. Global signal regression is one of the more contentious issues in the field of resting state fMRI, and entering into that debate is beyond the scope of this review. We only note that the objections raised to global signal regression are mainly based on results from low-dimensional simulations (Murphy et al., 2009; Saad et al., 2012), and that further work that determines the applicability of these arguments to empirical data would usefully inform decisions about using global signal regression as part of denoising strategies.

Censoring motion-contaminated data

At present, most regression techniques leave residual abnormalities in the data during at least some periods of motion (i.e., the denoising regressions fail to fully correct motion-induced artifact). These abnormalities are sometimes easily seen in time series or images (e.g., the upper panels of Figs. 4 or the slices in Fig. 7A), and sometimes they are more subtle and better identified via DVARS or other data quality

traces (e.g., even though the bottom panel of Fig. 4 is less obviously impacted by motion, DVARS traces still identify abnormalities during some movements, see (Power et al., 2014)).

In the face of incomplete artifact removal, another step that can be taken is to censor the problematic timepoints. Many groups have found that censoring improves the removal of motion artifact (Carp, 2013; Power et al., 2012, 2013, 2014; Satterthwaite et al., 2013a; Yan et al., 2013a,b). For example, a processing stream that incorporates global signal regression and censoring throughout processing nearly eliminates QC-RSFC correlations at all distances (though there may be a slight dependence at short distances, see Fig. 5D) and reduces group differences between motion-binned typical adults to chance levels (though differences are still slightly elevated from mean chance levels, see Fig. 5D).

Censoring introduces a tension between removing volumes to reduce motion artifact and a desire to not remove volumes in order to preserve a dataset. An important question therefore is what criteria should be set to censor data. As Fig. 2A indicates, the magnitude of motion artifact scales with the magnitude of motion (though not necessarily linearly). FD (and DVARS) traces, such as those shown in Figs. 1 and 2, exhibit low values throughout most of the scan, which we refer to as "floor" values in the traces. Practically speaking, we find that volumes that rise above this floor (e.g., $FD_{Power} > 0.2$ mm in Figs. 1 and 2) contain demonstrable motion artifact, and we censor these volumes in our own processing streams (Power et al., 2014). In our initial report, to characterize motion-contaminated volumes, we used a threshold of $FD_{Power} > 0.5$ mm (the top dotted lines in Fig. 2B). Lowering the scrubbing threshold from 0.5 mm to 0.2 mm (the bottom dotted lines in Fig. 2B), in order to censor as much motion-contaminated data as possible without impacting "floor" volumes, reveals and removes more distant-dependent motion artifact (Power et al., 2014).

Most analysis streams contain multiple linear regressions of nuisance variables (e.g., mean white matter signal). Censoring can be incorporated into multiple regressions by either deleting the unwanted timepoints via concatenation (Power et al., 2013, 2014) or adding single-timepoint nuisance regressors (delta functions) at the unwanted timepoints (Lemieux et al., 2007; Satterthwaite et al., 2013a; Yan et al., 2013a). These methods should yield identical regression fits.

Concerns about censoring

Several concerns about censoring have been raised by various groups. Investigators who censor motion-contaminated data need to be aware of several limitations of the technique.

One concern relates to degrees of freedom. At a given threshold, the number of volumes removed by censoring will scale with the amount of motion in the scan, leading to fewer degrees of freedom in subjects who move more. This could lead to biases that covary with factors of interest, such as less data in higher-moving clinical cohorts. One way to deal with this statistical bias is to censor all datasets to identical lengths, i.e., if the worst subject included in analysis only has 142 volumes remaining, to censor all other scans to include only 142 volumes as well. Alternatively, groups can be statistically matched for amount of data remaining. However, there are certain complications that should be considered. For example, contrast censoring the final 5 min of a 10 minute scan to censoring every other volume of a 10 minute scan. The data sum to 5 min in both cases, but due to hemodynamic autocorrelation, and the fact that lower frequencies contribute most of the power in resting state signal, in the latter scenario, much more than 5 min of signal properties are represented. In general, when the censored timepoints are many in succession, and/or temporally clustered, more information is lost about that scan than when an equivalent number of timepoints are censored sparsely throughout the scan, and the type of information lost may be specific to certain spectral or dynamic aspects of the data. Published examples of the duration of censored data portions are shown in Fig. S1. These histograms indicate that most censored data are short segments of 1 or a few TRs, though many successive TRs are sometimes removed (Fig. S1). Little attention has been given to the temporal "clustering" of censored timepoints and what effects this may have on data properties.

Another concern is that correlation estimates may become excessively noisy if data is removed. In general, as more timepoints contribute to a correlation, the correlation becomes more accurate and statistically reliable, and vice versa as timepoints are removed. For many years, the field has accepted ~5 min as an adequate resting state scan length (Van Dijk et al., 2010). To ameliorate concerns about noisy estimates, investigators may choose not to use subject data that is reduced below this limit by censoring. There are no established or agreed-upon limits on how much data must remain after censoring, what proportion of a subject's data may be censored, or even how to measure or report the amount of data that remains after censoring (e.g., to fully account for the considerations in the previous paragraph).

A variation of this concern is that the distribution of correlations will widen and is more likely to contain "extreme" values after censoring. From first principles, among random, uncorrelated signals, the distribution of correlations will narrow (around zero) as correlations are calculated from longer signals. Put differently, it is much easier to obtain a strong correlation (positive or negative) between noisy signals when only 8 timepoints are measured, as opposed to 80 or 800. Yan et al. (2013a) illustrated this issue in real data by successively removing timepoints from a ~7.8 min scan and displaying the median, maximal, and minimal correlations obtained in seed maps from the posterior cingulate (Fig. S1). When a third of the data are censored, bringing the scan to ~5 min, very little change is produced in the correlation distribution. Below this point, the most negative correlations observed gradually increase in magnitude, and eventually, when only 50 or 8 timepoints contribute to correlations, more obvious changes in the distribution are seen. These windows correspond to computing correlations on ~1.7 and ~0.25 min of data, respectively, well below the 5 min that is typically used in the field. Thus, when longer amounts of data are used, this concern is ameliorated, but when short amounts of data are used, this concern is amplified.

Another concern is whether censoring biases the sampling of brain "states". It is possible that transient motion-related brain states exist and that they become underrepresented by censoring. However, much or sometimes nearly all signal during motion is spurious (Friston et al., 1996; Kundu et al., 2013; Power et al., 2014; Satterthwaite et al., 2013a), spurious signal changes are of large amplitude and are found throughout the brain (i.e., not just in motor-related locations), and it is currently difficult to separate spurious signal from signal of interest. Future methods may enable better detection and characterization of "motion-related" states, but at the moment it is difficult to confidently characterize "states" derived from fMRI during motion.

Many of these concerns could be helpfully addressed by simulations. Little simulation of motion-related effects has been performed (though see, e.g. the supplemental materials of (Satterthwaite et al., 2013a)), in part due to the complicated appearance of signal changes. A more mechanistic understanding of motion-related signal changes, in concert with relevant simulations, will advance the field.

