Experimental Design and Data Analysis for fMRI

21

Geoffrey K. Aguirre

Introduction

Functional magnetic resonance imaging (fMRI) methods continue to evolve rapidly. Subtle experimental designs have been joined by more powerful data analysis methods to detect and interpret evoked changes in neural activity. Despite constant development, there are several core principles of fMRI methodology that can be used as a guide to understand the current state of the field and whatever advance awaits tomorrow. This chapter concerns itself primarily with this core understanding but considers several specific aspects of fMRI experiments. Along the way, the chapter also notes some of the specific challenges that exist in the fMRI studies of clinical populations, although a detailed consideration of these issues is contained in Chap. 25.

Topics will be raised roughly in the same order as they present themselves in the course of the conception and completion of an fMRI experiment. This order of presentation also moves from general issues in neuroimaging inference to more specific aspects of fMRI and finally to the idiosyncratic properties of blood-oxygen-level-dependent (BOLD) fMRI and their implication for experimental design and analysis. To start, the chapter considers two categories of neuroimaging experiments, each of which examines a different direction of the relationship between the brain and behavior. Next, different techniques of isolating and manipulating mental operations that might be used in the service of these experimental designs will be discussed. Regarding experimental design, the chapter considers the possible temporal ordering of stimuli within an fMRI experiment, including the paradigmatic "blocked" and "event-related" designs.

This section requires us to grapple with two critical properties of BOLD fMRI data: the hemodynamic response function and the temporal autocorrelation of the noise. We then turn our attention to related analysis issues. The steps of data preprocessing that prepare fMRI data for statistical analysis are reviewed, followed by a consideration of the univariate analysis of fMRI data.

G. K. Aguirre (✉)
Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA, USA
e-mail: aguirreg@upenn.edu; aguirreg@pennmedicine.upenn.edu

Basic Types of Neuroimaging Inference

Regardless of the particular neuroimaging methodology employed (e.g., fMRI, positron emission tomography [PET], and event-related potential [ERP]), there are three broad categories of experimental questions that might be asked. Each category probes a different direction of the relationship between the brain and behavior, and each one makes different assumptions for valid inference. Within these broad categories lie many different experimental designs, each with particular assumptions and inferences. Generally, placing a study within an inferential category can help organize one's thinking about the assumptions that underlie a particular experiment.

Forward Inference

One class of neuroimaging experiments is concerned with *forward inference*, which generally examines the anatomical and neural correlates of a given mental operation. One application of forward inference is to ask *localization* questions. For example, does perception of a face activates a particular area of the brain different from that evoked by perception of other stimuli? Does the cognitive process of "working memory" evoke neural activity within the frontal lobe or somewhere else? In general, these designs present a subject with a task designed to evoke a particular cognitive state of interest selectively, and the neuroimaging method identifies if and where the changes in neural activity accompany the cognitive process. Clearly, this type of experiment requires a way to manipulate the mental state of the subject, isolating the mental operation of interest from the other processes that

S. H. Faro, F. B. Mohamed (eds.), *Functional Neuroradiology*, https://doi.org/10.1007/978-3-031-10909-6_21

invariably are present (e.g., button pushing, preparing responses, etc.).

The next section discusses several methods that might be used to do so. If successful, a localization study allows one to conclude that a particular area of the brain is "activated" by a particular cognitive operation. Importantly, neuroimaging methods, in general, are severely restricted in their ability to make conclusions regarding the *necessity* of a region for a cognitive operation. In other words, the presence of focal activation for a particular mental operation does not imply that a lesion to that area of the brain would impair the subject's ability to perform that mental operation. The reasons for this are manifold. For example, multiple areas of activity might be found, any one of which (perhaps working in parallel or one serving as a "backup" for the other) would be capable of supporting the process of interest. In this case, the region still plays an interesting role in the cognitive process, although it is not strictly necessary. A second challenge is the lack of perfect control over the mental states of the subject we seek to study. Although stimuli and instructions designed to evoke a particular cognitive process can be presented, there can be no guarantee that the subject has entered that cognitive state and *no other*. The subject may unwittingly engage in confounding cognitive processes in addition to that intended by the experimenter or, alternatively, may fail to differentially engage in the process. This is the central challenge of interpreting most neuroimaging studies of localization—it is difficult to be certain that the experimental variable of interest has been properly manipulated.

Several applications of localization-type neuroimaging studies of patient populations can be conceived. The use of fMRI to identify "eloquent" (or otherwise functionally important) cortex for neurosurgical planning is one example. Importantly, the caveats expressed previously regarding the conclusion of "necessity" using a neuroimaging study are particularly relevant for this application (see Atlas et al. [1] and Chap. 31 for further details).

By contrast, *implementation* studies ask about the computational mechanisms of a cognitive process within a cortical region. This type of study begins with the assumption that a cortical region engages in computations that support a particular cognitive process. The purpose of the study is then to determine the parameters of neural activity that mediate the area's participation in that process. For example, does an area of the prefrontal cortex change its bulk level of neural activity as a function of increasing working memory load (i.e., remember four items instead of two)? Is the speed of motion encoded differently from the direction of motion within area MT? As is true for all forward inference studies, the key assumption is that the perception or behavior of the subject is controlled by the experiment, permitting conclusions that link neural states to internal mental states.

