Overview of fMRI analysis

S M SMITH, MA, DPhil

Oxford University Centre for Functional MRI of the Brain (FMRIB), John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

fMRI (functional magnetic resonance imaging) is a powerful non-invasive tool in the study of the function of the brain, used, for example, by psychologists, psychiatrists and neurologists. fMRI can give high quality visualization of the location of activity in the brain resulting from sensory stimulation or cognitive function. It therefore allows the study of how the healthy brain functions, how it is affected by different diseases, how it attempts to recover after damage and how drugs can modulate activity or post-damage recovery.

After an fMRI experiment has been designed and carried out, the resulting data must be passed through various analysis steps before the experimenter can get answers to questions about experimentally related activations at the individual or multi-subject level. This paper gives a brief overview of the most commonly used analysis pipeline: data pre-processing, temporal linear modelling and activation thresholding. For more details, see Jezzard et al [1].

fMRI Data

In a typical fMRI session a low-resolution functional volume is acquired every few seconds. (MR volumes are often also referred to as "images" or "scans"). Over the course of the experiment, 100 volumes or more are typically recorded. In the simplest possible experiment, some images will be taken whilst stimulation (for the remainder of this chapter, reference to "stimulation" should be taken to include also the carrying out of physical or cognitive activity) is applied, and some will be taken with the subject at rest. Because the images are taken using an MR sequence which is sensitive to changes in local blood oxygenation level (BOLD imaging; see Chapters 2 and 3 in Jezzard et al [1]), parts of the images taken during stimulation should show increased intensity, compared with those taken whilst at rest. The parts of these images that show increased intensity should correspond to the brain areas which are activated by the stimulation.

The goal of fMRI analysis is to detect, in a robust, sensitive and valid way, those parts of the brain that show increased intensity at the points in time that stimulation was applied.

A single volume is made up of individual cuboid elements called voxels (Figure 1). An fMRI data set from a single session can either be thought of as *t* volumes, one taken every few seconds, or as *v* voxels, each with an associated time series of *t* time points. It is important to be able to conceptualize both of these representations, as some analysis steps make more sense when thinking of the data in one way, and others make more sense the other way.

An example time-series from a single voxel is shown in Figure 2. Image intensity is shown on the *y* axis, and time (in scans) on the *x* axis. As described above, for some of the time points, stimulation was applied, (the higher intensity periods), and at some time points the subject was at rest. As well as the effect of the stimulation being clear, the high frequency noise is also apparent. The aim of fMRI analysis is to identify in which voxels' time-series the signal of interest is significantly greater than the noise level.

Preparing fMRI data for statistical analysis

Initially, a four-dimensional (4D) data set is pre-processed, *i.e.* prepared for statistical analysis. For more detail, see chapters 12 and 13 in Jezzard et al [1]. Once data has been acquired by the MR scanner, the pre-processing starts by reconstructing the raw "k-space" data into images that actually look like brains. Very often, the next step applied is slice-timing correction; because each slice in each volume is acquired at slightly different times, it necessary to adjust the data so that it appears that all voxels within one volume had been acquired at exactly the same time (all subsequent processing is far simpler if this is done). Each volume is now transformed (using rotation and translation) so that the image of the brain within each volume is aligned with that in every other volume; this is known as motion correction.

Most researchers now blur each volume spatially, principally to reduce noise, hopefully without significantly affecting the activation signal. After this, each volume's overall intensity level may be adjusted so that all volumes have the same mean intensity – this intensity normalization can help reduce the effect of global changes
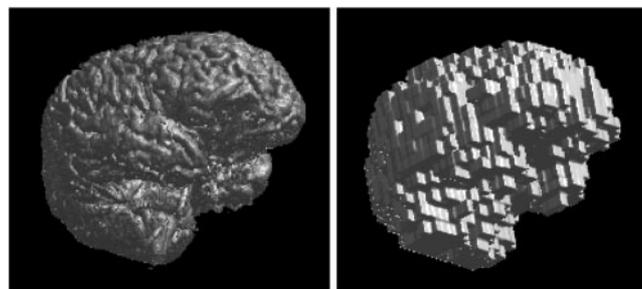

Figure 1. What are voxels? Shown here are surface renderings of 3D brain images. On the left is a high-resolution image, with small (1 mm × 1 mm × 1.5 mm) voxels; the voxels are too small to see. On the right is a low-resolution image of the same brain, with large (7 mm × 7 mm × 10 mm) voxels, clearly showing the voxels making up the image.
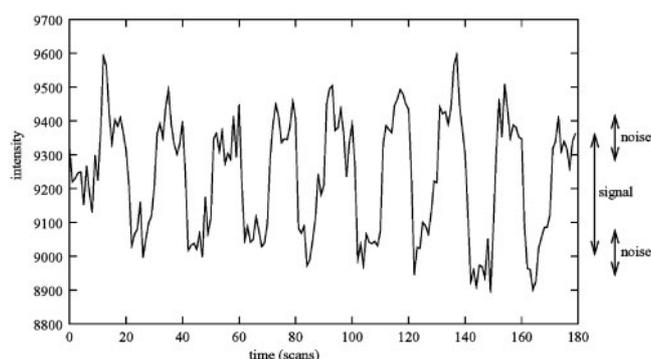
Figure 2. An example time series at a strongly activated voxel from a visual stimulation experiment. Here the signal is significantly larger than the noise level. Periods of stimulation are alternated with periods of rest – a complete stimulation-rest cycle lasts 20 scans.

in intensity over time. Reduction in low and high frequency noise is normally desired as a final step; each voxel's time series is filtered by linear or non-linear tools in order to achieve this.

The purpose of the pre-processing is to remove various kinds of artefacts in the data, and to condition the data, in order to maximize the sensitivity of later statistical analysis, and also, in some situations, to increase the statistical validity.

Statistical analysis of activation images

In this section we give a very brief overview of different approaches to obtaining activation maps, followed by a slightly more detailed introduction to analysis via the general linear model (GLM; currently the most popular statistical approach) and also various methods of thresholding the resulting statistics maps.