Frequency filtering

Many analysis streams also involve frequency filtering. Information on the frequency content of motion artifact is limited. One study found that high-motion subjects displayed increased power throughout the frequency spectrum typically examined in resting state studies (Satterthwaite et al., 2013a). This study also found that denoising regressions (including 24 motion parameters, white matter, ventricle, and global signals and censoring $FD_{FSL} > 0.25$ mm) removed lower-frequency disparities in power (below 0.08 Hz) but not higher-frequency disparities in power (above 0.08 Hz), suggesting that band- or low-pass filters set to remove frequencies over 0.08 Hz may aid removal of motion-related variance. Another study noted spatially-specific, motion-dependent modulation of power in particular frequency

bands, with effects in both fronto-parietal and default mode networks (Kim et al., 2014).

Usually, frequency filtering is performed by convolving time series with a filter kernel, a method that requires continuous time series. As noted by Carp (2013), these filters can spread a large-amplitude artifact from one motion-contaminated timepoint into temporally adjacent, non-contaminated timepoints. This poses a challenge for incorporating a censoring procedure into frequency filtering, a challenge that has been addressed in multiple ways. One way is to regress basis functions (e.g., sine and cosine waves) with the to-be-removed frequency content from the data, using delta functions to censor unwanted timepoints. This approach can be combined with nuisance regression in a single regression model (Hallquist et al., 2013; Jo et al., 2013). This approach can also be used to estimate the frequency content of discontinuous data (Power et al., 2014), and to generate frequency-based measures such as fractional amplitude of low frequency fluctuations (fALFF). Another effective way to reduce the spread of artifact during frequency filtering is to replace outlying to-be-censored values with less-outlying values via interpolation (discussed below; (Carp, 2013)).

One issue raised by several groups is that the frequency content of nuisance regressors needs to match the frequency content of the data undergoing regression (Hallquist et al., 2013; Jo et al., 2013; Patel et al., 2014). This frequency matching requirement means that 1) frequency filtering should be incorporated into the regression model (as described above), or 2) regression should precede frequency filtering (as described above), or 3) if regression follows frequency filtering, then all nuisance regressors should undergo the same filtering process as the data. In two of the first publications on distance-dependent motion artifact, unfiltered motion regressors were regressed from data that had been bandpass filtered, re-introducing high-frequency motion-related fluctuations into the data (Power et al., 2012; Van Dijk et al., 2012). Subsequent publications using frequency-congruent nuisance regressors (regression prior to frequency filtering) have reproduced the effects noted in the original papers (Power et al., 2014; Satterthwaite et al., 2012, 2013a; Yan et al., 2013a).

Interpolation

A practice conceptually linked to censoring is replacing abnormal data with synthetic data, usually via interpolation from surrounding timepoints. Data replacement raises many of the same issues as censoring (e.g., which data should be replaced, whether unintended biases may be created), but two additional and related issues are raised: what synthetic qualities are generated at interpolated timepoints, and how to treat the synthetic data in comparison to the real data.

The degree to which a dataset assumes the qualities of the (particular type of) synthetic data depends both on the amount of data replaced, and the sizes of the replaced portions of data. The qualities of any interpolation method should be similar over brief intervals, such as 1 or 2 TRs of replaced data. But the interpolated data will diverge across methods as the size of the interpolation increases. For example, a spline fit can be very different than a linear fit over gaps larger than a few TRs. In certain situations, such as during frequency filtering, the type of interpolation employed may have appreciable effects on the adjacent non-interpolated timepoints. In our own work, we have used estimates of the frequency content from the uncensored ("good") data to construct synthetic data at censored ("bad") timepoints, yielding a continuous signal which then undergoes frequency filtering, followed by re-censoring those "bad" time points (Power et al., 2014).

In addition to the qualities of the interpolation, investigators must decide how to treat the synthetic data in comparison to real data. Options include treating the data equivalently, weighting the different types of data, or ultimately censoring the synthetic data. As an example of the former approach, despiking procedures are sometimes used to replace outlying voxel intensities with less-outlying intensities, but the replaced timepoints may thereafter be treated no differently than the rest of the data. In contrast, as an example of the latter approach, when we replace abnormal timepoints, it is as an intermediate step in data processing (described above), and the synthetic data are ultimately censored (because they derive from uncensored volumes and do not add information to the data). When few timepoints are replaced, it may make little difference whether synthetic data are retained or censored, but the importance of this issue will scale with the proportion of data replaced.

Other post-hoc subject-level methods

In a recent publication (Scheinost et al., 2014), Scheinost and colleagues investigated interactions between rigid body realignment interpolations, image smoothness, and motion. These authors found that motion correlates with image smoothness, that cubic and spline interpolations resulted in less image smoothing than linear interpolations, and that image smoothness correlated with several connectivity measures. An iterative smoothing procedure designed to bring subjects within a group to similar levels of smoothness was proposed, and this procedure lessened the dependence of the examined connectivity measures on motion.

Some groups have, at the subject level, generated scrubbing plots similar to those shown in Fig. 3B, fitted those plots with polynomial curves, and regressed those curves from the subject-level data as a means of reducing motion-related effects (Fair et al., 2012).

Another approach to motion correction, proposed by Marko Wilke, involves deriving motion estimates and applying them to the first volume of a scan in order to create synthetic time series estimates of motion-related effects in the data, followed by derivation of signals from various portions of the synthetic dataset as nuisance regressors (Wilke, 2012). This approach can explain amounts of variance comparable to some motion regressions. However, the synthetic signals generated do not exhibit global modulations of signal, which are present empirically (Fig. 2), indicating that the model used to generate the synthetic data would benefit from further refinement.

Group-level covariates

Beyond subject-level corrections, group-level corrections for motion are also possible. These corrections are typically implemented by including in statistical models covariates of no interest across subjects such as mean FD (Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013a) or global levels of correlations (Saad et al., 2013; Yan et al., 2013b). A related technique is standardization, whereby a subject-level parameter is used to normalize some property of a subject, such as Z-standardization (Yan et al., 2013c). The merit of this approach is that regression of such covariates can reduce motion-related group differences, which is the primary aim of the technique. But as with any covariate of no interest (e.g., IQ, motion, etc.), if the covariate is collinear with an effect of interest (e.g., if motion covaries with age in a developmental study), across-subject regression of these covariates can also remove effects of interest. It is thus highly desirable to decouple motion covariates from anticipated variables of interest. Unfortunately, this is often difficult, since, as previously mentioned, motion tends to be higher in the young, the elderly, and clinical populations, who are often the focus of a study.

