NeuroImage 112 (2015) 278–287

Contents lists available at ScienceDirect

NeuroImage

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

Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI

Raimon H.R. Pruim [a,b,*], Maarten Mennes [a,b], Jan K. Buitelaar [a,b,c], Christian F. Beckmann [a,b,d]

[a] Radboudumc, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, The Netherlands
[b] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands
[c] Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands
[d] Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Accepted 25 February 2015
Available online 11 March 2015

Keywords:
Motion
Artifact
Independent component analysis
Functional MRI
Resting state
Connectivity

ABSTRACT

We proposed ICA-AROMA as a strategy for the removal of motion-related artifacts from fMRI data (Pruim et al., 2015). ICA-AROMA automatically identifies and subsequently removes data-driven derived components that represent motion-related artifacts. Here we present an extensive evaluation of ICA-AROMA by comparing our strategy to a range of alternative strategies for motion-related artifact removal: (i) no secondary motion correction, (ii) extensive nuisance regression utilizing 6 or (iii) 24 realignment parameters, (iv) spike regression (Satterthwaite et al., 2013a), (v) motion scrubbing (Power et al., 2012), (vi) aCompCor (Behzadi et al., 2007; Muschelli et al., 2014), (vii) SOCK (Bhaganagarapu et al., 2013), and (viii) ICA-FIX (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014), without re-training the classifier. Using three different functional connectivity analysis approaches and four different multi-subject resting-state fMRI datasets, we assessed all strategies regarding their potential to remove motion artifacts, ability to preserve signal of interest, and induced loss in temporal degrees of freedom (tDoF). Results demonstrated that ICA-AROMA, spike regression, scrubbing, and ICA-FIX similarly minimized the impact of motion on functional connectivity metrics. However, both ICA-AROMA and ICA-FIX resulted in significantly improved resting-state network reproducibility and decreased loss in tDoF compared to spike regression and scrubbing. In comparison to ICA-FIX, ICA-AROMA yielded improved preservation of signal of interest across all datasets. These results demonstrate that ICA-AROMA is an effective strategy for removing motion-related artifacts from rfMRI data. Our robust and generalizable strategy avoids the need for censoring fMRI data and reduces motion-induced signal variations in fMRI data, while preserving signal of interest and increasing the reproducibility of functional connectivity metrics. In addition, ICA-AROMA preserves the temporal non-artifactual time-series characteristics and limits the loss in tDoF, thereby increasing statistical power at both the subject- and the between-subject analysis level.

Introduction

Participant head motion during fMRI scanning can induce spurious temporal correlations between brain regions (Van Dijk et al., 2012; Power et al., 2012). Functional connectivity (FC) measures from resting state fMRI (rfMRI) data are especially vulnerable to the influence of such spurious correlations as signal of interest is related to the degree of temporal correlation between multiple voxel time series, and does not relate individual voxel time series to externally defined regressors. Motion-related fMRI artifacts can impact a wide range of FC measures (Satterthwaite et al., 2012; Yan et al., 2013) and can induce increases in local functional connectivity and decreases in long-distance functional connectivity (Van Dijk et al., 2012; Power et al., 2012; Satterthwaite et al., 2012). This effect is not due to motion per se, but an interaction between motion and the use of global signal regression (GSR) and equivalent regressors (Jo et al., 2013; Satterthwaite et al., 2013a). However, such spurious effects directly interfere with current hypotheses about local and long-distance functional connectivity in the context of neurodevelopment and Autism Spectrum Disorders (ASD; Courchesne and Pierce (2005), Fair et al. (2007, 2008, 2009), Kelly et al. (2009), Dosenbach et al. (2010), Power et al. (2010)). Structural between-subject differences in gross motion furthermore interfere with brain research into disorders which are associated with high motion (e.g. attention-deficit/hyperactivity disorder; ADHD). Moreover, by reducing sensitivity towards effects of interest, the presence of motion-related artifacts might not only increase the probability of false-positive but also of false-negative results.

To minimize the impact of head motion in (r-)fMRI analyses, various strategies have been proposed that complement initial volume-realignment by addressing removal of residual motion-induced signal variance. The most commonly used strategy for removal of motion-

* Corresponding author at: Kapittelweg 29, 6525 EN Nijmegen, The Netherlands.
E-mail address: r.pruim@fcdonders.ru.nl (R.H.R. Pruim).

http://dx.doi.org/10.1016/j.neuroimage.2015.02.063
1053-8119/

related artifacts from fMRI data is extensive nuisance regression (Friston et al., 1996; Satterthwaite et al., 2013a; Yan et al., 2013). Here, a set of parameters derived by the volume-realignment algorithm is exploited to model and regress out signal variance associated with motion. More recently, it was proposed to specifically target motion-affected volumes, either by regressing them out (spike regression; Lemieux et al. (2007), Satterthwaite et al. (2013a)) or by fully removing them from the fMRI data (scrubbing; Power et al. (2012)). While both spike regression and scrubbing remove motion-induced signal variations they have a significant impact on the temporal autocorrelation structure of fMRI data and can lead to a high and variable loss in temporal degrees of freedom (tDoF; Satterthwaite et al. (2013a), Yan et al. (2013), Power et al. (2014), Pruim et al. (2015)). Such variable loss in tDoF can induce between-group biases in cases where there is a between-group bias in gross head motion.

Alternative strategies for denoising fMRI data exploit Independent Component Analysis (ICA; Thomas et al. (2002), Kochiyama et al. (2005), De Martino et al. (2007), Perlbarg et al. (2007), Tohka et al. (2008), Kundu et al. (2012), Bhaganagarapu et al. (2013), Rummel et al. (2013), Storti et al. (2013), Salimi-Khorshidi et al. (2014)).ICA is a data-driven approach that decomposes fMRI data into signal (e.g. functional networks) and structured noise (e.g. motion-induced variance and cardiac pulsation; Beckmann et al. (2005)) components. Most ICA-based noise-removal strategies implement automatic classification of components representing noise. A recent example is 'FMRIB's ICA-based X-noiseifier' (ICA-FIX) that implements noise component classification using an extensive set of features and a multi-level classifier (Salimi-Khorshidi et al., 2014). Yet, the complexity of such classifiers hampers their generalizability across datasets. Accordingly, these classifiers typically require re-training for every new dataset (for an exception see Bhaganagarapu et al. (2013)). However, re-training is not trivial and entails manual component labeling in data of multiple participants that subsequently have to be excluded from further analyses.

To overcome drawbacks associated with current strategies for removal of secondary motion artifacts, we proposed an alternative strategy called 'ICA-based Automatic Removal Of Motion Artifacts', or ICA-AROMA (Pruim et al., 2015). ICA-AROMA identifies components representing motion-related artifacts by employing four theoretically motivated features embedded in a simple and robust classifier that avoids the need for classifier (re-)training across studies. The incorporated features evaluate the spatial structure of the component spatial maps with respect to overlaps with the edge of the brain and CSF within the brain. In addition, two temporal features evaluate the component time-course with respect to its tendency towards high-frequencies and its correlation with realignment parameters (RPs). Subsequent to their identification, components classified as motion-related are removed from the fMRI dataset by means of linear regression. Importantly, ICA-AROMA does not result in the removal of volumes but preserves the integrity of the fMRI time-courses. Accordingly, ICA-AROMA aims at preserving tDoF, increasing statistical power for any down-stream between-subject analysis.

Here, we provide an in-depth evaluation of ICA-AROMA using rfMRI data while comparing our strategy to alternative motion artifact removal strategies. Previous evaluations of strategies for motion artifact removal focused on their potential to remove motion artifacts (Power et al., 2012, 2014; Satterthwaite et al., 2013a; Yan et al., 2013), while the importance of retaining signal of interest is often disregarded. Importantly, we not only evaluate the techniques with respect to the ability to remove unwanted motion-induced signals of no interest in the data but also characterize the ability to preserve the identification of resting-state networks as signals of interest. Here, we operationalized the latter aspect by evaluating the ability to clearly identify resting-state networks, as well as by assessing the reproducibility of resting-state networks across random splits of the included data. Moreover, we evaluated the loss in tDoF induced by the different strategies. An additional concern is that currently available evaluations of motion artifact removal strategies focus on healthy controls and lack replication in clinical samples. Yet, clinical samples are often most affected by motion-related artifacts, potentially biasing group comparisons as denoising strategies might work differently across sub-groups, thereby introducing biases in the cross-group comparison. To mitigate these concerns, we assessed motion artifact removal, preservation of signal of interest and lost tDoF in rfMRI-based functional connectivity analyses across four datasets including one comprising participants with ADHD.

