◆ Human Brain Mapping 7:38–48(1999) ◆

Methods for Diagnosis and Treatment of Stimulus-Correlated Motion in Generic Brain Activation Studies Using fMRI

E.T. Bullmore,[1,2]* M.J. Brammer,[1] S. Rabe-Hesketh,[1] V.A. Curtis,[2] R.G. Morris,[3] S.C.R. Williams,[4] T. Sharma,[2] and P.K. McGuire[2]

Departments of Biostatistics & Computing,[1] Psychological Medicine,[2] Psychology[3] and Neurology,[4] Institute of Psychiatry, London, UK

Abstract: Movement-related effects in realigned fMRI timeseries can be corrected by regression on linear functions of estimated positional displacements of an individual subject's head during image acquisition. However, this entails biased (under)estimation of the experimental effect whenever subject motion is not independent of the experimental input function. Methods for diagnosing such stimulus-correlated motion (SCM) are illustrated by application to fMRI data acquired from 5 schizophrenics and 5 normal controls during periodic performance of a verbal fluency task. The schizophrenic group data were more severely affected by SCM than the control group data. Analysis of covariance (ANCOVA) was used, with a voxelwise measure of SCM as a covariate, to estimate between-group differences in power of periodic signal change while controlling for variability in SCM across groups. Failure to control for SCM in this way substantially exaggerated the number of voxels, apparently demonstrating a between-group difference in task response. *Hum. Brain Mapping* 7:38–48, 1999.

Key words: fMRI; movement correction; analysis of covariance; generic brain activation mapping

INTRODUCTION

Movement of the subject's head is a fundamental problem in functional magnetic resonance imaging studies of the brain [Hajnal et al., 1994; Friston et al., 1996]. The problem arises because the magnetic resonance signal is largely a function of two physical quantities, both of which may be locally variable, or spatially inhomogeneous: 1) the static magnetic field B_0, used to effect net longitudinal magnetisation of protons; and 2) the field of radio frequency (RF) electromagnetic radiation B_1, used to effect transient transverse magnetisation. It is possible to minimize B_0 field inhomogeneities by "shimming" prior to image data acquisition, but multislice data acquisition by slice-selective RF irradiation necessarily implies dramatic inhomogeneities in B_1 at the superior and inferior edges of an irradiated slice in the z direction. It follows that minimal motion of the subject's head between acquisition of an image at time t and acquisition of an image at time t + TR, where TR is the repeat time, may mean that the same volume of tissue is subject to very different static magnetic and / or electromagnetic fields on consecutive acquisitions.

Contract grant sponsor: Wellcome Trust; Contract grant sponsor: Bethlem Maudsley Research Fund.

*Correspondence to: E.T. Bullmore, Departments of Biostatistics & Computing and Psychological Medicine, Institute of Psychiatry, DeCrespigny Park, London SE5 8AF, UK.
E-mail: e.bullmore@iop.bpmf.ac.uk

Received for publication 1 December 1997; accepted 19 June 1998

A customary first step in dealing with this problem is to estimate post hoc the nature and extent of any motion in three dimensions {x, y, z} that might have occurred over the course of an fMRI experiment, typically lasting several minutes. Once the three-dimensional (3D) displacements of the "match" image volumes acquired at times {t} = {1TR, 2TR, 3TR, ..., pTR} are known relative to a "base" image volume, each match volume can be *realigned* or coregistered with the base volume. This step alone ensures that an individual time series of MR signal intensity values, at a given voxel of the realigned image, is consistently representative of a given volume of brain tissue. However, as Friston et al. [1996] were among the first to point out, realignment alone is insufficient to "undo" the movement-related effects on fMRI time series caused by having imaged a given brain volume inconsistently with respect to B_0 and/or B_1 field strength(s). They argued that these movement-related effects could be conceived as a linear (second-order polynomial) function of 3D positional displacement at times t and t−TR; and, as such, they could be removed by fitting an appropriate (dynamic) regression model to each realigned time series and considering the residuals. Note that this approach assumes that variance in the realigned time series can be partitioned into two orthogonal or independent components: 1) variance due to motion (of no interest); and 2) variance due to all other sources, including signal changes (of interest) induced by contrasting conditions in an activation study.

Whether or not this second step, i.e., regression after realignment, provides a final solution to the problem of subject motion will clearly depend on the extent to which the orthogonality assumption is justified; in our experience, this varies considerably from one image (or voxel) to another.

In a "good" image, the time series of estimated translations and rotations in {x, y, z} will indeed be independent of signal changes induced by experimental design. If the experimental input function consists of periodic alternation between two contrasting conditions, A and B, we can quantify this independence by saying there is more-or-less zero power in any of the movement parameter time series at the frequency of the input function. In this case, the effects of movement-correcting regression on our subsequent ability to detect signal changes of interest in the residual time series will be entirely beneficial. By removing movement-related variance, we will typically have reduced substantially the variance in an fMRI time series that is *not* experimentally determined; and this will enhance the sensitivity of any test for activated voxels that is based on a standardized statistic.

To put this more formally, in general we define the standardized test statistic for activation as $T = P/E$, where P is some estimate of the experimental effect and E is its standard error, proportional to the error variance of the fMRI time series. $T^{pre} = P^{pre}/E^{pre}$ is then the test statistic estimated after realignment but before (pre) movement-correcting regression; and $T^{post} = P^{post}/E^{post}$ is the test statistic estimated after (post) movement-correcting regression. If the orthogonality assumption is absolutely justified, $P^{post} = P^{pre}$; but E^{post} will almost certainly be reduced relative to E^{pre} by removal of (uncorrelated) movement effects that would otherwise contribute to error variance. The net effect is that T^{post} will be greater than T^{pre}, so that movement-correcting regression in a "good" voxel or image will lead to *decreased* probability of false-negative (type II) error in activation mapping.