Another consideration is whether covariates of no interest, which are usually removed via linear regression, exhibit linear relationships to outcomes of interest, such as RSFC correlations. In our data, relationships between FD_{Power} and RSFC correlations can be modeled linearly but they also exhibit some non-linearity: RSFC correlations increase most rapidly in the lower range of FD_{Power} but level off at higher ranges of FD_{Power} (Power et al., 2014). Such non-linear relations limit the utility of FD_{Power} as a covariate of no interest in group-level corrections. This nonlinearity is also a limitation for QC-RSFC analyses that measure linear relationships between RSFC and FD (see the discussion of ceiling

effects in the "Techniques to reveal distance-dependent artifact" section); this limitation is one reason why we advocate using multiple types of analyses to assess artifact removal. Whether other summary measures of data quality have linear relationships to outcomes of interest is unknown, though it is known that various FD measures correlate highly.

Partial correlations

One possible approach to removing motion artifact is to implement some denoising procedure that produces incompletely denoised time series (e.g., mean white matter signal regression), and then to use partial correlations rather than full correlations to study relationships among the time series. For example, given a set of N signals, for any 2 signals, the other N-2 signals can be used as regressors to derive the variance unique to those 2 signals, forming the basis of the partial correlation. If signals share artifact, then partial correlation approaches may reduce the contribution of artifact to computed relationships.

There is limited information on how well partial correlation approaches reduce motion artifact. One publication has examined the influence of motion on properties of 90-node networks formed of AAL atlas parcels, comparing networks derived by full correlation or partial correlation (derived using graphical lasso methods) (Yan et al., 2013b). In this report, using partial correlations did not fully suppress the dependence of network properties on motion, and using partial correlations was less effective in suppressing such dependencies than was global signal regression. One useful avenue of future research will be to replicate this finding, and to determine why partial correlation approaches are not more successful at reducing motion artifact.

Multi-echo acquisitions

One emerging technique for correcting motion is to acquire fMRI signal at multiple timepoints during a single TR (i.e., to have multiple echo times). The fMRI signal observed at any time during a TR is a function of the intensity to which the voxel was raised by the electromagnetic pulse at the beginning of the TR (initial intensity, S_0), the elapsed time (t), and the rate of exponential decay of the signal (the relaxation rate, R_2^*): $S(t) = S_0 \exp(-tR_2^*)$. Within a voxel, brain activity is measurable because neural activity increases blood flow, which increases the local ratio of oxy- to deoxyhemoglobin, which alters the magnetic properties of the blood, which ultimately changes the relaxation rate of the perfused tissue (R_2^*). Thus, signal intensity at a given time is a product of two factors (S_0 and R_2^*), only one of which reflects neural activity (R_2^*). As Fig. 6A illustrates, under ideal conditions in which S_0 is constant at all TRs, when signal is acquired at a single point during a TR, the relaxation rate can be inferred from measured signal intensity. However, in reality, S_0 is not constant but is instead modulated by factors including motion. As Fig. 6B shows, acquiring data at a single echo time cannot distinguish whether signal changes are caused by altered relaxation rates or altered S_0. Multi-echo methods, by acquiring data at multiple time points during a TR, can help to disambiguate S_0 changes from relaxation rate changes (Fig. 6C, see also (Kundu et al., 2012)).

Thus far, limited information exists on the efficacy of multi-echo methods to reduce or eliminate motion artifact. In one study that used 2 echo times (Bright and Murphy, 2013), signal from the 1st echo time (TE = 3.3 ms, the "short" TE) was regressed from the signal at the 2nd echo time (TE = 35 ms, a "typical" BOLD-weighted TE), helping to adjust the data for changes in S_0. This method explained motion-related variance beyond motion regressors, but did not eliminate the distance-dependent changes caused by motion artifact (Fig. 6D).

In another study (Kundu et al., 2013), data were acquired at 4 TEs (12, 28, 44, and 60 ms) in order to partition the data into ICA components whose signal reflects artifactual S_0 changes (such as those caused by motion) versus relaxation rate changes (such as those caused by neural activity or changes in pCO_2), building on the ME-ICA methods outlined in Kundu et al. (2012). Ultimately, the data were partitioned into components-of-no-interest (the artifactual

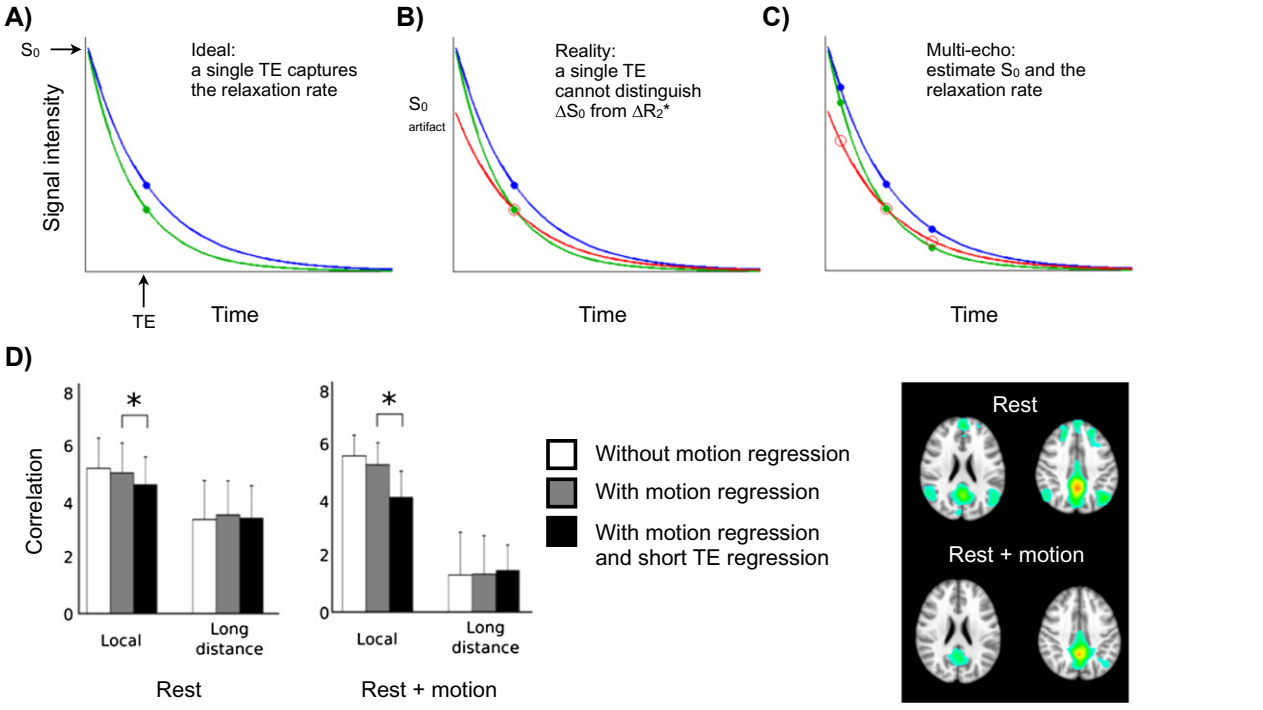

Fig. 6. Using multi-echo sequences to remove motion artifact. A–C) Schematic illustrations of the advantages of multi-echo sequences over single-echo sequences for identifying ΔS_0-related artifact. D) Data from the dual-echo study of Bright and colleagues. Bar plots show correlations between regions of the default mode network in conditions of normal rest and purposeful movement, without and with short-TE signal regression. Slices show seed correlation maps for the posterior cingulate after short-TE signal regression in conditions of low and high motion.
Modified from Bright and Murphy (2013).

non-BOLD-like components and any BOLD-like components that localize to venous structures) and components-of-interest (the remaining BOLD-like components).

