



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

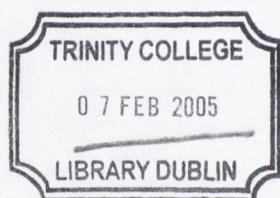
METHODOLOGICAL ISSUES IN EVENT-RELATED fMRI

Kevin Murphy

**A thesis submitted for the degree of
Doctor of Philosophy**

**University of Dublin
Trinity College**

2005

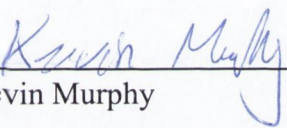


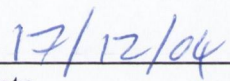
THOES
7467

Declaration

I declare that this work has not been submitted previously as an exercise for a degree at this or any other university and that it is entirely my own work. I give my permission for the Trinity College Library to lend or copy this thesis upon request.

Signed


Kevin Murphy


Date

Acknowledgements

Certainly, the first person I should thank is my supervisor Hugh Garavan. This thesis would not exist only for his brilliant ideas and advice. His ability to ask relevant and important questions has really influenced the direction of this research and I am very grateful for all the interest he took in my work.

Unlimited thanks go to Catherine and Rob whose friendship and helpful discussions throughout the last three years have really made my time here enjoyable and productive.

Next, I would like to thank all the other members of the lab throughout the years, Andrea, Clare, Erik, Cristina and Elena. I would never have learned so much in my three years, if you hadn't been there to ask questions.

Thanks to all my friends in the Department for their endless support and friendship, especially Jason, Deirdre, Mark, Slick, Clare, Mary, Fiadhait and my fellow banditeers Redmond and Paul.

I am deeply indebted to all my friends from the school days, Hickey, Fowler, Dave, Sean, Hynsey, Karen, Fiona and Grace for all the laughs over the years. I should also mention Dob here, who has proved himself one of the group and a great friend.

But for Orlaith's support at the beginning of this PhD, I don't think I'd have even started. Thanks for sharing all those years with me.

Apologies to anybody who deserves to be here but I've neglected to mention. I will make it up to you in pint format someday.

Love to all my family, Mam, Dad and Áine. Without your support and encouragement over the last eight years of college, none of this would have been possible.

Last but not least, I would like to thank Fiadhna for being there when I needed her and for making me happy.

Supported by USPHS grants DA14100-01, GCRC M01 RR00058, DA11326 and the Irish Research Council for Humanities and Social Sciences.

Summary

Functional magnetic resonance imaging (fMRI) is a very powerful tool for localising psychological functions to specific brain areas. Recently, there has been a surge of interest in determining the functional differences that might characterise certain populations. However, the results of event-related fMRI can be questionable due to the complexities of the techniques involved. A researcher must eliminate possible between-group confounds and optimally design experiments to increase confidence in findings. Addressing some of these methodological concerns is the focus of this thesis. Four questions have been chosen that are of particular importance in event-related fMRI, but have been poorly investigated in the literature. These questions span the entire field of fMRI, from physiology through experimental design to analytic techniques.

The first study addresses the concern of group difference confounds due to deviations in the haemodynamic properties of each group. Given that fMRI is an indirect measure of neuronal activity, to investigate with confidence differences between groups of interest, one must be sure that no confounding factors in haemodynamic properties that are independent of neuronal differences exist. By comparing three groups that abuse vascular-affecting drugs (cocaine, nicotine and THC) with controls, it was demonstrated that there is no significant alteration in the event-related haemodynamic response. Thus, fMRI activation differences reported between these drug groups can be more confidently interpreted as reflecting true neuronal differences. This result may extend to comparisons of other non-drug groups, since it

is unlikely that haemodynamic responses would be compromised to the same extent in these groups.

The second and third studies address questions that are closely related to each other: how many events are required to produce a stable single-subject activation map and how many subjects are required to produce a stable group activation map. By using three different analytic techniques on a fingertapping dataset containing 150 events, it was demonstrated that only 25 events are required to yield stable activation maps, provided the correct analytic method is chosen. Using a cohort of 58 subjects, the third study demonstrated that power and reproducibility, based on voxelwise overlap, are very poor with subject numbers in the typical range for fMRI studies (10-20). By using a more intuitive measure based on the centres-of-mass of activated areas, it was shown that the similarity between group maps does not increase beyond 20 subjects.