In a "bad" image (or voxel), on the other hand, movement-related effects will {*not*} be independent of experimentally induced signal changes. Motion, like the signal changes of interest, will be correlated with the input function of periodically alternating experimental conditions. We can quantify such stimulus-correlated motion in terms of substantial power in one or more of the movement parameter time series at the frequency of the input function. In this case, the use of regression to correct movement-related effects is problematic. P^{post} will be reduced relative to P^{pre}, and if this attenuation of the estimated effect is proportionally greater than the reduction in error variance (i.e., if $P^{pre}/P^{post} > E^{pre}/E^{post}$), then there will be a net reduction in the test statistic ($T^{post} < T^{pre}$), and consequently an *increased* probability of type II error in activation mapping. In short, when the orthogonality assumption is not justified, we risk "throwing out the baby with the bathwater" [Friston et al., 1996].

There are a number of possible responses to this problem, assuming that all reasonable steps to immobilize the subject's head in the scanner are routinely taken. One could simply avoid experimental designs that were likely to cause stimulus-correlated motion. However, this would probably rule out any experiment in which the subject was asked to provide an observable and "online" response by intentional movement, including speech. Alternatively, one could test images complicated by stimulus-correlated motion for brain activation at a more lenient threshold than usual, thus effectively reducing the type II error rate at the expense of increasing the type I (false-positive) error rate. Clearly, neither of these options is satisfactory;

perhaps particularly if one is interested in using fMRI to explore differences in cognitive activation between two groups of subjects, i.e., healthy volunteers and patients, that might differ systematically in terms of stimulus-correlated motion. For such clinical studies, we argue that it is essential at least to exclude the possibility of a group difference in stimulus-correlated motion before entertaining any other interpretation of a group difference in power of response.

The primary objective of this paper is to describe some simple methods that can be used to deal with this problem of comparing *patterns of brain activation between two groups that may differ systematically in terms of stimulus-correlated motion*. The issues are illustrated in the context of fMRI data acquired during performance of a verbal fluency task from a group of 5 healthy volunteers, and a group of 5 patients with schizophrenia, as described below. There follows a brief review of existing methods for movement estimation and correction in individual fMRI data sets, and of methods for brain activation mapping. However, we note that the problem posed by stimulus-correlated motion for comparative analysis of *grouped* or *generic* data is broadly independent of the precise algorithm used for realignment, or the exact specification of any regression model used to correct movement-related variance, in an *individual* data set. As long as these methods of individual image processing are good enough to identify and remove (at least some) signal variance induced by motion, then there will potentially be a problem in later generic analysis of these data. Later we describe methods for diagnosing stimulus-correlated motion (SCM) at the levels of a whole image and a single voxel, and show that these schizophrenic group data were more severely affected by SCM than the control group data. Finally, we present an analysis of covariance (ANCOVA) model to estimate between-group differences in amplitude of response after correcting for individual differences in stimulus-correlated motion at each voxel in a standard space, and demonstrate how this approach can reduce the confounding effect of SCM on between-group differences in task response.

DATA

Subjects

Two groups of subjects were scanned:

1. Five healthy male volunteers recruited from hospital staff, mean age 31.8 years (range 28–37 years), mean IQ 119.6.
2. Five male patients with schizophrenia diagnosed by DSM-IV criteria [American Psychiatric Association, 1987], mean age 29.6 years (range 25–45 years), mean IQ 116.8. All patients were clinically stable and had been on regular antipsychotic medication for at least 1 year.

There were no significant differences in age or IQ between groups.

Images

Two echoplanar imaging (EPI) data sets were acquired from each subject, in the same scanning session, using a GE Signa 1.5 T system (General Electric, Milwaukee, WI) retrofitted with Advanced NMR hardware and software (ANMR, Woburn, MA) at Maudsley Hospital, London. A quadrature birdcage coil was used for RF transmission and reception. Head movement was limited by foam padding within the head coil. The intercommissural (AC-PC) line was identified in conventional MR images (coronal and sagittal planes) prior to EPI data acquisition.

Functional

One hundred single-shot T_2*-weighted gradient echo echoplanar images depicting BOLD contrast [Ogawa et al., 1990] were acquired at each of 14 near-axial noncontiguous planes parallel to the AC-PC line: TE 40 msec, TR 3 sec, in-plane resolution 3 mm, slice thickness 7 mm, slice skip 0.7 mm, number of signal averages = 1, flip angle = 90°. Single-shot here (and below) means that the full k space matrix was sampled following a single RF excitation.

Structural

A higher-resolution, single-shot inversion recovery echoplanar image of the whole brain was acquired at 43 near-axial, noncontiguous planes parallel to the AC-PC line: TE 74 msec, TI 180 msec, TR 16 sec, in-plane resolution 1.5 mm, slice thickness 3 mm, slice skip 0.3 mm, number of signal averages = 8.

The structural EPI data set was later used to register the functional EPI data set in standard space [Brammer et al., 1997].

Experimental design

Two tasks were repeatedly presented in the order ABAB...; each task was presented for 30 sec, and the AB cycle was repeated 5 times over the course of a 5-min experiment.

Task A: Subjects heard the word "rest" spoken once every 1.5 sec and were asked covertly to repeat it to themselves (without speaking).

Task B: Subjects were aurally cued by a letter to generate a word beginning with that letter once every 1.5 sec. They were asked covertly to say the generated word to themselves.

This kind of design was previously used to activate a network of brain regions involved in covert verbal fluency [Grasby et al., 1995; Warburton et al., 1996].

IMAGE ANALYSIS

Movement estimation and correction

The nature and extent of rigid body motion in three spatial dimensions were estimated for each individual fMRI data set in the following way:

1. A "base" image volume of mean signal intensity over time was created by averaging (and then collating) the 100 images acquired in each plane.
2. The sum of absolute differences in gray-scale values between the "match" image volume acquired at each time point and the base image volume was computed.
3. A multidimensional search by the Fletcher-Davidon-Powell algorithm [Press et al., 1992] was used to find the translations and rotations in three dimensions $\{x, y, z\}$ which minimized the total absolute difference between each match volume and the base volume.
4. The match volumes were realigned relative to the base volume by tricubic spline interpolation. We note that theoretically the optimal method of realignment for band-limited imaging data is sinc interpolation [Hill et al., 1994]. However, in our experience, the incremental improvement in image realignment attained by sinc interpolation does not substantively affect the results of subsequent brain activation mapping, and demands considerably greater computational effort than tricubic spline interpolation. For these reasons, we prefer tricubic spline interpolation.

