PSYCHOMETRIKA—VOL. 78, NO. 3, 396–416
JULY 2013
DOI: 10.1007/S11336-012-9294-0

A SURVEY OF THE SOURCES OF NOISE IN FMRI

DOUGLAS N. GREVE

DEPARTMENT OF RADIOLOGY, MASSACHUSETTS GENERAL HOSPITAL

GREGORY G. BROWN

VA SAN DIEGO HEALTHCARE SYSTEM AND DEPARTMENT OF PSYCHIATRY, UNIVERSITY OF CALIFORNIA SAN DIEGO

BRYON A. MUELLER

DEPARTMENT OF PSYCHIATRY, UNIVERSITY OF MINNESOTA TWIN CITIES

GARY GLOVER

DEPARTMENT OF RADIOLOGY, STANFORD UNIVERSITY

THOMAS T. LIU

CENTER FOR FUNCTIONAL MRI, UNIVERSITY OF CALIFORNIA SAN DIEGO

FUNCTION BIOMEDICAL RESEARCH NETWORK

Functional magnetic resonance imaging (fMRI) is a noninvasive method for measuring brain function by correlating temporal changes in local cerebral blood oxygenation with behavioral measures. fMRI is used to study individuals at single time points, across multiple time points (with or without intervention), as well as to examine the variation of brain function across normal and ill populations. fMRI may be collected at multiple sites and then pooled into a single analysis. This paper describes how fMRI data is analyzed at each of these levels and describes the noise sources introduced at each level.

Key words: functional MRI, blood oxygen level dependent, first-level analysis, higher level analysis, sources of noise.

1. Introduction

Functional Magnetic Resonance Imaging (fMRI) is a tool for studying brain function, i.e., which neural systems are the brain substrate for a particular behavior. fMRI is noninvasive, meaning that neither surgery nor ionizing radiation are used to generate the images. It can also localize activity to a few millimeters. For these reasons, fMRI as a research area has exploded over the last 10 years. In task-based analysis,[1] the brain is treated as a black box to which a stimulus is applied and from which a response is measured. The fMRI measurement itself is related to the amount of deoxygenated hemoglobin in the blood and so it is called the Blood Oxygen Level Dependent (BOLD) signal. While dependent upon many other factors, the BOLD signal gives an indication of how much blood is flowing to a particular location in the brain at a particular

[1] Increasingly fMRI experiments are being performed without a task. These techniques are not extensively described in this paper as their methodology strongly overlaps with task-based analysis.

Requests for reprints should be sent to Douglas N. Greve, Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA. E-mail: greve@nmr.mgh.harvard.edu

time. Behavioral and thought processes cause neurons in a small area to become active, which increases blood flow to that area to supply the metabolic demand of the neural activity. The more intense the neural activation, the larger the increase in blood flow will be (though the relationship is not necessarily linear or straightforward). The analysis strategy is to correlate the BOLD waveform with the known time course of stimulation. Brain areas engaged in the task will have a BOLD signal that is correlated with the task; brain areas that are not engaged in the task will be uncorrelated. This allows the characterization of the neural substrate of behavior in terms of the locations in the brain that are engaged, the amount of brain tissue recruited, and the strength or amplitude of the response. This may be clinically relevant for an individual person for diagnosis, for treatment planning, for monitoring treatment response or for assessing disease course. We are also interested in how populations vary, e.g., how do the location, intensity, and size of neural activation change between those diagnosed with schizophrenia versus those who are deemed clinically healthy? This leads to potentially four levels of analysis: (1) time series analysis of an individual at a particular visit date, (2) longitudinal analysis (i.e., from visit to visit), (3) between subjects within a defined population, and (4) between defined populations. When studying large samples, it is often necessary to scan them at different study locations, which can add a fifth analysis level (site). Each of these levels introduces a set of noise that will increase variability and/or introduce bias to the final result. This paper describes how fMRI is analyzed at each of these levels, the sources of noise at each level, and ways to control the noise. We start with a case study to familiarize the reader with the terminology of fMRI as well as provide motivation for conducting a study. Next, we describe the analysis at the first and higher levels based on the general linear model (GLM). The biophysics of fMRI are then summarized to introduce the reader to the physics of the measurement and a foundation for understanding the sources of the noise. Finally, the sources of noise at the various levels are surveyed along with methods to reduce their impact.

2. A Case Study

To provide a concrete example of how fMRI analysis is performed, we present a design from one of the studies performed by the Function Biomedical Informatics Research Network (fBIRN). The study used a working memory paradigm to explore how emotionally disturbing images interfere with working memory and whether this interference differs between healthy individuals and persons with schizophrenia. Subjects were asked to perform a working memory task that was divided into three phases. During the Encode phase, the subject was shown a 16 second series of pictures each with a single drawing and asked to save these in memory for recall 16–32 seconds later. These were followed by a 16 second Distractor phase in which images were either emotionally neutral (e.g., a chair) or emotionally disturbing (e.g., a car accident). In the Probe phase, the subject was shown a 16 second series of images with two drawings and asked to respond with a button press as to which one was in the original series. This entire sequence (called a "block") lasted 48 seconds and gave five "conditions": (1) Encode, (2) Emotional Distractor (ED), (3) Probe following Emotional (PED), (4) Neutral Distractor (ND), and (5) Probe following Neutral (PND). We are interested in brain areas that respond to each of these conditions individually as well as differentially. During a "run," four such blocks were presented with 16 second between blocks during which scrambled, uninterpretable images were presented. The total run time was of 284 seconds. See Figure 1 for a diagram of the timing. This process was repeated seven more times for a total of eight runs during the visit. Multiple repetitions were used for averaging. All of this took place inside an MRI scanner. The subject lays on a bed inside of the bore of the scanner with his/her head inside of a coil apparatus used to detect the MRI signal. Images were presented to the subject either through goggles or projected onto a screen that the

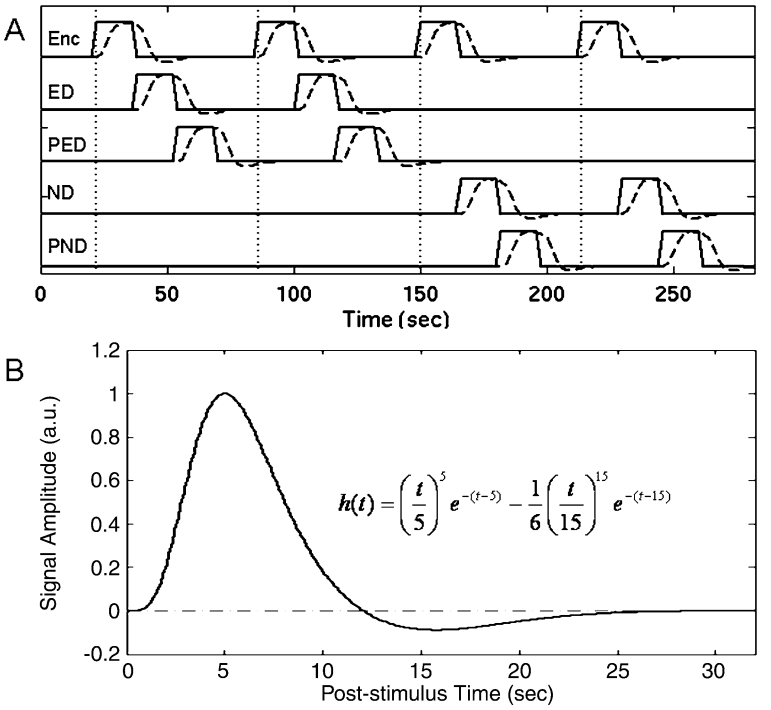

FIGURE 1.
Temporal waveforms used in the analysis of the working memory task. Panel **A**: Task timing waveforms (*solid lines*) and regressors (*dashed lines*) for each condition for a single run. Encode (*Enc*), Emotional Distractor (*ED*), Probe following Emotion Distractor (*PED*), Neutral Distractor (*ND*), and Probe following Neutral Distractor (*PND*). The regressor is formed by convolving the assumed HRF in Panel **B** with the paradigm timing waveform (*solid lines*) to form a biologically plausible waveform (*dashed lines*). **B**: Canonical shape to the HRF. This shape can be interpreted as an impulse response to a very brief stimulus. This shape and parameters are the default used by the SPM software package.

subject can see through a mirror mounted on the coil. Finger press responses were collected from a button box inside the scanner. The stimulus was controlled and button presses recorded by a single computer synchronized with the scanner.

During the time that images were being presented and button presses recorded, the MRI scanner was collecting images of the entire brain at a temporal sample rate of one every 2 seconds (called the "TR"). Each sample is referred to as a "time point" or "shot." For the task given above, each run lasted 284 seconds resulting in 142 time points; eight such runs resulted in a total of 1,136 time points. The brain volume collected at each time point consisted of a stack of 30 slices; each individual slice was a 64×64 image matrix of intensity values. Each value is a measurement (see Equation (6)) from a small $3.4 \times 3.4 \times 5$ mm box called a "voxel." The entire volume covered a field-of-view (FoV) of $220 \times 220 \times 150$ mm. A volume was acquired slice-by-slice (67 ms per slice) over the TR, so there was a delay of almost 2 sec between the time of the first slice and the last slice. See Figure 2 for an image of an fMRI slice (and corresponding anatomical). For the visit described above, there were a total of 122,880 ($30 \times 64 \times 64$) voxels in each time point and 1,136 time points for a total of over 139 million samples. Eighteen subjects were scanned at four sites; all subjects visited one of the sites twice making a total of five visits. For full documentation of this study, see (Brown, Mathalon, Stern, Ford, Mueller, Greve, McCarthy, Voyvodic, Glover, Diaz, Yetter, Ozyurt, Jorgensen, Wible, Turner, Thompson, & Potkin, 2011; Greve, Mueller, Liu, Turner, Voyvodic, Yetter, Diaz, McCarthy, Wallace, Roach, Ford, Mathalon, Calhoun, Wible, Brown, Potkin, & Glover, 2010).

