Statistical Improvements in Functional Magnetic Resonance Imaging Analyses Produced by Censoring High-Motion Data Points

Joshua S. Siegel,[1]* Jonathan D. Power,[1] Joseph W. Dubis,[1] Alecia C. Vogel,[1] Jessica A. Church,[1] Bradley L. Schlaggar,[1,2,3,4] and Steven E. Petersen[1,2,4,5,6,7]

[1]Department of Neurology, Washington University School of Medicine, St. Louis, Missouri
[2]Department of Radiology, Washington University School of Medicine, St. Louis, Missouri
[3]Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri
[4]Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri
[5]Department of Psychology, Washington University, Saint Louis, Missouri
[6]Department of Neurosurgery, Washington University School of Medicine, St. Louis, Missouri
[7]Department of Biomedical Engineering, Washington University, Saint Louis, Missouri

Abstract: Subject motion degrades the quality of task functional magnetic resonance imaging (fMRI) data. Here, we test two classes of methods to counteract the effects of motion in task fMRI data: (1) a variety of motion regressions and (2) motion censoring ("motion scrubbing"). In motion regression, various regressors based on realignment estimates were included as nuisance regressors in general linear model (GLM) estimation. In motion censoring, volumes in which head motion exceeded a threshold were withheld from GLM estimation. The effects of each method were explored in several task fMRI data sets and compared using indicators of data quality and signal-to-noise ratio. Motion censoring decreased variance in parameter estimates within- and across-subjects, reduced residual error in GLM estimation, and increased the magnitude of statistical effects. Motion censoring performed better than all forms of motion regression and also performed well across a variety of parameter spaces, in GLMs with assumed or unassumed response shapes. We conclude that motion censoring improves the quality of task fMRI data and can be a valuable processing step in studies involving populations with even mild amounts of head movement. *Hum Brain Mapp 00:000–000, 2013.* © **2013 Wiley Periodicals, Inc.**

Key words: fMRI; head movement; motion; general linear model; GLM; task; noise; data quality; scrubbing

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: The National Institutes of Health; Contract grant number: R21 NS61144; R01 NS26424; R01 ND057076; F30 MH940322; Contract grant sponsor: A McDonnell Foundation Collaborative Action award.

*Correspondence to: Joshua S. Siegel, Department of Neurology, Washington University School of Medicine, 660 S Euclid Ave, Box 8111, St. Louis, MO 63110, USA. E-mail: siegelj@wusm.wustl.edu

Received for publication 2 November 2012; Revised 20 February 2013; Accepted 2 April 2013.

DOI: 10.1002/hbm.22307

Published online in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Head motion is problematic in functional magnetic resonance imaging (fMRI) studies [Barch et al., 1999; Birn et al., 1998, 1999, 2004; Friston et al., 1996; Gopinath et al., 2009; Hutton et al., 2002; Jiang et al., 1995; Johnstone et al., 2006; Lemieux et al., 2007; Oakes et al., 2005; Wu et al., 1997; Yetkin et al., 1996]. Blood–oxygen-level-dependent (BOLD) signal acquisition depends on precise spatial and temporal placement of magnetic gradients on scales of millimeters and milliseconds. Head motion during scans causes image intensity to reflect not only blood oxygenation but also frank motion-related artifact. The more a data set is contaminated with such motion-related signal changes, the more difficult it becomes to detect neurophysiological events of interest. For investigators of subject populations that tend to move, methods to recover relatively high-quality data and results from relatively low-quality scans are of clear importance.

In task fMRI, head motion is often dealt with, first by mandatory data realignment and then by optional, additional measures to counter motion-related effects. It is a common practice to align the data throughout a scan by estimating the position of the head in space at each volume, followed by realignment using rigid body transforms. In such transforms, head position at each time point is described with six parameters: translational displacements along the *X*-, *Y*-, and *Z*-axes, and rotational displacements of pitch, yaw, and roll. Realignment is an essential part of data processing, but it cannot correct the signal alterations or image distortions that occur as a result of movement.

Further optional steps can be taken to counter movement-related disruption of BOLD signal in task fMRI [Birn et al., 2004]. Investigators have demonstrated the utility of a variety of methods, including interpolation over motion-corrupted voxels [Huang et al., 2008], weighting images by the inverse of their variance [Diedrichsen and Shadmehr, 2005], ignoring volumes containing stimulus-correlated motion [Birn et al., 2004], ignoring volumes during gross movement [Lemieux et al., 2007], monitoring and modeling physiologic and motion-related noise [Jones et al., 2008], and including motion-related nuisance regressors in general linear model (GLM) estimation [Friston et al., 1996]. Motion regression is perhaps the most widely used among these methods, and regressors have been shown to improve the reliability of GLM analyses in many cases [Lund et al., 2005; Morgan et al., 2007; Oakes et al., 2005].

GLM estimation in the presence of motion, or attempts to counter motion, is, however, not a simple matter. Movement-related signal changes tend to degrade the fit of parameter estimates to the data in a GLM, increasing the error term and reducing statistical significance. Additionally, movement that is correlated with tasks can produce spurious task-related activity [Bullmore et al., 1999; Field et al., 2000; Hajnal et al., 1994]. Motion regression is aimed at compensating for some of these effects, but motion regression cannot correct motion-induced signal drop-out or image distortion. Additionally, in instances where subject motion is not independent of task timing, motion regression can cause underestimation of true experimental effects [Bullmore et al., 1999; Johnstone et al., 2006]. Such considerations introduce uncertainty to the meaning of results in the presence of motion or attempts to counter motion.

Here, we evaluate an approach that mitigates effects of head motion in task fMRI analysis while avoiding some of the aforementioned ambiguities. We previously observed large-amplitude signal changes limited to the periods of subject movement in task-free resting-state functional connectivity (RSFC) MRI. These observations can be made even after motion regression is performed, indicating that at least some common regressions do not adequately remove motion-related changes in BOLD signal [Power et al., 2012]. Accordingly, we developed a "scrubbing" method to identify and remove completely high-motion data from our analyses (by censoring volumes using temporal masks). This procedure revealed clear and powerful effects of motion in our RSFC analyses despite having prepared our data with standard motion regressions [Power et al., 2012].

Here, we modify our "scrubbing" procedure, referred to here as motion censoring or simply censoring, to evaluate whether applying temporal masks to remove high-motion volumes can improve results in task fMRI. Similar censoring approaches have been used in task fMRI previously [Kennedy and Courchesne, 2008; Kirwan et al., 2009; Stark et al., 2010] and in RSFC [Lemieux et al., 2007] and are implemented in software packages such as AFNI. However, to our knowledge, the statistical benefit of these approaches has not been evaluated. We demonstrate the benefit of this motion censoring approach in three data sets, showing that censoring generally decreases variance across subjects in parameter estimates (i.e., time courses, or Level I analyses) and increases statistical power in ANOVAs and *t*-tests (Level II analyses). We propose methods to select unbiased regions of interest (ROIs) to test the effects of censoring in data sets, and we demonstrate how censoring parameters can be explored to yield increases in statistical power. We find that more stringent censoring criteria (removing more motion-contaminated data) produce increased statistical benefits up to some point, beyond which the cost of removing additional data points outweighs the benefit of censoring. We explore an additional parameter space associated with choosing volumes to censor (e.g., censoring additional volumes before and after periods of motion) and find little difference between choices. We demonstrate these benefits when the data are modeled without assumed response shapes and also with assumed response shapes. Finally, we demonstrate that motion censoring outperforms several varieties of motion regression in task fMRI.

METHODS

Subjects

Subjects were recruited from the Washington University in St. Louis campus and the surrounding community. The subjects were recruited for three separate studies, and are referred to as Cohorts 1, 2, and 3. For all three cohorts, individuals with metal implants, heart arrhythmias, claustrophobia, or a reported history of developmental delay were excluded. Individuals in Cohorts 1 and 3 reported no history of neurologic and psychiatric diagnoses and did not use psychotropic medications. Individuals in Cohort 2 were recruited as part of a study of Tourette Syndrome and were not excluded on the basis of any psychiatric diagnosis commonly comorbid with Tourette syndrome including attention-deficit/hyperactivity disorder and obsessive compulsive disorder, nor were they excluded for taking psychotropic medications. All subjects were native English speakers. All minors were brought in for a visit prior to scanning for the consenting process, an introduction to the scanning environment via a mock scanner, and neuropsychological testing. For all minor subjects, verbal assent and parental informed consent to the testing and scanning was acquired in accordance with the guidelines and approval of the Washington University Human Studies Committee. All adult subjects gave informed consent prior to scanning in accordance with the guidelines and approval of the Washington University Human Studies Committee. Subjects were compensated for their participation. All aspects of the studies were approved by the Institutional Review Board at Washington University School of Medicine.

Behavioral Paradigms and Data Collection

This study utilized three event-related task fMRI cohorts representing children, adolescents, and adults (Table I). We use these data not to focus on particular ages or patterns of activity, but to show across studies, conditions, and contrasts, the impact of various motion correction strategies.

