

# Statistical Issues in fMRI for Brain Imaging

Nicole A. Lazar <sup>\*</sup>    William F. Eddy <sup>†</sup>    Christopher R. Genovese <sup>‡</sup>  
Joel Welling <sup>§</sup>

July 13, 1999

## *Summary*

Functional magnetic resonance imaging is a technique developed in the last decade and used in the fields of cognitive psychology and neuroscience, among others, to study the processes underlying the working of the human brain. In this paper we examine some of the statistical issues in functional magnetic resonance imaging for brain research. We start by giving a brief introduction to the physics of magnetic resonance imaging. Using a psychological experiment as a case study, we then describe questions of design and statistical analysis. The data obtained from functional magnetic resonance imaging studies are of a highly complex nature, displaying both spatial and temporal correlation, as well as high levels of noise from different sources. Given this, the scope for statistics is vast, and is not limited to simple analysis of the data, once collected.

---

<sup>\*</sup>Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A. Partially supported by NSF grant DMS-9705034.

<sup>†</sup>Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A. Partially supported by NSF grants IBN-9418982, DMS-9505007, and DMS-9705034, NINDS grant P01-NS35949, NIDA grant R01-DA11469, and NICHD grant P01-HD35490.

<sup>‡</sup>Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A. Partially supported by NSF grants DMS-9505007 and DMS-9705034, and NINDS grants P01-NS35949 and R55-NS36556 (Shannon Award), and NIDA grant R01-DA11469.

<sup>§</sup>Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A. and Pittsburgh Supercomputing Center, Pittsburgh, PA 15213, U.S.A. Partially supported by NICHD grant P01-HD35490.

# 1 Introduction

The study of the human brain and its functioning has long been of interest to psychologists, from perspectives of development, cognition, psychopathology and others. However, it is only in relatively recent years that methods have been developed that allow scientists to address these questions directly, through observation of the working brain. Functional magnetic resonance imaging (fMRI) is one of the most recently-developed of those techniques and is the focus of this article.

Statistical issues arise because of the nature of the questions being asked – for instance, what regions of the brain are active (and how active are they) while a person is trying to read and understand a difficult sentence – and also from the complexity of the data themselves. Data are collected as images of the brain over the time course of an experiment, resulting in a large amount of available information on the timing and location of neuronal activity, the analysis of which is made more complicated by the presence of both spatial and temporal correlations. Furthermore, there is a high level of noise in the images, coming from sources as variable as the equipment used to perform the scans, the movement of the subject within the scanner, and the effects of respiration and heartbeat. Indeed, the signal we are interested in measuring is small in comparison to this noise, making it all the more important that we understand the processes affecting the data. In addition, there are pertinent questions relating to the design of experiments, from both a statistical and psychological perspective. The potential uses for statistics in this arena are therefore great, as are the challenges.

We attempt in this paper to outline some of the main questions and issues inherent in the use of fMRI for the study of human cognition. Along the way, we will highlight areas that are still unresolved, focusing on the role of the statistician as part of the MR team, involved in everything from planning the study, through collecting the data, and devising analysis schemes.

In the rest of this section, we review the basic physics of fMRI and introduce an example experiment that will be used as a case study throughout the paper, to demonstrate various of the issues that arise and how they may be dealt with. Section 2 discusses design questions, including a survey of the major imaging parameters and their impact on the quality of the images obtained. In Section 3, we elaborate on data collection, while Section 4 deals with data processing, presenting some of the packages which have been developed for the analysis of fMRI data and giving a detailed account of one of them. Sections 5 and 6 treat the questions of statistical analysis and challenges that face statistics as a field within the brave new imaging world.

## 1.1 Basics of fMRI

To understand the statistical issues inherent in fMRI, it is essential to first gain an understanding of how the imaging process itself is thought to work. In this section, we outline the physics and biophysics underlying fMRI image acquisition.

Atomic nuclei with an odd number of protons or an odd number of neutrons are affected

by magnetic fields. Exposing such nuclei to a strong magnetic field will cause them to try to align themselves parallel or anti-parallel to the field. The parallel orientation has a slightly lower energy than does the anti-parallel orientation. This causes more nuclei to align themselves in the parallel orientation, which, collectively, results in an overall magnetization of the object in the field. The alignment of the nuclei isn't perfect in either direction; instead, the atoms *precess* about the field at a fixed frequency. "Precession" refers to the revolution of the axis of rotation of the atoms. This is a quantum effect, but the analogy to physical precession (such as a spinning top) is acceptable for our purposes. Frequency of precession varies for each type of nucleus and is related linearly to the strength of the magnetic field. When radiofrequency energy at the frequency of precession of the nuclei is injected into the system, the level of energy increases temporarily, but then returns to equilibrium. The energy emitted in the return to the starting state is at the frequency of precession. Both the absorption and the emission of energy are therefore *selective*, in that only nuclei that are near the appropriate precession frequencies will be affected. This is the key aspect of *resonance*. It is the selective absorption and emission of energy that produce the MR (magnetic resonance) signal. Magnetic resonance imaging (MRI) involves gathering data on the precession of the atomic nuclei, resulting in high-resolution images. The strength of the signal is proportional to the number of nuclei of a specific type. Hence the method allows us to count nuclei with particular properties. MR images are typically three-dimensional, representing volumes. The images are divided into volume elements, or *voxels*; the amplitude of the recorded signal at each voxel in each image is the average nuclear density of the chosen species (usually hydrogen). See Section 1.3 below for more detail on the process of collecting fMRI data.

With functional MRI, we use a series of MR images collected over time to gather information about neuronal activation in the brain during the course of a scan. While the scan is being performed, subjects may be asked to carry out various cognitive processing tasks; the images will then convey information on which regions of the brain were active and hence involved in the particular task under study. The connection between neuronal activation and the MR images is believed to be as follows. When resting brain neurons become active, the rate of blood flow to the neighborhood of these neurons increases, as glucose is delivered to the regions in question. This is known as the *hemodynamic response*. As the rate of firing increases for the neurons, their metabolism also increases. The increase in metabolism results in an influx of oxygenated blood to the affected region. Oxygen levels rise in the nearby blood vessels, since active neurons do not consume much more oxygen than when at rest. The magnetic properties of oxygenated and deoxygenated hemoglobin differ (as demonstrated by Pauling in 1935), and this difference affects the measured MR signal (Thulborn, Waterton, Matthews and Radda, 1982), through what is called the Blood Oxygenation Level Dependent (or BOLD) effect. Hence, the MR signal from the neighborhood of a neuron should change as the concentration of oxygenated blood around the neuron changes. MR imaging is sensitive enough to detect these functionally induced changes in blood oxygenation in the human brain (Ogawa *et al*, 1990; Ogawa *et al*, 1992; Kwong *et al*, 1992).

Sentence type	Example
common-simple	<i>The writer attacked the king and admitted the mistake.</i>
common-hard	<i>The writer that the king attacked admitted the mistake.</i>
rare-simple	<i>The pundit attacked the regent and admitted the gaffe.</i>
rare-hard	<i>The pundit that the regent attacked admitted the gaffe.</i>

Table 1: Examples of different sentence types for language study

The idea that blood flow changes can be correlated to changes in brain function is an old one (Raichle, 1994), presented as early as the end of the last century by the British physiologists Roy and Sherrington (1890). They postulated the existence of an “automatic mechanism” that regulated the blood supply to the brain in a manner dependent on variations in activity. Subsequent research has confirmed this hypothesis, although the exact nature of the system is still unknown. Functional magnetic resonance imaging is a step in the further understanding of this process.

## 1.2 An Example Experiment

The data which will be used to demonstrate the various ideas and principles outlined in the rest of this paper come from a cognitive experiment carried out by researchers in the Psychology Department at Carnegie Mellon University and the MR Research Center at the University of Pittsburgh Medical Center (Keller, Just, Carpenter and Thulborn, 1998). This experiment examined the role of syntactic and lexical difficulty in the comprehension of sentences. Both variables had two levels, resulting in four experimental conditions in all. These were: common words (high lexical frequency) in sentences with simple syntax; common words in sentences with difficult syntax; rare words (low lexical frequency) in sentences with simple syntax; and rare words in sentences with difficult syntax. Examples of each type of sentence appear in Table 1.

Subjects were all native (American-)English speakers. Sentences were presented to the subjects visually, after which they were required to answer a true/false comprehension question regarding who-did-what-to-whom. Questions of interest to the researchers centered on the effect of difficulty manipulations on the amount of measured activity in the brain in the left temporal language area, known as Wernicke’s area (see Figure 1). This can be seen as a proxy for the real focus, namely, trying to learn about how humans comprehend language. How do we process complex sentences to glean meaning from them? By measuring the amount of brain activation we hope to further our understanding of this process, as well as others that can be studied using fMRI technology.