A related clinical application is the detection of dynamic pathology. The properties of this mode of inference derived from the use of fMRI to detect spontaneous patterns of neural activity that are unlike neural activity evoked by normal mental operations. The prototypical use is the detection of the cortical origin of seizure activity [2]. Unlike many of the other applications of fMRI discussed here, the localization of neural activity does not rely upon a behavioral or stimulus paradigm to create a particular pattern of neural activity but instead is designed to detect endogenous, pathological neural patterns. This may be done by relating neural imaging signals to the timing of the symptoms or another form of monitoring (e.g., simultaneously acquired scalp electroencephalogram [EEG]). Alternatively, the study may identify brain areas that demonstrate pathological patterns of neural activity, even when it cannot be specified when those events took place. To do so, it is necessary to specify signal parameters that can distinguish between normal and abnormal neural patterns (e.g., Esteller et al. [3]).

Reverse Inference

A *reverse inference* design probes a different direction of the relationship between brain and behavior and asks: What cognitive process does a given task evoke? This type of experiment leverages knowledge about the neural correlates of particular mental states to learn something about an imperfectly understood behavior. One begins by assuming that neural activity in a particular area of the brain is a marker of the presence of a particular mental state and no other. Depending on the application, this neural activity may be read from local or distributed neural locations.

The *Focal reverse inference* is based upon the presence of neural activity at a particular brain location. For example, the neural activity of a certain magnitude at a certain spot in the fusiform gyrus indicates that the subject has the visual perception of a face. The subject then performs a task that may or may not evoke the cognitive process of interest. For example, ambiguous stimuli are presented that can be perceived as a face or a vase. If the specified neural activity is seen, the conclusion is drawn that the subject saw a face at that moment in time. This type of design may therefore be used to test hypotheses regarding the engagement of cognitive processes during a behavioral state in which the cognitive processes need not be under experimental control. This type of experiment has been the basis of the rapidly growing fields of emotional, social, and economic neuroscience, in which activity at certain brain locations is taken as evidence of a particular emotional state or value judgment.

What provides the evidence that a particular region is uniquely activated by a specific cognitive process? Logically,

only an exhaustive neuroimaging examination of every possible cognitive process, under every possible circumstance, could provide the necessary evidence. This is obviously practically impossible, so a series of neuroimaging experiments that demonstrate activation of a particular region during a given cognitive process and no other usually suffices to support the assumption (a logical inference termed *enumerative induction*).

Within the clinical realm, this type of inference is used as a surrogate measure of a behavioral state. Some diseases produce symptoms that are subjective and can only be imperfectly measured by observation or patient report. The experience of pain is an example of this kind. In other cases, patients experience symptoms that they under-report, such as those with drug addiction who minimize their degree of drug craving. Some patients may have important internal cognitive states that are not evident to the clinician, such as patients who are "locked-in" from pontine lesions or those who are in a minimally conscious state following more extensive cortical damage [4]. Finally, there are patients who feign neurological deficit either due to psychopathology or due to hope for secondary gain. In each of these cases, fMRI might be used to measure an internal mental state of a subject that is not easily obtained through simple behavioral observation.

A related approach is *distributed reverse inference*. This is the inferential mode behind the enormously powerful and rapidly expanding techniques of multivoxel pattern analysis (MVPA) [5]. In these designs, the pattern of neural activity across voxels is used to classify behavioral or perceptual states. Critically, the process of "decoding" behavior from the neural activity is preceded by a training phase, in which the distributed neural response to a particular stimulus is measured. The patterns of response observed during the training period are then used to classify the novel behavioral or perceptual states during the test phase. Therefore, an MVPA experiment effectively includes both a forward inference study (during the training phase) and a reverse inference study (during the test or decoding phase). The result is a uniquely powerful design that provides an inferentially sound basis for prediction and classification.

Connectivity

The third class of experiments examines the relationship between brain states. These approaches obtain the time-varying signal from different brain regions and base inferences upon the relationship between these regions. The burgeoning field of resting-state connectivity [6] examines the distribution of phase-locked signal fluctuations across the brain during a putative rest state, to parcellate cortex and to classify mental states. A related approach quantifies the connectivity between different cortical regions—the extent to which one cortical region influences neural activity in another region [7]. These connectivity maps may then be compared across populations (clinical or otherwise) and between different behavioral states. Brain connectivity is detailed in-depth in Chap. 24.

Manipulation of the Cognitive Process

As was discussed, many neuroimaging experiments depend upon the isolated manipulation of a cognitive process for the study. In particular, localization experiments require that a cognitive process of interest be isolated from other mental operations so that the neural correlates of that process can be observed. In implementation experiments, some aspects of the stimulus or mental operation must be varied so that the neurocomputational correlate of its processing can be studied. Here, we consider several broad classes of experimental manipulation of a cognitive operation. Note that any of these techniques can be coupled with a particular temporal structure of design (e.g., event-related or blocked), as described in the next section.

Cognitive subtraction is the prototypical method of isolating a cognitive process and is the most problematic. Typically, one condition of an experiment is designed to engage a particular cognitive process, such as face perception, episodic encoding, or semantic recall. This "experimental" condition is contrasted with a "control" condition designed to evoke all of the cognitive processes present in the experimental period, except for the cognitive process of interest. Differences in neural activity between the two conditions can be attributed to the cognitive process of interest. In essence, a cognitive process is isolated in an "all or none" fashion. As discussed previously, there is no direct control over the mental states of the subject, so the danger is always present that the subject might engage in a confounding mental operation in addition to the one of interest. Additionally, cognitive subtraction relies upon the assumption that a cognitive process can be added to a preexisting set of cognitive processes without affecting them (an assumption termed pure insertion). This might fail if, for example, the act of pressing a button to signal a semantic judgment is different from pressing a button in response to a visual cue. Effects upon the imaging signal that result from this difference would be erroneously attributed to semantic judgment per se.