For all cohorts, data were acquired on a Siemens MAGNETOM Tim Trio 3.0T Scanner with a Siemens 12-channel Head Matrix Coil (Erlangen, Germany). A thermoplastic mask was individually fitted to each subject's head to limit head motion during data acquisition. A T1-weighted sagittal MP-RAGE structural image was obtained (echo time [TE] = 3.06 ms, repetition time [TR]-partition = 2.4 s, TI = 1,000 ms, flip angle = 8°, 176 slices with 1 × 1 × 1 mm voxels). A T2-weighted turbo spin-echo structural image (TE = 84 ms, TR = 6.8 s, 32 slices with 2 × 1 × 4 mm voxels) in the same anatomical plane as the BOLD images was also obtained to improve alignment to an atlas.

Cohort 1 consisted of 53 children of ages 7–8 years. Subjects in this cohort performed a string-matching task on two simultaneously presented strings of letters or letter-like forms. Five categories or strings (words, pseudowords, nonwords, consonant strings, and Amharic characters) were separated by run. Subjects were asked to make a visual matching decision via button press. Trials were arranged for analysis in a rapid event-related design. Intertrial intervals were randomly distributed between 1, 2, and 3 TRs. Functional images were obtained using a BOLD-contrast sensitive gradient-echo echo-planar sequence (TE = 27 ms, flip angle = 90°, in-plane resolution = 4 × 4 mm, 32 contiguous interleaved 4 mm axial slices, volume TR = 2.5 s). Five task runs each lasting 133 volumes (332.5 s) were obtained in each subject. In total, 15 conditions were modeled, each lasting seven time points (TRs). Subjects with <60% accuracy on the task, root-mean-square realignment estimates (RMS movement) exceeding 1.5 mm for the entire session, or the presence of any GLM variables with less than two data points contributing to its estimation following the framewise displacement (FD) = 0.9 motion censoring (described below) were excluded. In all, 20 out of 53 subjects remained after exclusion criteria were applied.

Cohort 2 consisted of 73 children and adolescents (9–15 years) with Tourette syndrome. Subjects in this cohort performed a cue-switching task drawing attention to either the color or the identity of cartoon characters. Cues (a single word presented in all capital letters) were presented for one TR, and Targets (a colorful cartoon character) were presented in the subsequent TR, with approximately 20% of trials having only a cue and not a target stimulus. Target judgments were made via button press. Trials were arranged for analysis in a rapid event-related design with complex trials (including separable cue and target trials).

TABLE I. Cohort properties. 361×270mm (150 × 150 DPI)

	N(M/F)	Age (years) (mean)	Study design	RMS movement (mm) mean (sd)	RMS movement FD (mm) mean (sd)	Main scrubbing settings	% data censored mean (s.d.)
Cohort 1: Children (typical)	20(10/10)	7–8 (8.0)	matching	0.78(0.31)	0.83(0.39)	FD > 0.9 mm	16(11)
Cohort 2: Adolescents (Tourette)	38(32/6)	9–15 (12.6)	rule switching	0.59(0.26)	0.55(0.29)	FD > 0.9 mm	8(7)
Cohort 3: Adults (typical)	30(14/16)	21–30 (24.4)	Posner task	0.38(0.20)	0.22(0.13)	FD > 0.5 mm	4(5)

Intertrial intervals were randomly distributed between 1, 2, and 3 TRs. Functional images were obtained using a BOLD-contrast sensitive gradient-echo echo-planar sequence (TE = 27 ms, flip angle = 90°, in-plane resolution= 4 × 4 mm, 32 contiguous interleaved 4 mm axial slices, volume TR = 2.0 s). Three to six task runs each lasting 144 volumes (288 s) were obtained in each subject. Four cue conditions and eight target conditions were modeled, each lasting nine time points (TRs). Subjects with <70% accuracy on the task or fewer than three task runs with RMS movement below 1.5 mm were excluded. In brief, 38 out of 73 subjects remained after exclusion criteria were applied.

Cohort 3 consisted of 35 adults of ages 21–30 years. Subjects in this cohort performed a visual attention (modified Posner) task. Target detection judgments were made via button press. Trials were arranged for analysis in a mixed block/event-related design with complex trials (cue and target). Intertrial intervals were randomly distributed between 1, 2, and 3 frames. Functional images were obtained using a BOLD-contrast sensitive gradient-echo echo-planar sequence (TE = 27 ms, flip angle = 90°, in-plane resolution = 4 × 4 mm, 32 contiguous interleaved 4 mm axial slices, volume TR = 2.5 s). Six to eight task runs each lasting 217 volumes (542.5 s) were obtained in each subject. Two cue conditions and 10 target conditions were modeled, each lasting seven time points (TRs). Subjects with <85% accuracy on the task, or fewer than four runs with RMS movement below 1.0 mm were excluded. In total, 30 out of 35 subjects remained after exclusion criteria were applied.

Head Realignment Estimate Calculations

Head motion estimation involved a series of rigid body transforms, T_i, where i indexes frame (volume) and T_i spatially registers volume i to a selected reference frame. Each transform was computed by minimizing the registration error $\epsilon_i = \langle (sI_i(T(\vec{x})) - I_o(\vec{x}))^2 \rangle$, where $I(\vec{x})$ is image intensity at locus $\vec{x}$, angle brackets denote the spatial average over the brain, subscript 0 denotes the reference frame (here, taken as the run midpoint) and s is a scalar factor that compensates for small changes in mean signal intensity. Each transform can be expressed as a combination of rotation and displacement components. Thus,

$$T_i = \begin{bmatrix} R_i & d_i \\ 0 & 1 \end{bmatrix}$$

where R_i is a 3 × 3 rotation matrix and d_i is a 3 × 1 column vector of displacements. R_i can be factored into three elementary rotations about each of the three axes. Thus, $R_i = R_{i\alpha} R_{i\beta} R_{i\gamma}$, where

$$R_{i\alpha} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & \cos\alpha_i & -\sin\alpha_i \\ 0 & \sin\alpha_i & \cos\alpha_i \end{bmatrix}, \; R_{i\beta} = \begin{bmatrix} \cos\beta_i & 0 & \sin\beta_i \\ 0 & 1 & 0 \\ -\sin\beta_i & 0 & \cos\beta_i \end{bmatrix},$$

$$\text{and } R_{i\gamma} = \begin{bmatrix} \cos\gamma_i & -\sin\gamma_i & 0 \\ \sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus, each rigid body transform is defined by six parameters.

FD Calculations

Differentiating head realignment parameters over time yields a six-dimensional time series that represents instantaneous head motion. To express instantaneous head motion as a scalar quantity, we used the empirical formula for FD, $FD_i = |\Delta d_{ix}| + |\Delta d_{iy}| + |\Delta d_{iz}| + |\Delta\alpha_i| + |\Delta\beta_i| + |\Delta\gamma_i|$, where $\Delta d_{ix} = d_{(i-1)x} - d_{ix}$, and similarly for the other rigid body parameters $[d_{ix}, d_{iy}, d_{iz}, \alpha_i, \beta_i, \gamma_i]$. FD_0 is set to zero. Rotational displacements were converted from radians to millimeters by calculating displacement on the surface of a sphere of radius 50 mm, which is approximately the mean distance from the cerebral cortex to the center of the head. This calculation is identical to that used in Power et al. [2012] for "scrubbing" RSFC-MRI data.

fMRI Preprocessing

Functional images were first processed to reduce artifacts [Miezin et al., 2000]. These steps included: (i) correction of odd versus even slice intensity differences attributable to interleaved acquisition without gaps, (ii) correction for head movement within- and across-runs, and (iii) within-run intensity normalization to a whole-brain mode value (across TRs and voxels) of 1,000.

Atlas transformation of the functional data was computed for each individual via the MP-RAGE scan. For Cohorts 1 and 2, the transformation was done by using an atlas-representative target composed of a mutually coregistered independent sample of 12 healthy adults and 12 healthy 7- to 8-year-old children, which was made to conform to the Talairach atlas using a spatial normalization method [Lancaster et al., 1995]. For Cohort 3, an atlas based on 12 healthy adults was used. Each run was then resampled in atlas space on an isotropic 3-mm grid combining movement correction and atlas transformation in a single interpolation. Data were resampled into 3-mm isotropic voxels for Cohorts 1 and 3 and into 2-mm isotropic voxels for Cohort 2. This discrepancy in voxel sizes arose incidentally but serves to demonstrate the generalizability of results beyond a single voxel size. The atlas-transformed image for each participant was checked against a reference average to ensure appropriate registration.

RMS movement was calculated from realignment parameters (rotational estimates converted to translational at radius of 50 mm). As previously mentioned, subjects were excluded from each study on the basis of study-specific RMS movement thresholds. This study thus documents the improvements that can be seen within "acceptable" subject populations. Excluded subjects are not reported in this study or in Table I.

Standard Processing/Uncensored Data

Some data were analyzed at this point with no further efforts to counter motion effects. Such data are said to have undergone Standard Processing and are referred to as "uncensored" data.