Materials & methods

To evaluate ICA-AROMA we compared our strategy to a range of alternative strategies that aim to remove motion-related artifacts from rfMRI data. All strategies were applied after primary motion correction by means of volume-realignment (see fMRI data preprocessing below). We compared the following nine strategies:

1. 'no MC': no secondary motion correction
2. '6RP regression': extensive nuisance regression, including 6 motion regressors
3. '24RP regression': extensive nuisance regression, including 24 motion regressors
4. 'spike regression': 24RP regression including additional spike regressors (Lemieux et al., 2007; Satterthwaite et al., 2013a)
5. 'scrubbing': 24RP regression followed by scrubbing (Power et al., 2012)
6. 'aCompCor': 24RP regression additionally including a set of principle components derived from WM and CSF signals (Behzadi et al., 2007)
7. 'SOCK': Spatially Organized Component Klassifikator (Bhaganagarapu et al., 2013)
8. 'ICA-FIX': FMRIB's ICA-based X-noiseifier (Salimi-Khorshidi et al., 2014)
9. 'ICA-AROMA': ICA-based Automatic Removal Of Motion Artifacts (Pruim et al., 2015).

Datasets

We conducted our statistical comparison in four different datasets. Table 1 provides an overview of the participant characteristics for each dataset. The first and second datasets are derived from the

Table 1
Participant characteristics of the four datasets included in the study. Each divided into three motion subsamples: low, medium and high motion. RMS-FD: root mean square of the frame-wise displacement time-series (Jenkinson et al., 2002).

	No. of participants	Age (years; mean ± SD)	Male (%)	RMS-FD (mm; mean ± SD)
NI Controls				
Total	100	16.9 ± 2.9	46	0.118 ± 0.090
Low motion	25	18.3 ± 2.0	28	0.046 ± 0.006
Medium motion	50	17.1 ± 2.7	46	0.090 ± 0.027
High motion	25	15.2 ± 3.3	64	0.245 ± 0.088
NI ADHD				
Total	100	17.4 ± 3.3	73	0.118 ± 0.090
Low motion	25	18.3 ± 3.3	76	0.047 ± 0.006
Medium motion	50	17.3 ± 3.5	66	0.090 ± 0.027
High motion	25	16.7 ± 2.9	84	0.245 ± 0.089
Power				
Total	69	15.3 ± 6.2	49	0.146 ± 0.090
Low motion	18	15.0 ± 5.8	50	0.067 ± 0.017
Medium motion	33	16.8 ± 6.3	42	0.119 ± 0.010
High motion	18	12.7 ± 5.7	61	0.276 ± 0.086
NYU				
Total	92	12.4 ± 3.1	51	0.077 ± 0.036
Low motion	23	13.8 ± 2.9	35	0.045 ± 0.005
Medium motion	46	12.4 ± 3.0	59	0.067 ± 0.010
High motion	23	10.9 ± 2.8	52	0.127 ± 0.034

NeuroIMAGE project, a large project aimed at studying attention-deficit/hyperactivity disorder (ADHD) in adolescence (von Rhein et al., 2015). As a third and fourth evaluation we respectively included the Power 2012 dataset and the healthy controls from the NYU Child Study Center ADHD200 dataset, both publically available through the International Neuroimaging Data-sharing Initiative (INDI; http://fcon_1000.projects.nitrc.org/).

Dataset 1 — 'NI Controls': 100 healthy controls selected from the NeuroIMAGE project. Participants were excluded if their data contained less than 125 volumes after scrubbing, as proposed by Power et al. (2012), or if they belonged to the 5% highest movers as determined by a motion summary score. As a motion summary score we defined the root mean square of the frame-wise displacement time-series (RMS-FD) derived from the realignment parameters (Jenkinson et al., 2002). Of note, the 30 randomly selected controls used for the development of ICA-AROMA (Pruim et al., 2015) were not included in the current selection.

Dataset 2 — 'NI ADHD': To replicate the findings in a clinical sample we also selected 100 ADHD participants from the NeuroIMAGE project. The participants were pair-wise matched to the participants included in the NI Controls dataset, using the RMS-FD motion summary scores. As for the NI Controls, participants were not selected if their data contained less than 125 volumes after scrubbing.

Dataset 3 — 'Power': This dataset was used for the development of the motion scrubbing strategy (Power et al., 2012). Most participants were scanned in multiple short runs. We conducted preprocessing and first-level analyses separately for every run, and combined results for every participant at the second-level. Runs within the 5% highest RMS-FD and/or with less than 50 volumes left after scrubbing were excluded. In addition, we excluded participants with less than 125 volumes left after scrubbing combined across runs. These criteria excluded eight participants, leaving 69 participants for further analyses.

Dataset 4 — 'NYU': The NYU Child Study Center data is part of the ADHD200 sample in INDI (The ADHD-200 Consortium, 2012). We only included healthy controls from this dataset. The NYU data included shorter rfMRI scanning sessions (175 or 176 volumes) and might inherently be less affected by motion-induced artifacts. We included this dataset as motion-correction strategies have to generalize across datasets with varying imaging parameters and thus potentially varying levels of motion artifacts. Participants within the 5% highest RMS-FD or with less than 125 volumes left after scrubbing were excluded. As a result 92 participants (out of 110) were selected for further analyses.

MRI acquisition

Table 2 provides an overview of the MRI acquisition parameters for each dataset, highlighting variability in TR, TE, voxel size, type of MR-sequence, slice acquisition order, and field strength. For more specific information on the scanning parameters of the NeuroIMAGE datasets we refer to von Rhein et al. (2015). Details on the Power and NYU MR acquisition parameters can be found on the INDI website (http://fcon_1000.projects.nitrc.org). Importantly, the NeuroIMAGE data was collected at two different scanner locations and the Power dataset consisted of multiple cohorts acquired with varying MRI protocols.

fMRI data preprocessing

Preprocessing of the rfMRI data was performed using the FMRIB Software Library (FSLv5.0, available at http://www.fmrib.ox.ac.uk/fsl; Smith et al. (2004), Woolrich et al. (2009), Jenkinson et al. (2012)). Standard preprocessing steps included (1) removal of the first five volumes to allow for signal equilibration, (2) head movement correction by volume-realignment to the middle volume using MCFLIRT, (3) global 4D mean intensity normalization, (4) spatial smoothing (6 mm FWHM), and (5) high-pass filtering (cut-off frequency of 0.01 Hz). Additional preprocessing differed across the applied motion artifact removal strategies and will be discussed below. We co-registered rfMRI datasets to the participant's structural image using affine boundary-based registration as implemented in FSL FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002; Greve and Fischl, 2009) and subsequently transformed them to MNI152 standard space with 4 mm isotropic resolution using non-linear registration through FSL FNIRT (Andersson et al., 2007).

In addition to these initial preprocessing steps, each motion artifact removal strategy required subsequent processing. Supplementary Fig. 1 specifies the preprocessing pipelines for each of the nine strategies. In the *no-MC* strategy we applied nuisance regression using three nuisance regressors: white matter (WM) signal, CSF signal and a linear trend. White matter and CSF time-series were derived by determining the mean time-series over voxels within a conservative predefined WM and CSF mask (see Supplementary Fig. 1). Importantly, nuisance regression was applied prior to high-pass filtering. In the *6RP regression* strategy we additionally included six rigid body parameters, resulting from the volume-realignment, in the nuisance regression model. The *24RP regression* strategy included a nuisance regression model including the six rigid body parameters, their derivatives, the squared version of the rigid body parameters, the squared derivatives, as well as the three standard nuisance regressors.