The *cognitive conjunction* design [8] has been proposed to reduce reliance upon the assumption of pure insertion. The method uses a set of paired cognitive subtractions, each of which need not completely isolate the cognitive process of interest. The imaging data are then analyzed to find areas that have a significant, consistent response across subtractions. The identification of the same region across multiple pairs of subtractions strengthens the conclusion that the area is acti-

vated by the cognitive process that is isolated in each of the subtraction pairs.

Parametric designs offer an attractive alternative to cognitive subtraction approaches. In a parametric design, the experimenter presents a range of different levels of some parameter and seeks to identify relationships (linear or otherwise) between the imaging signal and the values that the parameter assumes. This can be done to identify the neural correlates of straightforward changes in stimulus properties or manipulations of a cognitive process. As compared to cognitive subtraction methods, failure of the pure-insertion assumption is less plausible for parametric designs as the cognitive process is present during all conditions. This method can be further extended using *factorial* designs, where multiple parameters are manipulated to identify additive and interactive changes in neural activity [9].

Finally, *carryover* designs examine the effect of stimulus context and history upon neural response [10]. Sequential transitions between stimuli are controlled, and the effects examined. The most common application of these designs is for the measurement of neural adaptation or habituation [11]. The approach exploits the well-demonstrated repetition–suppression phenomenon, in which a set of neurons have a reduced response to the repeated presentation of a stimulus. For example, one might hypothesize that the fusiform face area encodes a viewpoint—an independent representation of a face. Reduced responses from this cortical area to the second presentation of the same face viewed from a different angle would support this assertion.

Properties of the BOLD fMRI System That Impact Experimental Design

The preceding sections have described properties of experimental design that might apply to any neuroimaging method. The next section discusses the ordering of experimental conditions in time and specifically contrast blocked and event-related designs. To understand the consequences of these experimental design choices, this chapter considers the idiosyncratic properties of one particular neuroimaging method: BOLD fMRI. The chapter focuses on two key properties of BOLD fMRI data that fundamentally impact the design of BOLD fMRI experiments: the hemodynamic response function and the presence of low-frequency noise.

As was described in Chap. 19, changes in neural activity give rise to a series of vascular and hemodynamic changes that ultimately result in changes in the BOLD fMRI signal. While many details of this relationship between neural activity and hemodynamic change are still under study, much of the messy details can be conveniently sidestepped by noting that the transformation of neural activity to BOLD fMRI signal is nearly *linear*. This implies, for example, that doubling the amplitude of neural activity results in doubling the amplitude of the BOLD fMRI signal, and so on. One important property of BOLD fMRI as a linear system is that it can be well characterized by the *hemodynamic response function* (HRF). This is the BOLD fMRI signal that results from a brief (<1 s), intense period of neural activity. Given the shape of the HRF, one can predict the BOLD fMRI signal change that would result from any arbitrary pattern of neural activity.

The HRF itself can be empirically measured from human subjects by studying the BOLD fMRI signal that is evoked by experimentally induced, brief periods of neural activity in known cortical areas (e.g., neural activity in the primary motor cortex in response to a button press). The shape of the HRF reflects its vascular origin and rises and falls smoothly over a period of about 16 s. While the shape of the HRF varies significantly across subjects, it is very consistent within a subject, even across days to months [12]. The stability of the shape of the HRF proves to be of value in the analysis of fMRI data, as it allows one to predict the pattern of BOLD fMRI signal that might result from an arbitrary pattern of neural activity. One difficulty, however, is that there is some evidence that the shape of the HRF varies from one region of the brain to another (perhaps from variations in neurovascular coupling). This is, however, a difficult notion to test as it is necessary to create evoked patterns of neural activity in disparate areas of the brain that can be guaranteed to be very similar. The further problem is that the properties of the HRF may differ between elderly and young subjects, perhaps due to vascular disease [13]. The consequences of misspecification of the shape of the HRF will vary depending upon the experimental design used, as elaborated later.

The temporal dynamics of neural activity are quite rapid, in the order of milliseconds, but changes in blood flow occur over the course of seconds. One consequence of this, as demonstrated by the smooth shape of the HRF, is that rapid changes in neural activity are not well represented in the BOLD fMRI signal. The "temporal blurring" induced by the HRF leads to many limitations placed on the types of experiments that can be conducted using BOLD fMRI. Specifically, the smooth shape of the HRF makes it difficult to discriminate closely spaced neural events. Despite this, it is still possible to detect the following: (1) brief periods of neural activity, (2) differences between neural events in a fixed order, spaced as closely as 4 s apart. (3) differences between neural events, *randomly* ordered, closely spaced (e.g., every second or less), and (4) neural-onset asynchronies in the order of 100 ms. The reason that these seemingly paradoxical experimental designs can work is that some patterns of events that occur rapidly or switch rapidly create a low frequency "envelope": a larger structure of pattern of alternation that can pass through the hemodynamic response function. The next section discusses several types of tempo-

ral structures for BOLD fMRI experiments and considers how the shape of the HRF dictates the properties of these designs.