Following estimation of the six parameters describing rigid body motion at each time point (step 3), movement of any voxel in the match volume relative to the base volume can be represented in terms of positional displacements in x, y, and z. The realigned time series at the ith voxel, S_i or $\{S_t\}_i$, t = 1, 2, 3, ..., 100 images, can be regressed on a second-order polynomial function of these positional displacements $\{\delta x_t, \delta y_t, \delta z_t\}$; and positional displacements at that voxel in the previous image $\{\delta x_{t-1}, \delta y_{t-1}, \delta z_{t-1}\}$:

$$\begin{aligned} S_t = {} & \beta_1\delta x_t + \beta_2(\delta x_t)^2 + \beta_3\delta y_t + \beta_4(\delta y_t)^2 + \beta_5\delta z_t \\ & + \beta_6(\delta z_t)^2 + \beta_7\delta x_{t-1} + \beta_8(\delta x_{t-1})^2 + \beta_9\delta y_{t-1} \\ & + \beta_{10}(\delta y_{t-1})^2 + \beta_{11}\delta z_{t-1} \\ & + \beta_{12}(\delta z_{t-1})^2 + \beta_0 + Y_t. \end{aligned} \quad (1)$$

After fitting this dynamic model by ordinary least squares (OLS), the residual time series $\{Y_t\}$ are uncorrelated with concomitant and lagged positional displacement in 3D at the cost of 1 degree of freedom in the residual time series.

By the rule of thumb that time series models should not generally specify more than one parameter to be estimated per 10 data points, this model for movement effects as a second-order polynomial function of lagged and concomitant positional displacement in three dimensions is slightly overparameterized (there are 13 parameters to be estimated). However, it is parsimonious compared to the 24-parameter model that must be fitted to account for movement effects by an equivalent function of 3 translations and 3 rotations [Friston et al., 1996].

GENERIC BRAIN ACTIVATION MAPPING

Estimation

The motion-corrected time series $\{Y_t\}$, t = 1, 2, 3, ..., 99 images, observed at each voxel in each image, was modelled by:

$$\begin{aligned} Y_t = {} & \gamma \sin(\omega t) + \delta \cos(\omega t) \\ & + \gamma' \sin(2\omega t) + \delta' \cos(2\omega t) \\ & + \gamma'' \sin(3\omega t) + \delta'' \cos(3\omega t) + \alpha + \beta t + \rho_t. \end{aligned} \quad (2)$$

Here ω is the (fundamental) frequency of alternation between A and B conditions = $2\pi/20$ radians per image in these data; $\alpha + \beta t$ is a linear trend; and ρ_t is the residual error at time point t. As is usually the case in time series regression, the errors are autocorrelated and model fitting by OLS would underestimate the parameter standard errors [Ostrum, 1978]. We

therefore fit the model by pseudogeneralized least squares (PGLS), modelling the OLS residual series $\{\rho_t\}$ as a first-order autoregressive process [Bullmore et al., 1996].

The power at fundamental frequency (fundamental power, FP) in each time series was estimated by the squared amplitude of an arbitrarily phase-shifted sine wave at fundamental frequency $FP = \gamma^2 + \delta^2$. Fundamental power was then divided by its standard error to yield the fundamental power quotient,

$$FPQ = \frac{\gamma^2 + \delta^2}{\sqrt{2(SE(\gamma)^4 + SE(\delta)^4)}}, \quad (3)$$

as a standardized estimator of the experimental effect.

Inference

Parametric maps or statistic images were constructed to represent FPQ at each intracerebral voxel of the observed images. We then randomly permuted the 99 images observed in each anatomical plane 10 times, estimated FPQ exactly as described above for each randomized image series, and so constructed 10 randomized parametric maps of FPQ at each plane of each data set. The observed and randomized FPQ maps generated from each individual data set were coregistered with a template image in the standard space of Talairach and Tournoux [1988] and were smoothed by a Gaussian filter with full width at half maximum (FWHM) = 7 mm. At each plane in standard anatomical space, this yielded 5 observed FPQ maps and 50 randomized FPQ maps for each group of subjects.

At every voxel position where the FPQ maps for each group overlapped sufficiently, we computed the median of 5 observed FPQ values $\text{med}(FPQ_{obs})$, and 10 medians of 5 randomized FPQ values $\text{med}(FPQ_{ran})$. The permutation distribution of $\text{med}(FPQ_{ran})$ sampled in this way over the whole standard space was used to test the null hypothesis that $\text{med}(FPQ_{obs})$ at each voxel was not determined by periodic experimental design. If the observed median exceeded the $(1 - \alpha)$ 100 percentile value of the sampled permutation distribution of $\text{med}(FPQ_{ran})$, then the null hypothesis was refuted at that voxel with one-tailed probability of type I error = α [Edgington, 1980; Manly, 1991]. Voxel locations at which the null hypothesis was refuted were colored on the gray-scale background of the template image to form a generic brain activation map or GBAM [Brammer et al., 1997].

Control and schizophrenic group activation maps

The generic brain activation maps obtained by within-group analyses of the control and schizophrenic group data are shown in Figure 1; for both maps, the voxelwise probability of type I error $\alpha = 4 \times 10^{-4}$. The controls demonstrated activation in the prefrontal and medial frontal cortex, medial and inferior parietal cortex, posterior cingulate cortex, and thalamus. The schizophrenics, in contrast, appeared to activate relatively few voxels in prefrontal and medial parietal cortices and the posterior cingulate gyrus [Curtis et al., 1998].