Scrubbing and *spike regression* require labeling fMRI volumes associated with high motion. To that end we used the frame-wise displacement time-series (FD; Jenkinson et al. (2002)) and labeled volumes with FD > 0.2 mm, as well as its preceding and the following volume. A single spike regressor was constructed for each of these frames and added to the 24RP regression model in the case of spike regression. In the case of scrubbing we deleted these frames from the residual fMRI data after 24RP regression and high-pass filtering.

aCompCor (Behzadi et al., 2007) is a nuisance regression strategy where realignment parameters are complemented by principle components derived from WM and CSF signals. As proposed by Muschelli et al. (2014), we constructed a nuisance model comprising the 24 realignment parameters complemented by the principle components that explained 50% of respectively the WM and CSF signals. We applied aCompCor before spatial smoothing and after temporal detrending.

We applied *ICA-AROMA* after spatial smoothing, and complemented the preprocessing with nuisance regression using WM, CSF, and a linear trend to regress out residual structured noise. After nuisance regression

Table 2
Resting-state fMRI acquisition characteristics of the datasets included in the study. NeuroIMAGE included both NI Controls and NI ADHD.

	No. of volumes per run	No. of runs	TR (s)	TE (ms)	Voxel size (mm)	MR-sequence	Slice acquisition	B_0-field strength
NeuroIMAGE	265/266	1	1.96	40	3.5 × 3.5 × 3.0	GE-EPI	Ascending	1.5 T
Power	76–164	1–6	2.0–2.5	27	4.0 × 4.0 × (4.0–6.8)	GE-EPI	Interleaved	3 T
NYU	175–176	1	2.0	15	3.0 × 3.0 × 4.0	ME-EPI	Interleaved	3 T

GE: gradient-echo and ME: multi-echo.

we applied high-pass filtering (Pruim et al., 2015). In contrast, *ICA-FIX* was applied after high-pass filtering and without subsequent nuisance regression (Salimi-Khorshidi et al., 2014). No preprocessing streamline was described for *SOCK* (Bhaganagarapu et al., 2013). SOCK implements a generic ICA-based classifier aimed at removing all types of structured noise, and additional nuisance regression was not specifically recommended. Therefore we implemented SOCK analogous to ICA-FIX.

Note that for ICA-FIX it is recommended to re-train the classifier when applying it to a new dataset (Salimi-Khorshidi et al., 2014). Re-training presents a formidable challenge, particularly in the case of multi-site studies as either one would need to define a balanced single training dataset (incorporating data and therefore noise features from all sites) or implement multiple training runs (one per site), thereby potentially inducing 'denoising bias' in any subsequent comparison. Since our aim is to evaluate ICA-AROMA, which was specifically designed to avoid re-training, we did not re-train ICA-FIX but applied the ICA-FIX classifier as trained for application to a standard fMRI dataset (Salimi-Khorshidi et al., 2014).

To ensure that the results of the different preprocessing strategies are comparable we embedded them within a single,'minimal' preprocessing steam. As such, we did not include extensions of WM and CSF signals (derivative or squared), GSR and/or band-pass filtering for any of the strategies. Naturally, the order of preprocessing steps could differ across strategies as required (e.g. aCompCor before spatial smoothing). Note that our embedding can differ from the processing pipeline as used in each strategy's respective manuscript. As GSR and band-pass filtering are frequently considered, yet actively debated, within rfMRI research, we added a supplementary evaluation of these processing steps (see Supplementary materials).

Statistical analyses

Fig. 1 outlines the different analyses that we conducted to assess the sensitivity and specificity of the nine motion artifact removal strategies. To that end we assessed the quality of motion artifact removal (sensitivity), and the preservation of signal of interest (specificity). In short, assessing the quality of the motion artifact removal focused on directly comparing participants with low motion to participants exhibiting high levels of movement during scanning. If secondary motion artifact removal strategies work, differences between participants exhibiting low and participants exhibiting high motion should be reduced to a minimum. In complementary analyses, we assessed whether the different strategies preserved signal of interest, i.e. potentially removed signal of interest along with motion-related artifacts. Both analyses were applied to different functional connectivity measures as described below. Additionally, we evaluated the strategies regarding their induced loss in tDoF.

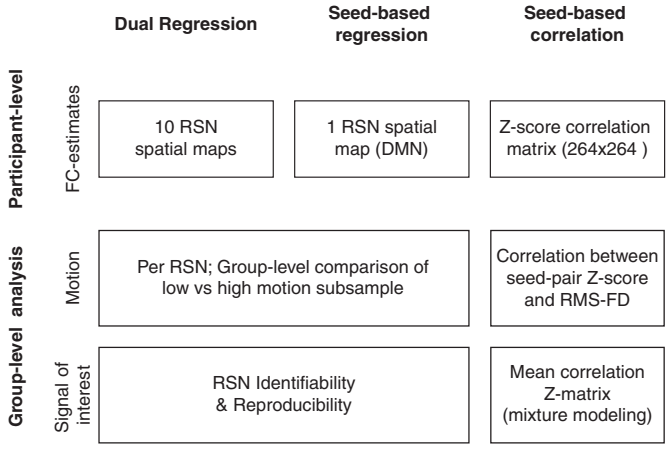

Fig. 1. Schematic overview of the analyses we conducted within our evaluation to assess the quality of motion artifact removal and preservation of signal of interest, using multiple functional connectivity measures.

Functional connectivity measures

Dual regression

For each participant we derived spatial maps of well-validated and commonly replicated resting-state networks (RSNs) using dual regression (Beckmann et al., 2009; Filippini et al., 2009). This analysis implements a multivariate spatial regression of a set of initial templates against the preprocessed fMRI data of every participant, and yields participant-specific time series for each template. Next, these time series are entered in a multivariate temporal regression against the same preprocessed data resulting in participant-level spatial representations, both parameter estimate (PE) and Z-statistical maps, of the initial templates. As templates, we used the 20 spatial maps as described by Smith et al. (2009). These templates include 10 RSNs which were found to correspond well to networks involved in task-related processing. The other 10 templates are artifactual or represent more complex networks. Dual regression was performed using the full set of components to optimally model the data. We selected the 10 RSNs for use in subsequent analyses. These RSNs include the visual-medial (Vis-med), visual-occipital (Vis-occ), visual-lateral (Vis-lat), default mode (DMN), cerebellum, sensorimotor, auditory, executive control, right fronto-parietal (right FP) and left frontal-parietal (left FP) networks.

Dual regression was applied to each preprocessed rfMRI dataset. Of note, the Power dataset contained multiple runs for most participants. We applied dual regression to each single run. Subsequently, we derived participant-specific PE-maps by means of a within-subject fixed effects analysis, and Z-statistical maps by averaging the obtained Z-statistical spatial maps across runs.

Seed-based functional connectivity

We further evaluated the performance of the motion artifact removal strategies in the context of seed-based (SB) functional connectivity analyses. We included two analyses adapted from Power et al. (2012). The first analysis comprised a whole brain seed-based regression using a medial parietal seed (MNI coordinates: $x = -7$ mm, $y = -55$ mm and $z = 27$ mm) to estimate a spatial map for the default mode network (DMN).

In addition we employed a seed-based correlation analysis using 264 ROIs (Dosenbach et al., 2010). The mean time-series of every seed was derived using a 10 mm sphere centered on the seed coordinates. All pair-wise correlation scores were converted to Z-scores by means of a Fisher r-to-Z transformation. In the Power dataset, correlation Z-scores were calculated for each run separately, and then averaged to obtain participant-specific fixed-effects estimates.

Assessing the quality of motion artifact removal

We assessed the ability of each strategy to remove motion artifacts by comparing the RSNs of participants exhibiting low head motion to RSNs of participants exhibiting high amounts of head motion. To that end we divided each dataset into three subsamples: low, medium and high motion subsamples representing respectively 25%, 50% and 25% of each sample. The subdivision was guided by a single summary motion score (RMS-FD; root mean square of the frame-wise displacement time-series) for every participant.