Analysis techniques comparing groups or conditions that vary in performance are open to a possible confound driven by those performance differences, if these errors are ignored. The final study demonstrated that by treating errors as corrects, many erroneous between-group and between-condition differences are produced. Consequently, brain-imaging investigations that do not accommodate error contributions to functional signals are at risk of misinterpreting activation patterns.

By advocating the use of event-related fMRI designs, addressing a number of confounding factors and making recommendations about fundamental experimental variables, this thesis contains essential knowledge to any researcher wanting to design accurate and reliable fMRI studies.

Publications

The following is a list of first- and co-author manuscripts that have been submitted for publication throughout the duration of this Ph.D.

Thesis chapters derived from the following articles:

Chapter 2:

Murphy, K., Dixon, V., LaGrave, K., Kaufman, J., Risinger, R., Bloom, A., & Garavan, H. *A validation of event-related fMRI comparisons between drug groups and controls.* American Journal of Psychiatry (Under Review).

Chapter 3:

Murphy, K., & Garavan, H. *Deriving the optimal number of events for an event-related fMRI study based on the spatial extent of activation.* NeuroImage (Under Review).

Chapter 4:

Murphy, K. & Garavan, H. (2004). *An empirical investigation into the number of subject required for an event-related fMRI study.* NeuroImage 22 (2), 879-885

Chapter 5:

Murphy, K. & Garavan, H. (2004). *Artifactual fMRI group and condition differences driven by performance confounds.* NeuroImage 21, 219-228.

Peer reviewed journal articles:

Burke, D., **Murphy, K.,** Garavan, H., & Reilly, R. (2004). *A Pattern Recognition Approach to the Detection of Single-Trial Event-Related fMRI.* Medical & Biological Engineering and Computing (In Press).

Fassbender, C., **Murphy, K.,** Foxe, J.J., Wylie, G., Javitt, D.C., Robertson, I.H., & Garavan, H. (2004). *A Topography of Executive Functions revealed by functional Magnetic Resonance Imaging.* Cognitive Brain Research 20(2), 132-143.

Hester, R. **Murphy, K.** & Garavan, H. (2004). *Beyond Common Resources: The Cortical Basis for Resolving Task Interference.* NeuroImage 23(1), 202-212.

Hester, R., **Murphy, K.,** Foxe, D.M., Foxe, J., & Garavan, H. (2003). *Predicting Success: The effect of pre-target cueing on inhibition performance.* Journal of Cognitive Neuroscience 16(5), 776-785.

Kübler, A., **Murphy, K.**, Kaufman, J., Stein, E.A., & Garavan, H. (2003). *Co-ordination within and between verbal and visuospatial working memory: Network modulation and anterior frontal recruitment*. *NeuroImage* 20, 1298-1308.

Garavan, H., Ross, T. J., **Murphy, K.**, Roche, R. A. P., & Stein, E. A. (2002). *Dissociable executive functions in the dynamic control of behaviour: Inhibition, error detection and correction*. *NeuroImage* 17, 1820-1829.

Kelly, A.M.C., Hester, R., **Murphy, K.**, Javitt, D.C., Foxe, J.J. & Garavan, H. *Prefrontal-Subcortical Dissociations Underlying Inhibitory Control Revealed by Event-Related fMRI*. *European Journal of Neuroscience* 19, 3105-3112.

Manuscripts currently under review:

Fassbender, C., **Murphy, K.**, Hester, R., Foxe, J. J., Foxe, D. M., Javitt, D. C., & Garavan, H. *Prefrontal and midline interactions mediating behavioural control*.

Fassbender, C., **Murphy, K.**, Hester, R., Meaney, J., Robertson, I. H., & Garavan, H. *The role of right fronto-parietal cortex in cognitive control: Common activations for "cues-to-action" and response inhibition*.

Kübler, A., **Murphy, K.**, & Garavan, H. *Effects of cocaine use on attentional control within and between verbal and visuospatial working memory*.

Garavan, H., Fassbender, C., **Murphy, K.**, Sweet, L. H., Rao, S., & Stein, E. A. *Quantitative Age-Related Differences In Functional Recruitment In A Working Memory Task*.

The abstracts for all articles in this section are reproduced in Appendix A.