These apparent between-group differences in activation might suggest a difference in amplitude of task response. However, it is also possible that they reflect systematic differences between groups in terms of stimulus-correlated motion: we will test this hypothesis below.

DIAGNOSIS OF STIMULUS-CORRELATED MOTION

Image level

A simple way to diagnose stimulus-correlated motion is to analyze the time series of 3D rotations and translations estimated for each image in the course of realignment. We fitted a sinusoidal regression model (Eq. 2) to these six time series for each subject, and quantified stimulus-correlated rotations and translations in terms of standardized power (FPQ) at the frequency of alternation between A and B conditions. Table 1 shows the results of this analysis for each of the 10 data sets in the study. It is clear that the data acquired from schizophrenic subjects are, in general, more severely affected by stimulus-correlated motion. In particular, it is notable that schizophrenics tend to demonstrate more stimulus-correlated rotations about the x and y axes (pitch and roll, respectively), which are especially likely to cause artifactual changes in signal intensity by displacing brain tissue relative to the superior and inferior edges of selectively irradiated slices in the z direction.

It is also interesting to compare patterns of movement between groups visually, by inspecting plots of positional displacement vs. time. To obtain a ''group average'' picture of stimulus-correlated rotation, we plotted the time series of rotational displacements for the images which had median values for stimulus-correlated rotation in each group (Fig. 2). The median time series for the schizophrenic group show x and y rotational displacements at the frequency of the input

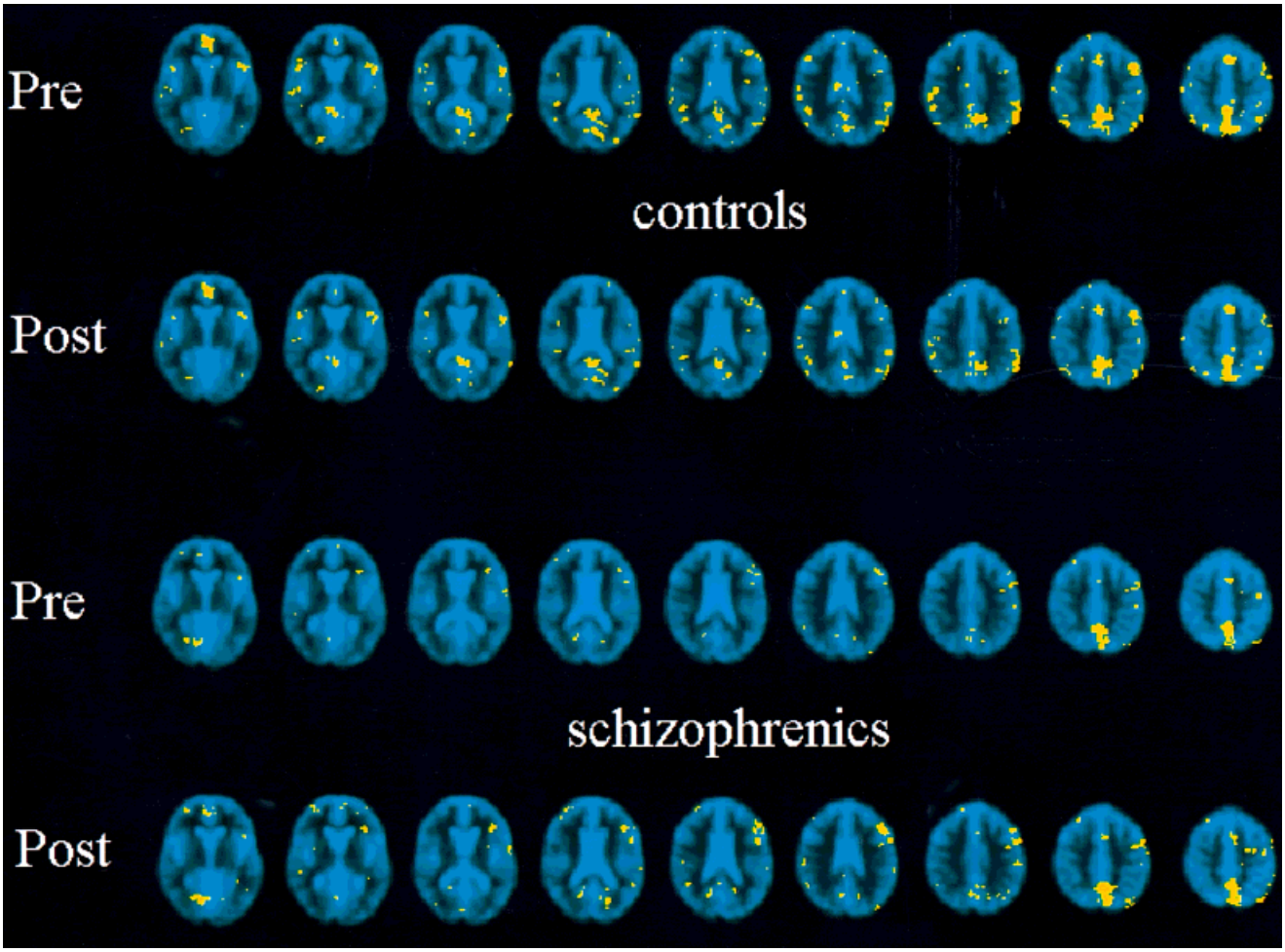

Figure 1.

Generic brain activation maps (GBAMs) for control and schizophrenic data, before and after correction for stimulus-correlated motion. Each row of nine 5.5-mm-thick axial slices represents the brain from +1 mm to +50 mm in standard space; the right side of each slice represents the left side of the brain. From top to bottom: control group GBAM before normalization to zero group mean ΔFPQ; control group GBAM after normalization; schizophrenic group GBAM before normalization; schizophrenic group GBAM after normalization. Note that the effect of normalization is to *increase* the number of generically activated voxels in the schizophrenic group GBAM and to *decrease* the number of generically activated voxels in the control group GBAM.

function, whereas the median time series for the control group are dominated by linear trends.