After the pre-processing steps, statistical analysis is carried out to determine which voxels are activated by the stimulation. This can be simple correlation analysis or more advanced modelling of the expected haemodynamic response to the stimulation. Various possible statistical corrections can be included, such as correction for smoothness of the measured time series at each voxel. The main output from this step is a statistical map, which indicates those points in the image where the brain has activated in response to the stimulus.

It is most common to analyse each voxel's time series independently ("univariate analysis"). For example, standard GLM analysis is univariate (although cluster-based thresholding, commonly used at the final inference stage, does use spatial neighbourhood information and is therefore not univariate). However, there are also "multivariate" methods [3] which process all the data together; these methods make more use of spatial relationships within the data than univariate analysis. Note that most model-free methods (see the following paragraph) are also multivariate.

There is also a distinction between model-based and model-free methods. In a model-based method [4], a model of the expected response is generated and compared with the data. In a model-free method [5, 6], effects or components of interest in the data are found on the basis of some specific criteria (for example, the spatial or

temporal components should be statistically independent of each other). This allows for "surprise" in the data, and also the analysis of data where it is difficult to generate a good model. There are also a few methods which lie between model-based and model-free, for example Clare et al [7], where the only "model" information given is the time of the beginning of each stimulation period (the actual time-course within each period is not pre-specified). A statistic map is generated by comparing the variance within periods with the variance across periods.

General linear model: overview

General linear modelling (more correctly, just "linear modelling") sets up a model (*i.e.* a general pattern which you expect to see in the data) and fits it to the data. If the model is derived from the timing of the stimulation that was applied to the subject in the MRI scanner, then a good fit between the model and the data means that the data were probably caused by the stimulation. As the GLM is normally used in a univariate way, the rest of this section considers one voxel only, and the fitting of models to a single voxel's time-course. Thus consider that the data of interest comprise a single 1D vector of intensity values.

A very simple example of linear modelling is:

$$y(t) = \beta \times x(t) + c + e(t) \qquad (1)$$

where $y(t)$ is the data, and is a 1D vector of intensity values – one for each time point, *i.e.* is a function of time. $x(t)$ is the model, and is also a 1D vector with one value for each time point. In the case of a square-wave block design, $x(t)$ might be a series of 1s and 0s; for example, 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 etc. β is the parameter estimate for $x(t)$, *i.e.* the value that the square wave (of height 1) must be multiplied by to fit the square wave component in the data. c is a constant, and in this example, would correspond to the baseline (rest) intensity value in the data. e is the error in the model fitting. Thus the model fitting involves adjusting the baseline level and the height of the square wave, to best fit the data; the error term accounts for the residual error between the fitted model and the data.

If there are two types of stimulus, the model would be:

$$y = \beta_1 \times x_1 + \beta_2 \times x_2 + c + e \qquad (2)$$

Thus there are now two different model waveforms corresponding to the two stimulus time-courses. There are also two interesting parameters to estimate, β_1 and β_2. Thus if a particular voxel responds strongly to model x_1 the model-fitting will find a large value for β_1; if the data instead looks more like the second model time-course, x_2, then the model-fitting will give β_2 a large value. Different model waveforms within a complex model are often referred to as explanatory variables (EVs), as they explain different processes in the data.

In order to get the best possible fit of the model to the data, the "stimulus function" (which is often a sharp on/off waveform) is convolved with the haemodynamic response function (HRF). This process mimics the effect that the brain's neurophysiology has on the input function
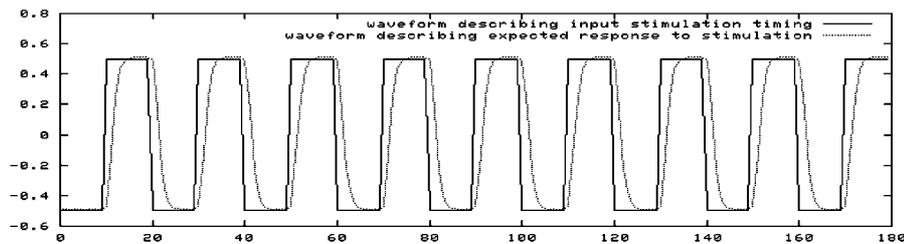
Figure 3. Model waveform formation: the square waveform describes the input stimulation timing; the smoothed waveform results from convolving the first with the haemodynamic response function, a transformation which leaves the model looking much more like the measured data.

(the stimulation). The brain's haemodynamic response is a delayed and blurred version of the input time-series, so a mathematical operation is applied to the stimulus function to take the square wave input and create a delayed and blurred version, which will better fit the data. For example, see Figure 3, showing the raw stimulation timing waveform and the HRF-convolved model – $x(t)$ – which will be used in the model fitting.

The GLM is often formulated in matrix notation. Thus all of the parameters are grouped together into a vector $\vec{\beta}$, and all of the model time-courses are grouped together into a matrix $\vec{X}$, often referred to as the design matrix. Figure 4 shows an example design matrix with two such model time-courses. Each column is a different part of the model. For example, in a particular experiment, both visual and auditory stimulations are applied, but with different timings; the left column (x_1 or EV 1) models the visual stimulation, and the right column (EV 2 or x_2) models the auditory stimulation.

As described above, when the model is fit separately to the data at each voxel, there will be found an estimate of the "goodness of fit", of each column in the model, to that voxel's time-course. In the example visual/auditory experiment, the first column will generate a high first parameter estimate in the visual cortex. However, the second column will generate a low second parameter estimate, as this part of the model will not fit the voxel's time-course well.