Motion Censoring

The motion censoring procedure entailed the following steps. (i) FD was calculated as mentioned above. (ii) All volumes whose FD exceeded a particular threshold were flagged to form a temporal mask (Table I and Fig. 1). (iii) An optional step of temporal mask augmentation flagged additional volumes preceding and/or following flagged volumes. (iv) The GLM ignored all flagged volumes during parameter estimation (equivalent to adding single-TR regressors at censored volumes). Data that underwent this process are called "censored" data.

Thresholds

The goal of this report is to document and explore the benefits of motion censoring in a task fMRI context. Accordingly, thresholds were chosen to remove modest portions of motion-contaminated data, not to remove all volumes with motion. A threshold of FD $>$ 0.90 mm is often used in this report (Figs. 1–3) though other thresholds are examined (Fig. 6).

Augmentation

The uncertainty of the precise timing of movement and the need to re-establish spin histories suggests that it may also be appropriate to flag volumes 1 back and at least 1 forward of any motion-flagged volume. Most figures display censoring using no augmentation. If augmentation is performed, "fX" is used to indicate temporal mask augmentation after flagged volumes, and "bX" indicates augmentation prior to flagged volumes (e.g., f0, b0 means forward zero, backward zero, i.e., no augmentation). For Figure 6, thresholds and augmentations were chosen to remove nearly identical amounts of data (i.e., relaxed thresholds with augmentation vs. stringent thresholds with no augmentation, each removing similar amounts of data). "Random censoring" was accomplished by examining a subject's temporal mask (in Fig. 6, FD $>$ 0.9 mm, f0, b0) and removing identical amounts of data (with identical

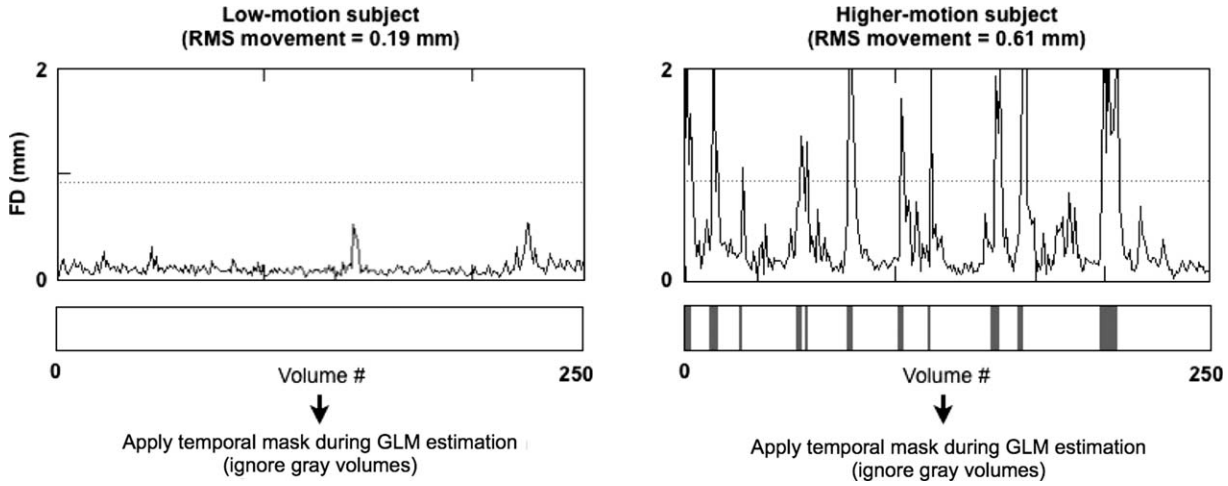

Figure 1.

Identification of high-motion volumes of BOLD data. Data from two subjects are presented to illustrate the methodology used in this manuscript. Two runs of BOLD data from each subject are used (2 $\times$ 125 frames), and a volume-to-volume index of head motion (FD) is plotted by summing at each frame the absolute values of the derivatives of the head realignment parameters used to realign the BOLD data. The subject at left moves very little, whereas the subject at right moves substantially at several points. A dotted line indicates a FD of 0.90 mm, the main threshold used in this manuscript. Volumes whose FD exceeded this threshold are shown below with vertical gray bars, forming a temporal mask that can be applied during GLM calculations to ignore volumes likely to contain artifactual BOLD signal changes caused by subject motion.

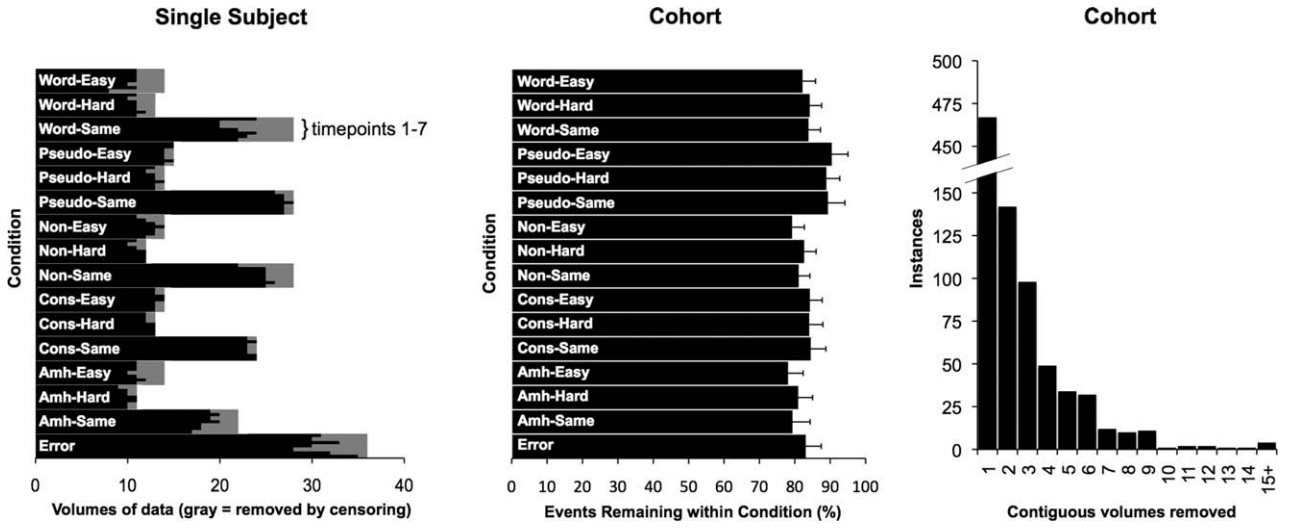

Figure 2.

An overview of how typical motion censoring settings impact Cohort 1. At left, a plot of data collected/censored from a single individual. The horizontal black bars represent the number of events collected for each condition. Within this, fine black bars show the number of data points used to model each condition at each time point (1–7) and gray bars are the number of data points removed by motion censoring. A two-factor ANOVA on the cohort reveals no effect of time point or condition on volumes removed. Middle, for the entire cohort, the average percentage of events remaining in each condition after censoring. At right, a histogram showing the distribution of sizes (in contiguous volumes) of the portions of data masked out by the standard temporal mask (FD $>$ 0.9 mm) used in this article. Many motion epochs are only last a single TR, but many are also considerably longer.

distribution of temporal lengths) from the subject but at random, not targeting periods of motion.

The motion censoring procedure used here differs from that described in Power et al. [2012] in one major respect, which is that the DVARS measure is not used to censor data. DVARS measures the change in signal across all voxels in the brain from volume to volume. As the nature of task fMRI is to evoke BOLD signal changes at particular time points, the use of such a measure would tend to target not only periods of movement but also task-related activity. The measure is therefore not used in this article.

Motion Regressions

To test motion regression as a means of countering movement-related effects, motion parameters were included as regressors of no interest in the calculation of the GLM. Five different combinations of motion regressors were tested:

1. FD_t: FD (1 regressor).
2. R_t: detrended rigid body realignment parameters (6 regressors).
3. R_t': temporal derivatives of R (6 regressors).
4. R_t and R_t' (12 regressors).
5. V_t, V_t^2, V_{t-1}, V_{t-1}^2: where Vs are the realignment parameters (24 regressors). This is the Volterra expansion proposed in Friston et al. [1996]. Values of 0 are used for V_{t-1} at a time point 0.

GLM Estimation

As individual volumes are being withheld from the data, it is worth defining all terms used to refer to the data. A single volume of data may be referred to as a frame (as in a movie) or as a TR of data. When a condition is modeled, the number of times the condition occurs is the number of events contributing to the condition. Each event lasts the number of modeled TRs. Thus, a single volume, if it contributes to the modeling of several conditions owing to a rapid event-related design, may represent multiple events. If such a volume was censored, each condition to which it contributed would have one less event at the time point represented by that volume.