Voxel level

Analyses of the movement-parameter time series can thus be used to diagnose stimulus-correlated motion at the level of an image, or group of images. However, these approaches are clearly limited:

1. They do not directly address the central question of how our test statistic for activation (FPQ = FP / SE(FP)) is likely to be affected by motion-correcting regression. If there is significant power in the movement-parameter time series at the frequency of AB alternation, one might expect that fundamental power (FP) in the image would be reduced overall by motion-correcting regression; but this fact by itself is not sufficient to predict an overall reduction in FPQ. To make that prediction, we need to know that motion-correcting regression *disproportionately* reduces FP, relative to SE(FP). In other words, we need to know that $FP^{pre} / FP^{post} > SE(FP)^{pre} / SE(FP)^{post}$, yet the movement-parameter time series are not informative about SE(FP) before or after regression.

TABLE I. Power of stimulus-correlated rotations and translations in functional MR images acquired from 5 control and 5 schizophrenic subjects*

Subject	Pitch	Roll	Yaw	X translation	Y translation	Z translation
C1	0.10	0.15	1.07	2.58	3.00	0.90
C2	0.96	0.57	0.15	0.03	1.75	1.82
C3	0.23	0.04	0.51	0.21	1.19	1.15
C4	0.24	0.26	0.42	1.13	2.54	0.96
C5	0.15	2.66	0.54	0.01	1.44	0.53
C medians	0.23	0.26	0.51	0.21	1.44	0.96
S1	0.025	0.47	0.06	0.40	1.47	0.08
S2	0.19	0.18	2.22	2.47	0.02	0.25
S3	2.47	2.38	0.32	10.51	0.55	1.13
S4	2.23	0.41	2.21	3.13	3.42	0.73
S5	1.76	2.92	1.66	2.21	0.37	3.08
S medians	1.76	0.47	1.66	2.47	0.55	0.73

* Given are the standardized power (FPQ) estimates derived by analysis of the six movement parameter time series in each subject. C, control; S, schizophrenic.

2. Even if we could predict the overall effect of motion-correction on FPQ by analyzing the movement-parameter time series, we should not expect this effect to be equivalent at every voxel in the image. For example, in an image complicated by stimulus-correlated pitch, the most extensive displacement through the z plane, and therefore the greatest impact of motion-correction on signal variance, is expected at the anterior and posterior poles of the brain. The effects on signal variance in more central brain regions, closer to the pivot of x rotation, will probably be less.

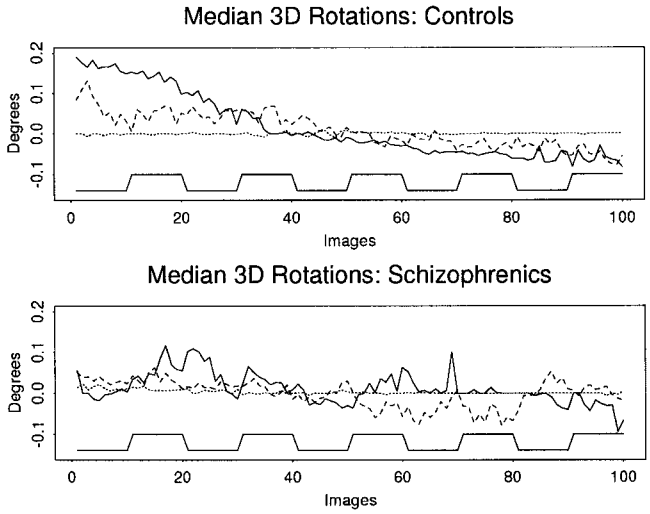

Figure 2.
Time series of median 3D rotational displacements (degrees) in control and schizophrenic groups: solid line, x rotation; dashed line, y rotation; dotted line, z rotation. The square wave indicates the experimental input function. Note that periodic x and y rotation at the frequency of the input function is more clearly evident in the schizophrenic group data than in the control group data.

One can assess the impact of corrected movement on FPQ at each voxel in the realigned image more directly by computing the difference in FPQ before and after motion-correction regression. To do this, we simply estimate FPQ at a given voxel in the realigned image before regression, and subtract FPQ estimated at the same voxel after regression. For each voxel in the image, this yields an estimate of $\Delta FPQ = FPQ^{pre} - FPQ^{post}$. At a "good" voxel, where the effect of motion correction has been to reduce SE(FP) without substantially attenuating FP, ΔFPQ will be less than zero; at a "bad" voxel, where the effect of motion-correction has been a disproportionate reduction in FP, ΔFPQ will be greater than zero.

The distributions of Δ FPQ over all voxels in each of the groups are summarized in Figure 3. Mean ΔFPQ is 0.15 for the schizophrenics and -0.07 for the controls. This shows that the schizophrenic images include many more voxels with positive ΔFPQ, i.e., many more voxels where the effect of motion-correction regression has been to diminish the size of the test statistic, rather than of the control images. It is therefore likely that at least some of the between-group differences evident in the GBAMs are indeed related to a systematic difference in subject movement. Next, we will describe and apply a number of techniques to compare patterns of