To convert a parameter estimate (PE, *i.e.* the estimated β value) into a useful statistic, its value is compared with the uncertainty in its estimation (resulting in what is known as a T value; T=PE/standard error (PE)). If the PE is low relative to its estimated uncertainty, the fit is not significant. Thus T is a good measure of whether the estimate of the PE value is significantly different from zero, *i.e.* whether there is believable activation. (Remember, all of this is carried out separately for each voxel.) To convert a T value into a p (probability) or Z statistic requires standard statistical transformations; however, T, p and Z all contain the same information; they describe how significantly the data are related to a particular part of the model (x_1 or x_2).

As well as producing images of Z values that describe how strongly each voxel is related to each EV (one image per EV), parameter estimates can be compared to test directly whether one EV is more "relevant" to the data than another. To do this, one PE is subtracted from another, the standard error for this new value is calculated, and a new T image is created. All of the above is controlled by setting up "contrasts". To compare two EVs (for example, to subtract one stimulus type – EV 1 – from another type – EV 2), set EV 1's contrast value to −1 and EV 2's to 1. This is often written as a contrast of

[−1 1], as the contrasted parameter estimate equals $-1 \times \beta_1 + 1 \times \beta_2$. A T statistic image will then be generated according to this request, answering the question "where is the response to stimulus 2 significantly greater than the response to stimulus 1?"

It is possible that the response to two different stimuli, when applied simultaneously, is greater than that predicted by adding up the responses to the stimuli when applied
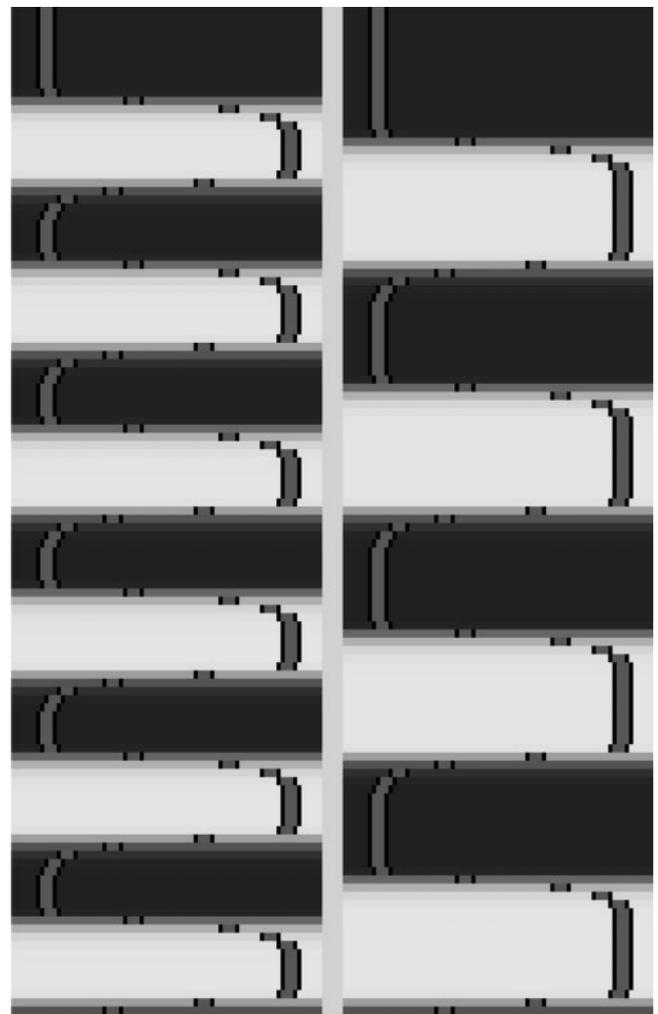

Figure 4. Example design matrix with two explanatory variables; two different stimulations are being applied. Because they have different timings, they are modelled separately (Equation 2). Time is on the y axis, pointing downwards. Note that a column has two representations of the model's value at each point in time: the underlying intensity encodes the model's value at a particular time point, and so does the line graph.

separately. If this is the case, then such "non-linear interactions" need to be allowed for in the model. The simplest way of doing this is to setup the two original EVs, and then add an interaction EV, which will only be "up" when both of the original EVs are "up", and "down" otherwise. In Figure 5, EV 1 could represent the application of a drug, and EV 2 could represent visual stimulation. EV 3 will model the extent to which the response to drug&visual is greater than the sum of drug-only and visual-only. A contrast of [0 0 1] will show this measure, whilst a contrast of [0 0 −1] shows where negative interaction is occurring.

Ideally, all of the EVs should be independent of each other. If any EV is close to being a sum (or weighted sum) of other EVs in the design, then the fitting of the model to the data does not work properly (the design matrix is not of "full rank"). A common mistake of this type is to model both rest and activation waveforms, making one an upside-down version of the other; in this case EV 2 is −1 times EV 1, and therefore linearly dependent on it. It is only necessary to model the activation waveform.

With "parametric designs", there are typically several different levels of stimulation, and it is common to estimate the response to each level separately. This requires a separate EV for each stimulation level. The different contrasts ask different questions about these responses: For example, for an experiment with three stimulation levels, [1 0 0] shows the response to stimulation level 1 versus rest (likewise [0 1 0] for level 2 *vs* rest and [0 0 1] for level 3). [−1 1 0] shows where the response to level 2 is greater than that for level 1. [−1 0 1] shows the general linear increase across all three levels. [1 −2 1]
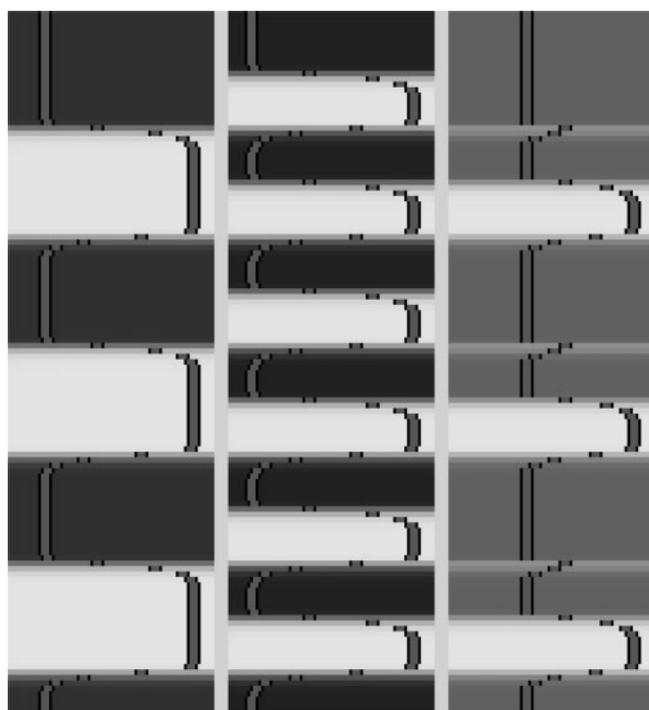