Analysis of an initial eight subjects revealed a main effect of lexical frequency, a main effect of syntactic complexity and an interaction between the two. The main effects were in the direction that might be expected – sentences using the rare words induce more ac-

Figure 1: Left hemisphere of the brain, Wernicke’s area displayed (with label “W”).

tivation in the language area, as do sentences with greater syntactic difficulty. Intuitively, this is a reasonable finding – we have to work harder in order to understand more complex information, whether this takes the form of words we rarely encounter, or sentences with a more difficult structure. The interaction between the two factors was not reliable in that it disappeared when additional subjects were added to the study (see Keller, Just, Carpenter and Thulborn, 1998, for the analysis of the original eight subjects).

This type of experiment is indicative of the sophistication we are now able to attain and is a far cry from the early fMRI paradigms, which involved, for instance, flashing checkerboards on a screen. No active cognitive task was used in these pioneering experiments. A gradual building on earlier studies has brought the field to the point where we are now capable of asking – and potentially finding the answers to – questions of much greater impact in terms of understanding human cognition.

### 1.3 Fourier Space Data

In this section, we will give more detail of the procedure by which images are created. Ideas from Fourier analysis are relevant here, since data collected from the MR signal are actually sampled in Fourier space, which is called  $k$ -space in this context. We present a simplified explanation of the acquisition process, but we believe it gives a useful picture of the main ideas. For a more complete, technical description, see, for instance, Hinshaw and Lent (1983). The raw data that are collected include information on the amplitude, frequency and phase of the precessing nuclei. The scanner in which subjects are placed during the

course of an imaging experiment generates a very strong, static magnetic field (3.0 Tesla is the field strength in the center of the magnet we use). It is desirable for the static field to be as homogeneous as possible. We will assume that the scanning takes place in a homogeneous static magnetic field,  $\vec{B}_0$ . When the subject is put into the scanner, the axes of rotation of the protons (hydrogen nuclei) orient themselves to the general direction of the magnet or in the opposite direction. The axis of rotation is parallel to  $\vec{B}_0$ ; conventionally, this is the  $z$  axis in a three-dimensional space. There will be a slight advantage given to protons oriented in the direction of the magnetic field, resulting in the overall magnetization of the tissue volume, denoted by  $\vec{M}_0$ , which is a vector with direction and length (magnitude). Each atom precesses in a small tight circle around the pole, with random orientation, or phase. At this point, the atoms are out of phase with each other. They are, however, all precessing at a frequency that is proportional to the strength of the magnetic field. Call this frequency  $\omega_0$ . Rate of precession is related to the strength of the field by the *Larmor equation*,

$$\omega_0 = \gamma B_0.$$

Here,  $\gamma$  is the *gyromagnetic ratio* of the nucleus of the atom and  $B_0 = |\vec{B}_0|$  is the strength of the static magnetic field. The gyromagnetic ratio differs for different chemical elements. For hydrogen nuclei – protons – on which we will focus,  $\gamma$  is equal to  $4.26 \times 10^7$  Hz per Tesla.

Now suppose we inject energy into the tissue with a radiofrequency (RF) pulse, of frequency  $\omega$ . Resonance results when  $\omega$  and  $\omega_0$  match; in other words, only protons that are precessing near the rate of the irradiation frequency applied in the RF pulse will be affected. The effect of the radiofrequency pulse is to “tip” the atoms. The affected protons are all aligned in a single angle, called the “flip angle”, causing the rotation of  $\vec{M}_0$  away from its equilibrium orientation, toward the transverse plane that is perpendicular to the axis of the original field. Atoms that are tipped all precess in phase, as opposed to spinning randomly in a tight circle in the direction of the magnetic pole, as previously. From a static field, we have obtained a dynamic field, which can be measured. This dynamic field generates current in the receiver coils that is proportional to the number of hydrogen atoms in the tissue, as the atoms emit energy and return to their equilibrium state.

The actual acquisition of images proceeds as follows. The body is parallel to the field of the magnet; recall that by convention, this is the  $z$  axis. By applying a gradient  $G_z$  in magnitude to the magnetic field in this direction, a slice of the brain is selected, say  $z_1$ ; frequency of precession varies linearly with  $z$ . The gradient causes protons in the particular region to precess at a rate that will match the frequency of oscillation of the RF wave, whereas nuclei in other slices are precessing either too slowly or too quickly to absorb the energy (*i.e.* to resonate). Figure 2 demonstrates the slice selection process for three slice planes,  $z_1$ ,  $z_2$  and  $z_3$ . As the gradient applied changes, we focus on different slices of the brain tissue. Application of the gradient adds to the energy already in the field, so that protons at location  $z_1$ , for instance, have a different resonant frequency than protons at location  $z_2$  ( $\omega_1$  and  $\omega_2$ , respectively).

Once a slice is chosen, we have to acquire data from tissue lying in the  $x$ - $y$  plane, which is perpendicular to  $\vec{B}_0$ . Think of the tissue in this slice as being divided into a matrix of rows

Figure 2: The slice selection process. When the gradient is applied, the total magnetic field to which a proton is exposed will depend on its location, according to the Larmor equation  $\omega(z_i) = \gamma(B_0 + G_z z_i)$ . At location  $z_i$ , tissue absorbs energy with frequency centered at  $\omega_i$ .

and columns. Immediately after the protons have been tipped, they are all in phase. Before beginning to record the current induced by the magnetic field, in gradient coils located in the scanner, we apply a *phase gradient* to the magnetic field, increasing it linearly as we go from the bottom row to the top row. This causes the atoms in the top row to precess slightly faster than those in the next-to-top row, and so on down. Within a row, protons are still in phase with each other, but different rows are in different relative phase positions (see Figure 3). As the data are being read, a frequency encoding gradient is added, increasing incrementally along the columns. Each column is now precessing at a different frequency and thus each element of the matrix, corresponding to a different location in the tissue, can be distinguished from the others, having a different phase (giving the  $y$ -coordinate) and frequency (giving the  $x$ -coordinate).

The effects of this may be seen by examining an expanded version of the basic Larmor equation, namely

$$\omega_i = \gamma(B_0 + \vec{G} \times \vec{r}_i),$$

where  $\omega_i$  is the frequency of the proton at position  $r_i$  and  $\vec{G}$  is a vector representing the amplitude and direction of the gradient (Hinshaw and Lent, 1983; Brown and Semelka, 1995). The implication of this equation is that each proton will resonate at a unique frequency in the presence of the gradient field, as also explained above. The magnetic resonance image is a frequency map of the protons as generated by the particular magnetic fields determined by the gradients applied at each point throughout the image.

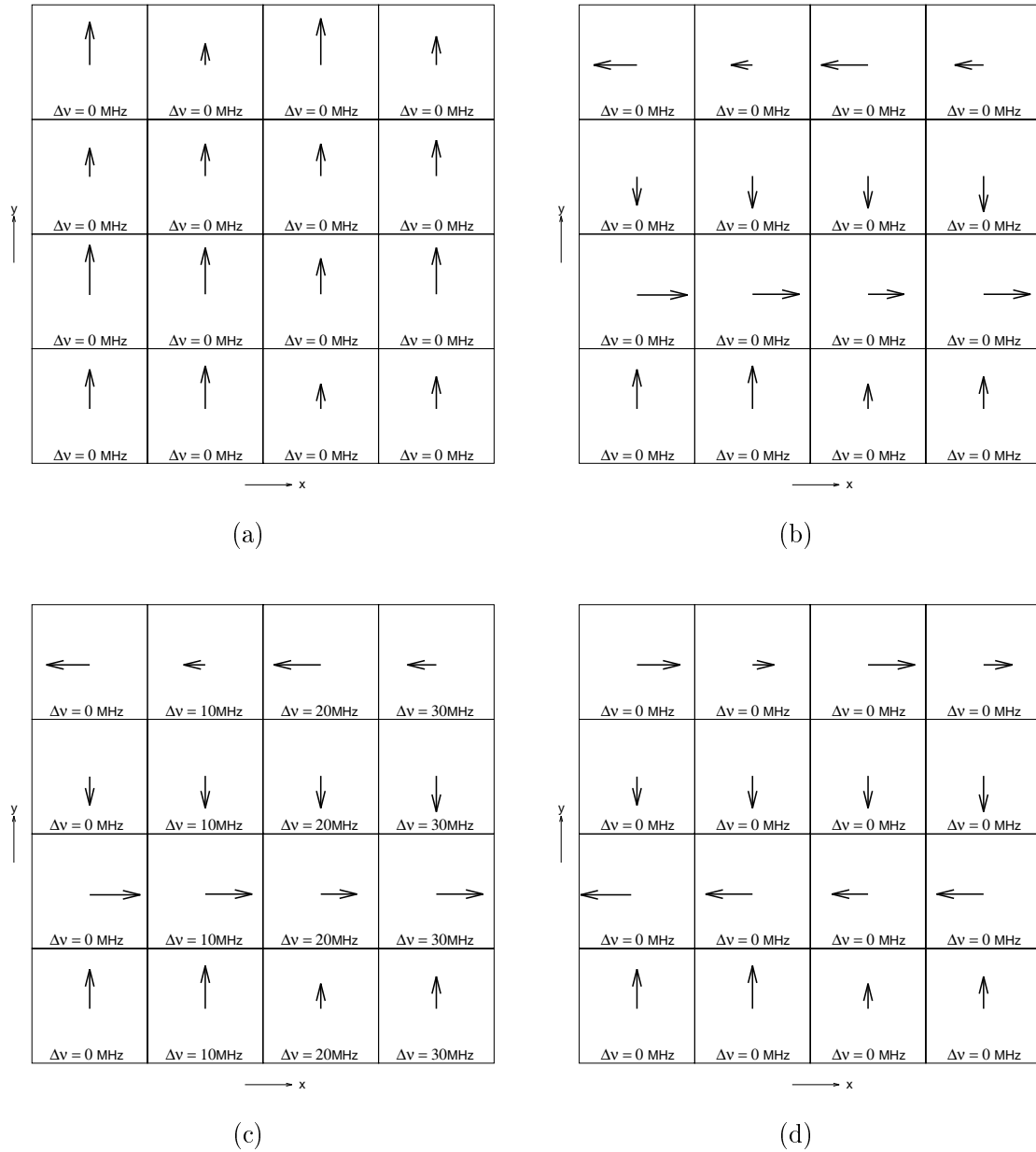


Figure 3: Schematic of the effect of phase and frequency gradients. Application of these gradients in sequence causes each location to be precessing at a distinctive phase and frequency. This allows us to select the areas for imaging. (Reprinted from Genovese and Sweeney (1999) with permission of Springer-Verlag New York, Inc.)