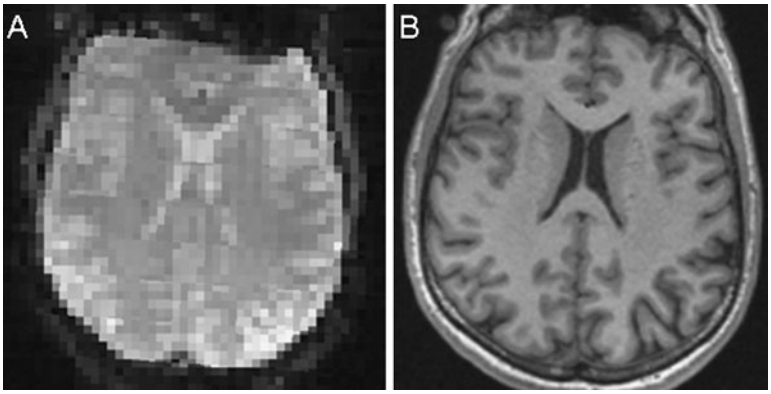

FIGURE 2.
Functional (Panel **A**) and anatomical (Panel **B**) images of the same subject. The functional image has a 3.4 mm resolution and a T_2^* weighting. The anatomical image has a 1 mm voxel resolution and a T_1 weighting.

3. fMRI Time Series (First-Level) Analysis

While there are many ways to analyze fMRI, we will describe a typical fMRI analysis (Friston, Jezzard, & Turner, 1994; Huettel, Song, & McCarthy, 2009; Jezzard, Matthews, & Smith, 2001; Monti, 2011; Woolrich, Ripley, Brady, & Smith, 2001). First-level analysis generally consists of several steps: preprocessing, constructing a design matrix as part of a general linear model (GLM), temporal whitening, constructing contrasts to test hypotheses, computing p-values, and correcting for multiple comparisons. The preprocessing operations, like motion correction and spatial smoothing, are designed to reduce a specific kind of noise, so they are discussed in detail as each source of noise is discussed in Section 6. This section describes the remaining analysis steps, how the signal is modeled and the noise is computed, and prepares the reader for tracing the propagation of noise to the higher level analysis.

Each voxel will have its own time course (the "Raw Signal" in Figure 3), and the analysis is performed at each of 122,880 voxels separately by correlating its time course with a biologically plausible waveform that represents the anticipated BOLD response to the task. This waveform is constructed based on two pieces of information. First, a square waveform (solid lines in Figure 1, Panel A) is created using the onset and offset of the task of interest (1 for "on" and 0 for "off"). The brain's hemodynamic response to stimulation, however, is known to be a smooth function that takes approximately 4 to 6 seconds to peak and 8 to 10 seconds to return to baseline levels (Figure 1, Panel B), rather than the square on/off waveform of the stimulation. For this reason, the on/off pattern of stimulation is convolved with a smooth kernel like the one in Figure 1, Panel B to transform it into a biologically plausible waveform of up- and down-regulation of the BOLD signal (dashed lines in Figure 1, Panel A). The kernel is referred to as the hemodynamic response function (HRF) and can be interpreted as an impulse response. HRF shapes have been derived empirically (Friston, Frith, Turner, & Frackowiak, 1995; Glover, 1999).

The time course for each voxel is analyzed using a GLM (Seber & Lee, 2003):

$$Fy = FX\beta + Fn \tag{1}$$

where y is a vector of measurements at each time point at a single voxel ("Raw Signal" in Figure 3), F is a temporal filter, X is the design matrix (independent of voxel), n is the noise at the voxel, β is the unknown amplitude of the HRF at the voxel. Note that y, X, and n are vectors of length N_t (the number of time points). The columns of X (regression vectors) are formed from the biologically plausible waveforms (dashed lines in Figure 1, Panel A).

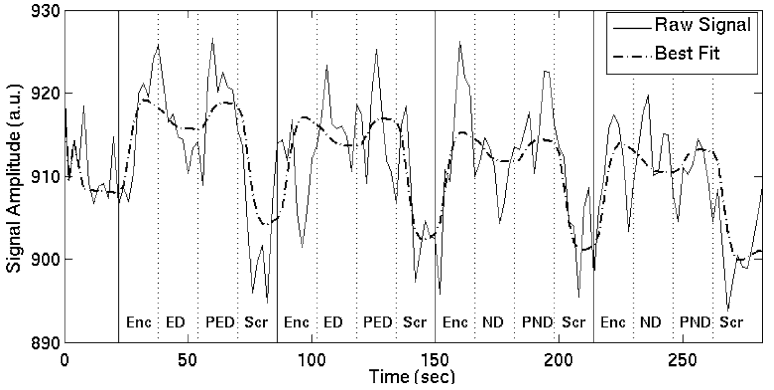

FIGURE 3.
Raw and best fit time courses for a single voxel over a single run using the regressors in Figure 1, Panel A. Scr = Scrambled images used as baseline. Other abbreviations are the same as in Figure 1, Panel A.

The quantity of interest is the amplitude of the hemodynamic response β because this relates to the neural activity at that particular location. An estimate of β can be computed using the least-mean-square (LMS) solution as

$$\hat{\beta} = \left(X'F'FX\right)^{-1}X'F'Fy \tag{2}$$

A raw time course and resulting fit ($\hat{y} = X\hat{\beta}$) are shown in Figure 3.

As explained below, the fMRI noise may be temporally correlated (or "colored"), i.e., one can do better than chance at predicting what the next noise value will be given past values, which is not true for temporally uncorrelated (or "white") noise. Temporal correlation in the noise is undesirable as it results in inaccurate computation of the p-values of the $\hat{\beta}$ estimates. Mathematically, the temporal relationship of the noise with itself is represented by the N_t-by-N_t temporal covariance matrix σV, where σ is the (scalar) variance of the noise, and V is the matrix of autocorrelation coefficients. The value at column i, row j of V indicates the correlation between the noise waveform and itself delayed by $i - j$ time points. By definition, the value along the diagonal is 1 (i.e., $V_{ii} = 1$). White noise is indicated by a diagonal matrix ($V = I$); nonzero off-diagonal terms indicate temporally correlated noise. Temporally correlated noise can be compensated for by appropriately selecting F, the temporal filter. To remove the temporal correlation, the filter should be set such that $FV^{-1}F' = I$ thereby converting the noise covariance into a diagonal matrix and producing temporal whiteness. The actual values of F can be computed as the square root of V^{-1} using the Cholesky decomposition. In practice, the temporal whitening is performed in two steps. In the first step, F is set to the identity, the GLM is solved, and an estimate of V is computed from the residuals. In the second pass, F is computed from the estimate of V, and the GLM is solved again (Burock & Dale, 2000; Woolrich et al., 2001).

If the task contains more than one condition, then X will have a column for each condition. The waveform for that condition is constructed from the part of the task waveform associated with only that condition (see Figure 1). Each condition will have an associated $\hat{\beta}$. In the case study above, there will be a column in X and corresponding $\hat{\beta}$ for each of the five conditions. This allows the computation of a contrast, e.g., the difference between Condition 2 (Emotional Distractor) and Condition 4 (Neutral Distractor). This contrast allows us to find brain areas that respond differentially to the emotional content of the distractor and ignore areas that respond to some other common aspect of the stimuli. This contrast is embodied in a contrast equation:

$$\hat{\gamma} = C\hat{\beta} \tag{3}$$

where C is the contrast matrix and $\hat{\gamma}$ is the contrast value. For the example contrast above, $C = [0\,1\,0\,-1\,0]$. We can then use a t-test to test the null hypothesis that $\hat{\gamma} = 0$ with

$$t = \frac{\hat{\gamma}}{\sigma_\gamma}, \qquad \sigma_\gamma = \sqrt{\frac{\sigma_r^2}{\xi}}, \qquad \xi = \frac{1}{\text{trace}(C(X'F'FX)^{-1}C')} \tag{4}$$

where σ_γ is the standard error of $\hat{\gamma}$, σ_r^2 is the variance of the residual error $e = F(y - X\hat{\beta})$, and ξ is the efficiency. The significance (p-value) at the voxel can be computed from t with degrees-of-freedom (DOF) equal to the rows of X minus the columns of X. From this, one draws conclusions about the effect of the task at this location in the brain for this individual on this particular occasion. If the contrast is multivariate (i.e., C has more than one row), then an F-statistic would be used. The noise, summarized by its variance or standard error, is obviously very important. Below, we describe the components that go into this noise as well as efforts made to compensate for them. Note that $\hat{\gamma}$ is not an absolute quantitative measure of neural activation or energy consumption and does not carry meaningful physiological units. This means that conclusions must be statements about the relative size of responses (i.e. "this condition has a larger response than that condition"). (See Liu, Glover, Mueller, Greve, & Brown, 2012, for more information about quantification of fMRI.) Note also that the efficiency depends on the stimulus schedule and can be optimized in advance (Dale, 1999; Liu & Frank, 2004; Maus, Van Breukelen, Goebel, & Berger, 2010).

After performing this analysis at each voxel separately, the results can be displayed as an image called a statistical parametric map (SPM) as shown in Figure 4, Panels A–C. These images show opaque color proportional to the t-value for voxels that have a t-value greater than 3.3 (i.e., $p < .001$). This collection of supra-threshold voxels is known as "activation." If the t-value is less than 3.3, then it transparently shows the gray-scale anatomical image. In this way, one can glance at an image to see if there are any significant voxels, where they are in the brain, and how much tissue they cover. The locus of activation is quantified based on the position of the supra-threshold voxels; the size of activation is quantified based on the number of supra-threshold voxels; and the intensity of activation quantified based on the $\hat{\gamma}$ in the activation region.