Several motion-related findings from this publication are noteworthy. First, though DVARS traces indicate that the bulk of motion artifact is partitioned into the components-of-no-interest, they also indicate that motion-related variance is partitioned into the components-of-interest. For example, in Subject 2, both red and blue DVARS traces contain many spikes that correspond to spikes in the FD trace (Fig. 7A). Images from this subject confirm the presence of abnormalities during motion in both BOLD-like and non-BOLD-like components (Fig. 7A). Second, ME-ICA identifies a significantly smaller number of BOLD-like components in high-motion subjects (~33) than in low-motion subjects (~40). Third, while distance-dependent artifact was not explicitly examined via scrubbing or QC-RSFC analyses, group differences among motion-binned subjects were examined in several seed maps (Fig. 7B), and the ME-ICA approach eliminated group differences in motion-binned subjects (which were seen under "conventional" processing). Conventional processing consisted of "despiking with a tanh function", regression of 12 motion parameters [R and R′], and temporal filtering.

Multi-echo sequences enable new denoising techniques, and future datasets could benefit from these sequences. However, the extent to which these techniques advance beyond existing denoising techniques, with respect to motion, is not presently clear. The dual-echo study by Bright and colleagues did not eliminate distance dependent artifact (under demanding conditions of purposeful movement), and distance dependence has not been examined in the multi-echo studies of Kundu and colleagues. Elimination of motion-related group differences was accomplished by ME-ICA, but this has also been accomplished using post-hoc methods, as described above. Given that both post-hoc and emerging multi-echo methods may adequately control motion artifact, but that multi-echo methods cannot be applied to the vast majority of existing (single-echo) datasets, a valuable line of investigation will be to relate the nuisance signals (and RSFC correlations) derived from post-hoc methods to the nuisance signals (and RSFC correlations) derived from the multi-echo approaches.

Other considerations

Further measures of denoising success

Many different analyses have been used to measure motion artifact and the success of denoising strategies. We have discussed scrubbing, QC-RSFC correlations, and the formation and elimination of group differences in typical subjects binned by motion. Another type of outcome measure used in many studies is the reduction of correlations between motion traces and fMRI signal (or signal quality measures) across stages of processing. In general, early in processing, spikes in motion traces (e.g., FD) are closely reflected in signal quality traces (e.g., DVARS), and successful processing ought to reduce such correlations across stages of processing. Similarly, correlations among DVARS traces can be followed through stages of processing to track decoupling of waveforms over processing.

This type of outcome measure is sensible, but our own experience, as well as data from the literature, indicates that these outcomes may not be easy to interpret. For example, we have found that volumes with high, outlying DVARS values early in processing may be reduced to "typical" DVARS values during processing, and yet still contain distance-dependent motion artifact (Power et al., 2014). In other

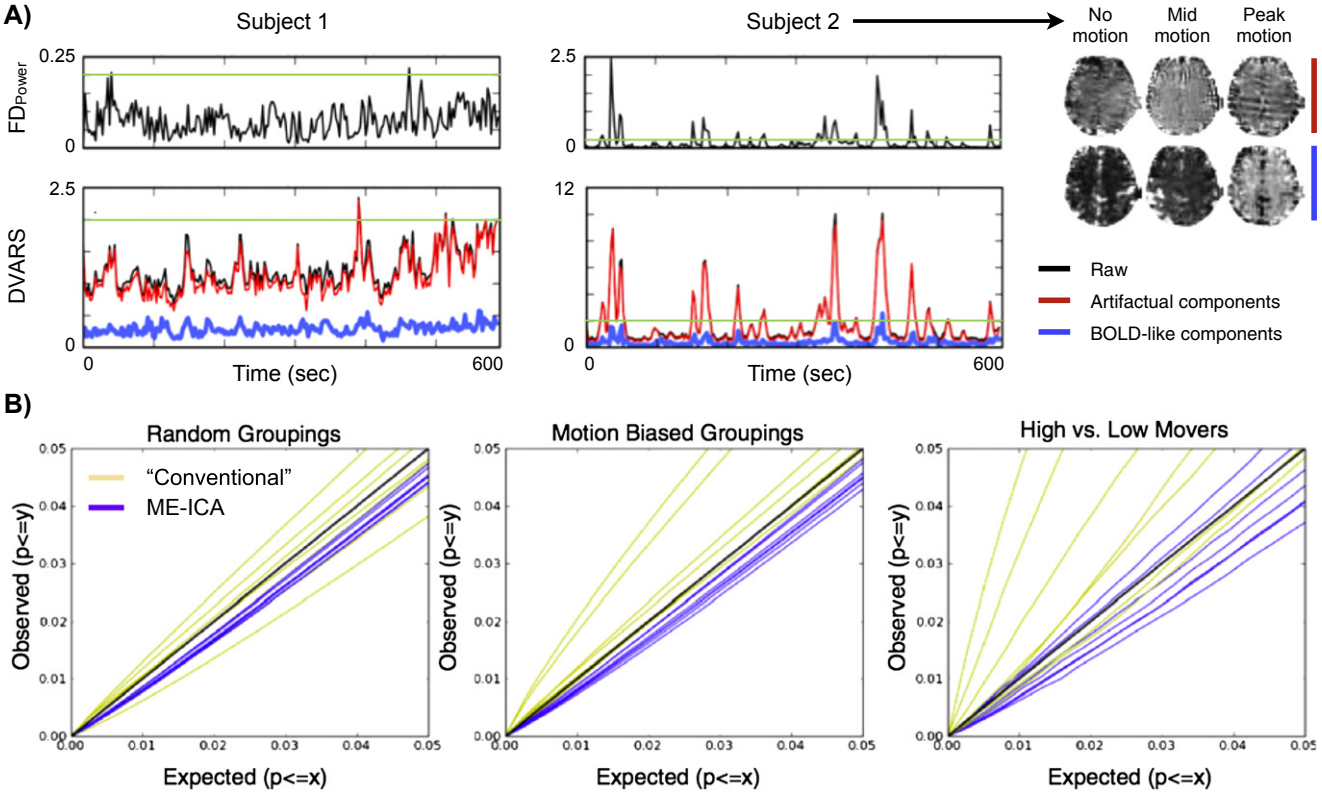

Fig. 7. Motion artifact reduction in the multi-echo ICA approach. A) FD and DVARS traces from 2 subjects. The y-axis scales differ between subjects, and green lines have been placed to aid comparison of the traces. Red and blue traces indicate "bad" (ΔS_0-related) and "good" (ΔR_2*-related) ICA components derived from ME-ICA. The slices at right illustrate the two types of components in Subject 2 at different timepoints in the scan. B) Observed vs. expected numbers of significant differences in several seed maps under "conventional" and multi-echo (ME-ICA) processing strategies. Each line represents a single seed map, where observed numbers of voxels surpassing a statistical threshold for group differences are plotted against null expectations for such differences. Lines above the black line ($y = x$) indicate that seed maps had greater-than-chance numbers of significantly different voxels.
Modified from Kundu et al. (2013).