Relaxation is the process by which protons release the absorbed energy and return to equilibrium, in an exponential decay. There are two types of relaxation of interest for MR imaging. The first,  $T_1$ , is called *longitudinal relaxation time* and is the time required for the  $z$  component of the excited field  $\vec{M}$  to return to 63% ( $1 - e^{-1}$ ) of its original value after an excitation pulse (Brown and Semelka, 1995). Another way to think about  $T_1$  relaxation is that it is the mechanism by which the protons release their absorbed energy and return to their original orientation. The second relaxation type is *transverse relaxation time*, denoted  $T_2$ , which is the time needed for the transverse component of  $\vec{M}$  to return to 37% ( $e^{-1}$ ) of its initial value immediately following the RF pulse. This relaxation is the result of the gradual loss of phase coherence; local irregularities move the precessing atoms out of phase with each other, reducing the signal. The transverse relaxation time,  $T_2$ , is typically much shorter than the longitudinal relaxation time,  $T_1$ , since phase coherence must be lost before equilibrium can be reached. The values of  $T_1$  and  $T_2$  differ for grey matter, white matter and cerebrospinal fluid; substances with longer  $T_2$  times will appear brighter (in  $T_2$ -weighted images) than substances with shorter  $T_2$  times. We can take advantage of this fact to weight the images to produce greater contrasts.

Now, recall that our goal is to extract information regarding the density of hydrogen nuclei at each location. Call this quantity  $\rho_{jk}$ , with  $j$  varying in the  $x$  direction and  $k$  in the  $y$  direction. Denote by  $\nu_0$  the known frequency of precession of protons in the main magnetic field. This value will be used in the slice selection process. Let  $\phi$  be the phase increment, which determines the voxel size, the gradient applied to the magnitude of the main field in the  $y$  direction, and the length of time for which the gradient is applied. Similarly, we have a frequency increment,  $\nu$ , which determines the voxel size, the gradient applied to the magnitude of the main field in the  $x$  direction, and the length of time for which the gradient is applied.

After one pass of slice selection and encoding for phase and frequency, the result is a complex combination of the quantities we wish to recover. We repeat this process so that all of Fourier space is sampled, according to the size of the image being taken. If, for instance, we were interested in a  $128 \times 64$  pixel image, which is a typical value (see below), then we could make one path through  $k$ -space, taking  $128 \times 64 = 8192$  samples. Since gradients move continuously – they can’t jump around from place to place in  $k$ -space – the type of sampling we have described imposes a particular pattern that will be followed by the sampling path. In fact, the velocity and acceleration of the gradients are limited by FDA regulation and this is one factor in determining the types of sampling paths that can be used. The trace of a particular path is given in Figure 4 and the resultant imaging method is called “echo planar imaging”. This gives us a matrix  $I_{s,t}$ , of the form

$$I_{s,t} = e^{2\pi i \nu_0 t} \sum_j \sum_k \rho_{jk} e^{2\pi i (j\nu t + k\phi s)},$$

whose elements are the Fourier transform of the matrix of  $\rho$  values. To recover  $\rho_{jk}$ , then, just apply the inverse Fourier transform to the signal as measured. Note that, because we used echo planar imaging, samples are taken on a rectangular grid and we can use the fast



Figure 4: Schematic of an echo planar imaging path in  $k$ -space.

Fourier transform algorithm, which results in a great reduction in processing time. For other sampling schemes, such as those that go through  $k$ -space on a spiral, the fast Fourier transform isn't directly applicable without a gridding step.

If we think of putting the slices together, to form a three-dimensional picture of the brain, then the data consist of “snapshots” of the working brain over time. For each voxel we therefore have a time series, representing the activation level in that location in the brain of the subject, over the course of the experiment. A common approach to analysis involves looking at these individual time series. Other approaches, such as “synthetic magnetic resonance imaging” (Glad and Sebastiani, 1995), are also in use. See the discussion in Section 5 below for more detail.

## 2 Design Issues

As in every scientific pursuit, there are issues of design that have to be addressed by both the experimenters and the collaborating statisticians. The purpose of this section is to highlight some of the questions that we face in designing an fMRI study, from the points of view of the psychologist, the statistician and the imaging technologist.

### 2.1 Psychological Design Issues

The types of issues we have to consider from the psychological point of view clearly depend on the experimental questions we are interested in addressing. One important point to consider is that of how the stimuli are to be presented to the subjects. For instance, if the experimental task is one of sentence comprehension, we might present the stimulus visually, on a screen inside the scanner, in complete sentences or one word at a time; or we might read

the sentence out loud and have the subject listen through headphones. Modality of stimulus presentation could have an effect on the results and this needs to be thought about prior to the start of an investigation. Furthermore, the fact of physically being in the scanner, as opposed to the natural environment of the individual, could impact psychological or cognitive functioning. This is especially relevant in studies of various psychopathologies, such as autism or schizophrenia. In any case, steps should be taken to make the scanning environment as comfortable and free of stress as possible.

An important issue that arises here, even for normal subjects in studies of cognition, is that of head motion. While the imaging is taking place, the person inside the scanner needs to remain as still as possible. The reason for this is the time-series nature of the data; we are interested in comparing voxel-wise performance over time, as the subject undertakes different cognitive tasks. In order to do this, we need to know how voxels correspond to locations in the brain. This will allow us to compare the level of activation in a particular area as the experiment unfolds. If the head is moving around, there is a blurring effect in the data, as the same voxel will now represent different locations. Since voxels are very small (see Section 3.2 below), even small amounts of head motion are enough to corrupt the data and make inference more difficult. As we will see later, there are statistical methods available for estimating and correcting the effects of head motion; however, it is better to try to reduce head motion during acquisition, rather than simply including it as a source of variability during data analysis. It is interesting to note that head motion can be correlated with the cognitive tasks being performed, such as when the subject shifts her head – rather than just her eyes – to focus on a visual stimulus (Hajnal *et al*, 1994; Hajnal, Bydder and Young, 1995). Additionally, there is anecdotal evidence that head motion increases with the difficulty of the presented task. Children, the elderly, and clinical patients also tend to exhibit significantly greater head motion than other subjects.

Yet another question of interest relates to the proper baseline condition against which to compare the experimental conditions. There are some unexpected subtleties here. For instance, in the Keller *et al* study which we are using as our prototypical example, the baseline condition consisted of subjects fixating on an  $\times$  in the middle of the screen, for a period of 24 seconds. The fixation symbol is a letter of the alphabet, specifically, the letter x. It is possible, then, that since the task at hand involves language processing, the “baseline” condition would actually cause activation in some of the same areas of the brain as would the experimental stimuli. Perhaps for this study, a + sign would have been a better fixation target, or, perhaps, it doesn’t matter.

Much of fMRI analysis (and indeed, psychological research in general) is based on the assumption that tasks being compared differ only along the dimension of interest, and hence that effects can be estimated by subtracting results. Price, Moore and Friston (1997) discuss this in the context of brain imaging. In reality, as pointed out there, it can be hard to find baseline tasks that activate in an appropriate manner. Another potential difficulty with the subtraction paradigm is that sometimes negative effects are found, implying, for instance, a deactivation in the experimental state relative to the baseline. In other words, the experimental condition induces less activation than the control. Such phenomena are

generally taken to be artificial, or random fluctuations. However, they could also underscore problems with the subtraction methodology itself. When using this approach, experimenters have to assume that they have a deep enough understanding of the process being studied to be able to introduce subtle changes between a control and an experimental condition, and that the difference is only along the dimension of interest. This is not necessarily going to be the case, and if the condition of interest differs from the baseline by more than just a small amount, the effect detected by subtraction could be confounded by other effects. The use of *graded stimuli* might be a way of reducing the possibility of new mechanisms entering into the picture, but even here there is a need to take care. For instance, in a memory task, it might be reasonable to think that there is a very precise, small-scale change that occurs when a subject is asked to remember a single digit as opposed to remembering two digits. A similar change might occur for escalating from two digits to three. However, we can't continue increasing the load on the working memory in a systematic fashion and expect that the same mechanism will hold. There are limits on how many bits of information human beings can hold in working memory; beyond that threshold, different mnemonic techniques and devices are needed. Remembering nine digits requires different mental strategies than remembering three and even graded increases in load, coupled with the subtraction methodology, might not be able to pinpoint this effect.