The presence of supra-threshold voxels needs to be interpreted with care. Each voxel represents a separate statistical test. If a p-threshold of $p < .01$ is used and there are 122,880 voxels, then one would expect 1,228 voxels to appear significant purely by chance, provided the statistical tests in each voxel are independent. In the MRI literature, this is referred to as "the problem of multiple comparisons." One solution is to apply a Bonferroni correction by dividing the p-threshold by the number of voxels. In the above example, this would change the threshold for significance from $p < 10^{-2}$ to $p < 10^{-7}$. While this controls false positives, it dramatically increases false negatives. Another solution is to only look for "clusters" (i.e., contiguous regions of supra-threshold activity). The basic idea being that false positives will be spatially random and several adjacent supra-threshold voxels would be uncommon by chance. Closed-form distribution functions of cluster sizes have been derived under random field theory (Worsley, Marrett, Neelin, Vandal, Friston, & Evans, 1996); this allows a p-value to be assigned to each cluster depending on its size. These p-values can also be computed using simulations under the null hypothesis (Hayasaka & Nichols, 2003). Another alternative is to threshold using false discovery rate (FDR) in which the number of false positives with respect to the total number of positives (instead of the total number of tests) is controlled (Genovese, Lazar, & Nichols, 2002).

4. Higher Level Analysis

In a full study, data on many (i.e., > 15) subjects are typically acquired, perhaps from different groups, perhaps at different visit dates, and perhaps with some sort of intervention between

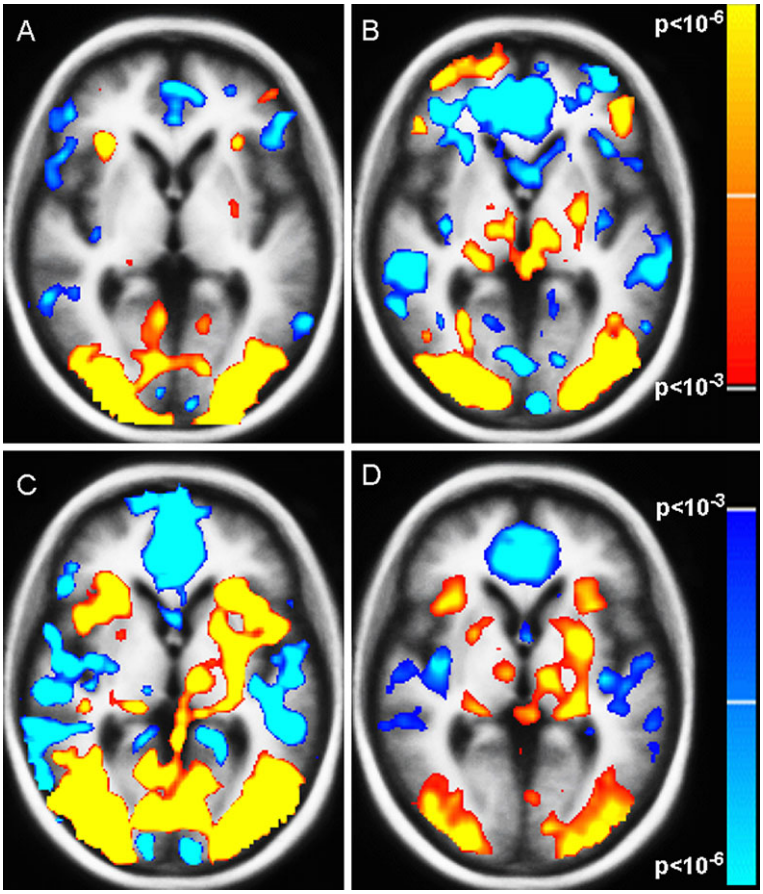

FIGURE 4.
Significance SPMs for the Probe-vs.-Scrambled Contrast. Panels **A**–**C** show activation maps for 3 individuals. Panel **D** shows the results of the random-effects test over 18 subjects. Thresholds are set at $t > 3.3$ ($p < .001$, uncorrected for multiple comparisons). *Red/yellow* indicates Probe > Scrambled; *blue/cyan* indicates Scrambled > Probe.

visits. The analysis at the higher levels is conceptually the same as the first-level in that the lower-level GLM contrast value $\hat{\gamma}$ (and its standard error σ_γ) are used as input to a new GLM in which a new design matrix and contrast matrices are constructed, a new fit is performed, and a new error term is computed. The new error term will be composed of low-level noise plus new sources of noise that may be specific to the level being analyzed. In this section, we summarize how the higher-level analysis is performed and introduce the problem of anatomical variability between subjects and registration to an atlas space.

The higher-level GLM can be written as

$$Hz = HG\varphi + Hu \tag{5}$$

where z is the vector of lower level $\hat{\gamma}$ at a given voxel, G is the group design matrix, φ is vector of regression coefficients to be estimated, u is the noise at the higher level, and H is a matrix that accounts for heteroscedasticity across the lower level measurements. Analyses are often performed using a "random effects model" in which $H = I$ (Friston, Holmes, & Worsley, 1999). Increasingly, mixed effects models that take into account the variance from the lower level are becoming popular (Beckmann, Jenkinson, & Smith, 2003). See Zhou, Konstorum, Duong, Tiue, Wells, Grown, Stern, & Shanbaba (2012) for a thorough review of Bayesian hierarchical analysis

of multilevel fMRI data, one form of mixed effects modeling. As with the first-level analysis, contrasts and p-values can be computed to create new SPMs (Figure 4, Panel D). At each of the higher levels, we can quantify the location, size, and amplitude of the response as described by the first-level.

Prior to comparison across subjects, one needs to make sure that one is comparing the appropriate values. If two subjects each have 100,000 voxels, which voxel in Subject 1 should be compared against which voxel in Subject 2? One solution is to align or register the brains of all subjects to an "atlas" (Collins, Neelin, Peters, & Evans, 1994). The human brain has a lot of structural variability, but there are some features that are consistent across all brains. These features are preserved in an atlas, which is typically formed from an average across a large sample of brain images from a representative population after the images have been rotated, translated, stretched, and sheared to make their features align better. More elaborate intersubject alignment schemes are becoming common (Ashburner, 2007; Fischl, Dale, Sereno, Tootell, & Rosen, 1998; Postelnicu, Zollei, & Fischl, 2009). Prior to group analysis, each subject's brain is transformed to align its features with those of the atlas. The voxel of analysis in Equation (5) then corresponds to a voxel in the atlas space (also known as "standard" space). The images in Figure 2 are in a standard space known as "MNI152" (Collins et al., 1994).

There are several publically available software packages that will perform fMRI analysis from first-level through higher levels. Some of the prominent ones are Statistical Parametric Mapping (SPM; www.fil.ion.ucl.ac.uk/spm), FMRIB Software Library (FSL; www.fmrib.ox.ac.uk/fsl), and Analysis of Functional NeuroImages (AFNI; afni.nimh.nih.gov/afni).

5. MRI Biophysics

To understand the sources of fMRI signal and noise, one needs to understand how the BOLD signal is formed. In this section, we describe the nature of the measurement, including the scanner acquisition parameters, tissue parameters, how these parameters interact to form signal, as well as scanner-related noise sources. Although a complete description of how individual nuclei interact with a magnetic field requires a quantum mechanical description, a classical physics approach is adequate to describe how groups of atoms generate the observable properties in most MR imaging (Hanson, 2008). We will focus our discussion on protons because most conventional fMRI imaging is dominated by signal from the nucleus of water-bound hydrogen atoms, which are protons, but our description is applicable to all MR observable nuclei. Our discussion is also necessarily brief; for in-depth descriptions of this material see (Brown, Perthen, Liu, & Buxton, 2007; Buxton, 2009; Liang and Lauterbur, 2000; Nishimura, 1996).

Protons exhibit a quantum mechanical property called "spin." This spin causes the proton to have a magnetic moment and an intrinsic angular momentum. Normally, the spin axes (and corresponding magnetic moments) point in uniformly random directions canceling each other out over the volume of a voxel. When placed in the strong static magnetic field of a scanner (called the B_0 field), the distribution becomes nonuniform with slightly more spins pointing along the axis of the scanner bore. This creates a net magnetization vector in the voxel, something like a compass needle pointing north. When the net magnetic moment is perturbed from its equilibrium position using a pulse of radio frequency (RF) energy, it exhibits resonance behavior. The RF pulse causes the vector to tip at a certain angle away from the B_0 field and precess around in a circle like a spinning top that has been disturbed. When the RF pulse is turned off, it continues to move in a circle but slowly realigns with the magnetic field (a process called "relaxation"). As it relaxes, it reemits RF energy, which is then received by a coil and detected by the scanner electronics. This is the fundamental quantity measured in all MRI; all other quantities are derived from this quantity.

When the RF pulse is first applied, magnetization vectors of neighboring voxels tend to precess in phase at the resonance frequency, which is proportional to the strength of the magnetic field. However, the exact frequency of precession depends on slight perturbations the local magnetic field caused by factors such as the chemical environment and the structure of the tissue, so, over time, nearby protons within the voxel will become out of phase. The rate at which this happens is called the T_2^* (pronounced TEE-TWO-STAR) relaxation rate. The rate at which the magnetization vector becomes realigned with the magnetic field is called the T_1 relaxation rate, which is also dependent upon the chemical/structural environment. Different tissues have different chemical/structural environments and so different T_1 and T_2^*. MRI works by creating an image where the signal intensity at each location is related to the T_1, T_2^*, or proton (or spin) density (PD) at that location, thereby revealing the different tissue types. Figure 2, Panel A shows a T_2^*-weighted image; Figure 2 Panel B shows the same slice in a T_1-weighted image.