Table of Contents

Declaration	i
Acknowledgements	ii
Summary	iii
Publications	v
Table of Contents	vii
Chapter 1	
Introduction	1
1.1 A brief explanation of MRI physics	3
1.2 Physiology of blood flow and fMRI	8
1.3 Task Designs in fMRI	10
1.4 Analysis of fMRI data	12
1.5 Optimising the experimental design	17
1.6 Topics addressed in this thesis	19
Chapter 2	
A validation of event-related fMRI comparisons between drug groups and controls	23
2.1 Introduction	24
2.1.1 Cocaine	24
2.1.2 Nicotine	25
2.1.3 THC	26
2.2 Methods	28
2.2.1 Participants and Task Design	28
2.2.2 Scanning Parameters	30
2.2.3 Image Analysis	31
2.3 Results	33
2.4 Discussion	36
2.5 Conclusions	39

Chapter 3

Deriving the optimal number of events for an event-related fMRI study based on the spatial extent of activation	40
3.1 Introduction	41
3.2 Materials and Methods	42
3.2.1 Participant and Task Design	42
3.2.2 Scanning Parameters	43
3.2.3 Image Processing	43
3.2.4 Analysis Techniques	44
3.2.5 Permutation Analyses	46
3.2.6 Group Analysis	47
3.3 Results and Discussion	48
3.3.1 Non-Overlap Analysis	48
3.3.2 Overlap Analysis	52
3.3.3 Group Comparison	55
3.4 Conclusions	56

Chapter 4

An empirical investigation into the number of subjects required for an event-related fMRI study	57
4.1 Introduction	58
4.2 Materials and Methods	60
4.2.1 Subjects and Task Design	60
4.2.2 Scanning Parameters	61
4.2.3 Image Analysis	61
4.2.4 Comparison of Activation Maps	63
4.3 Results and Discussion	65
4.3.1 “Gold Standard” Analyses	65
4.3.2 Independent Groups Analyses	69
4.4 Conclusions	74

Chapter 5	
Artifactual fMRI group and condition differences driven by performance confounds	76
5.1 Introduction	77
5.2 Methods	79
5.2.1 Between-group comparisons	79
5.2.2 Within-group comparisons	79
5.2.3 Simulations	80
5.2.4 Image Acquisition	80
5.2.5 Image Analysis	80
5.2.6 Old/Young Comparison	81
5.2.7 High/Low Conflict Comparison	84
5.2.8 Simulations	84
5.3 Results	85
5.3.1 Old/Young	85
5.3.2 High/Low Conflict	86
5.3.3 Simulations	87
5.4 Discussion	90
5.4.1 Error-Related Activation Confounding	90
5.4.2 Covariate Analysis	93
5.4.3 Predefined ROIs	94
5.4.3 Simulations	94
5.5 Conclusion	95
Chapter 6	
Discussion	96
References	101
Appendices	117
Appendix A: Abstracts of Publications	118

Chapter 1

Introduction

Functional magnetic resonance imaging (fMRI) has become a widely employed and useful technique in the field of brain imaging. Its superior spatial resolution allows a researcher to localise psychological functions to specific areas of the brain with a precision not possible with other modalities such as electroencephalography (EEG), magnetoencephalography (MEG) and positron emission tomography (PET). This technique has been used with great effect to investigate many primary sensory and higher cognitive centres in the brain, such as vision (Calhoun et al., 2001; Engel et al., 1994; Goodyear and Menon, 1998), language processing (Carpenter et al., 1999; Friederici et al., 2000; Gaillard et al., 2001), working memory (D'Esposito et al., 1999; Pollmann and von Cramon, 2000; Postle and D'Esposito, 1999) and inhibition (de Zubicaray et al., 2000; Garavan et al., 1999; Rubia et al., 2001).

Not only can fMRI localise functions, it can also give insights into how functionally defined areas contribute to the processing of a task. By varying a cognitive task parametrically, within-subject contrasts are possible. This can determine whether a brain region is required for all task levels or whether it is recruited at the higher levels to help facilitate processing. Comparisons between working and long-term memory (Braver et al., 2001), differing levels of n-back (Jansma et al., 2000) and varying difficulty in inhibition GO/NOGO tasks (Durstun et al., 2002) are examples of

parametrically varied tasks. With advances in analytic methods, it is also possible to determine how centres of psychological function are interconnected using techniques such as functional connectivity (Biswal et al., 1995; Lahaye et al., 2003; Worsley et al., 1998). Interruptions in these connections between functional regions of the brain may be the underlying cause of many clinical conditions.