## 2.2 Statistical Design Issues

From the statistical vantage point, planning an fMRI study involves all of the usual precautions, such as randomization, controlling for extraneous variables, and so forth. Two special questions relate to sample sizes and to time.

In terms of sample size, brain imaging studies are usually performed on a relatively small scale, six to eight subjects not being at all uncommon. Reported studies are often smaller than that and results from even a single subject also appear in the literature (see, for instance, Josephs, Turner and Friston, 1997). The cost and time required to carry out experiments are limiting factors in the scope of the studies and this needs to be accounted for at the design stage. On the other hand, a large amount of data is acquired for each subject, as we shall see in the next section.

Use of the scanner is quite expensive – about \$500 per hour, not counting the payment to the subjects and salaries for technicians. A typical experiment takes an hour and a half to run, per subject. Out of this, data might be collected for no more than thirty or forty-five minutes. The rest of the time is spent on setting up and calibrating the equipment, a cost (in both time and money) that is necessary at the outset but doesn't always carry over if consecutive subjects are scanned. Questions of greater complexity also require more data, necessitating both more subjects and longer scanning times. For instance, response to flashing checkerboards, a standard visual stimulus, is a reliable one and the location of the activated area is known. The resources needed to establish this response are small compared to a study that is trying to localize and quantify activation for a less well understood phenomenon, such as mental rotation of three-dimensional objects. It is important to design

experiments in such a way that the scientists can investigate particular questions of interest, while making efficient use of the time and equipment available. As statisticians, we can – and should – take an active role in this stage of the planning, to help ensure that the resultant study will be sound.

The sentence comprehension study initially used only eight subjects, all right-handed, all native speakers of English. Half of the subjects were female. The sentences had two levels of syntactic complexity crossed with the two levels of lexical frequency. There were four repetitions of each sentence-type; these were interleaved in a pseudo-random fashion with six repetitions of a 24-second baseline condition, during which subjects were required to fixate on an  $\times$  in the center of the screen. The size and design of this study is typical of those that we are now able to conduct. Using only right-handed subjects is also typical in language studies, due to hypothesized differences between the functional organization of the brain in left- and right-handed individuals.

## 2.3 Imaging Design Issues

The design questions from the point of view of the imaging process have to do with the various parameters that can be set by the operator of the scanner (the technologist). These parameters will determine temporal and spatial resolution and the general quality of the images that result. Now that we have covered the basics of image acquisition and experimental design, we turn to a discussion of some of the more important data acquisition parameters and their implications in terms of image quality.

“Imaging plane” refers to the plane in which the slices are taken. The standard choices of image plane are axial, coronal and sagittal. Axial slices are taken perpendicular to the longitudinal axis of the body; coronal (or frontal) slices are taken parallel to the front face of the body; sagittal slices are taken parallel to the center line of the skull. It is also possible to acquire in an oblique plane, in which the angle of the slice is tipped out of one of the standard orientations.

The size of the image in a direction (either frequency or phase encoded) is the “field of view”. These parameters specify the region from which the signals were sampled. By decreasing the FOV, higher resolution may be achieved. This comes at the expense of the signal-to-noise ratio (Brown and Semelka, 1995). “Sampling bandwidth” is the reciprocal of the elapsed time between successive samples of the MR signal.

Two other time-related parameters are “time to repetition” (TR) and “time to echo” (TE). Using the echo planar method, TR is the time between successive images. TE is the time elapsed between the original pulse and the peak of the signal, or echo, that is recorded as an image. Long TE times allow for more dephasing of the protons. Estimating the effect of changing TE and TR is an aim of the synthetic MRI mentioned in Section 1.3 above.

The excitation (or flip) angle sets the amount of rotation away from the equilibrium axis experienced by the field following the RF excitation pulse. The default value for the excitation angle is  $90^\circ$  in most scanners, as this gives the maximum magnetization in the transverse plane. Together with TR, this parameter helps determine the amount of  $T_1$ -

weighting.

Finally, there is a set of parameters that relates to the number and size of images collected – the number of phase and frequency encode steps, slice thickness and the number of excitation pulses which are used to generate each image. Phase and frequency encoding steps help determine the size of the image (what area of the brain is being scanned, for instance), whereas the other two parameters give the number of slices obtained from a particular series of scans. Slice gap is the space between slices. Leaving a gap between consecutive slices is a way of protecting against flaws in the radiofrequency pulse sequence in that it lessens the amount of overlap between successive slices.

In our case study example, the functional images were acquired using a 3 Tesla GE Signa scanner. An EPI pulse sequence was used, with TR=3000 ms, TE=25 ms and a flip angle of 90 degrees. Fourteen oblique axial slices were acquired; these had a slice thickness of 5 mm, a 1 mm slice gap, FOV of  $40 \times 20$  cm and a matrix of size  $128 \times 64$  voxels per slice. Values for all parameters are taken from Keller *et al*, 1998.

A description of the various tradeoffs in terms of image quality and signal-to-noise ratio resulting from varying the values of the parameters may be found in Brown and Semelka (1995, Chapter 6).

## 3 Data Collection

Having explored some of the scientific aspects of functional magnetic resonance imaging, we are now in a position to consider a number of the more practical facets of the endeavor. What are the data collection systems and how are they connected? What is the size of a typical data set from an fMRI experiment?

### 3.1 How Are the Data Collected?

Schematically, we can think of the system used to collect the data as comprising a number of subsystems. These are: the magnet system, the computer, the gradient system, the RF system and the data acquisition system (Brown and Semelka, 1995; National Research Council, Institute of Medicine, 1996). A rough diagram of these subsystems and the connections between them appears in Figure 5.

An MR scanner is built around a large, cylindrical superconducting magnet, immersed in liquid helium. This magnet, commonly made of niobium-titanium alloys, provides a very strong and uniform magnetic field, through the central core where the subject is placed. Indeed, the field strength and uniformity are key determinants of image quality. Since any body placed in the scanner distorts the main magnetic field, a delicate process called *shimming* is required to correct the field and restore image quality. The strength of the magnet requires that there be careful electromagnetic shields around the scanner room and strict safety procedures. The former protect external electronic equipment and people with, for example, pacemakers (aneurysm clips are equally bad); the latter prevent metallic objects from being drawn into the magnetic field and injuring or killing the subject. Gradient

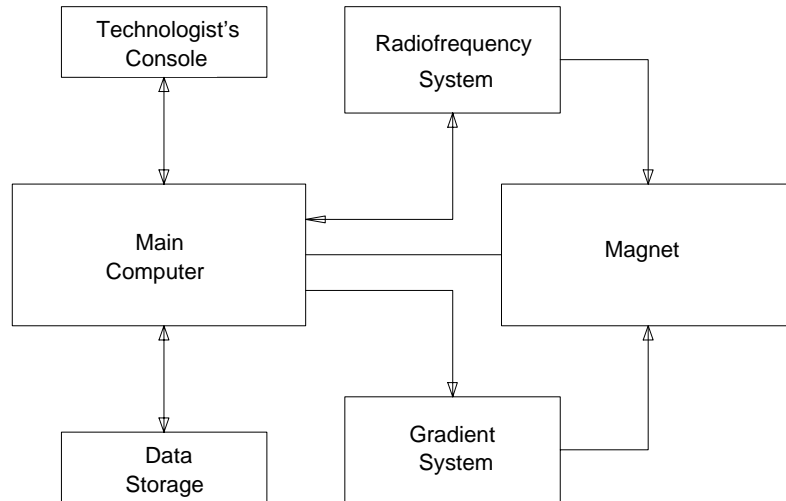


Figure 5: Subsystems comprising a typical MR system.

coils surrounding the magnet produce the  $x$ ,  $y$  and  $z$  field gradients used to encode spatial information into the MR signal. The subject is placed inside the scanner, with radiofrequency transmitter and receiver coils around the head. The scanner can also be equipped with a screen on which words or pictures can be projected, speakers or headphones through which sentences can be read and so forth, depending on the experiment in question, though all equipment must be non-magnetic and electrically shielded. Figure 6 is a picture of the scanner we use in our experiments, with our favorite technologist setting up the equipment, and gives some idea of the environment in which subjects would find themselves during an experiment.

A technologist configures and operates the scanner from an adjacent room. This involves the use of a complicated, computer-controlled, touch-screen console, with many layers of screens. From this console, the gradient and RF pulses are controlled and data acquisition schemes are determined. This includes choosing the parameters guiding the particular experiment, as described above; by directing current through the gradient coils, the requisite fields are produced. Moreover, since some subjects may become anxious while in the scanner (whereas others may fall asleep!), the technologist maintains an open line of communication with the subject.