To compare participants exhibiting low vs. high amounts of head motion we employed a between-group comparison within a General Linear Model (GLM), implemented through non-parametric permutation testing (5000 permutations). The three movement groups (low, medium, high) were included as separate group variables. Age and gender were included as covariates. Analyses on the NeuroIMAGE sample additionally included IQ and scan location as covariates. Contrasts of interest were 'low > high' and 'high > low'. Group-level

t-statistical maps were thresholded using threshold-free cluster enhancement and corrected for multiple comparisons using family wise error (FWE) correction with $p < 0.05$. This comparison was made for each RSN separately (10 dual regression-based RSNs and 1 seed-based RSN). As a supplementary analysis, we replicated the group-based motion comparison using a dimensional approach by including RMS-FD as an effect of interest instead of separate group variables.

The effect of motion on the 264 × 264 SB-correlation matrices was assessed qualitatively by calculating the absolute correlation across participants of every seed-pair Z-score (264 × 264) with the motion summary score RMS-FD. Subsequently, we converted these correlation scores to Z-scores by means of a Fisher r-to-Z transformation and assessed the distribution of the resulting values. The goal of the respective strategies is to avoid spurious correlations related to motion artifact. Accordingly, in case the SB-correlations are unrelated to motion, the distribution of their correlations with the RMS-FD score is expected to approximate zero with a small standard deviation. We further evaluated the impact of motion on the Z-score correlation matrices using a categorical and dimensional approach analogous to the analyses on the RSN spatial maps described above with the exception that the t-statistical maps were not thresholded using threshold-free cluster enhancement. Results of these analyses are presented in the Supplementary materials.

Testing the preservation of signal of interest

We assessed to what extent each strategy preserved, or even increased, sensitivity towards signal of interest by evaluating the ability to identify and reproduce the estimated RSNs. RSN identifiability was defined in terms of a Z-score ratio between the mean absolute Z-score inside and outside a RSN-mask. This ratio was calculated for every dual regression-based RSN at the participant-level, using the original templates thresholded at $|Z| > 2.3$ as masks. Similarly, for the seed-based DMN map, we calculated the ratio between the mean absolute Z-score inside and outside the DMN mask derived from the DMN template used in the dual regression analysis. By comparing signal within the RSN spatial map to noise outside the map, this score represents a signal to noise ratio of every participant-level RSN spatial map. As an example, a ratio of 1 would indicate that the RSN is indistinguishable from the surrounding voxels as both would have equal Z-scores. In contrast, a ratio > 1 indicates that the Z-scores within the mask were higher compared to the Z-scores of the surrounding voxels, suggesting increased RSN identifiability.

RSN reproducibility was investigated using split-half reproducibility. We randomly divided the total group of participants per sample into two equally sized groups. For both groups, we derived the average group-level spatial PE map across participants, for each of the 20 template ICs as estimated by dual regression. The group-level maps were masked to only include gray matter voxels using a MNI152 gray matter probability map thresholded at 50%. Subsequently, we determined the between-group spatial correlation for each spatial map, yielding a 20 × 20 correlation matrix. The correlation values on the diagonal of this matrix expressed the similarity (or reproducibility) of each IC between both groups (e.g., spatial correlation between the DMN of group 1 and the DMN of group 2). In contrast, the off diagonal correlations expressed the spatial correlation between non-matching IC templates (e.g., the spatial correlation between the DMN of group 1 and the sensorimotor network of group 2). We used the off diagonal correlations as a null-distribution to convert the spatial correlations for the 10 matching RSNs to pseudo Z-scores. Analogously, we calculated the spatial correlation between the seed-based DMN group-level maps obtained for each group. We used the null distribution as estimated for the 20 dual regression-based maps to convert the between-group seed-based DMN spatial correlation to a pseudo Z-score. To obtain average Z-scores and standard deviations we conducted the reproducibility analyses for 500 random group splits.

In addition to RSN identifiability and reproducibility estimates, we used the 264 × 264 seed-based correlation matrix to assess preservation of the correlation structure in the data. We derived the mean Z-score correlation matrix across participants per dataset and assessed the results qualitatively. In addition, we tested the correlation structure more specifically by implementing a mixture modeling approach to partition the seed-pairs respectively into three categories: anti-correlated, uncorrelated and correlated seed-pairs. Motion artifact removal strategies that effectively remove motion-related noise but preserve signal of interest will amplify differences between the three categories and therefore improve the detection of (anti-)correlated seed-pairs. We therefore compared the mixture modeling results with respect to the results of the no-MC strategy to evaluate changes in sensitivity towards significant correlation induced by the different strategies.

Of note, it should be clear that we evaluate all techniques using both inherently seed-based approaches as well as dual-regression-based characterization of ICA-derived networks, thus avoiding potential bias with respect to the implemented strategies. In addition, despite the fact that the spatial templates used for our evaluation were derived from an ICA decomposition (Smith et al., 2009), our evaluation metrics will not be biased towards 'good' results for ICA-based motion artifact removal strategies. This is because we do not evaluate the spatial quality of individual representations of each template. Instead we focus on the similarity of the obtained representations between datasets (whether they are a good representation of the initial template or not).

Assessing the loss in temporal degrees of freedom

We evaluated the loss in tDoF associated with the different strategies to investigate its potential impact on statistical power and between-group bias related to a variable loss in tDoF. For every participant we determined the number of lost tDoF associated with each of the strategies by regarding every nuisance regressor, component and/or fMRI volume which was regressed out or removed from the data as a single tDoF. To obtain comparable results between participants we expressed the number of lost tDoF as a fraction of the total available tDoF which we defined as the total number of volumes within the fMRI time-series. Note that these common estimates of loss in tDoF and available tDoF ignore temporal autocorrelation present in the data. The actual number of tDoF is therefore likely to be lower than these estimates. Finally, temporal filtering will also result in lost tDoF and should ideally be accounted for. We employed high-pass filtering by conducting a local Gaussian-weighted line-fitting procedure as implemented in fslmaths. However, as temporal filtering is equal for all strategies included in our evaluation we did not incorporate lost tDoF induced by temporal filtering into our tDoF calculations.

Results

Quality of motion artifact removal

As illustrated in Fig. 2, applying no specific correction for secondary motion artifacts (no-MC), yielded prominent differences between the low and high motion subsamples throughout the complete set of RSNs. A large portion of significant effects persisted after preprocessing the data using RP-based regression strategies, aCompCor or SOCK. Spike regression, scrubbing, ICA-FIX and ICA-AROMA on the other hand reduced the differences between participants exhibiting low and high amounts of head motion to a minimum. These results were consistent across datasets except in the NYU dataset where the initial differences between low and high movers were less prominent but clearly more resistant towards removal by the different strategies (see Fig. 2 and Supplementary Fig. 2). Results of the dimensional group-level motion