words, a spike in DVARS traces is practically eliminated, but the underlying distance-dependent artifact is not similarly eliminated. Such cosmetic improvement may result from correction of signal abnormalities at some voxels (which could reduce DVARS) and/or reduction of the amount of rate of signal change (since DVARS is a differential measure) without a full correction of the motion-induced abnormality. It is for this reason that we do not use reductions in FD:DVARS or DVARS:DVARS correlations to assess data improvement across stages of processing, and it is for this reason that we use initial (pre-denoising) DVARS values to identify data to censor (in addition to using FD for such purposes). It is possible that elimination of peaks in DVARS over processing steps are more veridical indices of improvement under processing streams other than our own. However, other reports also contain instances of reduced or eliminated FD:DVARS correlations with concomitant preservation of motion-related effects: in their report on CompCor, Muschelli and colleagues advocate a processing stream that reduced FD:DVARS correlations to being nearly zero-centered (Fig. 4 of (Muschelli et al., 2014)), but which produced data that still exhibited robust QC-RSFC correlations (Fig. S1 of (Muschelli et al., 2014), illustrated in Fig. 5B).

It is worth emphasizing that the scale of DVARS traces changes throughout processing steps (see Fig. 10 of (Power et al., 2014)), and will differ across scanners and sequences. For example, spatial blurring reduces DVARS magnitudes, and larger kernels produce larger reductions. Therefore, as was stated for FD traces earlier, if DVARS measures are to be used in some dataset, DVARS waveforms and magnitudes should be established (and reported) for that specific dataset, since magnitudes from the literature may not be appropriately scaled to the data at hand. In general, the magnitude of DVARS traces will become smaller with more thorough denoising, and spikes in the traces should be eliminated (but see above). However, a small spike in a smaller trace may be just as problematic as a large spike in a larger trace (see the example of the Kundu traces above). In our view, the greatest utility of DVARS lies in the waveform's ability to identify outlying volumes in a dataset, regardless of its magnitude. Thomas Nichols has proposed a standardization of DVARS[1] that might enable easier cross-site comparability of DVARS, but this measure is not widely used at the present time.

FD:BOLD signal correlations, often at the voxel level, are sometimes used to assess the presence of motion artifact and the success of denoising strategies. In our own data, where motion (increases in FD) can variably produce increases or decreases (or both) in fMRI signal, it is difficult to envision how FD:BOLD correlations can be reliably linked to data quality or data improvement. Reinforcing this view, in one of the first reports to use FD:BOLD correlations as outcome measures, of 6 cohorts studied, the 3 cohorts reported to have the highest motion had lower (closer to zero) FD:BOLD correlations than all the other cohorts (see Fig. S2, modified from (Yan et al., 2013a)).

It has been suggested that FD:BOLD correlations reveal motion-related neural activity due to positive FD:BOLD correlations in motor and other peri-Rolandic cortex (Pujol et al., 2014; Yan et al., 2013a,b). Several lines of data, in our view, suggest that that these correlations are probably not related to neural activity. First, it is unclear why FD:BOLD correlations would yield maps of neural activity when standard GLM-based approaches do not appear to yield motor-related activity during head motion but instead show signal decreases throughout the brain parenchyma (Satterthwaite et al., 2013a). Second, it is unclear why BOLD signal would instantly reflect motion-related neural activity, when multi-second delays are traditionally expected between neural activity and BOLD signal modulation. Third, in the report of Scheinost and colleagues, the iterative blurring approach to motion correction eliminates positive peri-Rolandic FD:BOLD correlations (Scheinost et al., 2014). Fourth, the positive FD:BOLD correlations span large portions of cortex beyond the motor strip, including much of the insula, much of the occipital cortex, and portions of prefrontal and temporal cortex — tissue that is not traditionally viewed as having roles in motor processes (see Fig. S2). Finally, positive correlations decrease with increasing motion and are seen in only the 2 lowest-motion datasets, making it difficult to establish such correlations as a neurophysiological correlate of motion; one must posit that motion artifact obscures these positive correlations in higher-motion cohorts to explain the data. Future work may reveal that FD:BOLD correlations are reflections of neural activity, but the existing data do not strongly support this account.

In sum, although FD:DVARS, DVARS:DVARS, and FD:BOLD correlations seem like sensible outcome measures, they can be complicated to interpret. More direct ways of identifying and measuring motion artifact, such as those mentioned at the beginning of this section, seem preferable outcome measures of denoising success in the view of the authors.

State versus trait considerations

Two studies of typical subjects who were scanned on multiple days indicate that within-subject motion is correlated across scanning sessions (Van Dijk et al., 2012; Zeng et al., 2014). That is, a subject who moves much one day is likely to move much on another day, and a subject who is still one day is likely to be still on another day. Such trait-like behavior may reflect aspects of physiology (e.g., a person who tends to breathe deeply, physical discomfort in the scanner, etc.) or perhaps cognitive traits (e.g., restlessness, impulsivity, etc.). An important question in interpreting findings related to motion is whether connectivity differences seen in high- and low-motion contexts reflect state effects alone (here, state meaning artifact) or state and trait effects.

In the first study to examine state versus trait motion characteristics of functional connectivity MRI, out of several thousand scans, Zeng and colleagues (Zeng et al., 2014) chose several dozen subjects who were scanned on 2 occasions but whose motion was relatively uncorrelated across scans (i.e., each subject had a low- and a high-motion scan, unlike most other subjects, whose motion correlated highly across scans). Another set of motion-matched scans, drawn from demographically matching subjects, was also created, but each subject only contributed one scan. These two sets of scans thus created a context to dissociate across-subject differences from within-subject differences in high- and low-motion scans. After denoising involving global signal regression and volume censoring, group differences in long-range connectivity were found in the across-subject comparisons, but not in the within-subject comparisons, suggesting trait-like effects beyond the state-like effects of motion artifact. In particular, the authors suggest that reduced long-distance connectivity, notably within the default mode network, is a neurobiological signature of people who tend to move their head (Fig. 8).

These findings are important and deserve replication and elaboration. If a neural trait-like predisposition to move manifests with a similar signature as the state-like effects of movement itself (here, this signature would be decreased long-distance connectivity, measured by long-distance degree), it will become difficult to determine appropriate strategies to measure and remove motion artifact. Three principal consequences will follow. First, it will become difficult to determine when state-like effects have been fully removed, since residual trait-like effects that resemble the state effects should remain in correctly denoised data. Second, it will become problematic to use the elimination of the state-like effects as a measure of denoising success. Third, it will become difficult to say whether residual group or individual differences consistent with motion artifact are not instead manifestations of a motion-predisposing "trait".