Figure 5. Example of modelling a non-linear interaction between stimuli. The first two EVs model the separate stimuli, whilst the third models the interaction, *i.e.* accounts for the "extra" response when both stimuli are applied together.

shows where the level dependence deviates, in an upwardly-curving way, from being linear.

Inference ("thresholding")

Thus we now have a statistic map (for example, *T* or *Z*). The next step is to threshold this, in order to decide, at a given level of significance, which parts of the brain were activated. There are a variety of ways of carrying out thresholding. These are now briefly outlined.

The simplest method of thresholding is to select a significance (*p*) threshold and apply this to every voxel in the statistic map (it is straightforward to convert a *p* level to a suitable threshold for the map type, *e.g.* a *T* threshold). A problem with this is that there are many tests being carried out, because there are so many voxels in the brain. If 20 000 voxels are tested for at a significance of $p<0.01$ then it is expected that 200 will activate by chance, even if no stimulation is applied. It is not ideal to blindly accept these as being activated! This "multiple-comparison problem" means that it is not valid to accept all activations reported by this method of thresholding; a correction is necessary to reduce the number of false positives. Typically a Bonferroni correction is used, where the significance level at each voxel is divided by the number of voxels; this corrects for the number of comparisons being made. However, this results in very stringent thresholding (*i.e.* in the case given above, the resulting *p* threshold is 0.01/20 000=0.0000005).

A refinement of the above voxel-wise thresholding is to use Gaussian random field (GRF) theory to threshold the image. The main difference is that this method takes into account the spatial smoothness of the statistic map (*i.e.* estimates the number of statistically independent voxels, which is smaller than the original number). This method is less "over-conservative" than simple voxel-wise thresholding with Bonferroni correction; typically the correction to *p*-values is reduced (compared with Bonferroni correction) by a factor of 2–20.

Finally, it is possible, again using GRF theory, to take into account spatial extent (*i.e.* size) of clusters of activations, before estimating significance. Thus instead of assigning a *p*-value to each voxel, clusters of voxels are created on the basis of an initial thresholding, and then each cluster is assigned a *p*-value, which may or may not pass the final significance test. It is often the case that this method is more sensitive to activation than the voxel-based methods. A limitation is the arbitrary nature of the initial thresholding, used to create the clusters.

Multi-subject statistics

Although so far we have only discussed single-session analyses, it is common to run an experiment several times, either on the same subject, or with several different subjects, or both. This can both increase the sensitivity of the overall experiment (as more data can lead to increased sensitivity to an effect) or allow the generalization of any conclusions to the whole population.

In order to combine statistics across different sessions or subjects, the first necessary step is to align the brain images from all sessions into some common space. This is typically done using generic registration tools and can

be carried out either on the raw data (*i.e.* before the single-session, or "first-level", analyses) or on the statistic maps created by the first-level analyses.

Once all the data are aligned, there are a variety of statistical methods for combining results across sessions or subjects, to either create a single result for a group of subjects, or to compare different groups of subjects (for example, placebo group versus drug treatment group). These methods include "fixed-effects" and "mixed-effects" analyses. Fixed-effects assumes that all subjects activate equally, and is only interested in within-session errors. Mixed-effects analysis additionally takes into account between-session errors, and therefore makes fewer assumptions about the data; its results are therefore valid for the whole population from which the group of subjects is drawn (whereas fixed-effects results are not). However, the mixed-effects analysis tends to give more "conservative" results.

How many subjects are required, in order for multi-subject statistics to be sensitive and robust? The answer is not simple, as it depends on many factors including the level of response to the stimulation, between-subject variability and scanner characteristics. (Furthermore, all things being equal, a smaller number of subjects is generally required for a fixed-effects group analysis than for a mixed-effects analysis.) However, it is generally accepted that groups containing less than 10 subjects are suboptimal.

For more detail on group modelling and estimation and group size, see Beckmann et al, Woolrich et al, Friston et al [8–10].

Registration, brain atlases and cortical flattening

As described above, registration (aligning different brain images) is often used when combining fMRI data from different sessions or subjects. It can also be used to align the low-resolution fMRI images to a high-resolution structural image, so that the activations can be viewed in the context of a good quality brain image; this aids in the interpretation of the activations.

A related issue is the use of templates and brain atlases. These consist of data which are transformed into some "standard brain space", for example the co-ordinate system specified by Talairach and Tournoux [11]. A template is typically an average of many brains, all registered into any given common co-ordinate system. An example is the MNI 305 average [12]. An atlas is also based in a common co-ordinate system, but contains more sophisticated information about the brain at each voxel, for example, information about tissue type, local brain structure or functional area. Atlases can inform interpretation of fMRI experiments in a variety of ways, helping the experimenter gain the maximum value from the data.

Finally, another related issue is that of cortical flattening. Here a high-resolution structural image is used to estimate the convoluted cortical surface (this can be considered to be the simple 2D surface, consisting of all cortical grey matter, which has been "crumpled" into the skull). Because cortical activation should only lie on this surface, new statistical constraints and pre-processing methodologies can be developed which use this spatial information, in order to improve on the simple 3D

processing most commonly used. Also, it can be of value to "project" estimated activations onto a flattened version of the surface, enabling interesting interpretation of the relative placements of different activations.