An image is *not* formed by simply sampling the signal intensity at each location in the brain. When the RF pulse is transmitted, it excites the spins across a slice[2] of the brain. These spins slowly reemit the excitation energy, so the measured signal is a composite from spins across the whole slice. Such a measurement actually represents a single complex point in the frequency space of the Fourier transform of the image of the slice. The frequency space is referred to as "k-space." Data samples are acquired at different points in k-space by making subtle manipulations to the magnetic field. When a full grid of k-space samples has been acquired, an inverse Fourier transform is applied to create an image of the slice. Typically, only the magnitude is used for fMRI. This Fourier-based image reconstruction implies that the data are sampled at known frequencies in k-space. However, small spatial variations in the B_0 field can cause this assumption to be violated, resulting in compression and stretching in parts of the image (referred to as B_0 distortion Jezzard & Balaban, 1995). This can also cause parts of the image to become darker than one would expect given the tissue there. Comparisons between Panels A and B in Figure 2 show a typical example of image distortion and signal loss by small spatial variations in the B_0 field.

The BOLD signal is created through a complex interaction between the MR physics and the living tissue. Active neurons expend energy which requires oxygen. Oxygen is delivered by hemoglobin in the blood vessels that run through the tissue. Energy usage by the tissue causes oxygen to be extracted from the blood, converting oxygenated hemoglobin (HbO) to deoxygenated hemoglobin (HbR). The BOLD signal is related to the amount of HbR present in the blood. The deoxygenated blood is pumped back to the lungs to acquire more oxygen so that the process can be repeated. The nearby vasculature react to the neural activation by dilating, producing an increase in blood flow and volume in the nearby tissue. This washes out the HbR, changing its concentration and so the BOLD signal. The key property that makes fMRI possible is that HbR has different magnetic properties than HbO. Specifically, blood with more HbR has a shorter (i.e., smaller) T_2^*. This change in blood properties causes the Blood Oxygen Level Dependent (BOLD) effect. BOLD-weighted images (e.g., Figure 2, Panel A) will then show the pattern of regional differences that depend upon the relative concentration of HbR, and from this the pattern of neural activity is inferred. This pattern is a snapshot because the blood flow and changing neural activation makes the pattern change over the course of a few seconds. The temporal relationship between the neural activity and the change in the BOLD signal is the hemodynamic response mentioned above (see also Figure 1, Panel B). This relationship is quite complicated and not yet well understood (Logothetis & Wandell, 2004). It may also depend on

[2]Prior to RF transmission, a third set of coils creates a controlled spatial gradient in the B_0 field so that only the spins in a given slice are on-resonance. The location of this slice is adjusted each shot so that all slices in the brain are imaged. Other types of MRI acquisitions will excite the spins across the entire brain.

many physiological variables not related to neural activation (Buxton, Uludag, Dubowitz, & Liu, 2004).

The value measured at a given voxel at a given time is governed by the following equation (Liang and Lauterbur, 2000):

$$y(v,t) = \frac{\mathrm{PD}(v) \cdot \sin(\alpha(v,t)) \cdot (1 - e^{\frac{-TR(v,t)}{T_1(v)}}) \cdot e^{\frac{-TE}{T_2^*(v,t)}}}{(1 - \cos(\alpha(v,t)) e^{\frac{-TR(v,t)}{T_1(v)}})} + w(v,t) \tag{6}$$

where v is the voxel index and t is the time of acquisition. The equation depends on three scanner acquisition parameters (TR, TE, and α) and three tissue MR parameters (T_2^*, T_1, and PD) while $w(v,t)$ is thermal noise (sometimes called background noise). Acquisition parameters can be selected to emphasize certain image contrast properties (something like setting the shutter speed on a camera). The flip angle α controls how far the RF pulse tips the spins from alignment with the magnetic field. The TR (mentioned above) is the repetition time which sets the rate of RF pulses applied to each slice. Since an image can only be measured after an RF pulse, the TR effectively controls the sampling rate. The TR is typically in the range of 2–3 sec for fMRI studies resulting in a Nyquist rate around 0.17–0.25 Hz. The TE is the echo time, which is the time delay between the RF pulse and the acquisition of the image; the TE controls the amount of T_2^* weighting. In the case study described in Section 2 $\alpha = 77°$, TE = 30 ms, and TR = 2 sec. The (unknown) tissue parameters indicate how the tissue interacts with the magnetic field.

An acquisition proceeds as follows: an RF pulse with flip angle α is applied to a slice of the brain, this tilts the spins by α; after the RF pulse ends, the spins slowly realign with the main magnetic field (time constant T_1) and dephase with nearby spins (time constant T_2^*); the measured signal, y, is dependant on the number of spins inside the voxel (PD) weighted by both the T_1 and T_2^* decay; another RF pulse is applied after TR seconds, and the process is repeated. Imaging takes place in the scanner while the subject is engaged in a behavioral task of known timing. The value from Equation (6) is used as input to Equation (1) which is solved to compute an estimate of the HRF amplitude for an assumed HRF shape. This estimate is subject to uncertainty because of the noise in the first-level measurement.

6. Sources of Time Series (First-Level) Noise

Time series noise (n in Equation (1)) manifests as any part of the measurement that does not fit the matrix model. It is summarized by its variance estimate σ_r^2 and temporal covariance V. It creates uncertainty in the estimate of the HRF amplitude ($\hat{\gamma}$ in Equation (2)); this uncertainty is quantified by the standard error of $\hat{\gamma}$ (σ_γ in Equation (4)). Equation (6) links the biophysics to both the HRF amplitude its uncertainty. Ideally, the measurement y would only change in response to neurally driven changes in T_2^*. However, changes in any of the seven parameters in Equation (6) (α, TR, TE, T_1, T_2^*, PD, w) will propagate to unmodeled changes in y; this includes fluctuations in T_2^* that are *not* linked to neural activation. In the remaining portion of Section 6, we trace these fluctuations first by describing scanner-related sources of noise and then move on to the more complicated subject-related sources. See Table 1 for a summary of sources of time series noise.

Scanner-Related Noise Scanner-related noise originates both from thermal noise and from instability in the scanning process itself. The thermal noise (w in Equation (6)) comes from Brownian motion of ions. Ions are charged particles, and the movement of charged particles is an electric current. This produces a fluctuating electromagnetic field, which is picked up by the electronics used to measure the MR signal. This electromagnetic noise is similar to the noise of a radio tuned to a frequency where there is no broadcast and is well modeled by spatially and temporally independent homoscedastic Gaussian noise.

TABLE 1.
Sources of first-level noise.

Source	Mitigation
Thermal	Optimize acquisition parameters; spatially smooth; eliminate external noise sources in the scanner room.
Scanner instability	Diagnose and minimize with QA protocol.
Head motion	Stabilize/restrain head; use prospective motion correction.
Spin history	Motion-related nuisance regressors.
Heartbeat	External monitor, nuisance regressors.
Heart rate variability	External monitor, nuisance regressors.
Respiration	External monitor, nuisance regressors.
Respiration variability	External monitor, nuisance regressors.
CO_2 variability	External monitor, nuisance regressors.
Resting-state/spontaneous Neural activation	Independent Component Analysis.
HRF model error	Include temporal derivative.

Thermal Noise Thermal noise can be reduced by spatial smoothing or by adjusting the acquisition parameters. Spatial smoothing is the averaging of the waveform at a voxel with those of nearby voxels. If the signals in nearby voxels exhibit greater spatial similarity with each other than with the noise (generally true), then smoothing will result in an increase in SNR at the cost of spatial resolution. Smoothing also assures that the data meet the requirements needed by random field theory in the correction of multiple comparisons (Hayasaka & Nichols, 2003). In addition, smoothing improves intersubject alignment (discussed below). Smoothing is an extremely common preprocessing step in fMRI, and few operations will have as dramatic effect on the results as spatial smoothing (Strother, La Conte, Hansen, Anderson, Zhang, Pulapura, & Rottenberg, 2004).

The acquisition parameters (TR, TE, α) can be selected to maximize the BOLD contrast with respect to the thermal noise. While this improves the overall SNR, the effect has diminishing returns (Kruger & Glover, 2001; Triantafyllou, Hoge, Krueger, Wiggins, Potthast, Wiggins, & Wald, 2005). To understand why, consider that Equation (6) is the sum of two components. The first component is the actual MR signal (dependent upon the tissue and acquisition parameters). The second component is the thermal noise. While the thermal noise is pure noise, the MR signal has both neurally driven changes (signal in Equation (1)) and changes due to fluctuations in the parameters which show up as noise in Equation (1). Thus, one can think of the BOLD measurement as having a single neurally driven signal component and two noise components, one thermal and one related to nonneural fluctuations in the MR signal. Adjusting the acquisition parameters can maximize the MR signal with respect to the thermal noise, but even if the relative thermal noise were reduced to zero, there would still be noise caused by the nonneurally driven changes in the MR signal. The actual contribution of thermal noise is hard to quantify exactly because it is highly dependent on acquisition parameters such as flip angle, TE, and voxel size. However, in the case study described in Section 2, the thermal noise generally accounted for only about 10–20 % of the total temporal noise variance in gray matter (Greve et al., 2010).