For these reasons, the study by Zeng and colleagues merits careful reading, replication, and extension. However, two comments seem appropriate to contextualize the existing findings. The first comment is that, in our own studies, our processing strategies generally eliminate differences across subjects that differ by motion, although we have not

[1] http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/scripts/fsl/standardizeddvars.pdf.

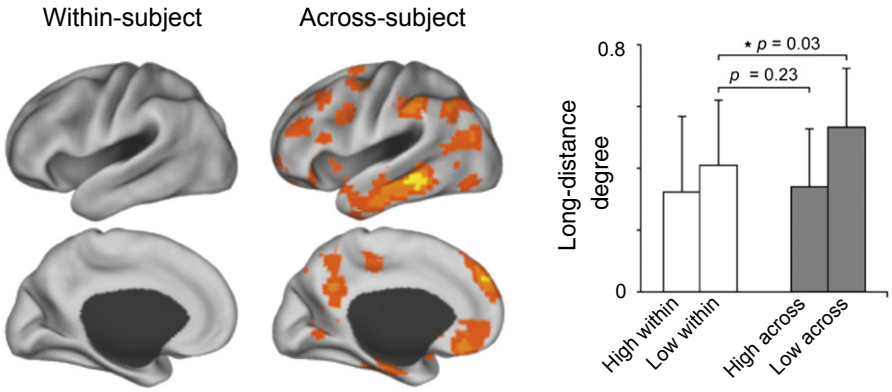

Fig. 8. Motion-related differences within and across subjects. At left, contrast images of long-distance degree. The within-subject contrast is between high- and low-motion scans within the same subjects, the across-subject contrast is between demographically and motion-matched scans that were acquired across subjects. Overall differences are quantified in the bar graph.
Modified from Zeng et al. (2014).

previously investigated long-distance degree. Understanding how the trait-like effect of long-distance degree manifests under various processing strategies will be an important question for future work. The second comment is that a case can be made that across-subject and within-subject contrasts are qualitatively different types of analyses, with different sensitivities to across-scan differences. This type of difference is, however, necessarily embedded in any experimental design aimed at discriminating state from trait characteristics. These caveats aside, the findings of trait-like signatures of motion is provocative and will stimulate much discussion (and work) on how to interpret differences in resting state correlations.

Future directions

It is likely that multiple independent methods will yield adequate control of motion-related variance. At the moment, investigators tend to independently examine and modify their own existing processing strategies. This approach has yielded improved and novel methods from several groups (e.g. (Beall and Lowe, 2014; Kundu et al., 2013; Patel et al., 2014; Power et al., 2014; Satterthwaite et al., 2013a; Scheinost et al., 2014)). In the future, it will be productive to determine whether particular approaches are more efficacious (e.g., if a white matter signal is to be used, is it equally productive to use ANATICOR or aCompCor?) or whether approaches can be productively combined (e.g., combining the iterative blurring method with other methods) to produce superior outcomes. It will also be useful to determine how well the volume-derived motion estimates, which underpin most techniques, correspond to external measures of subject motion, and to slice-specific measures of motion (e.g., those derived from SLOMOCO (Beall and Lowe, 2014)).

The motion-correcting methods being developed in resting state studies have relevance to task fMRI analyses. For example, two studies have investigated the impact of censoring high-motion volumes in task fMRI, both finding that statistical power can be increased by censoring (Christodoulou et al., 2013; Siegel et al., 2014). With the increased recognition of the prevalence and magnitude of motion-related effects in fMRI data, it may become more common to more aggressively denoise task fMRI datasets (e.g., (Kay et al., 2013)). As mentioned in the Introduction, increased scrutiny of motion has already begun to change both the analysis approach and neurobiological interpretation of some structural connectivity studies (Koldewyn et al., 2014; Yendiki et al., 2013). Whether scrutiny of small movements will reveal previously unrecognized artifacts in still other MRI modalities remains to be seen.

One issue that should to be mentioned before closing the paper is the substantial technical and conceptual difficulty of identifying a "ground truth" of what signals ought to be removed. Neural activity drives or reflects all aspects of motor function, including head motion, cardiac rate and contractility, blood flow, and respiration. At some level, complete removal of all signals related to these parameters must entail removal of neural signals (though such signals are usually considered "confounds" distinct from the processes of interest). Currently, we cannot achieve complete removal of all nuisance signals related to a nuisance parameter; we are limited both by our measurements of nuisance parameters and the models of how such parameters relate to fMRI signals. For example, our image-derived motion estimates are known to be imperfect, and regression of various sets of motion parameters has dampened but not eliminated the influence of motion in resting state data (Satterthwaite et al., 2013a; Yan et al., 2013a). Such imperfect motion regressors account for approximately 10–15% of the variance in resting state data (Hutton et al., 2011; Jo et al., 2010). As an additional example, recording and modeling physiological parameters such as heart rate, respiration, and end-tidal pCO_2, though a principled denoising strategy, explains only a small fraction of signal variance. Such physiologic regressors, in combination, tend to account for approximately 6–8% of variance in resting state data (Bianciardi et al., 2009; Jo et al., 2010; Shmueli et al., 2007). To provide a data-driven perspective on these numbers, in ~200-dimensional ICA analyses of Human Connectome Project data, "neural" components account for 4% of variance, "nuisance" components for 16%, motion regressors for 14%, and the rest is labeled unstructured noise (Marcus et al., 2013). There is, in general, a considerable gap between the variance we can currently explain using motion estimates or physiological recordings and the variance that is suspected to be non-neural. An objective for future research is to narrow this gap by improving our nuisance parameter measurements and the models of how such measures relate to nuisance signals. The multi-echo techniques discussed above may prove useful in refining estimates of neural versus non-neural variance.

An additional consideration is that although increasing the number of "confound" parameters during denoising should remove more of uninteresting signals, it also increases the chances of removing signals of interest, since it is always possible (and in some cases certain) that confound parameters will correlate with parameters of interest. For example, investigators may be wary of large numbers of motion regressors or ICA component regressors due to the possibility that they might also remove signals of interest. Denoising can therefore be viewed as a balance between maximizing the removal of uninteresting signal (minimizing false positives) and minimizing the loss of interesting signal (minimizing false negatives). In single-echo resting state data, it is difficult to discriminate between these classes of signals. Two approaches that can help guide discrimination of these signals (and development of denoising techniques) are the multi-echo approaches, which improve

identification of non-BOLD signals (e.g. (Bright and Murphy, 2013; Kundu et al., 2013)), and task fMRI, where an evoked signal of interest should be better isolated by better denoising (e.g., (Christodoulou et al., 2013; Siegel et al., 2014; Wilke, 2012)).

Detecting and reporting motion in a dataset

We close with a few recommendations about information that would be helpful to convey the quality and range in quality of a dataset, as well as the existence of motion artifact before and after processing. A single multi-panel supplemental figure can communicate many important motion-related aspects of the data. These data would be reported for each group in a study (e.g., separately for typical and clinical populations). We emphasize that these are recommendations, not requirements, and they are offered to illustrate the type of information we find helpful when evaluating a study.