For more detail on the structural analysis issues described in this section, see Chapter 15 of Jezzard et al [1].

GLM-based analysis example

This section describes briefly the practical and numerical details involved in the analysis of a particular fMRI experiment. Most of the analysis concepts referred to within this section are explained more fully in Jezzard et al [1].

The experiment attempted to compare the response to thermal stimulation between two groups of subjects. One group comprised 10 clinical patients and the second was a control group of 8 subjects. Within each session, brief periods of warm or hot thermal stimulation were applied briefly to the back of the subject's hand. For each session, 250 volumes were acquired, one every 3 s. A question of interest was how the relative response to hot and warm stimulation differed between the two groups.

All analysis was carried out within the fMRI analysis package FEAT (FMRIB's Easy Analysis Tool) [13].

First-level analysis

Each individual subject's session was processed independently, using the following analysis: Head motion correction was carried out using MCFLIRT [14], a tool that corrects for rigid-body (rotation and translation) motion, using an accurate and robust multiscale optimization strategy. Figure 6 shows the estimated rotation and translation parameters used to correct for the head motion. MCFLIRT is fully automated, *i.e.* requires no user-interaction.

Spatial filtering was carried out on every volume using a Gaussian profile filter of full-width-half maximum (FWHM) 5 mm. Example slices from an example volume, before and after spatial filtering, are shown in Figure 7.

Each session's data set was intensity normalized to have a mean within-brain intensity of 10 000 units. No volume-by-volume intensity normalization was carried out.

Linear (Gaussian-weighted running line detrending) high-pass filtering was applied, with FWHM 50 s. The high-pass cut-off was this low because this is a single-event design, which allows more aggressive high-pass filtering than block-design experiments. No low-pass filtering was applied, due to the nature of the later GLM procedure; as pre-whitening of each voxel's time series was used, low-pass filtering is inappropriate. An example time-course from a single (activated) voxel is shown in Figure 8, before and after high-pass filtering.

Next, GLM analysis was carried out on each subject's pre-processed data using FILM [15]: generalized least-squares multiple regression was used; the model was initially fit to the data and fitted activation subtracted from the data. The resulting first-pass residuals were then pre-whitened by estimating the autocorrelation structure
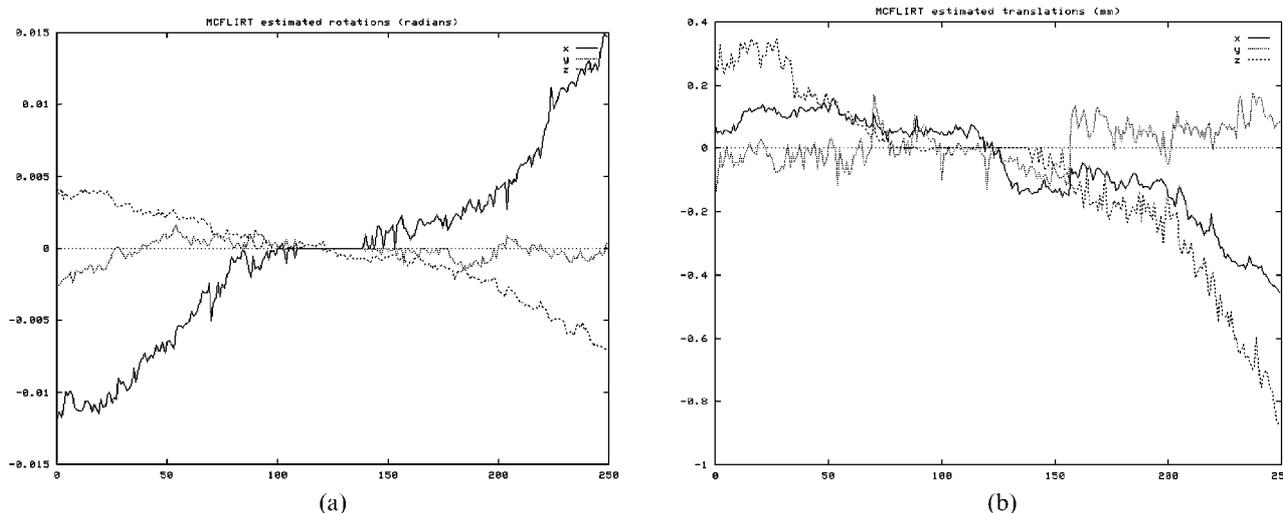
Figure 6. (a) Rotation and (b) translation parameters estimated by a rigid-body motion correction procedure, as a function of scan number.
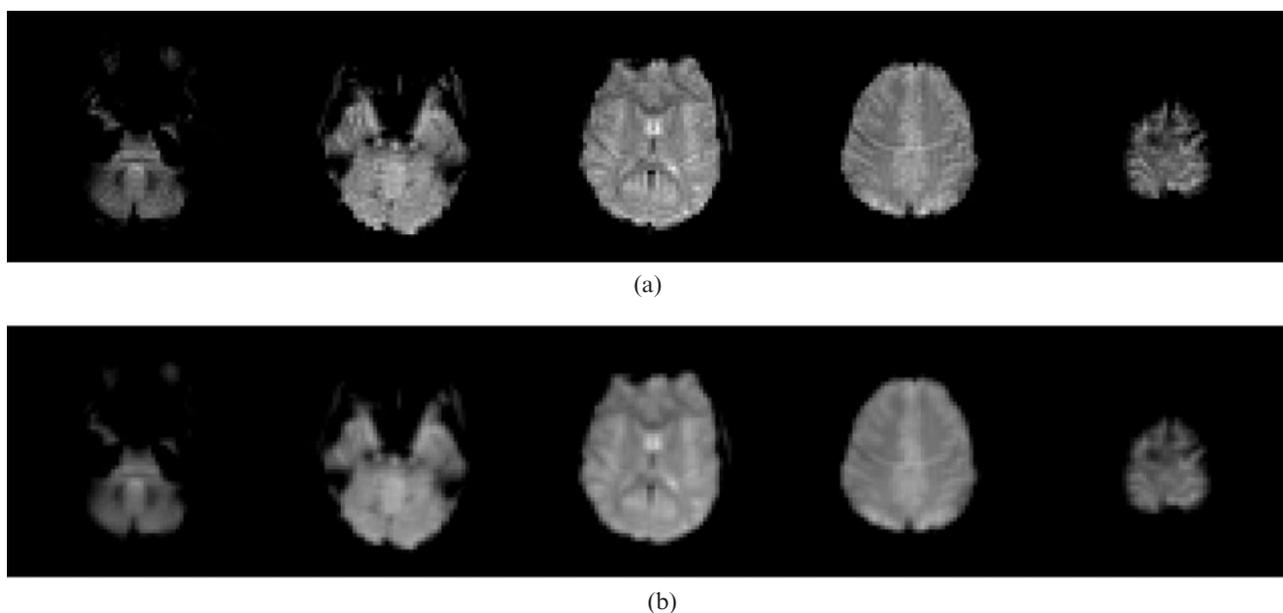