Another important property of BOLD fMRI data is that greater power is present at some temporal frequencies than others under the null hypothesis (i.e., data collected without any experimental intervention). The power spectrum (a frequency representation) of data composed of independent observations (i.e., white noise) should be "flat," with equal power at all frequencies. When calculated for BOLD fMRI, the average power spectrum contains ever-increasing power at ever-lower frequencies, often termed a 1/frequency distribution. This pattern of noise can also be called "pink," named for the color of light that would result if the corresponding amounts of red, green, blue, etc., of the visible light frequency spectrum, were combined. The presence of noise of this type within BOLD fMRI data has two primary consequences. First, traditional parametric and nonparametric statistical tests are invalid for the analysis of BOLD fMRI data, which is why much of the analysis of BOLD fMRI data is conducted using Keith Worsley and Karl Friston's "modified" general linear model [14] and its heirs, as instantiated in a statistical parametric mapping (SPM) and other statistical packages. The second impact is on the experimental design. Because of the greater noise at lower frequencies, slow changes in neural activity are more difficult to distinguish from noise.

Interestingly, the consequences for experimental design of the shape of the hemodynamic response function and the noise properties of BOLD fMRI are at odds. Specifically, the shape of the HRF would tend to favor experimental designs that induce slow changes in neural activity, while the presence of low-frequency noise would argue for experimental designs that produce more rapid alterations in neural activity. As it happens, knowledge of the shape of the HRF and the distribution of the noise is sufficient to provide a principled answer as to how best to balance these two conflicting forces.

It is worth noting that other neuroimaging methods have different data characteristics with different consequences for experimental design. For example, perfusion fMRI provides a noninvasive, quantifiable measure of local cerebral tissue perfusion [15]. Perfusion data do not suffer from the elevated, low-frequency noise present in BOLD, and as a result, perfusion fMRI can be used to detect extremely long timescale changes in neural activity (over minutes to hours to days) that would simply be indistinguishable from noise using BOLD fMRI [16]. This may prove to be very advantageous in studies of clinical populations. Functional changes in patient cognition, either improvement by functional recovery following focal lesions or decline in neurodegenerative disease, evolve over long time scales as well. Perfusion fMRI can be used to obtain stable measurements of evoked neural activity from this dynamic system.

Different Temporal Structures of BOLD fMRI Experiments

As BOLD fMRI generally includes multiple task conditions (prototypically, an "experimental" and "control" period), several ways of ordering the presentation of these conditions exist. Different terms are used to describe the pattern of alternation between experimental conditions over time and include such familiar labels as "blocked" or "event-related." While these are often perceived as rather concrete categories, the distinction between blocked, event-related, and other sorts of designs is fairly arbitrary. These may be better considered as extremes along a continuum of arrangements of stimulus order. Consider every period of time during an experiment as a particular experimental condition. This includes the "intertrial interval" or "baseline" periods between stimulus presentations. In this setting, blocked and event-related designs are viewed simply as different ways of arranging periods of "rest" (or no stimulus) with respect to other sorts of conditions. (For a more complete exploration of these concepts, see Friston et al. [17] and Liu [18]).

The prototypical fMRI experiment is a blocked approach in which two conditions alternate over the course of a scan. For most hypotheses of interest, these periods of time will not be utterly homogeneous but will consist of several trials of some kind presented together. For example, a given block might present a series of faces to be passively perceived, or a sequence of words to be remembered, or a series of pictures to which the subject must make a living/nonliving judgment and press a button to indicate his response. Blocked designs have the obvious difficulty that the subject can anticipate trial types, which may be undesirable in some settings (e.g., studies of recognition of novel vs. previously learned words). On the other hand, blocked designs have superior statistical power compared to all other experimental designs. This is because the fundamental frequency of the boxcar can be positioned at an optimal location with respect to the filtering properties of the hemodynamic response function and the low-frequency noise. For typical shapes of the HRF and distributions of temporal noise, this ideal balancing point occurs with epochs of about 20–30 s in duration.

Event-related designs model signal changes associated with individual trials, as opposed to blocks of trials. This makes it possible to ascribe changes in signal to particular events, allowing one to randomize stimuli, assess relationships between behavior and neural responses, and engage in the retrospective assignment of trials. Conceptually, the simplest type of event-related design to consider is one which uses only a single stimulus type and uses sufficient temporal spacing of trials to permit the complete rise and fall of the hemodynamic response to each trial; a briefly presented picture of a face once every 16 s for example. This is frequently termed a *sparse* event-related design. Importantly, while this

prototypical experiment has only one stimulus, it has *two* experimental conditions (the stimulus and the intertrial interval). More complex designs become possible if one is willing to abandon the fixed ordering and spacing of these conditions. For example, randomly ordered picture presentations and rest periods could be presented once a second (or even more rapidly). The ability to present rapid alternations between conditions initially seems counterintuitive, given the temporal smoothing effects of the hemodynamic response function. While BOLD fMRI is insensitive to the particular high-frequency alternation between one trial and the next, it is still sensitive to the low-frequency "envelope" of the design. In effect, with closely spaced, randomly ordered trials, one detects the low-frequency consequences of the random assortment of trial types. These rapid event-related designs are fairly sensitive to the accurate specification of the HRF for their success (unless a basis set is used for analysis; see below).

For experimental settings where one has an unlimited number of trials to present (e.g., flashes of light) but a limited period of scanning time, then rapid, randomly ordered designs are more statistically powerful than "sparse" designs. Alternatively, when the experiment is limited by the number of available trials (e.g., pictures of flightless birds), then maximal statistical power is obtained by presenting the available trials in a sparse manner and "stretching" the scanning period out as long as possible. In general, the provision of stimulus counterbalance is important for valid inference [10]. Counterbalance ensures that each type of stimulus appears equally often before another stimulus.