1. Report summary statistics for some measure of relative displacement, such as FD. An accompanying boxplot or scatterplot is helpful. Knowing the range of QC values is helpful for interpreting QC-RSFC correlations, as discussed earlier.

2. Whatever the processing strategy employed, it would be useful to present a QC-RSFC plot before and after denoising. Such plots will help accumulate a body of data that may eventually identify sequences or pre-processing approaches that yield reduced motion artifact, and will also communicate the efficacy of denoising in easily understood terms that are comparable across studies. If censoring is performed, because the QC range is typically very compressed after censoring, and because certain processing approaches may leave effects of censored (bad) data in the uncensored (good) data, for QC-RSFC purposes we recommend using the original QC measures derived from all volumes, not QC measures derived only from the post-censoring data.

3. If censoring is performed, include example QC traces of low, median, and high motion subjects so that it can be understood how the thresholds used identify outlying volumes. For reasons discussed earlier it is not especially helpful to simply apply thresholds from the literature. The point of the thresholds is to identify volumes that are not like most other volumes. Most subjects in our experience are usually still with readily identifiable QC excursions during movement. Showing traces allows readers to understand whether only especially bad data were censored, whether most motion-contaminated data were censored, and whether volumes with "floor" values were impacted by the threshold (which can happen when thresholds from the literature are applied to a dataset).

4. If censoring is performed, it is useful to know how many volumes each subject had initially, how many were removed, and what the proportion removed was. Again, summary statistics and scatterplots are helpful. This information communicates whether the post-censoring correlations are expected to be noisy due to aggressive censoring (e.g., if only 2 min of data remain per subject). It is useful to know whether the proportion of data censored covaries with any effects of interest (e.g., age).

Conclusions

Since the first reports in 2011 on motion artifact in resting state functional connectivity, much time and effort have been invested in understanding and removing effects of motion. Considerable progress has been made. Motion artifact is recognized to produce systematic decreases in signal, especially for larger movements, but it can also cause variable disruptions in fMRI signal. The duration of these disruptions can be brief or surprisingly long, impacting correlations for up to 10 s. Motion tends to increase all correlations, but proximal correlations are increased more than distal correlations, and the increases appear largest in correlations with a predominantly lateral orientation. Motion regressors are somewhat, but not adequately, effective in removing motion-related variance, even when voxel-specific or large sets of motion regressors are used. Mean white matter or mean ventricle signals are of modest utility as nuisance regressors. Fractionation of these signals via ANATICOR or aCompCor or other methods may provide additional benefit. An effective processing step is to regress the global signal. Many groups have found censoring to be helpful in reducing motion artifact. Newer methods, such as the multi-echo sequences, appear to hold promise for motion-related denoising. In sum, methods exist that can rid extant datasets of much motion artifact, and future datasets may benefit from new methods of motion denoising by dissociating BOLD effects from non-BOLD effects.

Acknowledgments

The authors thank Tim Laumann and Caterina Gratton for comments on the manuscript. We also thank our reviewers for suggestions that improved the manuscript. This work was supported by NIHF30 MH940322 (JDP), NIHNS046424 (SEP), a McDonnell Collaborative Activity Award (SEP), and by the Intellectual and Developmental Disabilities Research Center at Washington University (NIH/NICHDP30 HD062171).

Conflict of interest statement

The authors have no conflicts of interest to report.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.10.044.

References

Andrews-Hanna, J.R., Snyder, A.Z., Vincent, J.L., Lustig, C., Head, D., Raichle, M.E., Buckner, R.L., 2007. Disruption of large-scale brain systems in advanced aging. Neuron 56, 924–935.

Barch, D.M., Sabb, F.W., Carter, C.S., Braver, T.S., Noll, D.C., Cohen, J.D., 1999. Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. NeuroImage 10, 642–657.

Beall, E.B., Lowe, M.J., 2014. SimPACE: generating simulated motion corrupted BOLD data with synthetic-navigated acquisition for the development and evaluation of SLOMOCO: a new, highly effective slicewise motion correction. NeuroImage 101, 21–34.

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage 37, 90–101.

Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S.G., de Zwart, J.A., Shmueli, K., Duyn, J.H., 2009. Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: a 7 T study. Magn. Reson. Imaging 27, 1019–1029.

Bright, M.G., Murphy, K., 2013. Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. NeuroImage 64, 526–537.

Bullmore, E.T., Brammer, M.J., Rabe-Hesketh, S., Curtis, V.A., Morris, R.G., Williams, S.C., Sharma, T., McGuire, P.K., 1999. Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. Hum. Brain Mapp. 7, 38–48.

Carp, J., 2013. Optimizing the order of operations for movement scrubbing: comment on Power et al. NeuroImage 76, 436–438.

Christodoulou, A.G., Bauer, T.E., Kiehl, K.A., Feldstein Ewing, S.W., Bryan, A.D., Calhoun, V.D., 2013. A quality control method for detecting and suppressing uncorrected residual motion in fMRI studies. Magn. Reson. Imaging 31, 707–717.

Courchesne, E., Pierce, K., 2005. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. Curr. Opin. Neurobiol. 15, 225–230.

Fair, D.A., Dosenbach, N.U.F., Church, J.A., Cohen, A.L., Brahmbhatt, S., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., Schlaggar, B.L., 2007. Development of distinct control networks through segregation and integration. Proc. Natl. Acad. Sci. U. S. A. 104, 13507–13512.

Fair, D.A., Nigg, J.T., Iyer, S., Bathula, D., Mills, K.L., Dosenbach, N.U., Schlaggar, B.L., Mennes, M., Gutman, D., Bangaru, S., Buitelaar, J.K., Dickstein, D.P., Di Martino, A., Kennedy, D.N., Kelly, C., Luna, B., Schweitzer, J.B., Velanova, K., Wang, Y.F., Mostofsky, S., Castellanos, F.X., Milham, M.P., 2012. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. Front. Syst. Neurosci. 6, 80.

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magn. Reson. Med. 35, 346–355.