The radiofrequency system generates and transmits the energy that excites the protons. Coil shape has a strong effect on how the energy is dispersed spatially into the body and consequently on image quality. Indeed, the design of new coil systems is an active area of research (National Research Council, Institute of Medicine, 1996; Turner, 1993).

Finally, there is the data acquisition subsystem, which measures the signals emitted from the protons and digitizes them for processing. A receiver coil is used to detect the voltage induced from the protons after the RF pulse. Size and shape of the receiver coil varies, but the effective field should be perpendicular to  $\vec{B}_0$ . Smaller coils are more sensitive than larger ones. For brain imaging studies, the transmitter coil of the radiofrequency system



Figure 6: An MR scanner showing a head restraining cage and the table where a subject lies before entering the magnet.

usually also serves as the receiver coil. Since the signal produced by the protons is small, it has to be amplified. It is then demodulated, filtered and has the real and imaginary parts separated from each other. Following this process, the signal is converted by the analog-to-digital converter, which digitizes each analog signal at a rate dependent on the sampling time and the number of data points. The digitized data are stored, so that they may later be processed.

### 3.2 Size of the Data

The data sets coming out of a functional MRI study are typically quite large. One of the issues that comes into play here is the size of the voxels, which is in essence a question of the spatial resolution of the data. The height and width of a voxel in a typical study will be between 1 and 3  $mm$ ; slice thickness varies between 3 and 5  $mm$  in general. Therefore, voxel size can range from the order of 3  $mm^3$  to 40  $mm^3$ ; smaller voxels will result in a finer spatial resolution but bigger data sets and often lower signal-to-noise ratio, while with larger voxels the data sets are smaller and the spatial resolution is correspondingly poorer. A three-dimensional image will consist of a number of slices (typical values in our studies are currently between 7 and 14), each on a square array of voxels, of size  $n \times n$ , where  $n=64$ , 128 or 256 are common choices. Depending on the resolution, then, the number of

voxels accounted for in a study can vary, but will be on the order of tens to hundreds of thousands. For each voxel, we can now record time series of hundreds of images. We are therefore dealing with data sets that include millions of data points, for a single experiment on a single subject.

It should be pointed out that there is a certain amount of local variation in these numbers. That is, there are limitations imposed by the particular hardware used in any MR center, beyond the obvious physical limitations. The numbers we report are for the equipment we use in Pittsburgh, but are reasonably representative. For instance, the equipment available to us permits us to take up to five slices per second without overheating the equipment. It is necessary to take slices and not do direct three-dimensional imaging because the latter currently takes too long and the signal would decay before a complete sampling of  $k$ -space could be obtained. These type of local restrictions also determine the size of the data sets.

In our example, the voxel size is  $3 \times 3 \times 5mm$  and we have 300 images per voxel, taken over 14 slices, on a  $128 \times 64$  matrix, giving us 34,406,400 data points to be analyzed per subject.

## 4 Data Processing

### 4.1 Packages for the Analysis of fMRI Data

The recent explosion of fMRI as an experimental tool has led to a proliferation of computer softwares for the presentation and analysis of data. In this section we give a brief survey of some of the available packages. A comparison of the features found in different softwares may be found in Gold *et al* (1998). Computer packages are frequently updated and improved, as advances in this field are rapid. Several of the software packages discussed in Gold *et al* (1998) have gone through further releases since that paper appeared. Hence readers who want to investigate specific programs should check the relevant webpages, which we have included here for easy reference. Additionally, many researchers and laboratories develop their own software for analyzing fMRI data; these are not examined here.

AFNI (Analysis of Functional Neuroimages) (Cox, 1996; Cox and Hyde, 1997) is a flexible package that allows graphical display of image data and analysis using the correlation method, developed by Bandettini *et al*, among others (see below). The problem of multiple comparisons, owing to the large number of tests that are performed, is addressed using Bonferroni corrections. “Plug-in” modules are available to help users customize their analyses and extensive documentation on these and other aspects of the program can be found on the AFNI webpage, accessible from <http://varda.biophysics.mcg.edu/~cox/index.html>.

Statistical Parametric Mapping, or SPM (Friston, Frith, Liddle and Frackowiak, 1991), was originally developed for Positron Emission Tomography, another imaging technique, and was extended to fMRI. The approach used by this package is voxel based, assuming a parametric statistical model at each voxel. General linear models describe the variability in the data in terms of experimental and confounding effects and residual variability. At each voxel, hypotheses regarding the model parameters can be assessed and images can be created

based on the calculated test statistics. SPM addresses the issue of multiple comparisons using the theory of continuous random fields. There is extensive documentation about SPM at the site <http://www.fil.ion.ucl.ac.uk/spm/>, as well as an email discussion list where users can discuss questions relating to the software. Recently, a new toolbox was added to SPM, SnPM (Statistical nonParametric Mapping), which uses randomization and permutation tests to determine significance levels for each voxel.

Gaschler, Schindler and Scheich (1996) report on the post-processing system KHORFu, which was developed for the analysis of various kinds of neuroimaging data, including, but not limited to, fMRI. KHORFu is modular, incorporating pre-processing, statistical analysis (on a voxel by voxel basis) and post-processing options. Main features of this software package include an interactive graphical interface and an on-line post-processing system, which allows users to see statistical activation maps superimposed on anatomical images during the course of an experiment.

Another package that is in widespread use is FIASCO (Functional Imaging Analysis Software: Computational Olio) (Eddy, Fitzgerald, Genovese, Mockus and Noll, 1996), a modular system for the analysis of functional imaging data. We discuss FIASCO in greater detail in Sections 4.2 and 4.3.

As opposed to the other packages mentioned here, which are all freely available, MEDx (<http://www.sensor.com/>) is a commercial software system. SPM is included as a feature of MEDx, as is AIR (Automated Image Registration; Woods, Cherry and Mazziotta, 1992). As noted previously, head movement while subjects are in the scanner is a serious source of noise in fMRI data; AIR was developed to adjust for this effect. A variety of statistical methods are built in to the software, as are tools for easy visualization of the data.

This is only the tip of the iceberg as far as software for analyzing fMRI data is concerned. We again refer interested readers to the paper by Gold and colleagues for a discussion of some other packages, as well as to the web sites mentioned there and in the previous paragraphs for the most current versions.

## 4.2 Preprocessing of Data and FIASCO

We now turn to the details of the computer package we use to process and analyze fMRI data. As should be clear from the brief summary in the previous section, this represents one among many possible approaches to the problem of treating imaging data. Readers who are not interested in the technicalities of our software may skip to Section 5.

Analyzing the data from an imaging experiment is a daunting task. As technology improves and experience grows, the types of questions that cognitive scientists can consider become more complicated. Hardware advances also sometimes reveal difficulties that weren't apparent with more primitive equipment. Our data processing/statistical analysis system, FIASCO, was developed as a flexible, adaptive package. As such, it is built out of modules that allow the user a great deal of freedom in determining the order in which preprocessing and processing steps will be carried out and even which steps will be executed, according to the specifications and needs of the experiment in question. Various data collection modes,

such as echo planar imaging, spiral imaging and other forms, are available for the user.

The FIASCO homepage is at <http://lib.stat.cmu.edu/~fiasco>. In this section we describe the preprocessing done by the components of FIASCO and the problems they are designed to address. Preprocessing is useful prior to statistical data analysis because of the noisiness of the data, relative to which the signal is small, and because we understand some of the sources of noise. Ideally, we would like to include the preprocessing as part of the model fitting, but there are both statistical and computational barriers to this. The tools will be discussed roughly in the order in which they are used, so that we can demonstrate their effects on data from our experiment. Since not all steps have interesting, visually discernible effects on the output, we don't display all of them here. We give the names of the FIASCO modules in order to aid the reader who is interested in consulting the software's webpage.

The main steps in the preprocessing are: baseline correction, deghosting, mean correction, motion correction, outlier adjustment and detrending.

The first element in the preprocessing is *baseline correction*; this is achieved through the `baseline` step. This function corrects for the location of the data in Fourier space and reverses lines, which is needed for echo planar data collection (recall that echo planar data are obtained by traveling along  $k$ -space in the  $x$  direction, then moving down a little on the  $y$  axis, traversing now in the opposite direction of  $x$  and so on. That is, every other line, we reverse direction; see Figure 4). The baseline correction subtracts the mean of the Fourier coefficients in certain parts of  $k$ -space (specifically, the parts that are away from zero) from all of the coefficients in the image. Line reversal allows the user to specify which lines should be reversed – all, none, only the odd lines or only the even lines. Typically, axial and sagittal slices will undergo reversal of the odd lines, while reversal of the even lines is applied to coronal slices. Why is it necessary to do baseline correction? Having a non-zero mean would concentrate all of the information at the first Fourier coefficient, corresponding to (spatial) frequency zero, giving an image in which a single bright dot in the center overwhelms everything else. Subtracting out the mean allows the other coefficients to be visible.