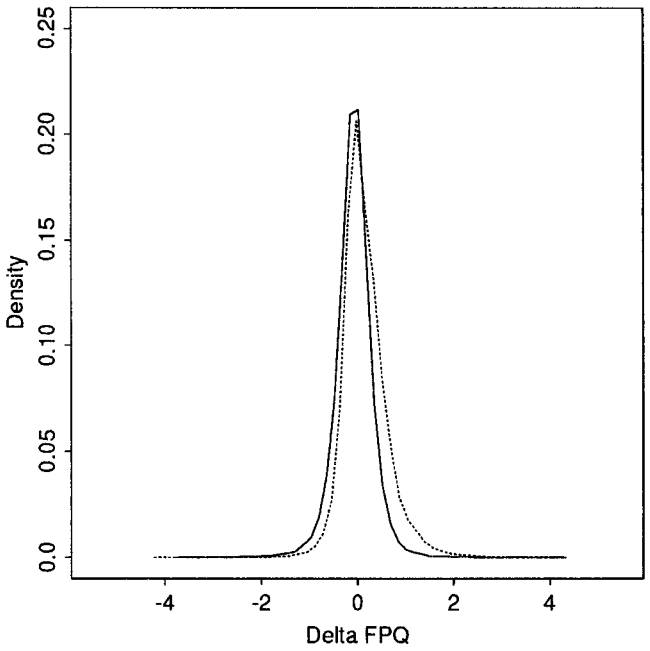

Figure 3.
Estimated densities of ΔFPQ estimated at each voxel over all images in the control and schizophrenic groups. The distribution for the schizophrenic group is shifted to the right relative to the control group, indicating that the schizophrenic data are generally more affected by stimulus-correlated motion than are the control data.

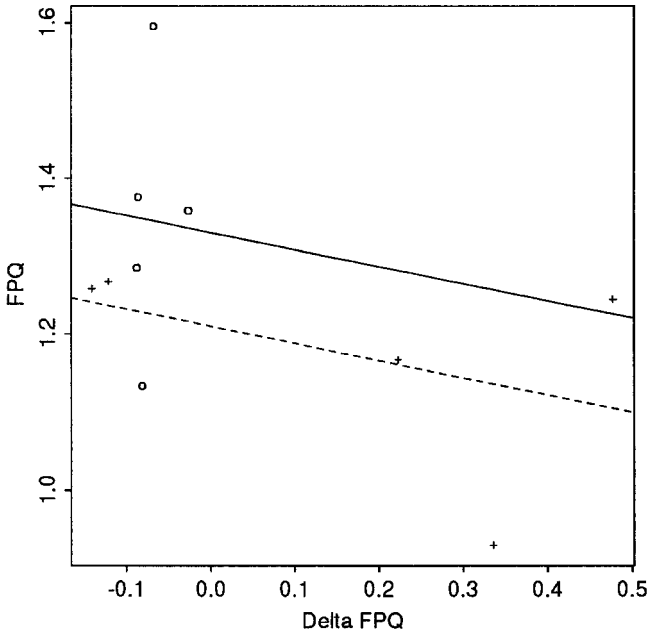

Figure 4.
Plot of image mean FPQ vs. image mean ΔFPQ for all 10 images. Control images are indicated by circles and schizophrenic images by crosses. Lines represent the result of fitting the ANCOVA model in Eq. 5: the solid line indicates the relationship between FPQ and ΔFPQ for control subjects; the dashed line indicates the relationship between FPQ and ΔFPQ for schizophrenic subjects.

brain activation between groups that will take this difference into account.

TREATMENT OF STIMULUS-CORRELATED MOTION

Image level

One crude way to correct the effects of stimulus-correlated motion within a group of images is to add group mean ΔFPQ to FPQ^{post} estimated at each voxel prior to generic brain activation mapping. More formally, the test statistic estimated at each voxel for the ith subject in the jth group (in this study, j = 1,2) is

$$FPQ_{i,j} = FPQ_{i,j}^{post} + \overline{\Delta FPQ_j}. \qquad (4)$$

The resulting GBAM may then be said to be normalized to zero group mean ΔFPQ. As shown in Figure 1, this operation *increases* the number of activated voxels in the schizophrenic GBAM (from 453 activated voxels before normalization to 658 voxels after normalization; a 45% increase). The same transformation applied to the control data *decreases* the number of activated voxels (from 1,845 to 1,173; a 36% reduction). The two GBAMs are now equivalent with respect to group mean stimulus-correlated motion. The remaining differences between the group maps can therefore be more confidently attributed to a between-group difference in task response.

To clarify the representation of differences between groups in FPQ, after controlling for variability in ΔFPQ, we can alternatively use analysis of covariance (ANCOVA). Consider a plot of mean FPQ vs. mean ΔFPQ for each of the 10 data sets in the study (Fig. 4). This shows that the controls tend to have greater mean FPQ than the schizophrenics, and that the schizophrenics tend to have greater mean ΔFPQ than the controls. In view of the results reported above, this should not be surprising. We can estimate the between-group difference in image mean FPQ, after controlling for the effects of variability in image mean ΔFPQ across groups, by fitting the following ANCOVA model:

$$\overline{FPQ_k^{post}} = \xi_0 + \xi_1 G_k + \xi_2 \overline{\Delta FPQ_k} + \epsilon_k. \qquad (5)$$

TABLE II. Parameters for ANCOVA and ANOVA models of image mean FPQ, and fitted group means

Model	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\xi}_2$	Control $\overline{FPQ}$	Schizophrenic $\overline{FPQ}$
ANCOVA	1.27	0.063	−0.22	1.333	1.207
ANOVA	1.26	0.088		1.348	1.172

Here, $\{\overline{FPQ_k^{post}}, \overline{\Delta FPQ_k}\}$ denote mean FPQ and mean ΔFPQ, respectively, in the kth image (k = 1, 2, 3, ..., 10); G_k is the kth element of the factor G which has value 1 for controls and −1 for schizophrenics; and ϵ_k is an error term. The estimated values of the three parameters $\{\xi\}$, and the fitted mean FPQs for both groups, are summarised in Table 2. Estimated ξ_2 is negative, indicating that over both groups relatively large positive values of $\overline{\Delta FPQ_k}$ (indicative of stimulus-correlated motion) are associated with relatively small values of $\overline{FPQ_k^{post}}$ (as expected). Estimated ξ_1 is positive, indicating that the control group mean FPQ is greater than the schizophrenic group mean FPQ (absolute difference in fitted group means = 0.126). In other words, these results suggest that mean FPQ in the control group is approximately 10% greater than in the schizophrenic group, after controlling for variability in stimulus-correlated motion between groups.