Statistical analyses of event-related fMRI data were based on the GLM as described previously [Brown et al., 2005; Miezin et al., 2000; Schlaggar et al., 2002] using in-house software programmed in the Interactive Data Language (ITT Visual Information Solutions, Boulder, CO) and C [Miezin et al., 2000; Ollinger et al., 2001]. Temporal masks (censoring) and motion regressors were incorporated into model estimation where indicated. GLM terms

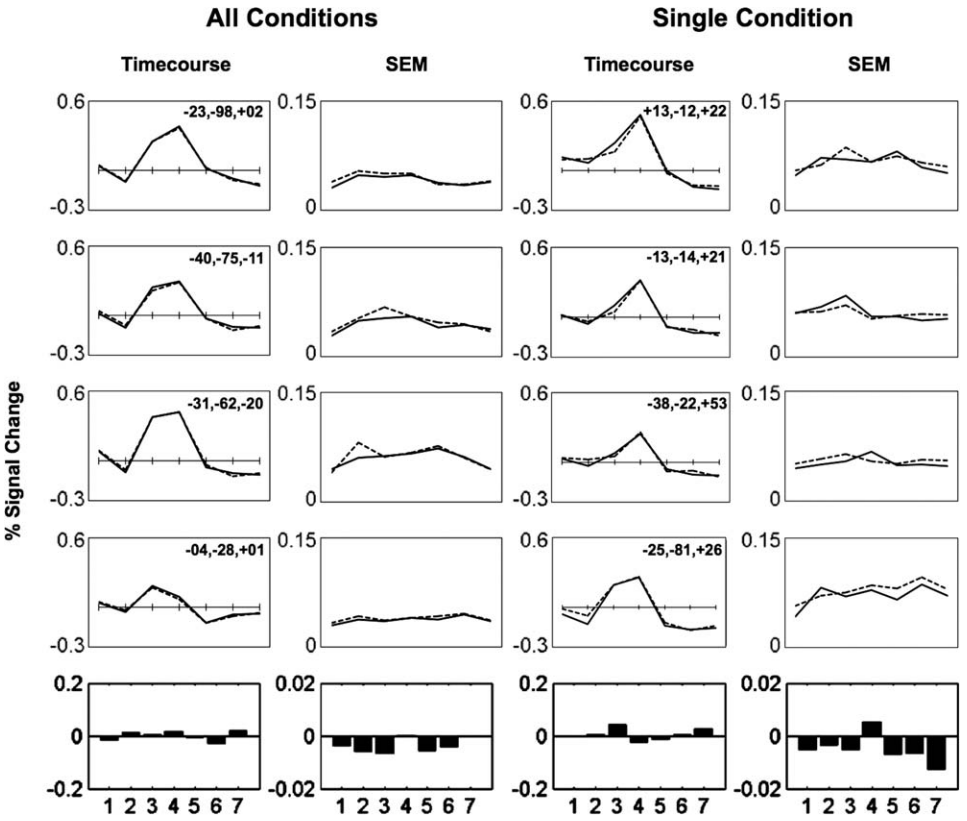

Figure 3.

The variance of parameter estimates across subjects typically decreases as a result of motion censoring. Uncensored (dotted line) and censored (solid line) time courses are shown for mean map ROIs identified by ANOVAs as significantly active in all conditions (left side) and a single condition (right side) (the same analyses shown in Figs. 4 and 5). Representative ROIs whose change in z-score is within one standard deviation of mean z-score change are used (the red points from the middle of the scatter plots in Fig. 5). To the right of each time course is a plot of standard error for each time point estimate. The bottom row shows mean differences (censored–uncensored) at each time point for the top 50 positive time courses and the mean change in SEMs. The magnitude of time course estimates changed little with censoring, but the variance of time course estimates across subjects typically decreased with censoring, providing a general mechanism for the increased z-scores seen in Figure 5.

included linear drift terms, baseline terms, and terms associated with each modeled condition. Unless specified, no assumptions were made regarding the shape of the hemodynamic response function (time course), only the durations were constrained for each condition (seven to nine time points depending on TR; ~18 s). This approach is similar to the FIR approaches available in packages such as FSL or SPM. ANOVA over time was used to assess significance of time series generated with unassumed response shapes. Some linear models were also calculated using a double gamma function as the assumed shape of the hemodynamic response. In this instance, *t*-tests of betas were used to assess significance of activation.

Typically, parameter estimates (time courses) for several conditions are modeled in a study (e.g., for Words, Nonwords, Pseudowords, Errors, etc., in Cohort 1). Most conditions are modeled over seven TRs (17.5 s in Cohort 1) and hence time courses will have seven time points. Motion censoring can result in a variable number of events (contributing volumes) at different time points in a condition. For example, prior to censoring, a subject may have 12 events in a condition, but if suprathreshold motion occurred during TRs 1–3 of one event in the condition, time points 1–3 would have one fewer event than time points 4–7 of the condition. Subjects for whom less than two events (two data points contributing to an estimate) remained for any time point in any condition following application of a temporal mask were excluded from further analysis (two subjects in Cohort 1, not included in Table I or any further analyses).

ROI Selection

Two methods for ROI definition are used in this article. In all cases, a peak detection algorithm was used to identify up to 50 ROIs from peaks of statistical images with $z > 3.5$ at a spacing of at least 10 mm. ROIs were modeled as 10-mm diameter spheres. Four ROIs that were clearly in white matter or ventricles were excluded in Cohort 2. Statistical images were generated from ANOVAs operating on time courses, and reported z-scores are the average values found within an ROI.

"Uncensored" ROIs were defined based on ANOVAs operating in uncensored (standard processed) data. These ROIs provided baseline z-scores to compare with the z-scores arising under various processing strategies.

"Mean map" ROIs were defined by separately running ANOVAs on censored and uncensored data and creating an averaged statistical map for a given effect or contrast. Peaks on the averaged map defined ROIs, providing an unbiased ROI set with which to compare censoring and standard processing. Hence, as long as the results are not known beforehand, this is an unbiased way of selecting ROIs to compare two methodologies. When this methodology was selected, we did not know the results. As it will be seen, censoring produces higher z-scores, in general, meaning that these ROIs are biased toward censored peaks. However, the same results are seen if "Uncensored" ROIs are used (Fig. 6), rendering discussion of such biases moot. Another possible way to select unbiased ROIs is to use conjunctions of the peak ROIs found from two methodologies.

Residual Signal Comparison

In Cohort 1, the uncensored ROIs were used to compare GLM residual signal between uncensored, censored, and randomly censored data. For a given ROI within a given subject, at each time point, residuals at each time point were calculated as the average residual across all voxels within the ROI. RMS values over time were then calculated for each ROI within each subject. Figure 4 shows these RMS values.

Analysis of Within-Subject Variance in Cohort 1

As described above, each subject in Cohort 1 completed five task runs. To obtain a measure of within-subject variance in parameter estimates for each individual subject, a separate linear model was computed for each individual task run to generate time courses. The following analysis focused on the time course generated for all correct trials across conditions in each run of each subject (the "All Conditions" of Fig. 4). For each subject, the variance of the five estimates was computed at each of seven time points for each of the 50 "mean map" ROIs, before and after censoring. Total variance across time points and ROIs within each subject was compared in censored versus uncensored data using a one-tailed paired t-test (Supporting Information Fig. S1).

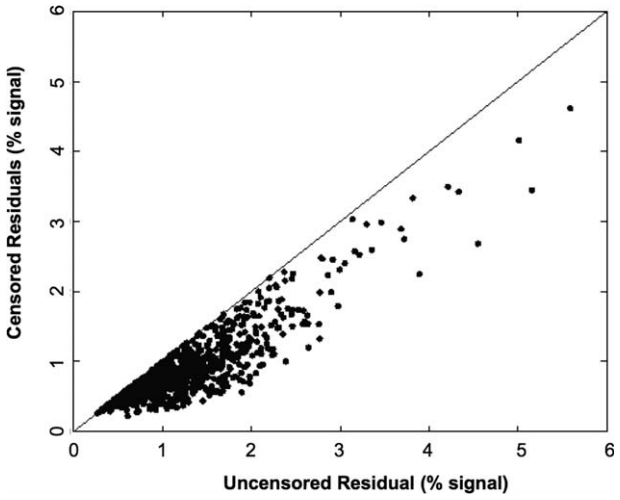

Figure 4.

Motion censoring reduces the signal assigned to error terms in GLMs. Fifty uncensored ROIs were defined across all correct trials. These ROIs were modeled as 10-mm diameter spheres and applied to censored and uncensored GLMs. The RMS residual value at each ROI in each subject was calculated across all time points. Each black dot on this plot compares the RMS residual of an ROI before and after the temporal mask is applied. Residuals almost always decrease upon application of the temporal mask, indicating that the GLM was better able to model the variance in the data when high-motion volumes were excluded. Such decreases were not seen with random censoring.

RESULTS

Three independent task fMRI studies are examined in this report. The studies varied by task, subject's age, clinical status, exclusion criteria, and average head motion estimates. Younger subjects exhibited greater amounts of movement (Table I). A variety of analyses are presented for the study involving children (Cohort 1). The studies involving adolescents and adults (Cohorts 2 and 3) are presented toward the end of the report as further demonstrations of the benefits of motion censoring.