Figure 7. Example slices from an example fMRI volume, (a) before and (b) after spatial filtering of 5 mm full width half maximum (FWHM).

(calculating the autocorrelation coefficients, then using Tukey tapering and within-tissue-type spatial regularization to improve estimation accuracy and robustness). The model was then re-fit to the pre-whitened data to give maximum efficiency estimation.

The model used is shown in Figure 9 (and an example fit of the model to one voxel's timecourse is shown in Figure 8). EV 1 models the periods of painful heat, EV 3 models "pain conditioning" (a light of a certain colour was shown just before the painful heat), EV 5 models warm thermal stimulation and EV 7 models "warm conditioning" (using a different coloured light). The even numbered EVs are simply the temporal derivatives of the odd numbered derivatives. These are included because they can correct for slight overall temporal shifts between the model and the data (adding, to an original signal, a small amount

of the temporal derivative of that signal, is equivalent to shifting the original signal slightly in time).

As the single-subject statistics were then to be fed into a multi-subject second-level analysis, thresholding was not in general carried out on the first-level results. However, an example single-session activation map was created by thresholding one contrast (pain-warm, *i.e.* a contrast of [1 0 0 0 −1 0 0 0]) from one subject's data with cluster-based thresholding; clusters were formed by thresholding at $Z > 2.3$, and then each cluster was tested for significance at $p < 0.01$. Also, as an example of using a single-subject's structural image to render activation onto, one subject's fMRI data were registered onto that subject's high-resolution image: first, the structural was brain-extracted (*i.e.* non-brain matter removed) using BET [16]. Then the fMRI image was registered to the
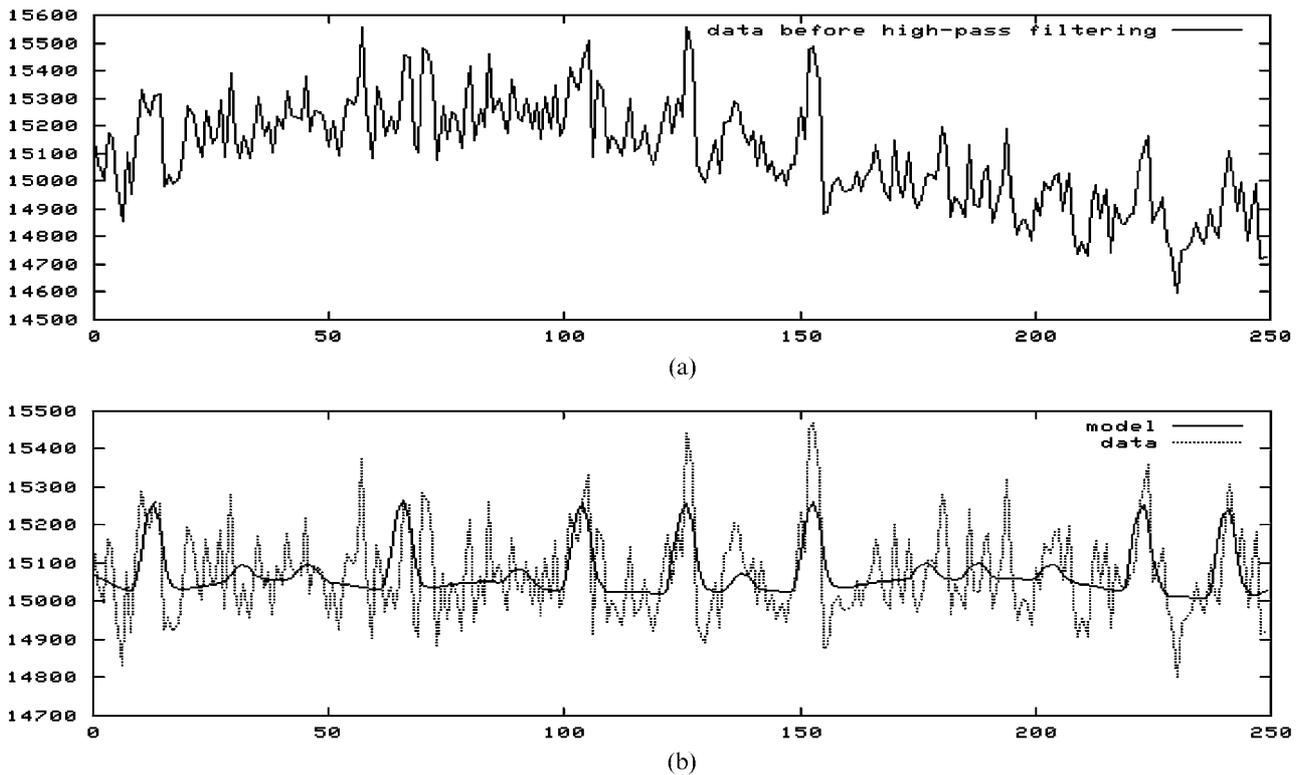
Figure 8. An example time-course from a single (activated) voxel. (a) Time-course before high-pass filtering. (b) After filtering, and plotted against the fitted model.
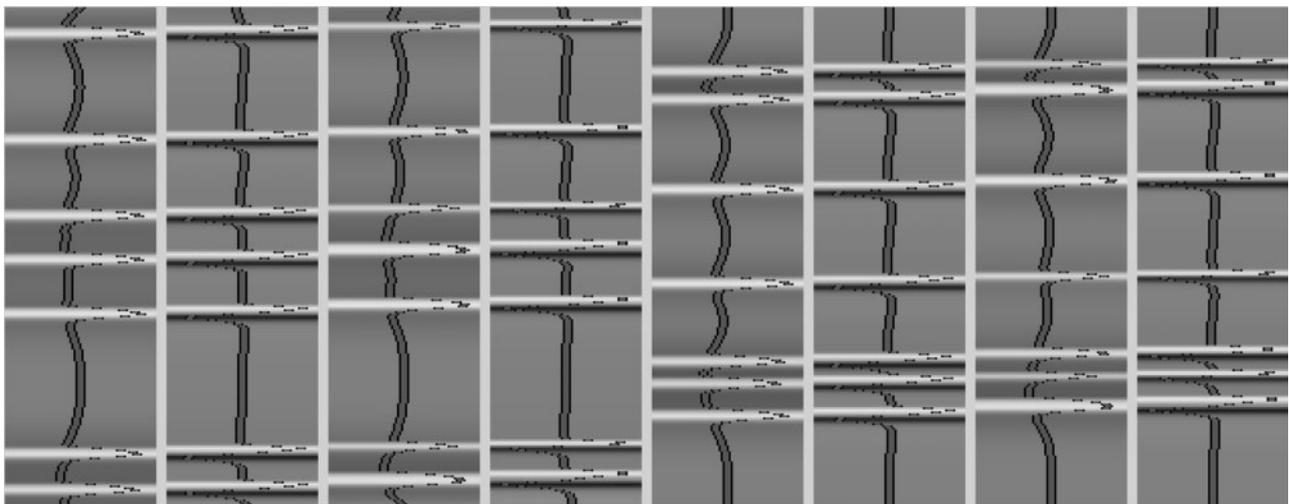