Next is `deghost`, which can come either before or after mean correction. We usually perform this step after baseline correction and prior to mean correction. The purpose of *deghosting* is to remove “ghosts” in the images. Where do these ghosts come from? When we do line reversal for echo planar imaging, since the zero spatial frequency in the  $x$  direction might not be exactly in the center of each line, there is a shifting of elements from where they should be. This is, effectively, a small mis-timing in the gradients. The “even” ghosts that result are simply faint images of the brain that appear in what should be the reconstructed image of the air surrounding the head in the scanner (there are also “odd” ghosts, but these are not currently well-understood).

In Figure 7 we show the effect of deghosting on the mean images produced by FIASCO. As can be seen by comparing the *before* (Figure 7(a)) and *after* (Figure 7(b)) mean images, `deghost` does actually succeed in improving the visual quality; specifically, the “ghosts” apparent in the *before* image are absent from the *after* image. We also present a difference image for the means prior to and following the use of the deghosting feature (Figure 7(c)).



Figure 7: (a) (b) (c)  
(a) Mean image before deghosting (b) Mean image after deghosting (c) Difference in mean images before and after deghosting. In panels (a) and (b), blacker values are larger and white is zero. In panel (c), blacker values are positive and whiter values are negative; grey is zero.

When the ghosts have been removed from the image, the signal is reduced outside the brain and increased inside.

*Mean correction*, achieved through the `meanc` component of FIASCO, adjusts each image to have roughly the same mean, for each image corresponding to a given slice (this translates to the  $z$ -coordinate of the data). This corrects for overall “drifts” in the intensity of the MR signal. To perform this correction, the mean value is calculated for the central quarter of each image and the reciprocal of the ratio of this mean to the mean of a fixed, predetermined image is applied as a *multiplicative* correction factor. Our purpose in doing the mean correction is to enable us to sensibly look at differences when performing motion correction.

After doing this basic preprocessing, the two-dimensional Fourier transform is applied to the images, using the `recon` module. Images are then *clipped* to remove empty space around the edges. Next, the amount of motion is estimated by `estireg`; this is accomplished by estimating shift and rotation parameters relative to a fixed image. This step involves a multi-dimensional optimization and is very time-consuming. The motion estimates are themselves subject to statistical noise. Under the assumption that the head moves slowly, we reduce this noise by smoothing the estimates using a kernel that is a few images in width. *Motion correction* follows the estimation of movement in the image. The two-dimensional motion correction, carried out by FIASCO’s `ireg` component, uses Fourier interpolation. Shifts in image space are achieved by applying a linear phase change to the  $k$ -space data. For rotations, the method again starts with a Fourier transform, then uses a location-varying linear phase change and returns via an inverse Fourier transform (Eddy, Fitzgerald and Noll, 1996); this is based on the factorization of a 2D-rotation matrix into the product of three shear matrices. Eddy and Young (1999) show that FIASCO motion correction has an information-preserving property shared by no other software. Motion estimation assumes mean correction, but motion correction does not. Hence it is possible to produce data that have been motion corrected, but not mean corrected, if this is desired.

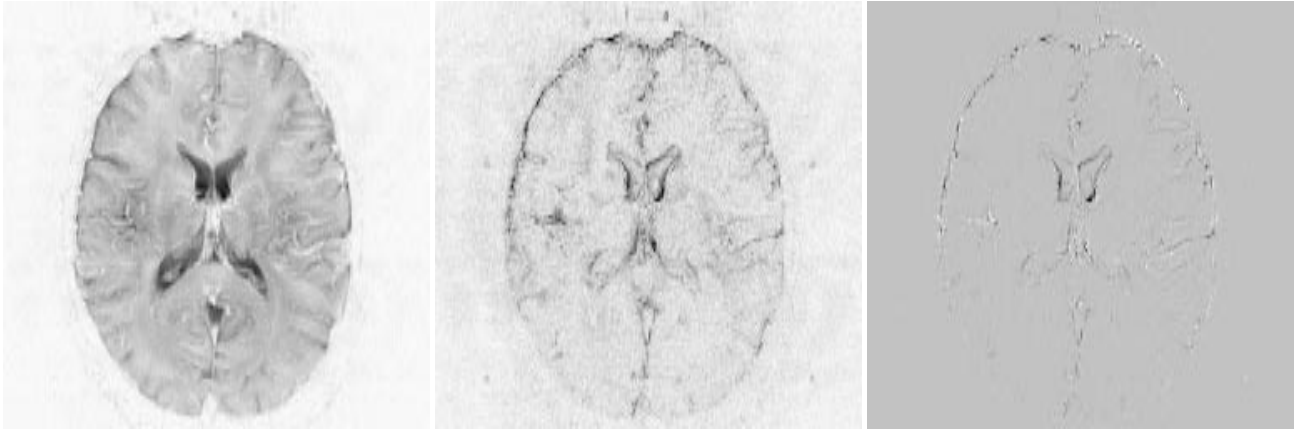


Figure 8: (a) (b) (c)  
(a) Mean image, no motion correction (b) Variance image, no motion correction (c) Difference in variance images, with and without motion correction. In panels (a) and (b), blacker values are larger and white is zero. In panel (c), blacker values are positive and whiter values are negative; grey is zero.

Effects of `ireg` are apparent from inspection of Figure 8. The mean isn't visually affected by this correction; the mean image without motion correction is in Figure 8(a). We give two variance images – the first shows the variance without motion correction (Figure 8(b)) and the second shows the difference in variance before and after image registration (Fig 8(c)). From Figure 8(c), we can see that the main impact of motion correction is to change the variance around the edge of the brain. Dark areas in Figure 8(a) in the middle of the brain contain cerebrospinal fluid, and so for our purposes the boundaries between these and the rest of the brain also constitute “edges”. While there is a fair amount of variance reduction, there is also evidence of *increases* in variance at some locations.

At the next stage, outliers are identified and brought into line with the rest of the image. FIASCO's default specifies an outlier to be any value that is 3.5 standard deviations from the mean of that pixel. Extreme pixels under this criterion have their values changed to the mean plus 3.5 standard deviations (for high outliers) or the mean minus 3.5 standard deviations (for low outliers). In fact, `outlier` can be applied at any point in the preprocessing, either in image space or in  $k$ -space. While the value 3.5 might seem large, remember that a typical study contains tens of millions of pixels and the distribution of values may not be Gaussian.

Linear, time-related trends are removed from the sequence of images, using `detrend`. Regressions of each pixel, using time as the covariate, are carried out, giving estimates of the slope and the intercept. The time series for each pixel is then corrected by removing the estimate of slope (multiplied by time) from each point in the data. It is interesting to note that the need for this module only became apparent when our research team switched from a 1.5 T magnet to a stronger, 3 T, magnet. FIASCO grows and changes to adapt to the needs of the researchers with whom we work.

It is apparent that the preprocessing done by FIASCO is very computationally intensive. Running all of the steps on a data set takes on the order of hours, even with fast machines. FIASCO is designed to run in parallel, if multiple processors are available, greatly speeding

data analysis. Using a 12 processor, Silicon Graphics Origin 2000, which we have devoted to fMRI data analysis, the calculations described here can be run on a dataset of the size described in previous parts, in roughly fifteen minutes, the time it takes the scanner to acquire the data. However, we anticipate that new scanner hardware (with a higher data rate) and new algorithms (such as three-dimensional motion correction) will significantly increase processing demands. We are constantly upgrading our equipment, as our knowledge and FIASCO progress, to be able to handle the types of data we are collecting and the questions we are now interested in addressing.

### 4.3 FIASCO and Data Analysis

In the previous section, we saw the steps that FIASCO uses to preprocess the data, even before any statistical data analysis is attempted. The effects of the preprocessing steps on our experimental data were also explored. The software has, in addition, the capability to perform some types of data summary and analysis.

Means and standard deviations for the various experimental conditions are the output of the `stats` module. It also calculates t-statistics. For each experimental condition, these are calculated on a pixel-wise basis. The module `mri_anova` carries out a pixel-wise analysis of variance. The module `mri_glm` fits a general linear model. We intend to use this to correct the data for physiological effects, but it could also be used for data analysis. Another useful inferential tool is found in the `overlay`. This function thresholds the statistical map produced by the program and then overlays the thresholded pixels onto an anatomical image of the brain. In this way it is possible to pick out which voxels are active during an experiment at a particular level of activation, as determined by, e.g., a t-statistic or F-statistic.

More sophisticated analysis can be implemented using the software package BRAIN (Bayesian Response Analysis and Inference for Neuroimaging), a standalone software that is compatible with the main package. BRAIN allows the user to fit hierarchical models to the times series for each voxel. For more detail on this, see Section 5.2 below, or Genovese (1998).