An Overview of How Motion Censoring Impacts Subjects in Cohort 1

Figures 1 and 2 show an overview of how one typical version of motion censoring impacted Cohort 1. This version of censoring is used throughout this report unless otherwise indicated. Figure 1 shows the FDs for two children (one still subject, one subject with intermittent movements) and how temporal masks are formed (gray volumes are

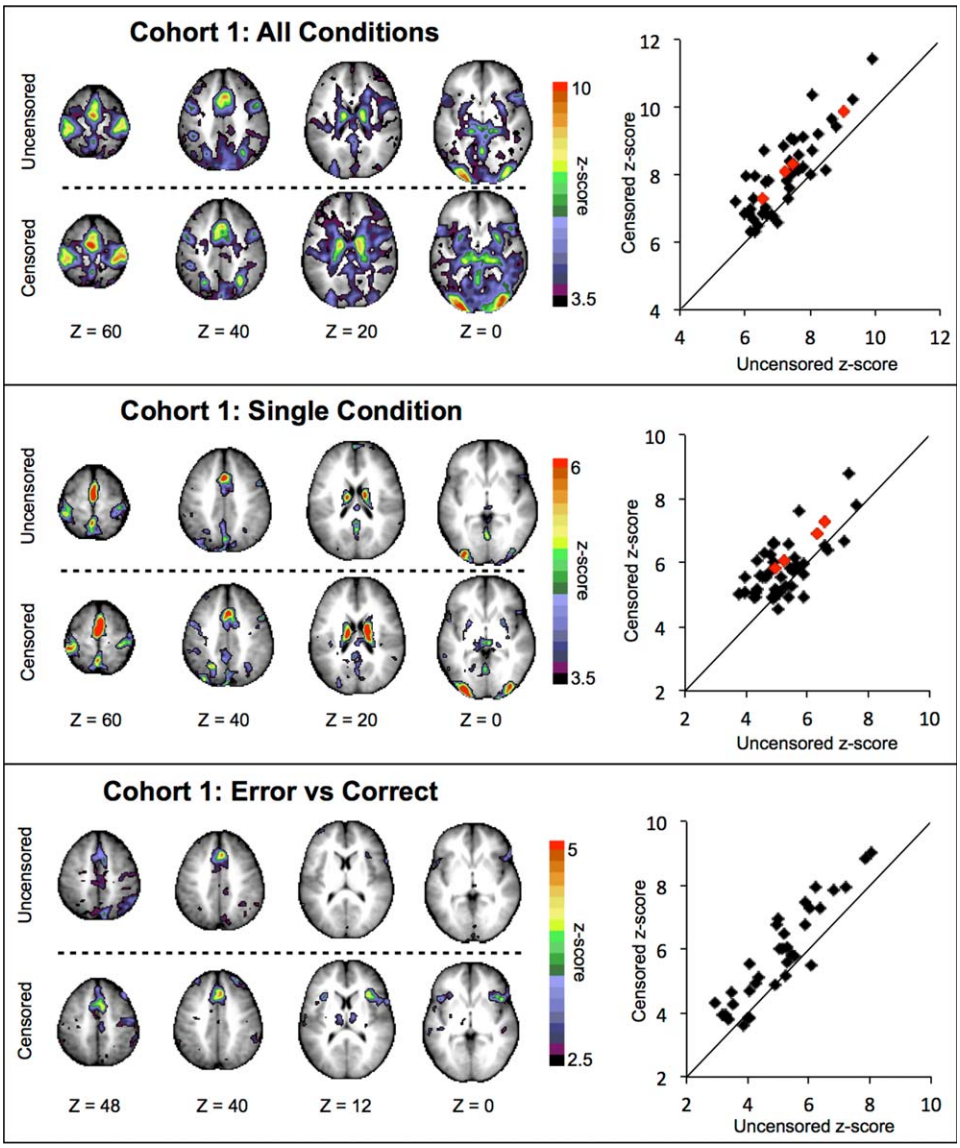

Figure 5.

z-Scores increase as a result of motion censoring in a variety of comparisons. Statistical maps from ANOVAs operating on all correct conditions (a main effect of time), a single condition (main effect of time), and a within-subject comparison of error versus correct trials are shown for uncensored and censored data. At left, the statistical maps with scale bars. At right, z-scores for mean map ROIs before and after censoring. Regions are "mean map" ROIs. All points above the black line in each scatter plot ($x = y$) demonstrate increased z-scores as a result of censoring. Time courses from representative ROIs (red points in the scatter plot) are shown in Figure 3.

ignored). For Cohort 1, a temporal mask defined by a FD > 0.9 mm with no augmentation (f0, b0) was chosen as a default setting for data censoring (Fig. 1). These temporal masks were chosen to remove a modest amount of high-motion data for proof-of-principle purposes. They are not designed to identify all time points when subjects moved but rather only periods of large movement.

Figure 2 shows how the resulting temporal masks impacted the design matrices of individual subjects and Cohort 1, in general. At left in Figure 2, the number of retained (black) and ignored (gray) events at each time point (fine bars) of each condition (big bars) modeled by the GLM is shown for a single subject. This subject was typical of the cohort in terms of motion; with an RMS

movement of 0.72 mm. Censoring removed 16 ± 11% of the data across subjects (range, 1–36%). The middle panel of Figure 2 shows the percentage of events removed from each condition across subjects (collapsed across time points within the condition). A two-way ANOVA (condition and time point) found no effect of time point or condition on percent of events removed in the cohort, meaning that motion was just as likely in all conditions, and at all time points in conditions.

One possible method of motion correction is to interpolate signal during instances of motion using data before and after movement [Huang et al., 2008]. The temporal masks generated by censoring indicate that this approach may not be practical for many instances of motion. The right panel of Figure 2 is a histogram of the number of temporally contiguous volumes excluded by the temporal masks across Cohort 1 (i.e., the sizes of the gray portions of the temporal masks in Fig. 1). Although many instances of motion are brief (1 volume), many are also extended. Of the 865 instances of motion identified across this cohort, 250 (29%) lasted three or more time points. Although gaps of one or two TRs can be interpolated meaningfully, interpolation over gaps of many TRs seems unlikely to reflect an underlying signal with much fidelity.

First-Level Analysis: Motion Censoring Reduces Variance in Parameter Estimates

GLMs were estimated in uncensored and censored versions of Cohort 1's data using the FD > 0.90 f0, b0 settings described above. Figures 3–5 show the impact of censoring on Level I (time course estimation and GLM fit) and Level II (statistical map) analyses. To compare parameter estimates before and after censoring, uncensored and censored statistical maps were averaged and peaks in this averaged map were selected as "mean map" ROIs to compare uncensored versus censored results (**METHODS**). Figure 3 shows representative time courses from mean map ROIs that showed effects over all correct conditions, and in a single condition (Pseudowords). Seven time points (17.5 s) are modeled for all conditions in these GLMs, and no assumptions are made about the shape of the hemodynamic response. Uncensored time courses are shown with dotted lines, and censored time courses are shown with solid lines. Time courses are relatively unchanged in shape and magnitude. However, the between-subject variance in parameter estimates generally decreases, suggesting that second-level analyses ought to gain statistical power.

An important question is whether within-subject variance in parameter estimates is reduced, in addition to between-subject variance. Answering this question is not straightforward in Cohort 1 because this data set is a rapid event-related design, not a widely spaced design, and hence the shape of each trial time course cannot be estimated individually. However, by splitting the data set into five parts (runs), for each subject, and obtaining parameter estimates within each subset under uncensored and censored processing, we were able to compare within-subject variance in parameter estimates before and after censoring. This comparison, shown in Supporting Information Figure S1, demonstrated a 14.2% reduction of within-subject variance in parameter estimates ($t[19] = 2.018$, $P = 0.029$).

First-Level Analysis: Motion Censoring Reduces the Error Term in GLM Estimation

Adding noise to data should reduce the fit of signal to parameters within GLMs, thereby increasing the signal left in the residual (error term). If censoring removes noise from the data, it should improve the fit of signal to parameters and reduce the residual signal. Results thus far indicate that censoring produces more uniform parameter estimates across subjects, consistent with the first prediction. To assess the second prediction, uncensored ROIs that were active across all correct conditions were identified, and the RMS residuals of those ROIs (across all time points) were computed for uncensored and censored GLMs, as well as for GLMs calculated using random censoring. Figure 4 shows the RMS values of these residuals for all subjects. The error term is uniformly decreased by the censoring procedure but unchanged by the random censoring procedure (Supporting Information Fig. S2), consistent with the removal of noise from data.