Figure 9. The model, or design matrix, used in the general linear model analysis of the heat-warm experiment. EVs 1 and 5 model pain and warm, EVs 3 and 7 model conditioning to pain and warm, and the even numbered EVs are simply the temporal derivatives of the odd numbers; these allow phase shifts in the fitting.

brain-extracted high-resolution image using FLIRT [17], and this transform was applied to the thresholded activation map. The various stages involved can be seen in Figure 10.

Second-level analysis

Second-level analysis was carried out to test for differences in the pain-warm contrast between the two

subject groups. The single-session data sets were registered into standard space using FLIRT in a two-step process. First the low-resolution fMRI data from each subject was registered to that subject's (brain-extracted) high-resolution structural MRI using a 7 degrees-of-freedom (DOF) linear fit. Then the high-resolution image was registered to a Talairach-space standard brain image (the MNI 152 image) using a 12 DOF linear fit. The average image from all 18 subjects is compared with the standard space target
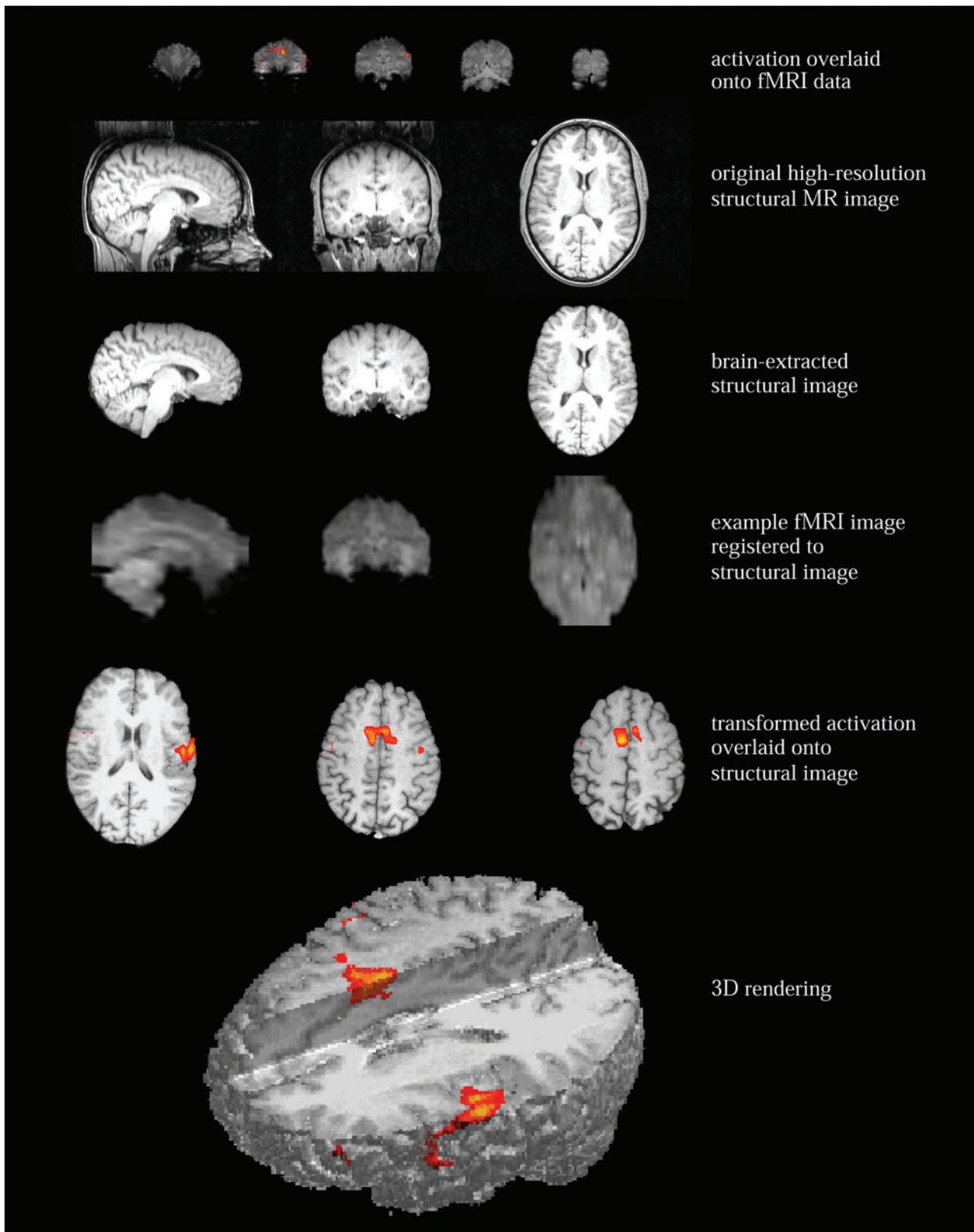
Figure 10. Various stages in the rendering of activation onto a high-resolution structural image.