Scanner Instability Noise Noise caused by scanner instability can come from several sources. The RF power used to flip the spins by angle α may fluctuate slightly from shot to shot. The electronics that control the location in k-space at which a sample is made may fluctuate from shot to shot which changes the location in k-space of a sample from the assumed frequency grid. This can cause a shot to shot change over the entire image when the inverse Fourier transform is applied. Its distribution is not well characterized, but it does have some temporal and

spatial correlation. In the well-operating scanners of the case study, the instability noise only accounted for a few percent of the total variance (Greve et al., 2010). Both thermal noise and instability are present when scanning a "phantom" (i.e., an inanimate object such as a bottle of water)—so no human needs to be present to characterize these sources of image noise.

Subject Motion-Related Noise When a human is placed in the scanner, noise from motion and physiological effects will be introduced. An assumption of Equation (1) is that all the data points in the measurement vector (y) come from the same location in the brain. This location is a voxel which encompasses a tiny box inside of the scanner, and the location of this box is fixed with respect to the scanner. Thus, if the subject moves during the run, then the time course will come from one brain region during the beginning of the scan and from a different part of the brain after the motion. This can cause a large error when fitting the task time course to the voxel time course. Efforts are made to restrain the subject's head during the MRI scan, but the head is not rigidly restrained and some motion may take place. Given that the typical fMRI voxel size is about 4 mm, even a small amount of motion can degrade the measured time course.

Motion can also interact with the MR physics in a complicated way to cause something called the "spin history effect." This causes some parts of the brain to brighten or darken dramatically for a few time points after the motion (Friston, Williams, Howard, Frackowiak, & Turner, 1996). These effects are unmodeled in Equation (1) and so increase the noise. They are also irreversible.

The noise caused by motion is nonnormal, heteroscedastic, nonstationary, and temporally correlated. It can also be correlated with the task if the subject moves with the task (e.g., flinching in response to a disturbing picture). The effects of motion can be reduced by applying a motion-correction (MC) algorithm (Cox & Jesmanowicz, 1999). MC attempts to adjust the translation and rotation at each time point so that each voxel represents a single location in the brain at all time points. This cannot undo all the effects of motion (Friston et al., 1996), and itself requires some interpolation. It is worth noting that there are many efforts to measure and compensate for motion prospectively, thus reducing the effects of motion greatly (Thesen, Heid, Mueller, & Schad, 2000; Tisdall, Hess, & van der Kouwe, 2010; Ward, Riederer, Grimm, Ehman, Felmlee, & Jack, 2000).

Physiological Noise All changes in blood flow affect the BOLD signal by changing the concentration of HbR (and so the T_2^*). Ideally, the flow would only change in response to neural activation. Unfortunately, changes in flow can be caused by other factors unrelated to neural activity, such as heartbeat and respiration. These factors are referred to as "physiological noise." The brain also has complicated autoregulatory mechanisms that attempt to keep blood flow constant (Payne, 2006). When the heart beats, it sends a surge of blood through the vasculature, and this can be detected in the BOLD signal. The heart rate is about 1 Hz, much faster than neurally driven changes in the BOLD signal. Unfortunately, the heart rate is also faster than the Nyquist rate in most fMRI studies, resulting in the heartbeat effects being aliased into the much lower task frequencies. The heartbeat can be measured at high sample rates using an external device synchronized with the scanner. This measurement can then be used to construct "nuisance" regressors that are appended as columns to the design matrix in Equation (1) to model these effects (Glover, Li, & Ress, 2000). Nuisance regressors can also be constructed based on heart rate variability—low frequency changes in heartbeat that appear to account for a large fraction of the BOLD noise (Chang, Cunningham, & Glover, 2009).

Respiration can have a number of effects on the BOLD signal. As the chest cavity expands and contracts, it actually changes the magnetic field strength in the brain by tiny amounts. Though small, this causes the entire image to stretch or compress slightly with the breath (Brosch, Talavage, Ulmer, & Nyenhuis, 2002). From a noise perspective, this acts like a type of motion artifact. Respiration rates are generally about 0.3 Hz. Like heartbeat, respiration is typically faster

than the BOLD Nyquist rate. The respiration rate also controls the amount of carbon dioxide (CO_2) in the blood. CO_2 causes blood vessels to dilate, which increases flow. As the respiration rate changes, the concentration of CO_2 changes. This changes the blood flow which changes the BOLD signal. As with heartbeat, nuisance regressors can be derived from a respiration monitor and included in the BOLD model design matrix (Birn, Diamond, Smith, & Bandettini, 2006). CO_2 concentration can also be tracked directly with an external device to create nuisance regressors (Wise, Ide, Poulin, & Tracey, 2004).

The brain also has endogenous neural activity that is not related to the task (Biswal, Yetkin, Haughton, & Hyde, 1995; Chang and Glover, 2009). This activity draws blood just like task-related activity, but since there is no model for it in the design matrix the effects of this endogenous activity go into the error term. These fluctuations happen in the absence of a task so they are sometimes called "resting state networks" (RSNs). In a "functional connectivity" study, the RSNs themselves are the object of investigation (Fox, Snyder, Vincent, Corbetta, Van Essen, & Raichle, 2005; Vincent, Snyder, Fox, Shannon, Andrews, Raichle, & Buckner, 2006). RSNs are currently a very active area of fMRI research. The use of model-free spatio-temporal analysis has become a prevalent tool for the study of RSNs. These include the use of independent component analysis (ICA), which decomposes the entire 4D data set into independent spatial patterns and corresponding temporal waveforms (Beckmann & Smith, 2004; Calhoun, Adali, Pearlson, & Pekar, 2001). Temporal waveforms that correspond to RSN noise, as well as noise from motion and physiological sources, can be determined and removed from the analysis. Since the noise is generally low frequency, polynomial nuisance regressors are often added to the design matrix (Worsley, Liao, Aston, Pere, Dunackn, Morales, & Evans, 2002).

Noise Associated with HRF Model Errors When analyzing the data, it is generally assumed that the neural activation follows the timing of the stimulus that the HRF takes a fixed, known shape, and that the response to each presentation is identical (e.g., as in Figure 1, Panel A). These assumptions can be violated in a number of ways. The shapes frequently used to model the HRF were derived from empirical data (Cohen, 1997; Friston et al., 1994; Glover, 1999), but it is well known that the shape changes across subject and brain region (Aguirre, Zarahn, & D'Esposito, 1998; Handwerker, Ollinger, & D'Esposito, 2004; Henson, Price, Rugg, Turner, & Friston, 2002; Hopfinger, Buchel, Holmes, & Friston, 2000; Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000). These HRF model errors add noise, although this is generally small fraction of the total noise. More importantly, they bias the estimate of the response amplitude by causing the estimate to be too low. This makes the activation more difficult to detect and can cause a bias between groups (see below). The effect of these errors tends to drop with stimulus duration (Huettel et al., 2009). It is quite common to model these errors by constructing additional regressors that correspond to the temporal derivative (i.e., first order Taylor series expansions) of the assumed HRF (Friston, Fletcher, Josephs, Holmes, Rugg, & Turner, 1998) or other basis sets (Woolrich, Behrens, & Smith, 2004a). While this will account for some of the error, it is not sufficient by itself to recover the true amplitude of the HRF. In the first-level analysis, the additional regressors can be used in an F-test (Liu & Frank, 2004). Since only a single value is passed to higher level analyses, these additional regressors essentially act as nuisance regressors when considered from the effect on the higher level. Bias due to small delay errors (< 1 s) can be reduced by computing a signed magnitude of the derivative and nonderivative regression coefficients (Calhoun, Stevens, Pearlson, & Kiehl, 2004), though this does introduce some nonlinearity into the process. There are some nonlinear methods that attempt to fit both the shape and the amplitude (e.g., Woolrich, Jenkinson, Brady, & Smith, 2004b), but these are computationally intensive and not in general use.

In addition to systematic shape errors, the responses to repeated identical stimuli may not be the same. For example, the amplitude of the response to the second of two closely spaced stimuli

TABLE 2.
Sources of within-subject/cross-visit level noise.

Source	Mitigation
Scanner changes	Track and reduce with QA protocol
Head placement	Use scanner auto-alignment; train technician
Subject wakefulness	Scan at same time of day; survey wakefulness
Caffeine/nicotine/etc.	Control/monitor subject chemical use

may be less due to refractory effects (Huettel & McCarthy, 2000). The subject may be presented with tens or hundreds of presentations of the same stimulus type (e.g., emotional distractors), and the response may change with repetition. In primary sensory areas, the response is stable across repeated presentations within visit (Miezin et al., 2000), but care must be taken to avoid practice effects (Kelly & Garavan, 2005).

Nuisance Regressors The nuisance regressors are added with the ultimate goal of reducing the uncertainty in the HRF amplitude contrast (quantified by its standard error σ_γ). However, they must be used with care. Adding regressors to the design matrix can reduce the residual variance σ_r^2, but this can also reduce the efficiency ξ. The final effect on σ_γ will be unknown. If σ_r^2 is reduced more than the efficiency, σ_γ will decrease and detectability will increase. However, it is possible for σ_r^2 to decrease and σ_γ to increase if the efficiency is sufficiently reduced, paradoxically causing a reduction in the amount of detected activation even though more noise is accounted for.

7. Source of Higher Level Noise

In this section, we describe the sources of noise in the higher level analysis, where the $\hat{\gamma}$ are combined across visit, subject, and group, and perhaps site.