To assess the impact of stimulus-correlated motion on between-group differences in mean FPQ, we can fit the following analysis of variance (ANOVA) model to the same data:

$$\overline{FPQ_k^{post}} = \xi_0 + \xi_1 G_k + \epsilon_k. \quad (6)$$

Note that this model is identical to Eq. 5, except for the omission of image mean ΔFPQ as a covariate. The estimated model parameters and fitted group means are shown in Table 2; absolute difference in fitted group means = 0.176, or 13% of control mean. It can be seen that failure to control for group differences in stimulus-correlated motion leads to a substantial overestimation of the between-group difference in mean FPQ.

Voxel level

This application of ANCOVA to estimate between-group differences in mean FPQ for each image, after controlling for imagewise variability in stimulus-correlated motion, can be extended to map between-group differences in FPQ at each voxel, after controlling for voxelwise variability in stimulus-correlated motion. To do this, we fit the following version of Eq. 5 at each intracerebral voxel in standard space:

$$FPQ_{i,j}^{post} = \xi_0 + \xi_1 G_j + \xi_2 \Delta FPQ_{i,j} + \epsilon_{i,j}. \quad (7)$$

At a voxel where the control mean FPQ is greater than the schizophrenic mean FPQ, ξ_1 will be positive; and where the schizophrenic mean FPQ is greater than the control mean, ξ_1 will be negative.

To map voxels where the value of ξ_1 is "significantly large," we use a nonparametric technique similar in principle to methods earlier described for brain activation mapping [Bullmore et al., 1996; Brammer et al., 1997]. Our null hypothesis is that observed values of ξ_1 are not determined by the diagnostic status of subjects constituting the two groups. To sample the distribution of ξ_1 under this null hypothesis, we randomly reassign each subject to one of the two groups (i.e., randomly permute the elements of G [Edgington, 1980]) and estimate ξ_1 by fitting Eq. 7 at each voxel after permutation. We repeat this process 10 times at each intracerebral voxel, and pool the resulting estimates over all voxels to sample the permutation distribution of ξ_1. For a two-tailed test of size α, the observed value of ξ_1 at each voxel is compared to critical values at the $100 \times (\alpha/2)$ and $100 \times (1 - \alpha/2)$ percentiles of this distribution. Voxels where the observed value of ξ_1 exceeds the upper or lower critical values are colored and overlaid on the gray-scale background of the template image, to form a map of brain regions that show a significant difference in size of experimental effect between groups.

The results of this analysis are shown in Figure 5. There are 328 colored voxels in total. Significantly large *positive* values of ξ_1, indicating significantly greater mean FPQ for controls compared to schizophrenics, were identified in prefrontal, retrosplenial, parahippocampal, and posterior parietal cortices. Fewer significantly large *negative* values of ξ_1, indicating significantly greater mean FPQ for schizophrenics compared to controls, were identified in medial parietal cortex (see Curtis et al. [1998] for more detailed discussion of these findings).

Also shown in Figure 5 is the equivalent map obtained by testing observed values of ξ_1 from a voxelwise version of the ANOVA model (Eq. 6):

$$FPQ_{i,j}^{post} = \xi_0 + \xi_1 G_j + \epsilon_{i,j}. \quad (8)$$

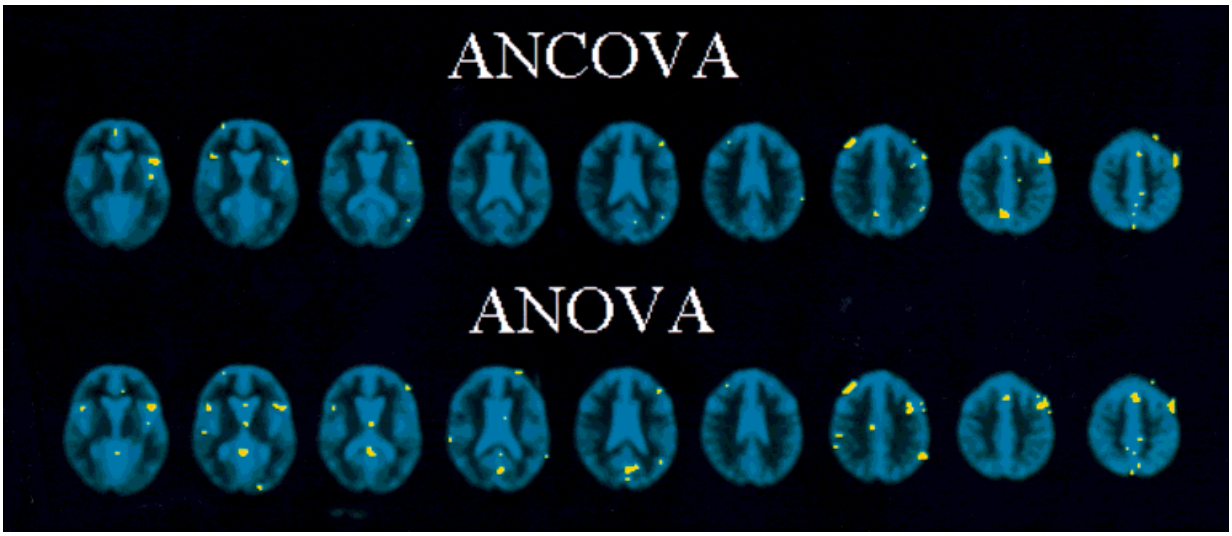

Figure 5.

Maps showing significant group mean differences in FPQ, with voxelwise probability of type I error $\alpha \leq 0.01$. Each row of nine 5.5 $\times$ mm-thick axial slices represents the brain from $+1$ mm to $+50$ mm in standard space; the right side of each slice represents the left side of the brain. **Top:** Voxels with significantly large absolute values of ξ_1, estimated by fitting the ANCOVA model (Eq. 7), are colored yellow. This shows that controls demonstrate a more powerful periodic response to the input function than schizophrenics in frontal and posterior parietal regions. **Bottom:** Voxels with significantly large absolute values for ξ_1, estimated by fitting the ANOVA model (Eq. 8), are colored yellow. Comparison of the ANOVA map to the ANCOVA map shows that failure to correct for between-group differences in stimulus-correlated motion, by analysis of covariance, exaggerates apparent between-group differences in regional brain activation.