Thus far, the discussion regarding event-related designs has assumed an ability to randomize perfectly the order of presentation of different event types. There are certain types of behavioral paradigms; however, they do not permit a random ordering of the events. For example, the delay period of a working memory experiment always follows the presentation of a stimulus to be remembered. In this case, the different events of the trial cannot be placed arbitrarily close together without risking the possibility of false-positive results that accrue from the hemodynamic response to one trial event (e.g., the stimulus presentation) being interpreted as resulting from neural activity in response to another event (e.g., the delay period). It turns out that, given the shape typically observed for hemodynamic responses, events within a trial as close together as 4 s can be reliably discriminated [19]. Thus, event-related designs can be used to examine directly, for example, the hypothesis that certain cortical areas increase their activity during the delay period of a working memory paradigm without requiring the problematic assumptions traditionally employed in blocked, subtractive designs.

A multiplicity of further designs might be considered that do not fall strictly within "blocked" or "event-related" categories. *Neural-onset asynchrony* designs [20, 21] are used to detect differences in the timing of neural activity evoked by different stimuli. Here, a "sparse" event-related design is used, along with exquisite coupling of the timing of stimulus presentation to image acquisition. A difference in the time of onset of the smooth, BOLD hemodynamic response evoked by two different stimuli within a cortical region is sought. *Traveling wave stimuli* are used to define topographic maps of cortical responses, the most familiar being the retinotopic organization of early visual areas [22]. These designs use stimuli that vary continuously across some sensory space (e.g., retinal eccentricity) and identify for any point within a cortical area the optimal position of the stimulus within the sensory space for the evocation of neural activity. These designs are often combined with cortical flat-map techniques to display results [23].

Data Preprocessing

In a perfect world, BOLD fMRI images would be acquired instantaneously from a stationary brain of uniform shape. Unfortunately, this is not the case, and a number of processing steps must be performed prior to the statistical analysis of fMRI data. These steps have two primary goals: (1) to reverse displacements of the data in time or space that may have occurred during acquisition and (2) to enhance the ability to detect spatially extended signals within or across subjects. This section briefly discusses several preprocessing steps that are commonplace in analyzing BOLD fMRI data.

Distortion Correction

BOLD fMRI data are typically acquired as echoplanar images and, as such, are likely to be distorted (stretched and pulled) in space to some extent due to static magnetic field inhomogeneities produced by the concentration of magnetic field lines at (for example) air–tissue interfaces. There are several methods to correct this spatial distortion, and in most cases, they use a "map" of the magnetic field within the bore of the magnet to correct distortion. In many cases, this correction is performed by the scanning system itself prior to writing out image data for analysis and does not enter into the routine preprocessing of fMRI data at some institutions.

Slice Acquisition Correction

A single volume of BOLD fMRI data, collected during one repetition time (TR), is assembled from multiple planar acquisitions (slices). One slice is collected at a time, either sequentially or in an interleaved fashion, with the result that

each slice samples a slightly different point in time. For a TR of 2 s and 20 axial slices, this would mean that one slice of the brain would be obtained 1 s later than another spatially adjacent slice within the same TR. As a consequence, a neural event that occurs simultaneously on multiple slices within the brain will appear as different, time-delayed BOLD fMRI responses in the data from different slices. Slice acquisition correction compensates for this staggering order of slice acquisition. The correction works by calculating (using sinc interpolation) the BOLD fMRI signal that would have been obtained for a given slice had that slice been acquired instead at the beginning of the TR. While not of great importance for low-temporal-frequency blocked designs, this preprocessing step is quite important for event-related designs.

Motion Correction

A variety of methods are used to minimize head motion during scanning. These include foam padding around the head, "bite bars," custom-designed, thermo-plastic face masks, and so on. Despite these efforts, subjects nonetheless move their heads during scanning. Therefore, a common data preprocessing step is to attempt to correct the effects of this motion. This is generally done by realigning the image of the brain obtained at each point in time back to the first image acquired at the start of the scanning session. Several methods exist to do so, but most use a "six-parameter" motion correction, in which the brain is treated as a rigid body, and the six possible movements (three translations and three rotations) are calculated at each point in time to minimize the image difference between the realigned brain and the brain in its original position. Importantly, motion correction of this kind does not completely remove the effects of movement upon the BOLD fMRI signal. This is because the movement of the brain within the exquisitely defined magnetic field gradients created during scanning alters the signal obtained at different points in the slice acquisition. As a result, movement-induced signal artifacts remain even after realigning the brain to its original position. As a consequence, statistical analysis of BOLD data will often include nuisance covariates that are themselves the six movement parameters measured during realignment. These covariates account for changes in the signal within voxels that are correlated with the movement of the head. (See the "Statistical Analysis" section for a further description of nuisance covariates.)