Second-Level Analysis: Motion Censoring Increases Statistical Effects

We next examined how censoring affects statistical power. ANOVAs were performed on uncensored and censored data to identify voxels with significant time courses across all conditions (a main effect of time), a single Pseudoword condition, and in a within-subject contrast of error versus correct responses. Statistical maps (z-scores from ANOVA over time) from these analyses are shown in Figure 5. Censoring produced clear increases in z-scores in each analysis. A modest threshold has been applied to the images to ease visualization. To quantify the statistical improvements produced by censoring, the z-scores of mean map ROIs before and after censoring were compared. In all cases, z-scores were significantly increased by censoring: all conditions ($\Delta z = 0.79$, $t[47] = 8.66$], $P = 2.7E-11$), single condition ($\Delta z = 0.59$, $t[45] = 5.69$, $P = 9.2E-7$), within-subject contrast of error versus correct trials ($\Delta z = 0.78$, $t[33] = 7.31$, $P = 2.18E-8$).

As mentioned previously, motion artifacts are known to cause greater spurious activation in particular locations of the brain. One such location is the frontal pole [Oakes et al., 2005; Wu et al., 1997]. In the single condition (Figure 5, middle), the uncensored slice at $z = 20$ displays a narrow strip of activation around the anterior cortical surface. This activation at the frontal pole is reduced upon

censoring, consistent with a reduction in spurious activation produced by motion artifact.

Exploring the Parameter Space of Temporal Mask Generation

The censoring settings used thus far were chosen to demonstrate the effects of removing a modest amount of motion-contaminated data. To explore how sensitive the censoring procedure is to choices of threshold and augmentation, other settings were tested using uncensored ROIs derived from an uncensored main effect of time ANOVA operating on all correct conditions.

Motion censoring was performed with a range of FD thresholds, from a lenient threshold of 1.3 mm down to a strict threshold of 0.3 mm (for reference, FD values in very still subjects range from 0 to 0.2 mm). At the threshold of FD $>$ 0.3 mm censoring reduced qualifying data to one or zero events for particular time points in particular events and *z*-scores decreased dramatically because too much data were discarded (data not shown). Figure 6 shows the results of the FD $>$ 0.5–1.3 mm analyses. Censoring increased *z*-scores at all thresholds examined, and the greatest increase was seen at FD $>$ 0.9 mm.

Various augmentations of the temporal mask were also tested. As spin histories are disrupted for many seconds by head motion [Friston et al., 1996], removing additional volumes after head movement may further improve data quality. Additionally, as realignment estimates integrate information from a full volume acquisition, it is uncertain exactly when movement occurred, and it may be

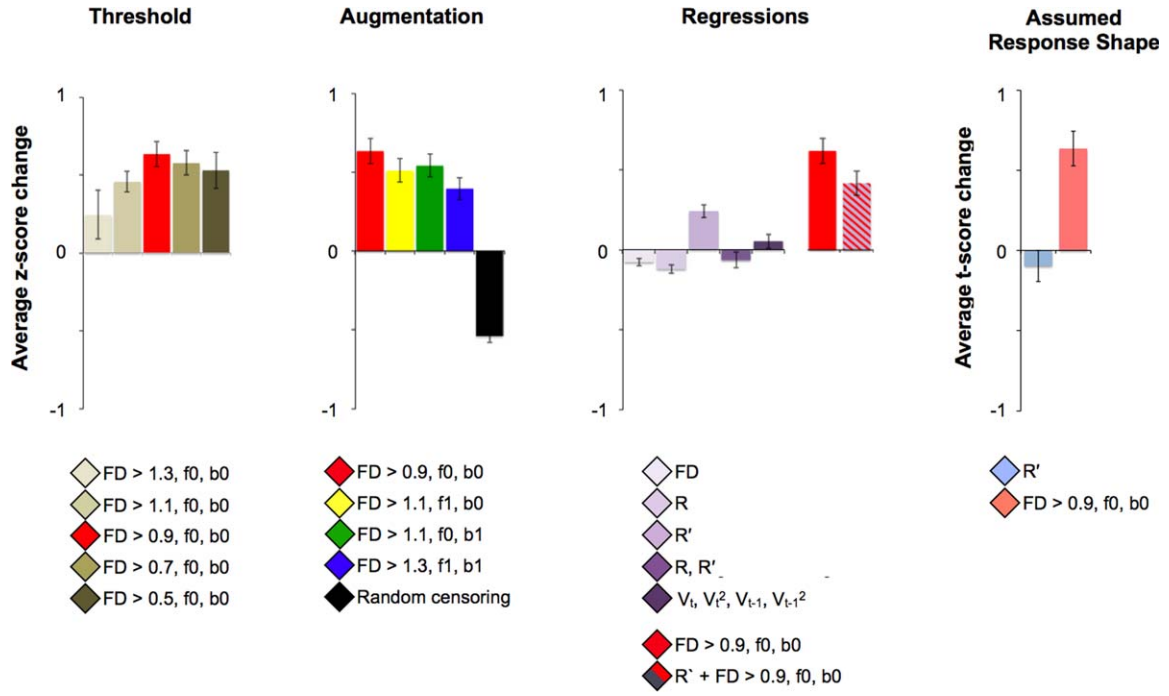

Figure 6.

Motion censoring outperforms regressions over a wide parameter space. For all comparisons, the same set of uncensored main effect of time ROIs is used. At left, the changes in *z*-scores produced by various threshold criteria (removing 9, 12, 16, 23, and 35% of the data, respectively). All thresholds increase *z*-scores, and the greatest increases are seen for FD $>$ 0.9 mm. At center, temporal mask augmentations are compared. FD $>$ 0.9 mm is the base comparator, and thresholds are relaxed to allow forward and backward augmentations that remove similar amounts of data. All approaches are effective, but none surpass the FD $>$ 0.9 mm results. At right, various regressors of no interest were included in GLM design (without censoring). FD denotes framewise displacement, *R* denotes detrended realignment estimates, and *V* denotes realignment estimates. The derivatives of detrended realignment estimates and the 24-parameter expansion of realignment estimates improve *z*-scores, but to a lesser extent than most censoring approaches. To the right, combined application of the best censoring threshold (FD $>$ 0.9) and the best regression (*R′*) results improves *z*-scores, but to a lesser extent than censoring alone. Far right, GLMs were generated using a double gamma assumed response function and the same set of uncensored main effect of time ROIs were used to compare the best regression (*R′*), and the best censoring threshold (FD $>$ 0.9). Error bars represent standard error of the mean change in *z*-scores between a given condition and the baseline "uncensored" condition.

advantageous to remove the volume immediately preceding motion. With these two considerations in mind, five censoring settings that removed equivalent amounts of data in different ways were tested: (1) FD > 0.9 mm only (f0, b0), (2) FD > 1.1 mm and 1 frame after flagged volumes (f1, b0), (3) FD > 1.1 mm and 1 volume before flagged volumes (f0, b1), (4) FD > 1.3 mm and 1 volume before and after flagged volumes (f1, b1), and (5) volumes removed at random in identically sized portions of data as the FD > 0.9 (f0, b0) temporal mask within each subject. As shown in Figure 6, all four censoring settings caused z-scores to increase, and the random censoring caused z-scores to drop, as expected. In a one-factor ANOVA, the FD > 0.9, (f0, b0) mask produced significantly higher z-scores than all of the other masks except FD > 1.1 mm (f0,b1) which was not significantly different. On the basis of these results, we do not recommend removing volumes proceeding or following high-motion volumes, and we do not do so in the other data sets used in this article.

Motion Censoring Outperforms Several Motion Regressions

Motion censoring was compared to a variety of motion regression techniques. Five GLMs were created, each incorporating a different combination of motion estimates as nuisance regressors: (1) the single measure of FD, (2) the six detrended rigid body realignment parameters, (3) the temporal derivatives of the six detrended rigid body realignment parameters, (4) a combination of the detrended realignment parameters and their derivatives (12 regressors), and (5) a 24-parameter Volterra expansion of the realignment parameters for each time point and the previous time point [Friston et al., 1996]. As shown in Figure 6, two of the regressions (derivatives of detrended realignment parameters and the 24-parameter expansion) increased z-scores, but the other regressions decreased z-scores.

Motion censoring generally outperformed motion regressions. Censoring at FD > 0.9 mm performed significantly better than the best regression ($t[49] = 4.51$, $P = 0.0073$). To see whether a combination of censoring and regression might most benefit the data, a GLM was created using the default censoring settings and regressions of the derivatives of realignment estimates. The changes in z-score produced by this GLM were not significantly different from censoring alone ($t[49] = 7.50$, $P = 0.34$).

Motion Censoring is Beneficial When Using Assumed Response Shapes

Finally, we tested the effects of regression (R') and censoring (FD > 0.9) when modeling the data with an assumed hemodynamic response function. The average t-score without censoring (of the top 50 ROIs from the uncensored data) was 6.16. Regression did not significantly improve t-scores ($\Delta t = -0.09$, $P = 0.83$) and censoring significantly raised t-scores ($\Delta t = 0.64$, $P = 3.4E-7$).

Similar Effects are Seen Within Additional Data Sets

Procedures similar to those just described were applied to a separate study conducted in an adolescent cohort with Tourette syndrome. In this cohort, cue and target were modeled separately (complex trials) using some cue-only partial trials [Ollinger et al., 2001].