image in Figure 11. Then, for each subject, the two transforms were combined mathematically to give a single transform taking the fMRI data into standard space; this was applied to the first-level statistic maps (activation parameter estimates and variance estimates) to take them into standard space. Thus all subjects' statistical data are now in alignment.

Second-level mixed-effects analysis was then carried out using the first-level statistic maps to test for differences in activation between the two subject groups. The resulting
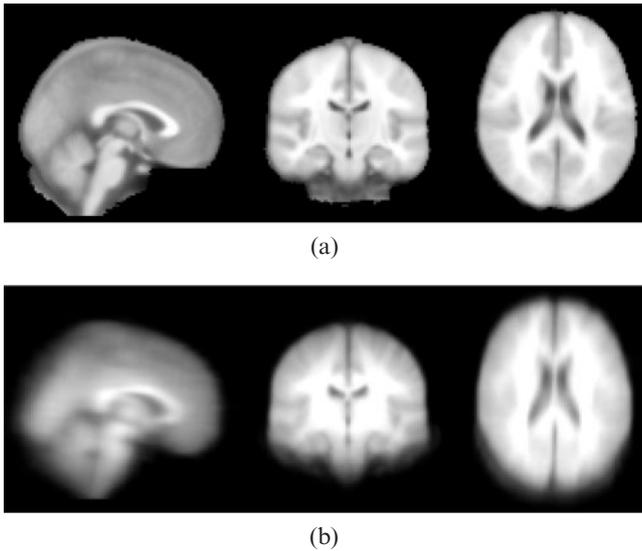
(a)

(b)

Figure 11. Comparison of (a) the standard space MNI 152 image and (b) the mean of the 18 subjects' high-resolution structural images after transformation into standard space. The mean of the subjects' MR ''high-resolution'' images is fairly blurred, due to the relatively low resolution of most of the structural scans taken for this study.
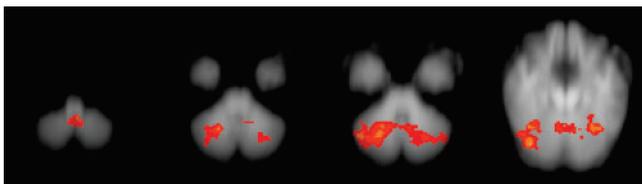

Figure 12. Significant differences between the two subject groups in the pain-warm contrast.

statistic maps were thresholded as above, *i.e.* using clusters formed by $Z>2:3$ and then tested with $p<0.01$. These results can be seen in Figure 12. These final results are in Talairach co-ordinate space, so meaningful co-ordinates can be given for the centres of activation. Thus the only remaining process is the final interpretation of this group comparison.

Acknowledgment

References

1. Jezzard P, Matthews P, Smith S, editors. Functional MRI: an introduction to methods. Oxford: OUP, 2001.
2. Friston K, Poline J-B, Holmes A, Frith C, Frackowiak R. A multivariate analysis of PET activation studies. Human Brain Mapping 1996;4:140–51.
3. Friston K, Holmes A, Worsley K, Poline J-B, Frith C, Frackowiak R. Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping 1995;2:189–210.
4. McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, et al. Analysis of fMRI data by blind separation into independent spatial components. Human Brain Mapping 1998;6:160–88.
5. Beckmann C, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 2004;23:137–52.
6. Clare S, Humberstone M, Hykin J, Blumhardt L, Bowtell R, Morris P. Detecting activations in event-related fMRI using analysis of variance. Magn Reson Med 1999;42:1117–22.
7. Beckmann C, Jenkinson M, Smith S. General multi-level linear modelling for group analysis in FMRI. NeuroImage 2003;20:1052–63.
8. Woolrich M, Behrens T, Beckmann C, Jenkinson M, Smith S. Multi-level linear modelling for FMRI group analysis using Bayesian inference. NeuroImage 2004;21:1732–47.
9. Friston K, Holmes A, Worsley K. How many subjects constitute a study? NeuroImage 1999;10:1–5.
10. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publisher Inc., 1988.
11. Collins D, Neelin P, Peters T, Evans A. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 1994;18:192–205.
12. Smith SM, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 2004;23:S208–19.
13. Jenkinson M, Bannister P, Brady J, Smith S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 2002;17:825–41.
14. Woolrich M, Ripley B, Brady J, Smith S. Temporal autocorrelation in univariate linear modelling of FMRI data. NeuroImage 2001;14:1370–86.
15. Smith S. Fast robust automated brain extraction. Human Brain Mapping 2002;17:143–55.
16. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Medical Image Analysis 2001;5:143–56.