Spatial Normalization

If one wishes to test a hypothesis regarding a certain area of the brain within a population, then it is first necessary to identify that same area of the brain across subjects. This is frequently done by computationally "warping" the anatomical structure of the brain of one subject to match a template brain within a standard, defined space. While there are various sophisticated methods available for registering and aligning the brains of different subjects into a standard space, there are theoretical limits to what such an alignment can achieve. First, there may be intersubject variability in anatomy that cannot be overcome by warping brains to a standard space. For example, the arrangement of the sulci may vary between subjects. Thus, while two subjects may have neural responses at the same "true" cytoarchitectonic location, the position of this site with respect to other landmarks in the brain may differ between subjects, leading to the spread of these locations when data are converted to a standard space. Second, even given the rigid alignment of anatomy across subjects, there may be variability in the structure–function relationships between subjects. For example, two subjects may truly have distinct face-selective neural regions, but these may be located in different sections of a cortical area due to differences in experience. Again, this variability in location will obscure functional dissociations when normalized to a standard space. An alternative to anatomical registration is functional identification. The approach here is first to identify a region across subjects by its functional responses. For example, one might identify a region that responds more to pictures of faces than general objects. Then, hypotheses regarding the response of this functionally defined region to other types of stimuli can be independently tested across subjects within this area. This powerful approach allows one to make inferences across subjects regarding the responses of some functional area (e.g., the fusiform face area) at the expense of making statements regarding some particular position in a standardized anatomical space.

Spatial Smoothing

It is common practice to digitally smooth BOLD fMRI data in space prior to statistical analysis. There are several reasons for this. First, BOLD fMRI data are typically composed of time-series information from many thousands of individual voxels. Statistical analysis of this data involves the application of a statistical test (e.g., *t*-test) at each of these voxels. As there are, therefore, thousands of individual statistical tests being performed, control of the false-positive rate requires a fairly large *t*-result to exceed the chance that random noise will produce a "significant" result in one or more of those thousands of voxels. By smoothing the data in space, one reduces the number of independent statistical tests that are being performed, thus allowing less-stringent control over what *t*-value is considered a significant result. Another motivation for smoothing is that, when analyzing data across

a population, spatial smoothing helps to overcome residual differences in anatomy between subjects that might otherwise render common areas of activation nonoverlapping. The amount of spatial smoothing to perform can be difficult to determine, as smoothing too much will decrease statistical sensitivity for small focal areas of activation, while smoothing too little will have the same deleterious effects upon large areas of signal change. A reasonable balance between these two extremes can be obtained by smoothing data with a filter that has a width roughly equal to the size (in voxels) of predicted areas of activity.

Statistical Analysis

Several methods exist to analyze BOLD fMRI data. Some of these are described as "multivariate" techniques, in which latent patterns of spatially coherent activity are identified automatically by the method (e.g., Partial Least Squares, McIntosh et al. [24]). This chapter focuses upon the more commonly implemented "univariate" techniques, in which a statistical model is applied to each voxel independently within a dataset. The discussion concerns, in particular, the details of the creation of a statistical model for analysis. Some of these details are handled automatically in many software packages, such as SPM (statistical parametric mapping). The purpose of this section is to provide an understanding of what is going on "under the hood."

The centerpiece of the analysis of neuroimaging data is the construction of a model that is composed of one or more "covariates." In general, covariates are predictions regarding patterns of variability in the data, expressed as changes in the BOLD fMRI signal over time within a single voxel. The better the predictions, the more valid and powerful the statistical model becomes. Covariates can be broadly divided into two categories. Covariates of interest describe changes in the signal that are (typically) the result of experimental manipulations and are subject to hypothesis testing. Covariates of no interest describe, instead, changes in the signal that are unintended or undesired and are not typically the focus of a hypothesis test.

Covariates of interest might be generated in one of two ways. First, covariates might be created to model the expected the shape in time of evoked BOLD fMRI signal changes. A principled way to create covariates of this kind is to begin with a prediction regarding the pattern of neural activity that might be evoked by the experiment in a single voxel. For example, a simple "blocked" experimental design might be predicted to produce a uniformly greater amount of bulk neural activity during an experimental condition as compared to a control condition. The anticipated BOLD fMRI signal under these circumstances can be obtained by applying a model of the hemodynamic response function (HRF) to the predicted pattern of neural activity. As mentioned earlier, knowledge of the HRF is sufficient to predict the BOLD fMRI signal that will result from any arbitrary pattern of neural activity, through the mathematical process of convolution. This prediction of BOLD fMRI signal change is then suitable for use as a covariate of interest. The model of the HRF that is used might be obtained from the subject himself during a preliminary experiment (e.g., Aguirre [12]), or an average representation of an HRF across subjects might be employed, as is the case in the SPM and other analysis packages.

Alternatively, covariates of interest cannot represent a specific pattern of BOLD fMRI response but instead have the property of flexibly fitting a family of possible responses that might occur. This approach uses a "basis set," which is a collection of covariates that can be scaled and combined to fit any pattern of BOLD fMRI response that might be evoked within a set period of time by a particular experimental condition. In general, the basis set will be composed of multiple covariates, with as many elements as there are points in time to be modeled. For example, in a sparse event-related experiment, in which a stimulus is presented every eight TRs (16 s at a TR of two), then a basis set of eight covariates will be needed to model, in effect, the average evoked BOLD fMRI response across trials. Typically, there is no clear interpretation of any one element of a basis set. Instead, one interprets the explanatory power of the set en-mass using an F-test. Basis set approaches provide the advantage of flexibility in that one is sensitive to any pattern of response (or difference between two trial types) that might take place. The price of this flexibility is reduced inferential power. For example, one can no longer say that a given response was greater in amplitude than another or longer in duration. Instead, one can only say that some consistent response was present.