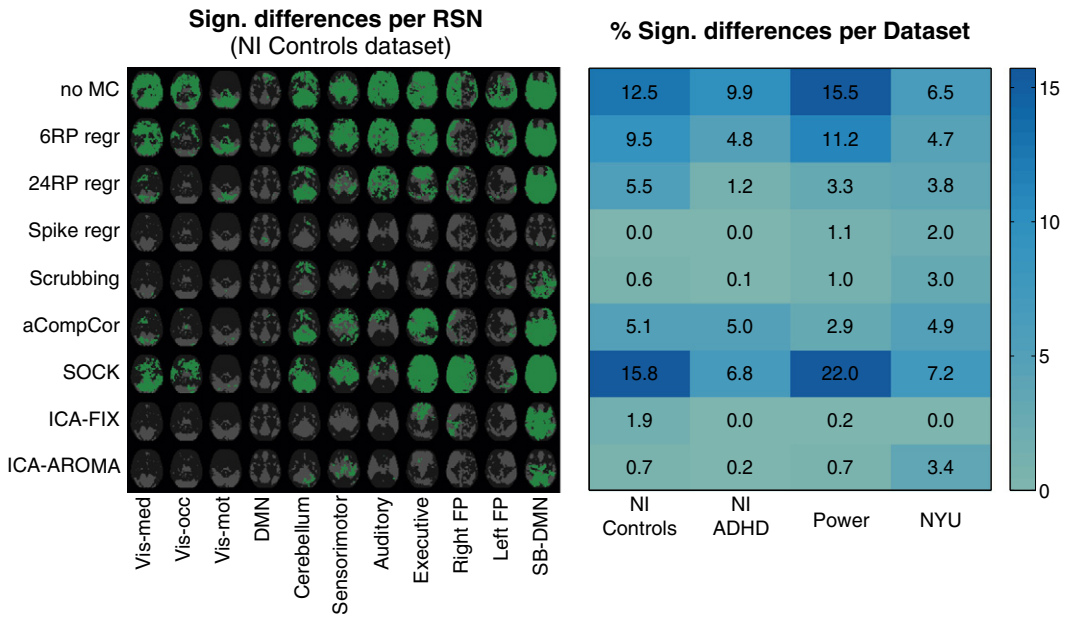

Fig. 2. Significant group-level differences between participants exhibiting low and high head motion (threshold-free cluster enhanced, $p < 0.05$, FWE-corrected). Left panel: Results of the between group-comparison in the NI Controls dataset, for all tested RSN spatial maps and preprocessing strategies. Each single map illustrates a RSN mask (gray) overlapped with a map of the significant between-group differences (green). The original 3D spatial maps are summed and binarized over all axial slices, i.e. a green pixel reflects that at least one voxel along that z-axis showed a significant residual motion-induced effect. See Supplementary Fig. 2 for similar results for all datasets. Right panel: Percentage of voxels exhibiting a significant group-level difference in each of the included datasets. This percentage was calculated for each dataset across all 11 RSNs as shown in the left panel (columns). See Supplementary Fig. 3 for similar results when employing a dimensional instead of categorical analysis, and for replication of these analyses on seed-based Z-score correlation matrices.

analysis were in line with the categorical results presented here (see Supplementary Fig. 3).

The assessment of motion artifact removal by means of evaluating the correlation between RMS-FD and seed-based correlation functional connectivity estimates yielded comparable results. RP-based regression strategies, aCompCor and SOCK resulted in substantial correlation between seed-pair correlation Z-scores and RMS-FD. In contrast, spike regression, scrubbing, ICA-AROMA and to a lesser extent ICA-FIX decreased such correlation towards zero with small standard deviation across seed pairs (Fig. 3, upper triangles and bottom right panel; see Supplementary Fig. 4 for the results on individual datasets). In addition, Supplementary Fig. 3 illustrates the results of a categorical and dimensional analysis of the group-level effects of motion. Results confirmed the effects observed in Fig. 3.

Preservation of signal of interest

RSN identifiability scores were highly consistent across motion artifact removal strategies, datasets and RSNs (Fig. 4; see Supplementary Fig. 5 for the results on individual datasets and RSNs). An important exception was ICA-FIX, which often resulted in decreased identifiability scores, especially in the NYU dataset. An identifiability score of 1

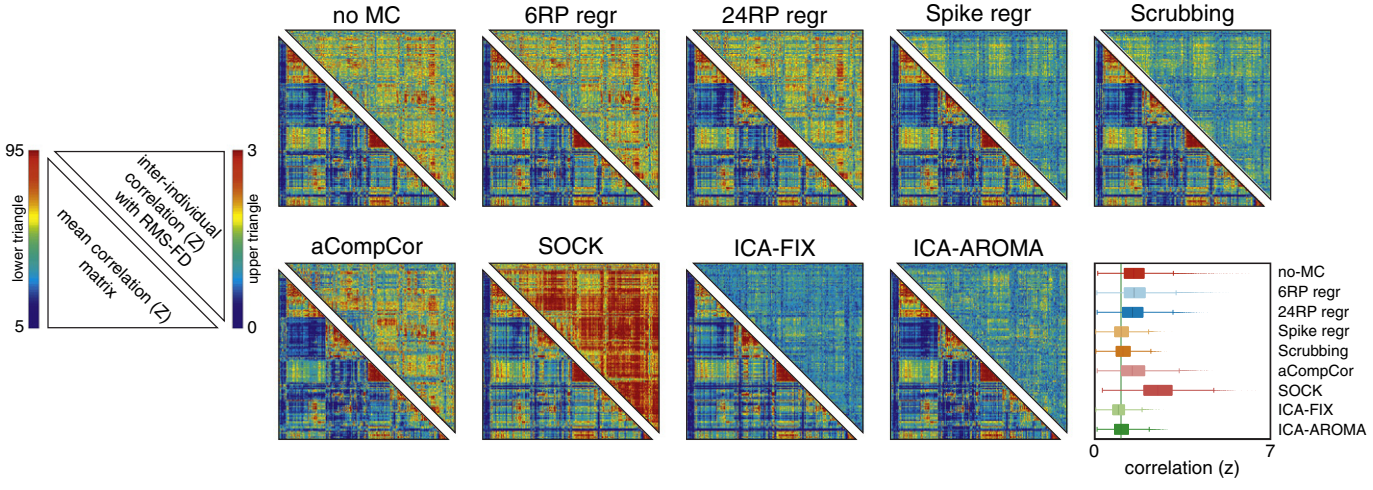

Fig. 3. Motion artifact removal and preservation of signal of interest in seed-based correlation functional connectivity analyses (264 ROIs). The upper triangles illustrate the correlation (Fisher r-to-Z transformed) across participants of seed-pair correlation Z-scores with RMS-FD (mean 264 × 264 matrix across the four datasets). The lower-right panel presents the distributions of these scores. The lower triangles illustrate the mean seed-pair correlation Z-scores over all participants across datasets, with upper and lower limits to the 5th and 95th percentiles of the total range to account for scaling differences between strategies. Strategies that effectively remove motion artifacts while preserving signal of interest are considered to present correlation scores in the upper matrices towards zero with a small standard deviation (i.e. no association between RMS-FD and seed-pair correlation Z-scores), while retaining or accentuating the correlation structure as presented in the lower triangles. Supplementary Fig. 6 and Supplementary Table 1 present the results obtained after conducting mixture modeling on the mean seed-pair Z-score correlation matrices, showing that ICA-AROMA yielded increased and profoundly more consistent detection of correlated seed-pairs compared to all alternative strategies.

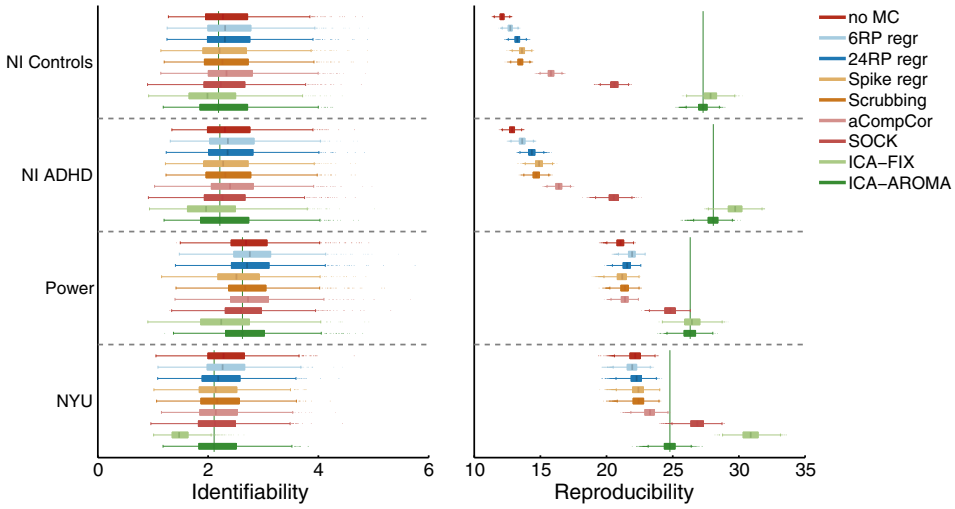

Fig. 4. RSN identifiability and RSN reproducibility. RSN identifiability is quantified using a Z-score ratio expressing the Z-scores within a RSN relative to Z-scores outside the RSN. RSN reproducibility scores reflect the spatial correlation between group-level RSN maps (i.e. consistency) for random splits of the samples, normalized to pseudo Z-scores. Results are shown for all processing strategies in each dataset, combined across all 11 RSNs (10 dual regression + 1 seed-based).

indicates no difference between Z-scores within and Z-scores outside the RSN mask, or, in other words, no contrast between signal and noise. Such decreased ratios suggest substantial removal of signal of interest during the artifact removal procedure. Note, however, that here we evaluate ICA-FIX without explicit re-training of the classifier, i.e. using the default classification criteria derived from a separate rfMRI study (Salimi-Khorshidi et al., 2014).