## 5 Statistical Analysis

### 5.1 General Issues

As should be clear from the previous discussion of the data acquisition process and the types of questions that are of interest in a functional MRI experiment, the statistical issues are quite varied and complex. We have to deal with questions of image reconstruction, noise, massive amounts of data, and all of this before we even begin the formal statistical analysis! Further complications arise because of potentially high correlations among the observations; there is spatial correlation, in that voxels located near each other can be expected to behave in similar fashions (as can some voxels that are far away from each other, because of communication leading to connections between different parts of the brain) and temporal correlation, due to

the repeated imaging of the brain over time. There is a great deal of structure inherent in the data, again stemming from the fact that they represent images of the functioning brain. Any statistical analysis should attempt to take into account these aspects of the problem. This isn't always feasible, as a result of computing limitations and the sheer difficulty of the data analysis task itself. There are still many questions regarding how this type of data should best be analyzed. The vast array of tools available in the statistical literature has not yet made its way into the world of fMRI; since many classical statistical methods make assumptions, such as independence and identical distribution of observations, that are not warranted here, their applicability might in any case be limited and the need for new methodologies arises. Progress is being made on this front, and in the next section we discuss common techniques currently in use for analyzing fMRI data.

Noise is a major challenge in fMRI data, underlying much of the difficulty in identifying regions of activation and in understanding how these regions interact (Eddy, Fitzgerald, Genovese, Lazar, Mockus and Welling, 1998). In our view, there are two ways of dealing with this noise – one is to try to remove the source of the noise; the other is to model it statistically. We pursue both of these approaches and believe that they are both essential. Head motion is one of the primary sources of noise, as we have already discussed, but there are others. Another fundamental source of noise is the thermally driven vibration of the atomic nuclei in the imaged material, which depends on the anatomical structure and the function which we are trying to detect. This noise can only be eliminated by reducing the temperature toward absolute zero, which unfortunately has been found to compromise the ability of subjects to perform cognitive tasks. Machine-based causes of noise include mechanical vibration and field inhomogeneities. Maintenance technicians work to limit the effect of these sources.

Physiological processes such as breathing and the beating of the heart while the subject is in the scanner also contribute to the noise. Since these processes can be measured, we should be able to model them as well. Work along these lines is in progress. A more complete understanding of the effects of these processes on the quality of the image would require integrating the physiological data obtained from a subject being scanned with the imaging data acquired at the same time.

There is also variation among subjects, another area that has yet to be fully explored. Sometimes this variation is quite surprisingly large. Experiments still tend to be concentrated on single subjects, but there is a lot to be gained from comparing individuals for regions and levels of activation. This question warrants further study and indeed is now one focus of our research. A foreseeable complication in this endeavor is the need to make anatomical comparisons across subjects; coordinate systems and mappings have been developed to permit this, but these are somewhat unsatisfactory. One popular method, for example, involves “morphing” each subject's brain onto a “standard” brain, based on a limited number of landmarks (Talairach and Tournoux, 1988).

## 5.2 Current Statistical Techniques

Methods for analyzing fMRI data are increasing in sophistication, although there is still great possibility for improvement. Many current techniques are based on the idea of classification, that is, deciding which voxels are active and which are not. This involves, for example, calculating  $t$ -tests on individual voxels, for comparing, as in the simplest study paradigms, a baseline condition with an experimental one and classifying as active all those voxels with a  $t$ -statistic above a certain threshold. While this is useful for detecting areas of activity in the brain, it doesn't take account of the spatial correlations between voxels that are close to each other. It also relies on adjustments for the multiple comparisons that are made, not even considering statistical questions about testing in general. However, since the signal is small relative to the noise, the use of standard multiple comparison techniques can unnecessarily eliminate activity. As a result, voxels which perhaps should be classified as active may appear to be inactive.

Various model-oriented approaches have been proposed that refine the classification available from the simple  $t$ -tests. One such is that of Worsley and colleagues (see for example Worsley, 1994; Worsley, Evans, Marrett and Neelin, 1992), which uses random fields of different types (Gaussian,  $t$ ,  $\chi^2$ ) to model the observed patterns of activation in the brain. This method incorporates some of the spatial correlations in the data and determines regions that are active, according to thresholding determined by "excursion sets". These sets are areas in the random field where a given threshold value is exceeded. The observed excursion sets can then be compared to theoretical values from the random field, to set significance levels. Exact probabilities are hard to calculate, involving as they do the  $p$  values of the maximum of a random field, but reasonable approximations have been found (Adler, 1981). Worsley (1996) summarizes these results and demonstrates the main ideas on studies of both the structure of the universe and of the human brain. This approach is a step forward, in that it allows for the spatial correlation in the data. However it is very sensitive to the assumptions of normality underlying the random field calculations. Excursion sets are by their nature extreme points, lying in the tails of the distributions and there is no statistical reason to believe that the Gaussian assumptions do in fact hold. There is also potential sensitivity to the assumptions on the spatial and temporal covariance structure.

An alternative way of including spatial correlations is presented by Forman *et al* (1995), who suggest assessing significant changes by looking at clusters of contiguous voxels. Areas where there is interesting activity are not likely to be restricted to a single voxel, but instead are more probable to be comprised of several neighboring voxels. Thresholds for significance can be calculated to take account of the amount of spatial correlation and the size of the clusters we wish to detect.

As with all work in this area, research on these modeling approaches is constantly evolving and becoming more refined with the accumulation of knowledge about the response and more complex experimental designs. One recent example (Worsley, Poline, Friston and Evans, 1998), uses canonical variates analysis of least squares estimates from a multivariate linear model. The method aims to evaluate the relationship between scans of an experiment and a set of predictors, such as type or intensity of stimulus response. Multivariate linear

models allow researchers to study the effects of temporal correlation; test statistics are calculated for individual voxels and then combined to make a global test for the influence of the predictors on cerebral blood flow. In a graded manner, it is possible to test for the necessity of including increasing numbers of combinations of predictors, obtained via principal components analysis. An advantage of this procedure over others based on random fields is that inference can be based on standard multivariate statistics, instead of resorting to simulations or approximations to get  $p$  values.

Another model-based approach is that of Bandettini *et al* (1993), who developed the “correlation method”. This technique takes a reference function,  $r(t)$ , and correlates it with each voxel data time series  $v(t)$ . At each voxel, the correlation coefficient is calculated; this can be used as a threshold to select which voxels are to be deemed “active”. Different types of reference function can be used;  $t$ -tests, for example, correspond to a simple square wave. More sophisticated analyses replace square waves with sine waves (Bandettini *et al*), or parametric functions.

An example of this approach is Lange and Zeger (1997), which presents a nonlinear Fourier-based method. The goal of this work is to extend existing time-series models to allow the hemodynamic response function to vary spatially, with the time course for each voxel modeled individually. Furthermore, the model is cast in the frequency domain, which facilitates accommodation of the autocorrelations, both spatial and temporal, that are present in the data. There are a number of drawbacks to the implementation in Lange and Zeger (1997). Chief among them is a limitation of applicability. The algorithm detailed in that paper assumes a very simple data structure – two conditions, a baseline and a stimulus, that alternate over fixed intervals. We are currently using much richer designs for fMRI data acquisition; the “on/off” experiment is too restrictive. In addition, Lange and Zeger use a simple Gamma density to model the response. Again, this is too restrictive. Many important features of the response are not captured by this curve.

All of the methods discussed up to this point have been refinements on the basic idea of classification. A more flexible and general system is proposed by Genovese (1998). Here, a nonlinear hierarchical model is described for analyzing fMRI data with greater flexibility than was previously possible. In particular, it allows the investigator to model directly some of the sources of noise in the data, such as drift and baseline signal, instead of resorting to preprocessing in FIASCO (or any other similar software package). This model also provides a framework for coping with the complex spatial and temporal relations among voxels. A simple version is the following, where  $Y(t)$  is the observed signal at time  $t = 0, \Delta, \dots, T\Delta$  for a given voxel;  $\Delta$  represents the sampling interval:

$$Y(t) = \mu + d(t) + a(t|\mu, \gamma, \theta) + \epsilon(t).$$

Here  $\mu$ ,  $\gamma$ ,  $\theta$  and  $d(\cdot)$  are parameters of the model and  $\epsilon$  is the noise process. The four components in the model specified above correspond to the baseline signal ( $\mu$ ), the drift ( $d(t)$ ), the activation ( $a(t|\mu, \gamma, \theta)$ ) and the noise ( $\epsilon(t)$ ). Each of the components has an associated prior distribution, which permits the investigator to incorporate knowledge available from previous experience into the hierarchical model. For more details on the parameterization

and the forms of the prior distributions, see Genovese (1998).

It is also worth noting another feature of this formulation, namely, the manner in which the activation profile is modeled. The hemodynamic response is realized as a disturbance in the MR signal. For one period of activity, the course of the activation can be described. Specifically, the signal begins at a baseline level and stays there for up to two or three seconds following the start of the task performance. It isn't clear what this delay represents. After some time the neuronal activity causes the MR signal to increase, eventually reaching a plateau, with height that is related to the amount of activity in the voxel. As long as the task is being performed, the plateau is maintained. When task performance ends, the signal begins to decay back to baseline, often even dipping below the baseline level before coming back up.

Using piecewise polynomial functions, the shape of the response curve, with all of its features, can be more effectively modeled. As new information about the hemodynamic response comes to light, this, too, can be incorporated into the system, either through the polynomials or through the priors which are a part of the hierarchical approach. This is a much more complex and comprehensive attack on the problem of analyzing fMRI data than has previously been attempted. We have given here only the barest details; Genovese (1998) discusses these and many more issues relevant to the modeling process.