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C.F., Auerbach, E.J., Douaud, G., Sexton, C.E., Zsoldos, E., Ebmeier, K.P., Filippini, N., Mackay, C.E., Moeller, S., Xu, J., Yacoub, E., Baselli, G., Ugurbil, K., Miller, K.L., Smith, S.M., 2014. ICA-based artefact removal and

accelerated fMRI acquisition for improved resting state network imaging. NeuroImage 95, 232–247.
Hajnal, J.V., Myers, R., Oatridge, A., Schwieso, J.E., Young, I.R., Bydder, G.M., 1994. Artifacts due to stimulus correlated motion in functional imaging of the brain. Magn. Reson. Med. 31, 283–291.
Hallquist, M.N., Hwang, K., Luna, B., 2013. 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.
Hutton, C., Josephs, O., Stadler, J., Featherstone, E., Reid, A., Speck, O., Bernarding, J., Weiskopf, N., 2011. The impact of physiological noise correction on fMRI at 7 T. NeuroImage 57, 101–112.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.
Jo, H.J., Saad, Z.S., Simmons, W.K., Milbury, L.A., Cox, R.W., 2010. Mapping sources of correlation in resting state FMRI, with artifact detection and removal. NeuroImage 52, 571–582.
Jo, H.J., Gotts, S.J., Reynolds, R.C., Bandettini, P.A., Martin, A., Cox, R.W., Saad, Z.S., 2013. Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state FMRI. J. Appl. Math. 2013.
Kay, K.N., Rokem, A., Winawer, J., Dougherty, R.F., Wandell, B.A., 2013. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Front. Neurosci. 7, 247.
Kennedy, D.P., Courchesne, E., 2008. The intrinsic functional organization of the brain is altered in autism. NeuroImage 39, 1877–1885.
Kim, J., Van Dijk, K.R.A., Libby, A., Napadow, V., 2014. Frequency-dependent relationship between resting-state functional magnetic resonance imaging signal power and head motion is localized within distributed association networks. Brain Connect. 4, 30–39.
Koldewyn, K., Yendiki, A., Weigelt, S., Gweon, H., Julian, J., Richardson, H., Malloy, C., Saxe, R., Fischl, B., Kanwisher, N., 2014. Differences in the right inferior longitudinal fasciculus but no general disruption of white matter tracts in children with autism spectrum disorder. Proc. Natl. Acad. Sci. U. S. A. 111, 1981–1986.
Kundu, P., Inati, S.J., Evans, J.W., Luh, W.-M., Bandettini, P.A., 2012. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. NeuroImage 60, 1759–1770.
Kundu, P., Brenowitz, N.D., Voon, V., Worbe, Y., Vértes, P.E., Inati, S.J., Saad, Z.S., Bandettini, P.A., Bullmore, E.T., 2013. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. Proc. Natl. Acad. Sci. U. S. A. 110, 16187–16192.
Lemieux, L., Salek-Haddadi, A., Lund, T.E., Laufs, H., Carmichael, D., 2007. Modelling large motion events in fMRI studies of patients with epilepsy. Magn. Reson. Imaging 25, 894–901.
Marcus, D.S., Harms, M.P., Snyder, A.Z., Jenkinson, M., Wilson, J.A., Glasser, M.F., Barch, D.M., Archie, K.A., Burgess, G.C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., Coalson, T., Schindler, J., Elam, J.S., Curtiss, S.W., Van Essen, D.C., 2013. Human Connectome Project informatics: quality control, database services, and data visualization. NeuroImage 80, 202–219.
Mowinckel, A.M., Espeseth, T., Westlye, L.T., 2012. Network-specific effects of age and in-scanner subject motion: a resting-state fMRI study of 238 healthy adults. NeuroImage 63, 1364–1373.
Murphy, K., Birn, R.M., Handwerker, D.A., Jones, T.B., Bandettini, P.A., 2009. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? NeuroImage 44, 893–905.
Muschelli, J., Nebel, M.B., Caffo, B.S., Barber, A.D., Pekar, J.J., Mostofsky, S.H., 2014. Reduction of motion-related artifacts in resting state fMRI using aCompCor. NeuroImage 96C, 22–35.
Patel, A.X., Kundu, P., Rubinov, M., Jones, P.S., Vertes, P.E., Ersche, K.D., Suckling, J., Bullmore, E.T., 2014. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. NeuroImage 95, 287–304.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59, 2142–2154.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2013. Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. NeuroImage 76, 439–441.
Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage 84, 320–341.
Pujol, J., Macià, D., Blanco-Hinojo, L., Martínez-Vilavella, G., Sunyer, J., de la Torre, R., Caixàs, A., Martín-Santos, R., Deus, J., Harrison, B.J., 2014. Does motion-related brain functional connectivity reflect both artifacts and genuine neural activity? NeuroImage 101, 87–95.
Saad, Z.S., Gotts, S.J., Murphy, K., Chen, G., Jo, H.J., Martin, A., Cox, R.W., 2012. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain Connect. 2, 25–32.
Saad, Z.S., Reynolds, R.C., Jo, H.J., Gotts, S.J., Chen, G., Martin, A., Cox, R.W., 2013. Correcting brain-wide correlation differences in resting-state FMRI. Brain Connect. 3, 339–352.
Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. NeuroImage 90, 449–468.
Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. NeuroImage 60, 623–632.
Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., Wolf, D.H., 2013a. 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.
Satterthwaite, T.D., Wolf, D.H., Ruparel, K., Erus, G., Elliott, M.A., Eickhoff, S.B., Gennatas, E.D., Jackson, C., Prabhakaran, K., Smith, A., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R.E., Gur, R.C., 2013b. Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. NeuroImage 83, 45–57.
Scheinost, D., Papademetris, X., Constable, R.T., 2014. The impact of image smoothness on intrinsic functional connectivity and head motion confounds. NeuroImage 95, 13–21.
Shmueli, K., van Gelderen, P., de Zwart, J.A., Horovitz, S.G., Fukunaga, M., Jansma, J.M., Duyn, J.H., 2007. Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal. NeuroImage 38, 306–320.
Siegel, J.S., Power, J.D., Dubis, J.W., Vogel, A.C., Church, J.A., Schlaggar, B.L., Petersen, S.E., 2014. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Hum. Brain Mapp. 35, 1981–1996.
Smyser, C.D., Inder, T.E., Shimony, J.S., Hill, J.E., Degnan, A.J., Snyder, A.Z., Neil, J.J., 2010. Longitudinal analysis of neural network development in preterm infants. Cereb. Cortex 20, 2852–2862.
Tyszka, J.M., Kennedy, D.P., Paul, L.K., Adolphs, R., 2014. Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. Cereb. Cortex 24, 1894–1905.
Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L., 2010. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J. Neurophysiol. 103, 297–321.
Van Dijk, K.R., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. NeuroImage 59, 431–438.
Wilke, M., 2012. An alternative approach towards assessing and accounting for individual motion in fMRI timeseries. NeuroImage 59, 2062–2072.
Yan, C.G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R.C., Di Martino, A., Li, Q., Zuo, X.N., Castellanos, F.X., Milham, M.P., 2013a. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. NeuroImage 76, 183–201.
Yan, C.G., Craddock, R.C., He, Y., Milham, M.P., 2013b. Addressing head motion dependencies for small-world topologies in functional connectomics. Front. Hum. Neurosci. 7, 910.
Yan, C.G., Craddock, R.C., Zuo, X.N., Zang, Y.F., Milham, M.P., 2013c. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. NeuroImage 80, 246–262.
Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., Fischl, B., 2013. Spurious group differences due to head motion in a diffusion MRI study. NeuroImage 88C, 79–90.
Zeng, L.L., Wang, D., Fox, M.D., Sabuncu, M., Hu, D., Ge, M., Buckner, R.L., Liu, H., 2014. Neurobiological basis of head motion in brain imaging. Proc. Natl. Acad. Sci. U. S. A. 111, 6058–6062.