As was mentioned, covariates of the no-interest model changes in the BOLD fMRI signal are not thought to be the result of experimental influence. For example, if one was aware of the influence of the room temperature upon the BOLD fMRI signal, and if the pattern of fluctuations of the room temperature were known, a representation of temperature could be included as a covariate to explain variations in the signal that are attributable to temperature fluctuations. Note that some covariates that are not of interest model changes in neural activity, and some do not. For example, the experiment may present "instruction" screens occasionally, which would be expected to elicit transient changes in neural activity that are not the subject of any hypothesis. For these types of covariates, which model expected neural effects of no-interest, one would want to convolve the representation of neural change by the hemodynamic response function. Convolution is not indicated for other covariates that are not derived from neural activity (e.g., a measure of subject head motion, or cardiopulmonary variation; Glover et al. [25]).

One can further classify covariates of no-interest as "nuisance" covariates or "confounds." A nuisance covariate is defined as a covariate the inclusion of which is expected to alter only the magnitude of the error term but not the relationship between the data and covariates of interest. When covariates of no-interest are correlated with covariates about which one wishes to test a hypothesis, they are termed "confounds," and their inclusion will be expected to alter the behavior of the covariates of interest. Under some circumstances, the sign of the relationship between the covariate of interest and the data can be reversed! An example of a covariate that frequently acts as a confounding factor is a *global signal* covariate. A global signal is an average signal change over time across the entire brain, obtained by taking a simple average of the voxel-wise time series. It is common to include a measure of the global signal as a covariate of no-interest (or to scale the data prior to analysis by this measure) to remove changes in blood oxygenation that impact the entire brain (resulting from, for example, changes in heart rate or respiration), which would otherwise obscure regional changes in neural activity. Because of the way in which it is measured, however, the global signal is expected to have some positive correlation with any experimentally evoked signal changes (as the average of all brain voxels will include those voxels responding to the task). As a result, correction for global signal changes can have a confounding effect on covariates of interest and greatly change the interpretation of evoked signal changes [26].

The resulting statistical model, composed of covariates of interest and those of no-interest, is then used to evaluate the time-series data from each voxel within the brain. The resulting weights upon the covariates (termed beta values) can then be evaluated alone or in combinations using *t* and *F*-statistics. The product is a *statistical map* in which every voxel in the brain contains a corresponding statistical value for the contrast of the covariates of interest. The final step of analysis involves assigning a level of statistical significance to those values. If the dataset were composed of a single voxel, then this would be a straightforward enterprise: a *t*-value of greater than 1.96 would be significant at a $p = 0.05$ level (presuming lots of degrees of freedom and a two-tailed test). As there are many voxels, however, corrections must be made for the likelihood that noise alone might render one *t*-value significant if many are tested. Such a correction attempts to control the false-positive rate at a *map-wise* level, meaning that if 20 statistical maps were produced under null-hypothesis conditions (i.e., in the absence of any actual experimental treatment), only one would on average be expected to contain even a single false-positive voxel. Solutions to perform this correction in the face of spatial smoothness within the statistical map (which yields statistical tests in adjacent voxels that are not fully independent) exist within Gaussian Random Field Theory [14] or through permutation [27].

Performing the appropriate, map-wise correction to control the false-positive rate can frequently yield a rather stringent statistical value necessary to label any result as significant. This, in turn, raises concerns about "false-negative" results, in which true experimental effects might be missed because the experiment is underpowered. There are several responses to this concern. Beyond the flippant call for more data, one might choose to relax the *p*-value that will be accepted as significant. Note that stating that the data were evaluated at a $p = 0.1$ level (corrected for multiple comparisons) is more intellectually honest than reporting results at a $p = 0.00023$ level (uncorrected), the latter tending to lull statistically naïve readers into a false sense of security. It is also preferable, whenever possible, to anatomically narrow one's hypothesis test. Using a predefined region of interest to test hypotheses can greatly reduce the number of independent statistical tests for which correction is required, thus improving power. In the limit, the number of tests can be reduced to one by taking the average signal within a region and performing the statistical test upon this representative data. Several methods are available for the definition of regions of interest. They might be defined anatomically, based upon gyral or cytoarchitectonic boundaries, or based on the previously reported lesion or functional neuroimaging studies. Regions of interest might also be defined functionally. For example, subjects might participate in an initial scan, the purpose of which is to define a region of the cortex that is maximally responsive to faces. Data obtained from this putative "face region" in subsequent experiments could then be studied with the benefits of focusing on the hypothesis test. Finally, regions might be defined using a "main effect" contrast, with subsequent, orthogonal "interaction" contrasts tested within the region. For example, an experiment might present pictures of upright and inverted faces. A region would be defined as the area that responds more to pictures of faces in either orientation than a third baseline condition. Within the defined region, the difference in response between upright and inverted faces could be assessed, without loss of statistical rigor, as the result of the test used to define the region does not prejudice the result of the subsequent orthogonal test of the effect of orientation of the stimulus. Of course, there are inferential consequences (such as loss of generality) of testing hypotheses only within predefined regions of interest. This might be countered by performing the same experiment-focused hypothesis tests within regions of interest, followed by more "exploratory" analyses that evaluate the data from the remainder of the brain, using appropriate map-wise correction for the increased number of voxels.