Within-Subject/Cross-Visit Level Noise Sources (Table 2) Subjects are often scanned longitudinally to track disease process or evaluate an intervention. When a subject returns to be scanned for a second time, differences may appear between the BOLD signal in the two visits that have little to do with changes in neural activity. These differences have several origins. First, the subject may not be placed in the scanner in exactly the same position which can change the B_0 distortion. This can be reduced by using an on-line automatic slice positioning method (van der Kouwe, Benner, Fischl, Schmitt, Salat, Harder, Sorensen, & Dale, 2005) and/or by proper training of study staff to position the subject's head in the coil and position the slices. It is also possible that the scanner itself has changed between visits. This may take the form of degradation in scanner hardware over time, recalibration of the scanner, and/or hardware or software upgrades. For this reason, it is recommended that a quality assurance (QA) protocol (Friedman & Glover, 2006; Greve et al., 2010) be implemented to assure consistent scanner performance. Subjects may also have changed in ways that do not relate to the task performance but can affect the HRF amplitude. These factors include the amount of sleep the previous night (Thomas and Kwong, 2006), caffeination level (Liu, Behzadi, Restom, Uludag, Lu, Buracas, Dubowitz, & Buxton, 2004), and recent alcohol (Levin, Ross, Mendelson, Kaufman, Lange, Maas, Mello, Cohen, & Renshaw, 1998) and nicotine (Kumari, Gray, ffytche, Mitterschiffthaler, Das, Zachariah, Vythelingum, Williams, Simmons, & Sharma, 2003) consumption. These can be reduced by instructing or monitoring subjects' sleep patterns and/or use of chemicals.

TABLE 3.
Sources of cross-subject and cross-group level noise.

Source	Mitigation
Anatomical variability	Use high-DOF intersubject registration methods
HRF model error	Include temporal derivatives
Head motion variation across group	Use weighted-least squares/mixed effects models

Within-Group/Cross-Subject Level Noise Sources (Table 3) A group is a cohort of individuals who have been classified together by some criteria, e.g., diagnosis (Alzheimer's, Schizophrenics, healthy), handedness, gender, age, etc. These subjects will have natural variation in their HRF amplitude across the population that is purely related to functional/neural differences. However, some of the differences may be due to HRF model error mentioned above, i.e., individuals may have the same HRF amplitude but different shapes, and the different shapes cause the estimate of the HRF amplitude to differ. If the HRF error is delay-related, the effect can be reduced by adding a temporal derivative of the HRF to the design matrix then passing up to the next level a signed magnitude of the main and derivative regression coefficients as discussed above.

Anatomical variability is a major source of noise when comparing across subject. As with all analyses, the assumption is that all the values in the input data vector (z in Equation (5)) come from the same place in the standard atlas brain, but individuals differ considerably in both brain structure and how function maps to that structure, and it is not clear exactly how this mapping between individuals should be performed. If the registration procedure does not align functional units properly, then there will be differences between the subjects because the $\hat{\gamma}$ being compared come from different brain regions. This will increase the variability within the sample, though the effect of registration errors can be reduced by spatial smoothing or by using high-DOF registration methods.

Cross-Group Level Noise Sources (Table 3) Comparing across groups (e.g., Alzheimer's patients versus age-matched controls) adds several more sources of noise. The anatomical variability between subjects can be even more important between groups because different groups may have systematic differences in brain structure that cause their registration to standard space to be systematically different (Shen, Sterr, & Szameitat, 2005). This means that at a given point in the standard brain, the HRF amplitudes from one group could come from one place in the brain while the amplitudes from a second group could come from a different place. While group correlated registration errors can lead to added variability, they can also generate a bias in the group effect. For this reason, the use higher order registration methods are recommended. There may also be systematic differences in the shape of the HRF between groups (D'Esposito, Zarahn, Aguirre, & Rypma, 1999), which can bias the amplitude comparisons, though these can also be reduced by passing the signed magnitude of the nonderivative and derivative components to the higher level as mentioned previously. Finally, different groups may have differences in first-level noise (Maxim, Sendur, Fadili, Suckling, Gould, Howard, & Bullmore, 2005) which can create heteroscedasticity at the higher level. This potential noise source should be controllable by the use of a mixed effects modeling in the higher level analysis. There can also be differences in vascular physiology between groups that can affect the BOLD response. For example, the baseline cerebral blood flow (CBF) will affect the amplitude of the HRF. The CBF can change with age, the use of caffeine or other drugs, and disease state (Fleisher, Podraza, Bangen, Taylor, Sherzai, Sidhar, Liu, Dale, & Buxton, 2009). It might be possible to reduce this confound by measuring CBF using an additional MRI acquisition called arterial spin labeling (ASL) (Alsop & Detre, 1996).

TABLE 4.
Sources of variation across site.

Source	Mitigation
Field strength	Choose scanners of the same strength
Manufacturer	Analyze with B_0 distortion correction; to the extent possible standardize protocol across vendors
RF noise	Use weighted-least squares/mixed effects models
Head restraint	Use same head restraint method
Stimulus delivery	Match stimulus delivery devices
Personnel/screening	Training and documentation; protocol standardization
Acquisition parameters	Match acquisition parameters as closely as possible
Head coil	Match head coils

Noise Sources in Multiple Site Studies (Table 4) Scanning different subjects at multiple sites can greatly increase the number and diversity of the sample pool. However, it also introduces new sources of noise related to different hardware, research protocol, and personnel (Glover, Mueller, Turner, van Erp, Liu, Greve, Voyvodic, Rasmussen, Brown, Keator, Calhoun, Lee, Ford, Mathalon, Diaz, O'Leary, Gadde, Preda, Lim, Wible, Stern, Belger, McCarthy, Ozyurt, & Potkin, 2012). Scanner hardware can differ in many ways, with the two most prominent being field strength and manufacturer. Field strength directly affects the T_2^* and so systematically affects the size of the measured BOLD hemodynamic response. Even when the scanners have the same field strength, they may have been manufactured by different companies. There are details in the ways that the scanners are constructed, configured, and programmed that can cause differences in the results (Friedman, Stern, Brown, Mathalon, Turner, Glover, Gollub, Lauriello, Lim, Cannon, Greve, Bockholt, Belger, Mueller, Doty, He, Wells, Smyth, Pieper, Kim, Kubicki, Vangel, & Potkin, 2008). Many of these effects can be mitigated by appropriately selecting the scanner parameters to be as similar as possible across the scanners (Brown et al., 2011; Suckling, Ohlssen, Andrew, Johnson, Williams, Graves, Chen, Spiegelhalter, & Bullmore, 2008). Sites may have different RF noise environments (Greve et al., 2010) due to differences in room shielding and/or the presence of electronics (such as a projector) inside the room.

The subjects will likely have different experiences at different sites, and this difference may introduce systematic changes across site. For example, different sites may use different head restraints which can affect the amount of motion and subject comfort. Sites may also differ in the way they deliver stimuli to the subject (e.g., back projection vs. goggles) or record subject responses (e.g., keyboard vs. button box). Each site will also have its own personnel responsible for executing the study. Different personnel mean that recruiting and screening practices may be different resulting in a different group sample for each site. Finally, the locale of each site may differ in terms of access to chemicals that affect the hemodynamic response (e.g., caffeine, nicotine, alcohol, etc.).

Some site differences, such as magnet vendor, may have the same differential effect on all subjects. Other site differences, such as the ease of access to nicotine or caffeine, may cause the magnitude of between-subject differences to vary by site, producing site-by-subject interactions. All of this suggests that site be taken into account when performing the higher level analysis with data pooled from multiple sites. For a small number of sites (less than 6), modeling site as a fixed effect is recommended; for a larger number, a random effects model is possible (Glover et al., 2012). It is also suggested that the experimental design be balanced across site. For a thorough review of recommendations for all aspects of multisite fMRI, see work from the Function Biomedical Informatics Research Network (Glover et al., 2012).

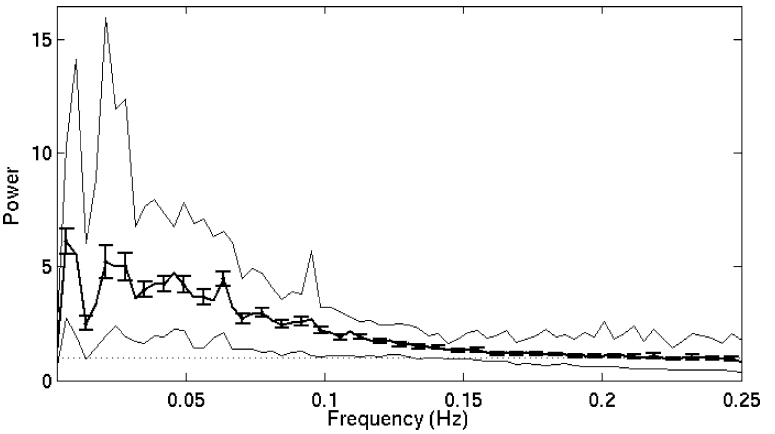

FIGURE 5.
Residual noise power spectrum averaged over gray matter. The *center dark line* is the mean across 18 subjects (with standard *error bars*). The *lines above* and *below* are the maximum and minimum across the 18 subjects. The values have been scaled so that the mean in the high frequency (white noise) range is 1. The prominent dip in the spectrum at 0.015 Hz is due to the fitting of the task waveforms shown in Figure 1, Panel A. The time series were demeaned resulting in zero power at DC.