Figure 7 shows statistical maps and mean map ROI z-scores for uncensored and censored data for ANOVAs operating on all correct cue and target conditions using the censoring settings of FD > 0.9 mm (f0, b0). z-Scores for mean map ROIs consistently increase in both cases with censoring (Cue: $\Delta z = 0.81$, $t[47] = 7.09$, $P = 6.1E-9$. Target: $\Delta z = 0.70$, $t[48] = 8.45$, $P = 4.7E-11$). This suggests that motion censoring did not interfere adversely with complex trial modeling.

An additional study conducted in healthy adults was examined to compare censoring and regression in a different population. These subjects moved much less than the other cohorts (Table I), enabling stricter censoring settings. Settings of FD > 0.5 mm, f0, and f0 were used to censor this data, removing 2% of the data. To see whether regression alone might be just as good as censoring in this low-movement population, we calculated a GLM using R' as a coregressor of no interest. To compare between options, we used z-scores from top 50 uncensored ROIs. As shown in Figure 8, even this small change caused uncensored ROI z-scores for all correct trials to significantly increase ($\Delta z = 0.15$, $t[49] = 6.03$, $P = 1.7E-8$). Including regressors of no interest in the uncensored GLM produced a decrease in z-scores ($\Delta z = -0.20$, $t[49] = -6.73$, $P = 2.1E-7$).

DISCUSSION

This report compares two methods to counter effects of motion in task fMRI data: (1) motion censoring and (2) motion regression. We first explored the effects of motion censoring on a healthy pediatric data set with relatively high-movement estimates. As measured by variance in time course estimates across subjects, GLM residuals, and the statistical significance of activation, censoring improves modeling of the BOLD signal in nearly every region examined (and decreased activation in frontal pole regions known to exhibit artifactual motion-induced activity). We compared a range of parameters for motion regression and motion censoring and found that motion censoring produces sizeable increases in z-scores across most choices of parameters (the exception, predictably, is that z-scores decrease when too much data have been removed from a data set). Next, we compared z-scores from standard processing (realignment only) to those found after motion regression or motion censoring. Results indicated that motion censoring performs significantly better

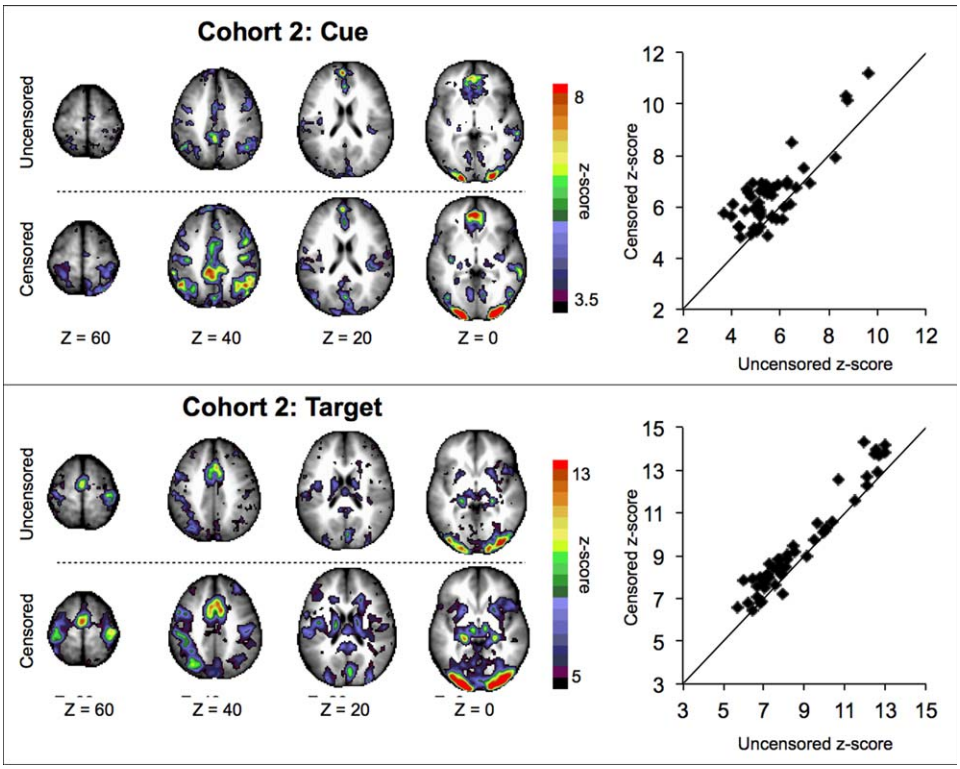

Figure 7.

Motion censoring improves z-scores in a clinical data set. Statistical images and mean map ROI z-scores are shown from ANOVA main effect of time in cue conditions and ANOVA main effect of time in target conditions before and after censoring in Cohort 2.

than a variety of motion regressions. Similar advantages for censoring over regression were observed when data were modeled with an assumed hemodynamic response shape. We then examined the impact of motion censoring in two additional data sets: a pediatric Tourette cohort and a healthy adult cohort (Table I). The Tourette syndrome group showed improvements similar to those of the healthy children. The healthy adults showed smaller, but still significant, increases in statistical significance of task-evoked activation and no improvement with motion regression.

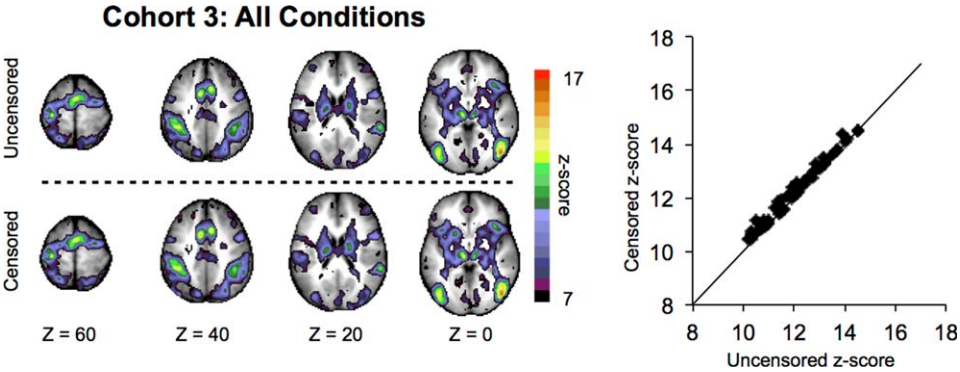

Figure 8.

Motion censoring produces modest improvements in adults who move little. Statistical images and mean map ROI z-scores are shown for an ANOVA main effect of time across all correct conditions before and after censoring in Cohort 3.

These data range in age, clinical status, amount of motion, study design, and task. They are, however, all from the same site, acquired on the same scanner, and use similar pulse sequences. It is possible that our findings may not generalize beyond these populations (e.g., to the elderly) or acquisition parameters (e.g., other flip angles), and that faster TRs and other advances in methodology will diminish the utility of the types of corrections used in this report. However, given the growing number of reports on incomplete removal of motion-related artifact (using various scanners, pulse sequences, denoising strategies, etc.) [Bright and Murphy, 2013; Satterthwaite et al., 2013; Van Dijk et al., 2012], these findings are likely relevant for many groups and existing data sets. Our findings may also be less pronounced when analyses are performed using other strategies, such as using other assumed response shapes or when using group level analyses that take into account the error variance of lower-level estimates (e.g., FLAME in FSL).

A natural issue that arises is the extent to which data should be censored. We have no definitive solution to this important issue. Our experiences with RSFC indicate that any and all head motion produces artifactual changes in BOLD signal [Power et al., 2012], suggesting that optimal task fMRI results would be obtained with stringent motion censoring criteria that rigorously exclude volumes during which even modest movements occurred. As shown in Figure 1, a floor does exist for FD values (which rarely exceed 0.2 mm in still subjects), and stringent thresholds could be set just above this floor. Such a threshold could conceivably identify and remove almost all data contaminated with motion related effects. However, as the removal of noise by censoring entails a reduction in data available for analysis, the impact of censoring is entwined with study design and subject movement. As censoring settings are made more stringent, fewer trials remain, and the accuracy of GLM estimation diminishes. Figure 6 shows this pattern quite clearly: z-scores improve as FD thresholds eliminate increasing amount of motion-contaminated data from thresholds of FD $>$ 1.3 mm down to FD $>$ 0.9 mm, but below FD $>$ 0.9 mm improvements diminish until z-scores frankly decrease at thresholds of FD $>$ 0.3 mm. In cohorts with substantial motion, ideal levels of censoring are probably unobtainable. Nevertheless, even modest amounts of censoring can produce notable improvements in data quality.

Hence, we are unable to offer blanket prescriptions for censoring settings. We have demonstrated methods to select and test ROIs for improvements under a variety of motion censoring thresholds. Within our laboratory, we currently explore several FD thresholds, seeking the inflection point at which statistical improvements become overshadowed by the deleterious effects of losing data (e.g., between FD $<$ 0.9 and FD $<$ 0.7 for Cohort 1, Fig. 6). The position of this inflection point varies by cohort and is dependent on study design, subject motion, and so forth, leaving us unable to make universal recommendations.