However, another approach that has gained popularity is the use of a "false-discovery rate" (FDR) statistical threshold

Table 21.1 Key articles exploring data analysis and experimental designs for fMRI

Authors	Article Title	Significance
Friston K, et al., 1995	The analysis of fMRI time-series revisited. Neuroimage. 1995 Mar;2(1):45–53	A method for detecting activations in fMRI time-series based on the general linear model and a heuristic analysis
Aguirre GK, et al., 2002.	Experimental design and the relative sensitivity of BOLD and perfusion fMRI. Neuroimage. 2002 Mar;15(3):488–500	This paper compares the statistical power of BOLD and arterial spin labeling perfusion fMRI for a variety of experimental designs within and across subjects
Friston K, et al., 2010.	Computational and dynamic models in neuroimaging. Neuroimage. 2010 Sep;52(3):752–65	This article reviews the substantial impact computational neuroscience has had on neuroimaging over the past years

[28]. Instead of controlling the false-positive rate at a map-wise level (allowing, for example, only 1 in 20 maps to have a single false-positive voxel), the FDR method controls the *proportion* of false-positive voxels present within a single map. For example, an FDR threshold of 5% implies that, of the voxels identified as significant within a statistical map, 5% of them are expected to be false positives on average. This is neither better nor worse than traditional map-wise control of the statistical significance but is instead a different stance with regard to inference. FDR methods will likely be of considerable use in clinical applications. For example, it may be desirable to express confidence in the results of functional mapping for surgical planning in terms of the specificity of the population of voxels identified.

Conclusion

This chapter summarizes the basic principles underlying the analysis of functional neuroimaging data. It also describes the methods that are currently being used in analyzing the functional neuroimaging data as well as new methods that are being developed in this field. (see Table 21.1 for a summary of key articles exploring data analysis and experimental designs for fMRI).

References

1. Atlas SW, Howard RS II, Maldjian J, Alsop D, Detre JA, Listerud J, et al. Functional magnetic resonance imaging of regional brain activity in patients with intracerebral gliomas: findings and implications for clinical management. Neurosurgery. 1996;38(2):329–38.
2. Detre JA, Alsop DC, Aguirre GK, Sperling MR. Coupling of cortical and thalamic ictal activity in human partial epilepsy: demonstration by functional magnetic resonance imaging. Epilepsia. 1996;37(7):657–61.
3. Esteller R, Echauz J, D'Alessandro M, Worrell G, Cranstoun S, Vachtsevanos G, et al. Continuous energy variation during the seizure cycle: towards an on-line accumulated energy. Clin Neurophysiol. 2005;116:517–26.
4. Schiff ND, Rodriguez-Moreno D, Kamal A, Kim KH, Giacino JT, Plum F, et al. fMRI reveals large-scale network activation in minimally conscious patients. Neurology. 2005;64(3):514–23.
5. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001;293:2425–30.
6. Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci. 2007;8:700–11.
7. Buchel C, Friston KJ. Assessing interactions among neuronal systems using functional neuroimaging. Neural Netw. 2000;13:871–82.
8. Price CJ, Friston KJ. Cognitive conjunctions: a new experimental design for fMRI. NeuroImage. 1997;5:261–70.
9. Sternberg S. Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. Acta Psychol. 2001;106:147–246.
10. Aguirre GK, Mattar MG, Magis-Weinberg L. de Bruijn cycles for neural decoding. NeuroImage. 2011;56:1293–300.
11. Grill-Spector K, Malach R. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol. 2001;107:293–321.
12. Aguirre GK, Zarahn E, D'Esposito M. The variability of human BOLD hemodynamic responses. NeuroImage. 1998;8:360–9.
13. D'Esposito M, Zarahn E, Aguirre GK, Rypma B. The effect of normal aging on coupling of neural activity to the BOLD hemodynamic response. NeuroImage. 1999;10(1):6–14.
14. Worsley KJ, Friston KJ. The analysis of fMRI time-series revisited again. NeuroImage. 1995;2:173–82.
15. Detre JA, Alsop DC. Perfusion fMRI with arterial spin labeling. In: Bandettini PA, Moonen C, editors. Functional MRI. Berlin: Springer; 1999. p. 47–62.
16. Aguirre GK, Detre JA, Zarahn E, Alsop DC. Experimental design and the relative sensitivity of BOLD and perfusion fMRI. NeuroImage. 2002;15:488–500.
17. Friston KJ, Zarahn E, Josephs O, Henson RN, Dale AM. Stochastic designs in event-related fMRI. NeuroImage. 1999;10:607–19.
18. Liu T. Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: design of experiments. Neuroimage. 2004;21:401–13.
19. Zarahn E, Aguirre GK, D'Esposito M. A trial-based experimental design for fMRI. Neuroimage. 1997;6(2):122–38.
20. Menon RS, Luknowsky DC, Gati JS. Mental chronometry using latency-resolved functional MRI. Proc Nat Acad Sci U S A. 1998;95:10902–7.
21. Henson RNA, Price CJ, Rugg MD, Turner R, Friston KJ. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage. 2002;15:83–97.
22. Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, et al. fMRI of human visual cortex [letter] [published erratum appears in Nature 1994 Jul 14;370(6485):106]. Nature. 1994;369(6481):525.
23. Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, et al. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging [see comments]. Science. 1995;268(5212):889–93.
24. McIntosh AR, Bookstein FL, Haxby JV, Grady CL, et al. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage. 1996;3:143–57.

25. Glover GH, Li TQ, Ress D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med. 2000;44:162–7.
26. Aguirre GK, Zarahn E, D'Esposito M. The inferential impact of global signal covariates in functional neuroimaging analyses. NeuroImage. 1998;8(3):302–6.
27. Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 2002;15:1–25.
28. Nichols T, Hayasaka K. Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat Methods Med Res. 2003;12(5):419–46.