RSN reproducibility on the other hand, varied across strategies (Fig. 4; see Supplementary Fig. 5 for individual datasets and RSNs). Extensive nuisance regression strategies, scrubbing and spike regression resulted in the lowest reproducibility scores, aCompCor exhibited slightly increased scores, while ICA-based strategies SOCK, ICA-FIX, and ICA-AROMA yielded the highest reproducibility scores. ICA-AROMA exhibited the most consistent pseudo Z-scores across the four datasets (see Supplementary Fig. 5). Of note, it is evident from Fig. 4 that all strategies resulted in high reproducibility Z-scores (i.e., the minimum observed pseudo Z-scores was 10.8). Yet, a clear distinction could be made between the RP-based and ICA-based motion artifact removal strategies.

Finally, all motion artifact removal strategies largely preserved the correlation structure observed between 264 ROIs (Fig. 3, lower triangles). Supplementary Fig. 4 illustrates the correlation structure for each dataset, highlighting that ICA-FIX completely destroyed the correlation structure in the NYU dataset, corroborating the low RSN identifiability observed for ICA-FIX in the NYU dataset and suggesting that dataset-specific classifier training would be necessary for acceptable denoising performance to be achieved. Mixture modeling analyses on the mean seed-based correlation matrices yielded increased and profoundly more consistent detection of correlated seed-pairs after applying ICA-AROMA compared to all alternative strategies (see Supplementary Fig. 6 and Supplementary Table 1). Additionally detected connections were primarily located within clusters already detected when employing the no-MC strategy; therefore these additional connections likely reflect true-positives.

Loss in temporal degrees of freedom

The fractional loss in tDoF strongly differed between the nine strategies (see Fig. 5). No secondary motion correction and 6RP regression consistently resulted in the lowest and least variable loss in tDoF whereas spike regression and scrubbing resulted in the highest and most variable loss in tDoF. For the ICA-based strategies, SOCK preserved

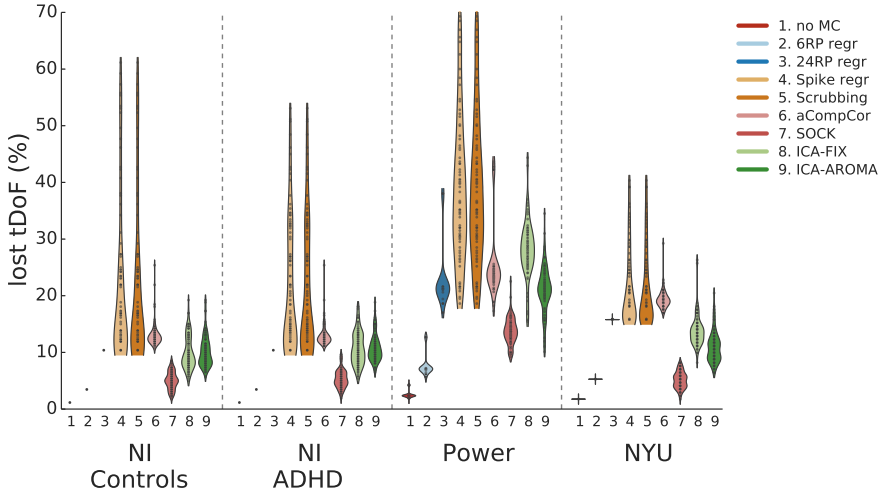

Fig. 5. Lost temporal degrees of freedom (tDoF), as a percentage of the total available tDoF (implemented here as the available number of time points), for all participants within each dataset. Plotted for every motion artifact removal strategy, per dataset.

most tDoF across all datasets whereas ICA-AROMA and ICA-FIX yielded comparable results across samples. However, in the Power and NYU datasets the loss in tDoF associated with ICA-AROMA was significantly lower compared to ICA-FIX (paired t-test, $p < 0.01$). Likewise, the difference between ICA-AROMA and 24RP regression was non-significant for the NI-samples ($p = 0.38$ and $p = 0.12$) and the Power dataset ($p = 0.91$), but significantly lower for the NYU datasets ($p < 0.01$).

Discussion

We evaluated ICA-AROMA and alternative strategies aiming at removal of motion artifacts from rfMRI data. Importantly, we not only focused on the ability of each strategy to remove motion-related artifacts, but additionally evaluated how well each strategy preserved signal of interest. Results were replicated across four datasets, including one clinical sample. ICA-AROMA performed equally well as spike regression, scrubbing and ICA-FIX in removing motion-related artifacts. ICA-AROMA and ICA-FIX both yielded improved RSN reproducibility and decreased loss in tDoF compared to spike regression and scrubbing. However, without re-training the classifier, ICA-FIX decreased the level of signal of interest whereas ICA-AROMA successfully preserved signal of interest across all datasets.

Corroborating previous reports we observed that head motion during rfMRI scanning has a large impact on functional connectivity estimates (Van Dijk et al., 2012; Power et al., 2012; Satterthwaite et al., 2012; Yan et al., 2013). Across RSNs we observed widespread differences between participants exhibiting low and high amounts of head motion.

Nuisance regression including realignment parameters was unsuccessful in addressing such effects of secondary motion artifacts. The 6RP and 24RP regression strategies minimally reduced the impact of motion on RSNs and SB-correlation measures. This finding is consistent with previous results (Van Dijk et al., 2012; Power et al., 2012; Satterthwaite et al., 2012; Yan et al., 2013) and further highlights the limited applicability of realignment parameters to model the full complexity of secondary motion-induced variance (e.g. spin history effects). Extension of the nuisance model by principle components derived from WM and CSF as implemented in aCompCor had no benefit regarding the reduction of group-level motion effect. In contrast, spike regression, scrubbing, ICA-FIX and ICA-AROMA were successful in reducing significant effects of head motion. A notable exception was SOCK, which exhibited limited ability to reduce motion-related group-level differences and generally showed large amounts of variability with respect to its denoising ability.

Additionally we investigated how the different strategies affected group-level variability of RSN spatial maps. To that end we assessed RSN reproducibility, and found that all ICA-based strategies resulted in profoundly increased reproducibility scores relative to the alternative strategies or no additional motion denoising. This suggests that that ICA-based strategies remove structured noise from fMRI data more efficiently, decreasing group-level variability of RSN spatial maps. In contrast, nuisance regression, spike regression and scrubbing exhibited reproducibility scores that were similar to when no additional motion denoising was applied, suggesting that these artifact removal strategies are not only specific to motion artifact, but equally affect signal of interest. This suggestion is corroborated by evidence showing that fMRI data contained substantially higher levels of motion-related and generic noise, relative to signal of interest, after extensive nuisance regression and spike regression compared to ICA-AROMA (Pruim et al., 2015). Notably, whereas aCompCor did not have added value regarding the removal of group-level motion effects it did slightly increase RSN reproducibility compared to the nuisance regression strategies. This result suggests that the set of WM and CSF regressors derived by PCA particularly contributes to the removal of structured noise other than motion artifacts.

ICA-FIX increased RSN reproducibility and has been demonstrated to have the potential for accurate component classification and denoising of fMRI data (Salimi-Khorshidi et al., 2014; Griffanti et al., 2014). However, without re-training, we found that the ICA-FIX classifier was not sufficiently generalizable across datasets and resulted in decreased levels of signal of interest. In contrast, SOCK preserved signal of interest across datasets but at the cost of poor removal of motion-related noise. Such decreased sensitivity for SOCK has been previously documented (Sochat et al., 2014). The variability between ICA-based strategies directly reflects the trade-off that such strategies have to overcome regarding the sensitivity and specificity of the classifier that they implement. By implementing a small set of specific, standardized and theoretically motivated features, ICA-AROMA was designed to allow high sensitivity and specificity while achieving robust classification performance. Indeed, ICA-AROMA was the only ICA-based strategy within our evaluation that retained signal of interest while reducing motion artifacts across datasets.