8. Discussion

Noise is important to understand and quantify because it creates uncertainty and can lead to the drawing of both false positive and false negative conclusions. With the exception of model error and possibly motion, the first-level noise sources will not bias the expected value of the estimate of the HRF amplitude. They will, of course, make it more difficult to detect activation (i.e., increased false negatives). Aside from the thermal noise, the fMRI noise is generally heteroscedastic, temporally correlated, and nonstationary (Turner & Twieg, 2005). A power spectrum of the residual time series noise from the case study described in Section 2 is shown in Figure 5; the noise is clearly nonwhite, with much power in the lower frequencies. These nonideal statistical properties bias the estimates of the standard error σ_γ, and so bias the first-level t- and F-statistics by generally making them larger than they should be (false positives) due to the low-frequency nature of the noise (Bullmore, Long, Suckling, Fadili, Calvert, Zelaya, Carpenter, & Brammer, 2001). This biases estimates of the volume of activation computed by counting the number of voxels above threshold. The nonthermal noise generally accounts for about 80–90 % of the variance in fMRI time series (Greve et al., 2010); this makes thresholded statistical maps of an individual subject highly variable (McGonigle, Howseman, Athwal, Friston, Frackowiak, & Holmes, 2000). To some extent, this can be corrected by adding appropriate regressors or using of temporal whitening as discussed previously, but we recommend that the HRF amplitude estimate (or contrasts thereof) be used in higher level analyses (Smith, Beckmann, Ramnani, Woolrich, Bannister, Jenkinson, Matthews, & McGonigle, 2005) rather than supra-threshold voxel counts or other noise-dependent statistics. The magnitude of σ_γ may be affected by visit and group membership (e.g., schizophrenics may move more than healthy controls). This can result in heteroscedasticity at higher levels. To some extent, this can be reduced by including σ_γ in a mixed effect model.

At the higher levels, the sources of noise are also very complicated with poorly understood distributions. Permutation methods (Hayasaka & Nichols, 2003) are becoming more common to control for this. At the higher level, bias is more of a concern because of systematic differences between populations that can affect the HRF amplitude estimate in ways that do not relate to underlying changes in neural activation (e.g., HRF shape variability and anatomical variability).

Methodology to reduce the impact of these sources of noise is a very active area of research. While one always wants to reduce noise, one needs to consider the cost of reduction and weigh it against the potential benefit. For example, at the first-level, the noise is dominated by physiological effects and subject motion with thermal and instability noise playing progressively smaller roles (Greve et al., 2010; Triantafyllou et al., 2005). This suggests that measures to reduce the relative contribution of thermal noise (e.g., increasing field strength) may have very little effect in the final fMRI analysis (though increasing the field strength may be very beneficial for other MRI applications). Likewise, at the level where subjects are combined, the intersubject variance may account for 50–90 % of the total noise (Brown et al., 2011; Smith et al., 2005; Suckling et al., 2008), depending upon the number of stimulus presentations and other factors. Thus, efforts to reduce first-level noise of any kind, even when successful, might have very little impact at the higher levels.

9. Conclusion

The purpose of fMRI is to draw conclusions about neural activation in an individual at a single time point or across time points, across individuals, or across groups. Noise is introduced at each of these levels and causes uncertainty and/or bias in the final conclusions. At the lowest level, fMRI reveals changes in neural activation based on changes in the deoxygenation of blood (the BOLD effect). The analysis is performed by fitting the (known) stimulation time course with the BOLD signal at each point in the brain. This yields an estimate of the amplitude of the BOLD signal in response to the stimulus. This amplitude is then used as a surrogate of neural activation, though the relationship between the two is complicated and not well understood. Noise at this level manifests itself as error in the fit. This noise may be due to the scanner (thermal noise or instability), subject motion, deviations between the assumed and actual HRF shape, or any physiological effect that changes blood flow, but is not related to neural activation (e.g., respiration). Across days or weeks, variation can be caused by changes in the scanner, subject wakefulness level, and consumption of vaso-active chemicals such as caffeine. When comparing across subjects, the subjects' brains must be registered to a common space. Errors in this registration can cause error at the higher level analysis. Group differences in the underlying shape of the HRF can create differences in the higher level analysis even when no amplitude difference exists. In a multisite study, vendor and field-strength differences can also be a source of noise.

The field of fMRI is very rich in methodology, both in terms of acquisition and analysis, and this survey has only touched on the most general of these methods. fMRI data is also rich in noise sources, not all of which have been reviewed here, and some of which are still being discovered and understood. While noise in fMRI does represent a substantial challenge to its practical use, fMRI has been successfully used in a large number of studies to map the functions of the human brain.

Acknowledgements

Support for this research was provided in part by the National Center for Research Resources (P41-RR14075, R01 RR16594, P41-009874, the NCRR BIRN Morphometric Project BIRN002, and Functional Imaging Biomedical Informatics Research Network (FBIRN) U24 RR021382), the National Institute for Biomedical Imaging and Bioengineering (R01 EB001550, R01EB006758), as well as by the Department of Energy (DE-F02-99ER62764-A012) to the Mind Research Network (previously known as the MIND Institute).

References

Aguirre, G.K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*, *8*(4), 360–369.

Alsop, D.C., & Detre, J.A. (1996). Reduced transit-time sensitivity in noninvasive magnetic resonance imaging of human cerebral blood flow. *Journal of Cerebral Blood Flow and Metabolism*, *16*(6), 1236–1249.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113.

Beckmann, C., Jenkinson, M., & Smith, S. (2003). General multi-level linear modelling for group analysis in FMRI. *NeuroImage*, *20*, 1052–1063.

Beckmann, C.F., & Smith, S.M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, *23*(2), 137–152.

Birn, R.M., Diamond, J.B., Smith, M.A., & Bandettini, P.A. (2006). Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage*, *31*(4), 1536–1548.

Biswal, B., Yetkin, F., Haughton, V., & Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar imaging. *Magnetic Resonance in Medicine*, *34*, 537–541.

Brosch, J.R., Talavage, T.M., Ulmer, J.L., & Nyenhuis, J.A. (2002). Simulation of human respiration in fMRI with a mechanical model. *IEEE Transactions on Biomedical Engineering*, *49*(7), 700–707.

Brown, G.G., Mathalon, D.H., Stern, H., Ford, J., Mueller, B., Greve, D.N., McCarthy, G., Voyvodic, J., Glover, G., Diaz, M., Yetter, E., Ozyurt, I.B., Jorgensen, K.W., Wible, C.G., Turner, J.A., Thompson, W.K., & Potkin, S.G. (2011). Multisite reliability of cognitive BOLD data. *NeuroImage*, *54*(3), 2163–2175.

Brown, G.G., Perthen, J.E., Liu, T.T., & Buxton, R.B. (2007). A primer on functional magnetic resonance imaging. *Neuropsychology Review*, *17*(2), 107–125.

Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., Carpenter, T.A., & Brammer, M. (2001). Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Human Brain Mapping*, *12*(2), 61–78.

Burock, M.A., & Dale, A.M. (2000). Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. *Human Brain Mapping*, *11*, 249–260.

Buxton, R.B. (2009). *Introduction to functional magnetic resonance imaging: principles and techniques* (2nd ed.). New York: Cambridge University Press.

Buxton, R.B., Uludag, K., Dubowitz, D.J., & Liu, T.T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, *23*(Suppl 1), S220–S233.

Calhoun, V.D., Adali, T., Pearlson, G.D., & Pekar, J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, *14*(3), 140–151.

Calhoun, V.D., Stevens, M.C., Pearlson, G.D., & Kiehl, K.A. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *NeuroImage*, *22*(1), 252–257.

Chang, C., Cunningham, J.P., & Glover, G.H. (2009). Influence of heart rate on the BOLD signal: the cardiac response function. *NeuroImage*, *44*(3), 857–869.

Chang, C., & Glover, G.H. (2009). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *NeuroImage*, *47*(4), 1448–1459.

Cohen, M.S. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, *6*(2), 93–103.

Collins, D.L., Neelin, P., Peters, T.M., & Evans, A.C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*(2), 192–205.

Cox, R., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance Imaging*, *42*, 1014–1018.

D'Esposito, M., Zarahn, E., Aguirre, G.K., & Rypma, B. (1999). The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. *NeuroImage*, *10*(1), 6–14.

Dale, A.M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*, 109–114.

Fischl, B., Dale, A.M., Sereno, M.I., Tootell, R.B.H., & Rosen, B.R. (1998). A coordinate system for the cortical surface. *NeuroImage*, *7*(4), S740.

Fleisher, A.S., Podraza, K.M., Bangen, K.J., Taylor, C., Sherzai, A., Sidhar, K., Liu, T.T., Dale, A.M., & Buxton, R.B. (2009). Cerebral perfusion and oxygenation differences in Alzheimer's disease risk. *Neurobiology of Aging*, *30*(11), 1737–1748.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., & Raichle, M.E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678.

Friedman, L., & Glover, G.H. (2006). Report on a multicenter fMRI quality assurance protocol. *Journal of Magnetic Resonance Imaging*, *23*(6), 827–839.

Friedman, L., Stern, H., Brown, G.G., Mathalon, D.H., Turner, J., Glover, G.H., Gollub, R.L., Lauriello, J., Lim, K.O., Cannon, T., Greve, D.N., Bockholt, H.J., Belger, A., Mueller, B., Doty, M.J., He, J., Wells, W., Smyth, P., Pieper, S., Kim, S., Kubicki, M., Vangel, M., & Potkin, S.G. (2008). Test-retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, *29*(8), 958–972.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage*, *7*(1), 30–40.

Friston, K.J., Frith, C.D., Turner, R., & Frackowiak, R.S. (1995). Characterizing evoked hemodynamics with fMRI. *NeuroImage*, *2*(2), 157–165.