There are 735 colored voxels in this map, suggesting that failure to control for variability in stimulus-correlated motion in these data substantially exaggerates (more than doubles) the number of voxels demonstrating a significant between-group difference in response to the experimental input function.

DISCUSSION AND CONCLUSIONS

Functional MRI is potentially a very valuable tool for investigation of psychiatric and neurological disorders. However, it is also exquisitely sensitive to the effects of subject motion during image acquisition, and neurological or psychiatric patients may not always move in the same way, or to the same extent, as "control" subjects. In particular, there may be systematic differences between patient and control groups in stimulus-correlated motion. This means that clinically trivial, but physically important, differences in the way patients move during a study may bias assessment of differences in brain activation.

We have described a number of diagnostic tests for stimulus-correlated motion that can be applied to any fMRI data set at the level of the whole image or a single voxel. In our view, the most informative of these tests involves measuring the voxelwise difference in *any* standardized test statistic for activation before (pre) and after (post) movement-correcting regression. Here, since we used standardized power at the (fundamental) frequency of the input function (FPQ) as our test statistic [Bullmore et al., 1996], we measured the impact of stimulus-correlated motion in terms of the voxelwise difference $\Delta\text{FPQ} = \text{FPQ}^{\text{pre}} - \text{FPQ}^{\text{post}}$.

By this measure and others, we diagnosed a systematic difference in stimulus-correlated motion between two groups of fMRI data acquired from 5 schizophrenics and 5 controls. The schizophrenic group demonstrated more stimulus-correlated motion than controls; this complicated interpretation of evident differences between the generic brain activation maps (GBAMs) separately computed for the two groups.

Finally, we described some simple methods to allow comparison of activation between groups while controlling for the confounding effects of between-group differences in stimulus-correlated motion. Our preferred method of treatment is to use analysis of covariance (ANCOVA) to estimate the difference in group mean FPQ^{post} at each voxel in standard space

after controlling for variability in ΔFPQ across groups. Voxels that demonstrate a significant (and relatively unbiased) between-group difference in amplitude of experimentally determined response may then be identified by a nonparametric test of the estimated group factor coefficient.

It is possible that the schizophrenic group data we have used to illustrate these issues and methods are particularly ''bad,'' in the technical sense of being complicated by stimulus-correlated motion. In other words, it might be argued that we are here concerned with a problem that may not arise very often, or that may not often cause serious problems in data interpretation. However, these were calm and cooperative patients performing a covert task. It is easy to imagine that the problem could be much worse in data acquired from agitated or aroused patients, or from subjects performing a task which demanded speech or hand movement. We suggest that the only way to be sure that a given study is not complicated by stimulus-correlated motion is explicitly to exclude the possibility that it is; the diagnostic techniques here reported have potential value in that respect.

The use of analysis of covariance to estimate a group effect (of interest) while controlling for the possibly confounding effects of one or more covariates (of no interest) could well have future use in comparing two or more groups of fMRI data, even when stimulus-correlated motion has not been identified as an important confound. The method could, for example, be generalized to control for between-group differences in task performance, or image quality, that might otherwise bias estimation of one or more factorial effects of interest.

ACKNOWLEDGMENTS

E.T.B. was supported by the Wellcome Trust. Data acquisition was funded by a grant from the Bethlem Maudsley Research Fund. Thanks are due to Dr. Paul Grasby for advice on experimental design, and to Dr. John Suckling for image processing expertise.

REFERENCES

American Psychiatric Association (1987): Diagnostic and Statistical Manual of Mental Disorders, Revised 3rd Edition. Washington, DC: American Psychiatric Association.

Brammer MJ, Bullmore ET, Simmons A, Williams SCR, Grasby PM, Howard RJ, Woodruff PWR, Rabe-Hesketh S (1997): Generic brain activation mapping in fMRI: A nonparametric approach. Magn Reson Imaging 15:763–770.

Bullmore ET, Brammer MJ, Williams SCR, Rabe-Hesketh S, Janot N, David AS, Mellers JDC, Howard R, Sham P (1996: Statistical methods of estimation and inference for functional MR image analysis. Magn Reson Med 35:261–277.

Curtis VA, Bullmore ET, Brammer MJ, Sharma TS, Morris RG, Murray RM, McGuire PK (1998): Attenuated frontal activation during verbal fluency in schizophrenia. Am J Psychiatry 155:1056–1063.

Edgington ES (1980): Randomisation Tests. New York: Marcel Dekker.

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

Grasby PM, Williams SCR, Bullmore ET, Brammer MJ, Checkley SA (1995): A functional MRI study of covert verbal fluency in normal volunteers. Proc Soc Magn Reson 3:1337.

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.

Hill DLG, Hawkes DJ, Studholme C, Summers PE, Taylor MG (1994): Accurate registration and transformation of temporal image sequences. Proc Soc Magn Reson 1:820.

Manly BJF (1991): Randomisation and Monte Carlo Methods in Biology. London: Chapman and Hall.

Ogawa S, Lee TM, Kay AR, Tank DW (1990): Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Nat Acad Sci USA 87:9868–9872.

Ostrum CW (1978): Time Series Analysis: Regression Techniques. Beverly Hills: Sage Publications.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992): Numerical Recipes in C: The Art of Scientific Computing. Cambridge: Cambridge University Press.

Talairach J, Tournoux P (1988): A Coplanar Stereotactic Atlas of the Human Brain. Stuttgart: Thieme Verlag.

Warburton E, Wise RJS, Price CJ, Weiller C, Hadar U, Ramsay S, Frackowiak RSJ (1996): Noun and verb retrieval by normal subjects. Studies with PET. Brain 119:159–179.