Moreover, ICA-AROMA showed very consistent results across the four datasets. Most notably, ICA-AROMA resulted in consistent RSN reproducibility scores (see Fig. 4) and detection of functional connections using seed-based analysis (see Supplementary Fig. 6). Such consistency was achieved despite the large variance in MRI acquisition parameters across datasets. This not only demonstrates the applicability of ICA-AROMA to new datasets without re-training, but it additionally indicates its potential to improve the reliability of FC estimates. Functional connectivity metrics derived from rfMRI data have already been shown to be reliable across participants, scan sequences, imaging sites, and time (Van De Ven et al., 2004; Damoiseaux et al., 2006; Shehzad et al., 2009; Biswal et al., 2010; Van Dijk et al., 2010; Zuo et al., 2010; Wisner et al., 2013). However, there is still considerable improvement possible in the reliability of these metrics (Wisner et al., 2013; Zuo and Xing, 2014). Although the impact of ICA-AROMA on within-participant test–retest reliability requires additional investigation, the increased between-participant reproducibility illustrates that ICA-AROMA can possibly facilitate such improved reliability. By improving the consistency of FC estimates over scanner-sites and datasets ICA-AROMA furthermore enhances the potential of multi-site studies and data-sharing initiatives (e.g. FCON1000, ADHD200, ABIDE) which typically suffer from variability across the total sample due to different scanner sites and MRI acquisition protocols. We furthermore showed that ICA-AROMA generalizes to clinical datasets such as those provided in the ADHD200 (http://fcon_1000.projects.nitrc.org/indi/adhd200/) or ABIDE sample (http://fcon_1000.projects.nitrc.org/indi/abide/) by replicating our findings in the NYU dataset and the NI ADHD dataset.

Despite controversy about their usage, global signal regression (GSR) and band-pass filtering are often considered in rfMRI research. While band-pass filtering exhibited some positive effects in reducing significant group-level motion-related effects, we observed no added benefit of GSR in addition to 24RP regression, scrubbing or spike-regression (see Supplementary Fig. 7). Importantly, both GSR and specifically band-pass filtering decreased signal of interest as reflected by decreased RSN identifiability. These findings are not surprising when considering that the global signal is a superposition of both signal and noise components, and that higher frequencies contain signal of interest (Niazy et al., 2011; Liao et al., 2013; Kalcher et al., 2014).

We did not specifically evaluate the usefulness of combining multiple strategies that could potentially complement each other (e.g. 24RP-regression and SOCK) and did not include all currently available strategies for motion artifact removal, e.g. wavelet-despiking (Patel et al., 2014), SLOMOCO (Beall and Lowe, 2014), RDI (Spisák et al., 2014) or alternative ICA-based strategies (Thomas et al., 2002; Kochiyama et al., 2005; De Martino et al., 2007; Perlbarg et al., 2007; Tohka et al., 2008; Kundu et al., 2012; Rummel et al., 2013; Storti et al., 2013; Sochat et al., 2014). In our evaluation we included a set of strategies that varied in their underlying principles (e.g. volume-removal, RP-model regression, component-regression), are

currently most considered within the research field, are applicable across datasets, and are representative for the wide spectrum of available strategies.

Similar to the alternative strategies, ICA-AROMA did not achieve full removal of significant differences between participants exhibiting low versus high head motion (see Fig. 2). The question remains whether these effects reflect motion artifact or neurobiological correlates which are directly (e.g. motor control) or indirectly (e.g. age) related to motion. Several studies suggest that such neurobiological correlates exist (Van Dijk et al., 2012; Couvy-Duchesne et al., 2014; Kong et al., 2014; Pujol et al., 2014; Zeng et al., 2014). In this regard, we note that the residual significant differences found within the resting-state networks appear to be consistent across datasets, comprising the cerebellum, sensorimotor, auditory and executive control networks. The cerebellum and sensorimotor network are involved in motor control, potentially suggesting that these findings might be related to neurobiological underpinnings of head motion. This hypothesis is particularly supported by the analysis of the NYU dataset which almost exclusively resulted in significant group-level effects within the four previously listed networks, and which were more resistant to removal by the different strategies (disregarding the seed-based results). These results are striking since the NYU dataset comprises participants with profoundly lower levels of head motion (see Table 2). However, these participants were scanned at a substantially earlier stage within their neurodevelopmental trajectory (12.4 ± 3.1 years) compared to the participants in the other datasets, providing support for the idea that the observed effects might be related to neurodevelopment, explaining variance not captured by the age covariates. Yet, additional research is required to reliably disentangle potential neurobiological effects from motion-related artifacts. It is for instance important to note that most denoising procedures discussed in the current manuscript implement artifact removal through linear regression whereas head-motion is likely to induce highly non-linear effects. Although such non-linear effects can be approximated by a set of components in data-driven strategies, any residual motion-related group effects can be related to inadequate modeling of the non-linear motion-related effects and hence sub-optimal denoising. Similarly, volume-removal strategies depend on a binary decision on a volume being affected by head motion and inherently will not remove the full (non-linear) dynamics associated with motion artifacts.

Residual motion-related effects as discussed above are of particular importance when considering group-level co-varying using summary motion scores. Though this might prevent false positives, it might also reduce sensitivity towards any effect of interest sharing variance with motion scores, e.g. age, gender, motion-related traits, or neural activity (Van Dijk et al., 2012; Couvy-Duchesne et al., 2014; Kong et al., 2014; Pujol et al., 2014; Zeng et al., 2014). In case residual differences between participants exhibiting low and high head motion are attributable to artifacts, the question remains whether and how such residual noise impacts future analyses. We note that our group comparison compared two highly distinct samples comprising the 25% lowest versus the 25% highest movers, without any overlap in motion summary scores. Such extreme group differences are unlikely to appear in typical fMRI research. Therefore we do not expect such residual noise to profoundly bias typical fMRI between-group comparisons. However, with increasing sample sizes or when testing continuous variables that potentially correlate with motion levels (e.g. age or hyperactivity), investigators should interpret results carefully. To rule out spurious effects related to residual motion artifacts, one could validate the results in motion-matched groups (Satterthwaite et al., 2013b) or investigate the robustness of the results when incorporating group-level motion covariates (Yan et al., 2013).

Conclusion

We provided an extensive evaluation of currently available strategies for motion-artifact removal from rfMRI data, by means of comparing our strategy, ICA-AROMA, to extensive nuisance regression utilizing 6 or 24 realignment parameters, spike regression, scrubbing, aCompCor, SOCK, and ICA-FIX. Our results indicated that scrubbing, spike-regression, ICA-FIX and ICA-AROMA minimized the impact of motion artifacts to a large extent, in contrast to extensive nuisance regression, aCompCor, and SOCK which fail at reducing these effects. Despite spike regression and scrubbing being comparable with respect to removing motion artifacts, ICA-FIX and ICA-AROMA resulted in increased reproducibility of resting-state networks while limiting the loss of tDoF and preserving the temporal autocorrelation structure. However, without re-training its classifier, ICA-FIX can have profound impact on signal of interest. ICA-AROMA on the other hand retained signal of interest and resulted in highly consistent results across functional connectivity metrics and datasets, endorsing its robustness and potential added value for multi-center studies.

Acknowledgments

The NeuroIMAGE project was supported by NWO Large Investment Grant 1750102007010, ZonMW Grant 60-60600-97-193, and NWO Brain and Cognition grants 433-09-242 and 056-13-015 to Dr. Buitelaar the EU FP7 grant TACTICS (grantno. 278948), and grants from Radboud University Medical Center, University Medical Center Groningen and Accare, and VU University Amsterdam. Dr. Mennes has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 327340. Dr. Beckmann is supported by the Netherlands Organisation for Scientific Research (NWO-Vidi 864-12-003) and we gratefully acknowledge the funding from the Wellcome Trust UK Strategic Award [098369/Z/12/Z].

Conflict of interest

The authors declare that they have no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.02.063.

References

Andersson, J.L., Jenkinson, M., Smith, S., 2007. Non-linear registration, aka spatial normalisation. Technical Report. FMRIB Centre, Oxford, United Kingdom.

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.

Beckmann, C.F., DeLuca, M., Devlin, J.T., Smith, S.M., 2005. Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360, 1001–1013.