Friston, K.J., Holmes, A.P., & Worsley, K.J. (1999). How many subjects constitute a study? *NeuroImage*, *10*(1), 1–5.
Friston, K.J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, *1*, 153–171.
Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346–355.
Genovese, C.R., Lazar, N.A., & Nichols, T.E. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*, 870–878.
Glover, G.H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*(4), 416–429.
Glover, G.H., Li, T.Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*, *44*(1), 162–167.
Glover, G.H., Mueller, B.A., Turner, J.A., van Erp, T.G., Liu, T.T., Greve, D.N., Voyvodic, J.T., Rasmussen, J., Brown, G.G., Keator, D.B., Calhoun, V.D., Lee, H.J., Ford, J.M., Mathalon, D.H., Diaz, M., O'Leary, D.S., Gadde, S., Preda, A., Lim, K.O., Wible, C.G., Stern, H.S., Belger, A., McCarthy, G., Ozyurt, B., & Potkin, S.G. (2012). Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *Journal of Magnetic Resonance Imaging*, *36*(1), 39–54.
Greve, D.N., Mueller, B.A., Liu, T., Turner, J.A., Voyvodic, J., Yetter, E., Diaz, M., McCarthy, G., Wallace, S., Roach, B.J., Ford, J.M., Mathalon, D.H., Calhoun, V.D., Wible, C.G., Brown, G.G., Potkin, S.G., & Glover, G. (2010). A novel method for quantifying scanner instability in fMRI. *Magnetic Resonance in Medicine*, *65*(4), 1053–1061.
Handwerker, D.A., Ollinger, J.M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639–1651.
Hanson, L. (2008). Is quantum mechanics necessary for understanding magnetic resonance? *Concepts in Magnetic Resonance. Part A 32A*(5), 329–340.
Hayasaka, S., & Nichols, T.E. (2003). Validating cluster size inference: random field and permutation methods. *NeuroImage*, *20*(4), 2343–2356.
Henson, R.N., Price, C.J., Rugg, M.D., Turner, R., & Friston, K.J. (2002). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *NeuroImage*, *15*(1), 83–97.
Hopfinger, J.B., Buchel, C., Holmes, A.P., & Friston, K.J. (2000). A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. *NeuroImage*, *11*(4), 326–333.
Huettel, S.A., Song, A.W., & McCarthy, G. (2009). *Functional magnetic resonance imaging* (2nd ed.). Sunderland: Sinauer Associates.
Huettel, S.A., & McCarthy, G. (2000). Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. *NeuroImage*, *11*(5), 547–553.
Jezzard, P., & Balaban, R.S. (1995). Correction for geometric distortion in echo planar images from Bo field variations. *Magnetic Resonance in Medicine*, *34*, 65–73.
Jezzard, P., Matthews, P., & Smith, S.M. (2001). *Functional MRI: an introduction to methods*. Oxford: Oxford University Press.
Kelly, A.M., & Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, *15*(8), 1089–1102.
Kruger, G., & Glover, G.H. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*(4), 631–637.
Kumari, V., Gray, J.A., ffytche, D.H., Mitterschiffthaler, M.T., Das, M., Zachariah, E., Vythelingum, G.N., Williams, S.C., Simmons, A., & Sharma, T. (2003). Cognitive effects of nicotine in humans: an fMRI study. *NeuroImage*, *19*(3), 1002–1013.
Levin, J.M., Ross, M.H., Mendelson, J.H., Kaufman, M.J., Lange, N., Maas, L.C., Mello, N.K., Cohen, B.M., & Renshaw, P.F. (1998). Reduction in BOLD fMRI response to primary visual stimulation following alcohol ingestion. *Psychiatry Research*, *82*(3), 135–146.
Liang, Z., & Lauterbur, P.C. (2000). *Principles of magnetic resonance imaging: a signal processing perspective*. Piscataway: IEEE Press.
Liu, T.T., Glover, G.H., Mueller, B.A., Greve, D.N., & Brown, G.G. (2012). An introduction to normalization and calibration methods in functional MRI. *Psychometrika*.
Liu, T.T., Behzadi, Y., Restom, K., Uludag, K., Lu, K., Buracas, G.T., Dubowitz, D.J., & Buxton, R.B. (2004). Caffeine alters the temporal dynamics of the visual BOLD response. *NeuroImage*, *23*(4), 1402–1413.
Liu, T.T., & Frank, L.R. (2004). Efficiency, power, and entropy in event-related FMRI with multiple trial types. Part I: theory. *NeuroImage*, *21*(1), 387–400.
Logothetis, N.K., & Wandell, B.A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, *66*, 735–769.
Maus, B., Van Breukelen, G.J., Goebel, R., & Berger, M.P. (2010). Optimization of blocked designs in fMRI studies. *Psychometrika*, *56*, 1338–1352.
Maxim, V., Sendur, L., Fadili, J., Suckling, J., Gould, R., Howard, R., & Bullmore, E. (2005). Fractional Gaussian noise, functional MRI and Alzheimer's disease. *NeuroImage*, *25*(1), 141–158.
McGonigle, D.J., Howseman, A.M., Athwal, B.S., Friston, K.J., Frackowiak, R.S., & Holmes, A.P. (2000). Variability in fMRI: an examination of intersession differences. *NeuroImage*, *11(6 Pt 1)*, 708–734.
Miezin, F.M., Maccotta, L., Ollinger, J.M., Petersen, S.E., & Buckner, R.L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage*, *11*(6 Pt 1), 735–759.
Monti, M.M. (2011). Statistical analysis of fMRI time-series: a critical review of the GLM approach. *Frontiers in Human Neuroscience*, *5*, 28.

Nishimura, D.G. (1996). *Principles of magnetic resonance imaging*. Stanford: Standford University.
Payne, S.J. (2006). A model of the interaction between autoregulation and neural activation in the brain. *Mathematical Biosciences*, *204(2)*, 260–281.
Postelnicu, G., Zollei, L., & Fischl, B. (2009). Combined volumetric and surface registration. *IEEE Transactions on Medical Imaging*, *28*(4), 508–522.
Seber, G.A.F., & Lee, A.J. (2003). *Linear regression analysis* (2nd ed.). Hoboken: Wiley.
Shen, S., Sterr, A., & Szameitat, A. (2005). A template effect study on voxel-based morphometry in statistic parametric mapping. *Conference Proceedings (IEEE Engineering in Medicine and Biology Society. Conf.)*, *3*, 3051–3054.
Smith, S.M., Beckmann, C.F., Ramnani, N., Woolrich, M.W., Bannister, P.R., Jenkinson, M., Matthews, P.M., & McGonigle, D.J. (2005). Variability in fMRI: a re-examination of inter-session differences. *Human Brain Mapping*, *24*(3), 248–257.
Strother, S., La Conte, S., Hansen, L.K., Anderson, J., Zhang, J., Pulapura, S., & Rottenberg, D.A. (2004). Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. *NeuroImage*, *23*, S196–S207.
Suckling, J., Ohlssen, D., Andrew, C., Johnson, G., Williams, S.C., Graves, M., Chen, C.H., Spiegelhalter, D., & Bullmore, E. (2008). Components of variance in a multicentre functional MRI study and implications for calculation of statistical power. *Human Brain Mapping*, *29*(10), 1111–1122.
Thesen, S., Heid, O., Mueller, E., & Schad, L.R. (2000). Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magnetic Resonance in Medicine*, *44*(3), 457–465.
Thomas, R.J., & Kwong, K. (2006). Modafinil activates cortical and subcortical sites in the sleep-deprived state. *Sleep*, *29*(11), 1471–1481.
Tisdall, M., Hess, A., & van der Kouwe, A. (2010). *Prospective motion correction of anatomical brain sequences using EPI navigators*. Paper presented at the ISMRM Workshop on Current Concepts of Motion Correction for MRI & MRS.
Triantafyllou, C., Hoge, R.D., Krueger, G., Wiggins, C.J., Potthast, A., Wiggins, G.C., & Wald, L.L. (2005). Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *NeuroImage*, *26*(1), 243–250.
Turner, G.H., & Twieg, D.B. (2005). Study of temporal stationarity and spatial consistency of fMRI noise using independent component analysis. *IEEE Transactions on Medical Imaging*, *24*(6), 712–718.
van der Kouwe, A.J., Benner, T., Fischl, B., Schmitt, F., Salat, D.H., Harder, M., Sorensen, A.G., & Dale, A.M. (2005). On-line automatic slice positioning for brain MR imaging. *NeuroImage*, *27*(1), 222–230.
Vincent, J.L., Snyder, A.Z., Fox, M.D., Shannon, B.J., Andrews, J.R., Raichle, M.E., & Buckner, R.L. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of Neurophysiology*, *96*(6), 3517–3531.
Ward, H.A., Riederer, S.J., Grimm, R.C., Ehman, R.L., Felmlee, J.P., & Jack, C.R. Jr. (2000). Prospective multiaxial motion correction for fMRI. *Magnetic Resonance in Medicine*, *43*(3), 459–469.
Wise, R.G., Ide, K., Poulin, M.J., & Tracey, I. (2004). Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *NeuroImage*, *21*(4), 1652–1664.
Woolrich, M., Ripley, B., Brady, J., & Smith, S. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *NeuroImage*, *14*(6), 1370–1386.
Woolrich, M.W., Behrens, T.E., & Smith, S.M. (2004a). Constrained linear basis sets for HRF modelling using variational Bayes. *NeuroImage*, *21*(4), 1748–1761.
Woolrich, M.W., Jenkinson, M., Brady, J.M., & Smith, S.M. (2004b). Fully Bayesian spatio-temporal modeling of FMRI data. *IEEE Transactions on Medical Imaging*, *23*(2), 213–231.
Worsley, K.J., Liao, C.H., Aston, J., Pere, V., Dunackn, G.H., Morales, F., & Evans, A.C. (2002). A general statistical analysis for fMRI data. *NeuroImage*, *15*, 1–15.
Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., & Evans, A.C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*(1), 58–73.
Zhou, B., Konstorum, A., Duong, T., Tiue, K.H., Wells, W.M., Grown, G.G., Stern, H.S., & Shanbaba, B. (2012). A hierarchical modeling approach to data analysis and study desgin in a multi-site experimental FMRI study. *Psychometrika*.

Manuscript Received: 2 AUG 2011
Final Version Received: 7 MAR 2012
Published Online Date: 14 NOV 2012