## 6 Other Challenges and Uses for Statistics

It should be clear that the scope for using statistics in the analysis of MR data is vast. There is much of a statistical nature that can be done, though, even before formal data analysis starts. The complex nature of the data, the fact that the noise is large relative to the signal being measured and other issues addressed above, all contribute to the ability of the statistician to become involved from the very early stages of a study, to truly become part of the team. Here, at the very least, it is easy to say that “there is more to life than just t-tests”!

As statisticians, we can contribute to fMRI studies by helping to design experiments, by asking questions of the neuroscientists and psychologists, by improving image reconstruction techniques, by finding more efficient ways of dealing with the noise in the images, either via models or through improved computational methods and of course, by coming up with more sophisticated tools for data analysis. Steps in all of these directions have been made; however, there is still room for progress. In particular, different approaches to modeling could be taken, corresponding to different theories of brain function. Possible advances include the use of clustering algorithms to point out connections between various brain systems and to model the nonlinear network that most likely underlies cognitive processing of information, and extending the hierarchical models to incorporate data obtained from different sources (neurophysiological or psychological evaluations, for example) or imaging modalities (for

instance positron emission tomography).

## Résumé

La résonance magnétique fonctionnelle, une technique d'imagerie développée au cours des dix dernières années, permet l'étude des processus qui gouvernent le fonctionnement du cerveau humain, dans les domaines des sciences neurologiques et de la psychologie cognitive. Dans cet article, nous examinons les propriétés statistiques de cette méthode. Nous commençons par une brève description du fonctionnement de la méthode. Nous abordons ensuite les questions de planification et d'analyse d'expériences, que nous illustrons à l'aide de données issues du domaine de la psychologie.

Les données obtenues par l'imagerie de résonance magnétique fonctionnelle sont complexes – sujettes, en outre, à des dépendances spatiales et temporelles, et à plusieurs sources d'erreurs – donc peuvent se prêter à de diverses analyses statistiques.

**ACKNOWLEDGMENTS:** We would like to thank Marcel Just, Patricia Carpenter and Tim Keller, of Carnegie Mellon's Department of Psychology and the Center for Cognitive Brain Imaging, and Keith Thulborn, of the University of Pittsburgh Medical Center's Magnetic Resonance Research Center, for providing the data used in this article, and for many helpful conversations. The comments of an anonymous referee and the editor are also much appreciated.

## REFERENCES

- Adler, R.J. (1981) *The Geometry of Random Fields*, John Wiley, New York.
- Bandettini, P.A., Jesmanowicz, A., Wong, E.C. and Hyde, J.S. (1993) Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine* **30**, 161–173.
- Brown, M.A. and Semelka, R.C. (1995) *MRI: Basic Principles and Applications*, Wiley-Liss, New York.
- Cox, R.W. (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research* **29** 162–173.
- Cox, R.W. and Hyde, J.S. (1997) Software tools for analysis and visualization of FMRI Data. *NMR in Biomedicine* **10** 171–178.
- Eddy, W.F., Fitzgerald, M., Genovese, C., Lazar, N., Mockus, A. and Welling, J. (1998) The challenge of functional magnetic resonance imaging. To appear in *The Journal of Computational and Graphical Statistics*.
- Eddy, W.F., Fitzgerald, M., Genovese, C., Mockus, A. and Noll, D.C. (1996) Functional Imaging Analysis Software – Computational Olio. *COMPSTAT Proceedings in Compu-*

- tational Statistics* (ed Prat, A.), Physica-Verlag, Heidelberg, 39–49.
- Eddy, W.F., Fitzgerald, M. and Noll, D.C. (1996) Improved image registration using Fourier interpolation. *Magnetic Resonance in Medicine* **36**, 923–931.
- Eddy, W.F. and Young, T.K. (1999) Optimizing MR resampling. To appear in *Handbook of Medical Image Processing* (ed Bankman, I.), Academic Press, San Diego.
- Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A. and Noll, D.C. (1995) Improved assessment of significant change in functional magnetic resonance imaging (fMRI): Use of a cluster size threshold. *Magnetic Resonance in Medicine* **33**, 636–647.
- Friston, K.J., Frith, C.D., Liddle, P.F. and Frackowiak, R.S.J. (1991) Comparing functional (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism* **11**, 690–699.
- Gaschler, B., Schindler, F., and Scheich, H. (1996). KHORFu: A KHOROS-Based functional image post processing system. A statistical software package for functional magnetic resonance imaging and other neuroimage data sets. *COMPSTAT Proceedings in Computational Statistics* (eds Prat, A. and Ripoll, E.), Physica-Verlag, Heidelberg, 57–58.
- Genovese, C.R. (1998) Statistical inference in functional magnetic resonance imaging. Submitted.
- Genovese, C. R. and Sweeney, J. A. (1999). Functional connectivity in the cortical regions subserving eye movements (with discussion). *Case Studies in Bayesian Statistics*, Volume 4 (eds Kass, R. E., Carlin, B. P., Carriquiry, A. L., Gatsonis, C., Gelman, A., Verdinelli, I., and West, M.), Springer Verlag, New York, 59–132.
- Glad, I.K. and Sebastiani, G. (1995) A Bayesian approach to synthetic magnetic resonance imaging. *Biometrika* **82**, 237–250.
- Gold, S., Christian, B., Arndt, S., Zeien, G., Cizadlo, T., Johnson, D.L., Flaum, M. and Andreasen, N.C. (1998) Functional MRI statistical software packages: A comparative analysis. *Human Brain Mapping* **6**, 73–84.
- Hajnal, J.V., Bydder, G.M. and Young, I.R. (1995) FMRI: Does correlation imply activation? *NMR Biomed.* **8**, 97–100.
- Hajnal, J.V., Myers, R., Oatridge, A., Schwieso, J.E., Young, I.R. and Bydder, G.M. (1994) Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magnetic Resonance in Medicine* **31**, 283–291.
- Hinshaw, W.S. and Lent, A.H. (1983) An introduction to NMR imaging: From the Bloch equation to the imaging equation. *Proceedings of the IEEE* **71** 338–350.
- Josephs, O., Turner, R. and Friston, K. (1997) Event-related fMRI. *Human Brain Mapping* **5**, 243–248.
- Keller, T.A., Just, M.A., Carpenter, P.A. and Thulborn, K.R. (1998) Lexical and syntactic processing in sentence comprehension. *NeuroImage* **7**, S187.

- Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., Cheng, H.M., Brady, T.J. and Rosen, B.R. (1992) Dynamic magnetic–resonance–imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Science of the United States of America* **89**, 5675–5679.
- Lange, N. and Zeger, S.L. (1997) Non–linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Applied Statistics* **46**, 1–29.
- National Research Council, Institute of Medicine (1996) *Mathematics and Physics of Emerging Biomedical Imaging*, National Academy Press, Washington D.C.
- Ogawa, S., Lee, T.M., Kay, A.R. and Tank, D.W. (1990) Brain magnetic–resonance–imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Science of the United States of America* **87**, 9868–9872.
- Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H. and Ugurbil, K. (1992) Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic–resonance–imaging. *Proceedings of the National Academy of Science of the United States of America* **89**, 5951–5955.
- Price, C.J., Moore, C.J. and Friston, K.J. (1997) Subtractions, conjunctions, and interactions in experimental design of activation studies. *Human Brain Mapping* **5**, 264–272.
- Raichle, M. (1994) Visualizing the mind. *Scientific American* **270:4**, 58–64.
- Roy, C.S. and Sherrington, C.S. (1890) The regulation of the blood–supply of the brain. *Journal of Physiology, London* **11**, 85–108.
- Talairach, J. and Tournoux, P. (1988) *Co–planar Stereotaxic Atlas of the Human Brain: 3–Dimensional Proportional System: An Approach to Cerebral Imaging*, Georg Thieme Verlag, Stuttgart.
- Thulborn, K.R., Waterton, J.C., Matthews, P.M. and Radda, G.K. (1982) Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochimica et Biophysica Acta* **714**, 265–270.
- Turner, R. (1993) Gradient coil design: A review of methods. *Magnetic Resonance Imaging* **11**, 903–920.
- Woods R.P., Cherry S.R. and Mazziotta J.C. (1992) Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography* **16**, 620–633.
- Worsley, K.J. (1994) Local maxima and the expected Euler characteristic of excursion sets of  $\chi^2$ ,  $F$  and  $t$  fields. *Advances in Applied Probability* **26**, 13–42.
- Worsley, K.J. (1996) The geometry of random images. *Chance* **9**, 27–40.
- Worsley, K.J., Evans, A.C., Marrett, S. and Neelin, P. (1992) A three–dimensional statistical analysis for rCBF activation studies in the human brain. *Journal of Cerebral Blood Flow and Metabolism* **12**, 900–918.

Worsley, K.J., Poline, J.B., Friston, K.J. and Evans, A.C. (1998) Characterizing the response of PET and fMRI data using multivariate linear models. *NeuroImage* **6**, 305–319.