Beckmann, C., Mackay, C., Filippini, N., Smith, S., 2009. Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. NeuroImage 47, S148.

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.

Bhaganagarapu, K., Jackson, G.D., Abbott, D.F., 2013. An automated method for identifying artifact in independent component analysis of resting-state FMRI. Front. Hum. Neurosci. 7, 343.

Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.J., Lin, C.P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A.R.B., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.F., Zhang, H.Y., Castellanos, F.X., Milham, M.P., 2010. Toward discovery science of human brain function. Proc. Natl. Acad. Sci. U. S. A. 107, 4734–4739.

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.

Couvy-Duchesne, B., Blokland, G.A.M., Hickie, I.B., Thompson, P.M., Martin, N.G., de Zubicaray, G.I., McMahon, K.L., Wright, M.J., 2014. Heritability of head motion during resting state functional MRI in 462 healthy twins. NeuroImage 102 (Pt 2), 424–434.
Damoiseaux, J.S., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proc. Natl. Acad. Sci. U. S. A. 103, 13848–13853.
De Martino, F., Gentile, F., Esposito, F., Balsi, M., Di Salle, F., Goebel, R., Formisano, E., 2007. Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. NeuroImage 34, 177–194.
Dosenbach, N.U.F., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., Barnes, K.A., Dubis, J.W., Feczko, E., Coalson, R.S., Pruett, J.R., Barch, D.M., Petersen, S.E., Schlaggar, B.L., 2010. Prediction of individual brain maturity using fMRI. Science (New York, N.Y.) 329, 1358–1361.
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., Cohen, A.L., Dosenbach, N.U.F., Church, J.A., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., Schlaggar, B.L., 2008. The maturing architecture of the brain's default network. Proc. Natl. Acad. Sci. U. S. A. 105, 4028–4032.
Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach, N.U.F., Church, J.A., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2009. Functional brain networks develop from a "local to distributed" organization. PLoS Comput. Biol. 5, e1000381.
Filippini, N., MacIntosh, B.J., Hough, M.G., Goodwin, G.M., Frisoni, G.B., Smith, S.M., Matthews, P.M., Beckmann, C.F., Mackay, C.E., 2009. Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. Proc. Natl. Acad. Sci. U. S. A. 106, 7209–7214.
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.
Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. NeuroImage 48, 63–72.
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.
Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.
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.
Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage 62, 782–790.
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, 1–17.
Kalcher, K., Boubela, R.N., Huf, W., Bartova, L., Kronnerwetter, C., Derntl, B., Pezawas, L., Filzmoser, P., Nasel, C., Moser, E., 2014. The spectral diversity of resting-state fluctuations in the human brain. PLoS One 9, e93375.
Kelly, A.M.C., Di Martino, A., Uddin, L.Q., Shehzad, Z., Gee, D.G., Reiss, P.T., Margulies, D.S., Castellanos, F.X., Milham, M.P., 2009. Development of anterior cingulate functional connectivity from late childhood to early adulthood. Cereb Cortex (New York, N.Y.: 1991) 19, 640–657.
Kochiyama, T., Morita, T., Okada, T., Yonekura, Y., Matsumura, M., Sadato, N., 2005. Removing the effects of task-related motion using independent-component analysis. NeuroImage 25, 802–814.
Kong, X.Z., Zhen, Z., Li, X., Lu, H.H., Wang, R., Liu, L., He, Y., Zang, Y., Liu, J., 2014. Individual differences in impulsivity predict head motion during magnetic resonance imaging. PLoS One 9, e104989.
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.
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.
Liao, X.H., Xia, M.R., Xu, T., Dai, Z.J., Cao, X.Y., Niu, H.J., Zuo, X.N., Zang, Y.F., He, Y., 2013. Functional brain hubs and their test–retest reliability: a multiband resting-state functional MRI study. NeuroImage 83, 969–982.
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 96, 22–35.
Niazy, R.K., Xie, J., Miller, K., Beckmann, C.F., Smith, S.M., 2011. Spectral characteristics of resting state networks. Prog. Brain Res. 193, 259–276.
Patel, A.X., Kundu, P., Rubinov, M., Jones, P.S., Vértes, 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.
Perlbarg, V., Bellec, P., Anton, J.L., Pélégrini-Issac, M., Doyon, J., Benali, H., 2007. CORSICA: correction of structured noise in fMRI by automatic identification of ICA components. Magn. Reson. Imaging 25, 35–46.
Power, J.D., Fair, D.A., Schlaggar, B.L., Petersen, S.E., 2010. The development of human functional brain networks. Neuron 67, 735–748.
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., 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.
Pruim, R.H.R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage. http://dx.doi.org/10.1016/j.neuroimage.2015.02.064 (in press).
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.
Rummel, C., Verma, R.K., Schöpf, V., Abela, E., Hauf, M., Berruecos, J.F.Z., Wiest, R., 2013. Time course based artifact identification for independent components of resting-state FMRI. Front. Hum. Neurosci. 7, 214.
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.
Shehzad, Z., Kelly, A.M.C., Reiss, P.T., Gee, D.G., Gotimer, K., Uddin, L.Q., Lee, S.H., Margulies, D.S., Roy, A.K., Biswal, B.B., Petkova, E., Castellanos, F.X., Milham, M.P., 2009. The resting brain: unconstrained yet reliable. Cereb. Cortex 19, 2209–2229.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (Suppl. 1), S208–S219.
Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc. Natl. Acad. Sci. U. S. A. 106, 13040–13045.
Sochat, V., Supekar, K., Bustillo, J., Calhoun, V., Turner, J.A., Rubin, D.L., 2014. A robust classifier to distinguish noise from fMRI independent components. PLoS One 9, e95493.
Spisák, T., Jakab, A., Kis, S.A., Opposits, G., Aranyi, C., Berényi, E., Emri, M., 2014. Voxel-wise motion artifacts in population-level whole-brain connectivity analysis of resting-state fMRI. PLoS One 9, e104947.
Storti, S.F., Formaggio, E., Nordio, R., Manganotti, P., Fiaschi, A., Bertoldo, A., Toffolo, G.M., 2013. Automatic selection of resting-state networks with functional magnetic resonance imaging. Front. Neurosci. 7, 72.
The ADHD-200 Consortium, 2012. The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Front. Syst. Neurosci. 6, 62.
Thomas, C.G., Harshman, R.A., Menon, R.S., 2002. Noise reduction in BOLD-based fMRI using component analysis. NeuroImage 17, 1521–1537.
Tohka, J., Foerde, K., Aron, A.R., Tom, S.M., Toga, A.W., Poldrack, R.A., 2008. Automatic independent component labeling for artifact removal in fMRI. NeuroImage 39, 1227–1245.
Van De Ven, V.G., Formisano, E., Prvulovic, D., Roeder, C.H., Linden, D.E.J., 2004. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. Hum. Brain Mapp. 22, 165–178.
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.A., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. NeuroImage 59, 431–438.
von Rhein, D., Mennes, M., van Ewijk, H., Groenman, A.P., Zwiers, M.P., Oosterlaan, J., Heslenfeld, D., Franke, B., Hoekstra, P.J., Faraone, S.V., Hartman, C., Buitelaar, J., 2015. The NeuroIMAGE study: a prospective phenotypic, cognitive, genetic and MRI study in children with attention-deficit/hyperactivity disorder. Design and descriptives. Eur. Child Adolesc. Psychiatry 24, 265–281.
Wisner, K.M., Atluri, G., Lim, K.O., Macdonald, A.W., 2013. Neurometrics of intrinsic connectivity networks at rest using fMRI: retest reliability and cross-validation using a meta-level method. NeuroImage 76, 236–251.
Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. NeuroImage 45, S173–S186.
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., 2013. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. NeuroImage 76, 183–201.
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.
Zuo, X.N., Xing, X.X., 2014. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. Neurosci. Biobehav. Rev. 45, 100–118.
Zuo, X.N., Kelly, C., Adelstein, J.S., Klein, D.F., Castellanos, F.X., Milham, M.P., 2010. Reliable intrinsic connectivity networks: test–retest evaluation using ICA and dual regression approach. NeuroImage 49, 2163–2177.