A related issue is that in data sets limited by data quantity (e.g., numbers of events within a particular condition), gentler motion correction tools might be preferable to censoring techniques. This concern is certainly reasonable, but even in the most limited condition considered in this article (the Single Condition of Fig. 5 had $\sim$25 events per subject), removal of 11% of the data produced considerable improvement in the reliability of estimates.

In general, motion censoring appears to improve data quality more than many motion regression approaches. Only regression of derivatives of realignment estimates and the 24-parameter expansion produced improvements in data quality, and these improvements were significantly less than those produced by almost any version of censoring. It is possible that voxel-specific regressors or regressors built using more elaborate methods may perform better than the brain-wide regressors used here (though, see Satterthwaite et al., 2013). Previous studies have found benefit to regression [Morgan et al., 2007; Oakes et al., 2005], and we are unable to account for their relative lack of benefit in this study other than to note that there are, of course, many differences in acquisition parameters, data processing, tools, and so forth, between such studies and this study. For example, we perform a single atlas transformation step in preprocessing and then compute a GLM with regressors on data in atlas space, whereas other studies compute a GLM with regressors in native space and then transform to atlas space. It is difficult to say if this would significantly alter the effect of motion regressors.

However, though motion regression can be beneficial [Morgan et al., 2007; Oakes et al., 2005], there are also drawbacks to regression. If motion is correlated with behavioral condition in the data, motion regressors can cause task-related activity to be modeled as an effect of motion, resulting in underestimation of the effect of task [Bullmore et al., 1999]. This concern is particularly pertinent in block design experiments in which motion often does correlate with condition [Johnstone et al., 2006] (though, see Birn et al., 2004 for a discussion of optimizing study design in the presence of motion). As motion censoring avoids such problems and produces substantial improvements in data quality, we see little reason to regress rather than censor fMRI data.

Investigators planning to use censoring techniques may wish to alter study design in several ways. In populations that tend to move, investigators may overpower study design to accommodate the loss of some data. Modest censoring in our pediatric and adolescent cohorts removed 15–20% of the data, and caused a small number of subjects to be excluded from analysis because sufficient trials no longer remained in particular conditions. Simpler designs with increased numbers of trials will best tolerate strict censoring settings. Another modification of study design that may further improve data quality would be the use of optical recording techniques to measure subject motion [Dold et al., 2006]. Current motion estimates are derived

from MRI acquisitions spaced by 2–3 s, and movements occurring under the Nyquist limit could still impact data quality but pass undetected. Optical motion measurements have finer temporal and spatial resolution and should be capable of forming very precise temporal masks.

CONCLUSIONS

This article presents a simple way to reduce the effects of subject motion in task fMRI data. This method reduces between-subject variance in parameter estimates, reduces the error term in GLM calculations, and boosts statistical power in several data sets. The method is ad hoc but effective and can already be implemented in a variety of analysis platforms such as AFNI. In populations that tend to move, such as pediatric or clinical populations, motion censoring can substantially increase the quality, sensitivity, and accuracy of fMRI studies.

ACKNOWLEDGMENT

The authors report no conflicts of interest.

REFERENCES

Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC, Cohen JD (1999): Overt verbal responding during fMRI scanning: Empirical investigations of problems and potential solutions. NeuroImage 10:642–657.

Birn RM, Bandettini PA, Cox RW, Jesmanowicz A, Shaker R (1998): Magnetic field changes in the human brain due to swallowing or speaking. Magn Reson Med 40:55–60.

Birn RM, Bandettini PA, Cox RW, Shaker R (1999): Event-related fMRI of tasks involving brief motion. Hum Brain Mapp 7:106–114.

Birn RM, Cox RW, Bandettini PA (2004): Experimental designs and processing strategies for fMRI studies involving overt verbal responses. NeuroImage 23:1046–1058.

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

Brown TT, Lugar HM, Coalson RS, Miezin FM, Petersen SE, Schlaggar BL (2005): Developmental changes in human cerebral functional organization for word generation. Cereb Cortex 15:275–290.

Bullmore ET, Brammer MJ, Rabe-Hesketh S, Curtis VA, Morris RG, Williams SC, Sharma T, McGuire PK (1999): Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. Hum Brain Mapp 7:38–48.

Diedrichsen J, Shadmehr R (2005): Detecting and adjusting for artifacts in fMRI time series data. NeuroImage 27:624–634.

Dold C, Zaitsev M, Speck O, Firle EA, Hennig J, Sakas G (2006): Advantages and limitations of prospective head motion compensation for MRI using an optical motion tracking device. Acad Radiol 13:1093–1103.

Field AS, Yen YF, Burdette JH, Elster AD (2000): False cerebral activation on BOLD functional MR images: Study of low-amplitude motion weakly correlated to stimulus. Am J Neuroradiol 21:1388–1396.

Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R (1996): Movement-related effects in fMRI time-series. Magn Reson Med 35:346–355.

Gopinath K, Crosson B, McGregor K, Peck K, Chang Y-L, Moore A, Sherod M, Cavanagh C, Wabnitz A, Wierenga C, White K, Cheshkov S, Krishnamurthy V, Briggs RW (2009): Selective detrending method for reducing task-correlated motion artifact during speech in event-related FMRI. Hum Brain Mapp 30:1105–1119.

Hajnal JV, Myers R, Oatridge A, Schwieso JE, Young IR, Bydder GM (1994): Artifacts due to stimulus correlated motion in functional imaging of the brain. Magn Reson Med 31:283–291.

Huang J, Francis AP, Carr TH (2008): Studying overt word reading and speech production with event-related fMRI: A method for detecting, assessing, and correcting articulation-induced signal changes and for measuring onset time and duration of articulation. Brain Lang 104:10–23.

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002): Image distortion correction in fMRI: A quantitative evaluation. NeuroImage 16:217–240.

Jiang A, Kennedy DN, Baker JR, Weisskoff RM, Tootell RBH, Woods RP, Benson RR, Kwong KK, Brady TJ, Rosen BR, Belliveau JW (1995): Motion detection and correction in functional MR imaging. Hum Brain Mapp 3:224–235.

Johnstone T, Ores Walsh KS, Greischar LL, Alexander AL, Fox AS, Davidson RJ, Oakes TR (2006): Motion correction and the use of motion covariates in multiple-subject fMRI analysis. Hum Brain Mapp 27:779–788.

Jones TB, Bandettini PA, Birn RM (2008): Integration of motion correction and physiological noise regression in fMRI. NeuroImage 42:582–590.

Kennedy DP, Courchesne E (2008): The intrinsic functional organization of the brain is altered in autism. NeuroImage 39:1877–1885.

Kirwan CB, Shrager Y, Squire LR (2009): Medial temporal lobe activity can distinguish between old and new stimuli independently of overt behavioral choice. Proc Natl Acad Sci USA 106:14617–14621.

Lancaster JL, Glass TG, Lankipalli BR, Downs H, Mayberg H, Fox PT (1995): A modality-independent approach to spatial normalization of tomographic images of the human brain. Hum Brain Mapp 3:209–223.

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

Lund TE, Nørgaard MD, Rostrup E, Rowe JB, Paulson OB (2005): Motion or activity: Their role in intra- and inter-subject variation in fMRI. NeuroImage 26:960–964.

Miezin FM, Maccotta L, Ollinger JM, Petersen SE, Buckner RL (2000): Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. NeuroImage 11:735–759.

Morgan VL, Dawant BM, Li Y, Pickens DR (2007): Comparison of fMRI statistical software packages and strategies for analysis of images containing random and stimulus-correlated motion. Comp Med Imaging Graph 31:436–446.

Oakes TR, Johnstone T, Ores Walsh KS, Greischar LL, Alexander AL, Fox AS, Davidson RJ (2005): Comparison of fMRI motion correction software tools. NeuroImage 28:529–543.

Ollinger JM, Shulman GL, Corbetta M (2001): Separating processes within a trial in event-related functional MRI. NeuroImage 13:210–217.

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

Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, Hakonarson H, Gur RC, Gur RE, Wolf DH (2013): 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.

Schlaggar BL, Brown TT, Lugar HM, Visscher KM, Miezin FM, Petersen SE (2002): Functional neuroanatomical differences between adults and school-age children in the processing of single words. Science 296:1476–1479.

Stark CEL, Okado Y, Loftus EF (2010): Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms. Learn Mem 17:485–488.

Van Dijk KRA, Sabuncu MR, Buckner RL (2012): The influence of head motion on intrinsic functional connectivity MRI. NeuroImage 59:431–438.

Wu DH, Lewin JS, Duerk JL (1997): Inadequacy of motion correction algorithms in functional MRI: Role of susceptibility-induced artifacts. J Magn Reson Imaging 7:365–370.

Yetkin FZ, Haughton VM, Cox RW, Hyde J, Birn RM, Wong EC, Prost R (1996): Effect of motion outside the field of view on functional MR. Am J Neuroradiol 17:1005–1009.