Driven by this progress in fMRI techniques, there has recently been a surge of interest in determining the functional differences that might characterise certain populations of interest. Comparisons between neurologically impaired groups with clinical syndromes, such as schizophrenia (Braus et al., 2000; Menon et al., 2001; Mueller et al., 2002; Surguladze et al., 2001), ADHD (Rubia et al., 1999; Smith et al., 2002), addiction (Pfefferbaum et al., 2001; Wexler et al., 2001) and controls, are abundant in the literature. Comparisons of demographic differences, such as sex differences (Cowan et al., 2000) and age differences (Grossman et al., 2002) are also frequently investigated.

However, fMRI results are inherently noisy due to the natural variations in the biological systems under scrutiny and the complexities of the techniques, with their many adjustable parameters, associated with the collection and analysis of fMRI data. To optimally perform group comparisons with this technique, a researcher must first be sure that no confounds exist in the biological properties of the groups that would render the results suspect. They must then optimise the experimental design to guarantee an acceptable degree of power in each group. Finally, they must utilise the most advantageous analytic techniques to ensure that the group differences uncovered

are due to neuronal differences between the groups and are not due to particular aspects of the analytic methods used.

The focus of this thesis is on these user-driven methodological questions in fMRI, specifically event-related fMRI. Recommendations are provided throughout that will greatly enhance the accuracy of fMRI experiments. To give a better understanding of the issues addressed, the following sections provide the reader with a basic knowledge of the physics that underpins the fMRI technique, the physiology that measurements are based on, the advantages of different types of experimental design, common analytic methods used and efforts to date that have tried to optimise task design. Finally, a brief introduction to the specific topics of this thesis is provided.

1.1 A brief explanation of MRI physics

Due to the quantum mechanical property called *spin*, protons by themselves are slightly magnetic. When an external magnetic field is applied, protons tend to align with this field. As a result, water (H_2O), which is not normally magnetic (since the

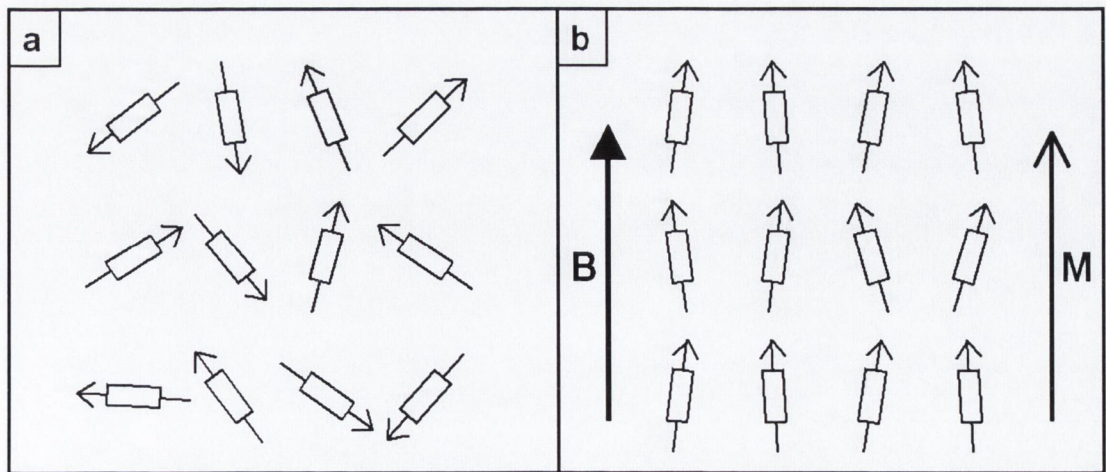


Figure 1.1: Protons are slightly magnetic. A) When no external magnetic is present, the protons point in random directions resulting in a net magnetisation of zero. B) When a magnetic field **B** is present the protons tend to align with this field producing a net magnetisation **M** in the same direction.

protons are pointing randomly in all directions), becomes magnetised when placed in a magnetic field (see Figure 1.1).

Thus, when a human body (which contains water) is placed into a large magnetic field, the net magnetisation $\mathbf{M}$ of the water lines up with this external field $\mathbf{B}$. This alignment, however, does not follow a simple path. When $\mathbf{M}$ is not parallel to $\mathbf{B}$, the forces on $\mathbf{M}$ cause it to precess (i.e. rotate) around the $\mathbf{B}$ -direction at a frequency proportional to the magnetic field strength B . This frequency is called the *Larmor frequency* and is given by the equation: $f = \gamma B$ (where γ is the *gyromagnetic ratio*). In this way, when $\mathbf{M}$ aligns itself with $\mathbf{B}$, a process called *relaxation*, it does so in a spiral fashion (see Figure 1.2).

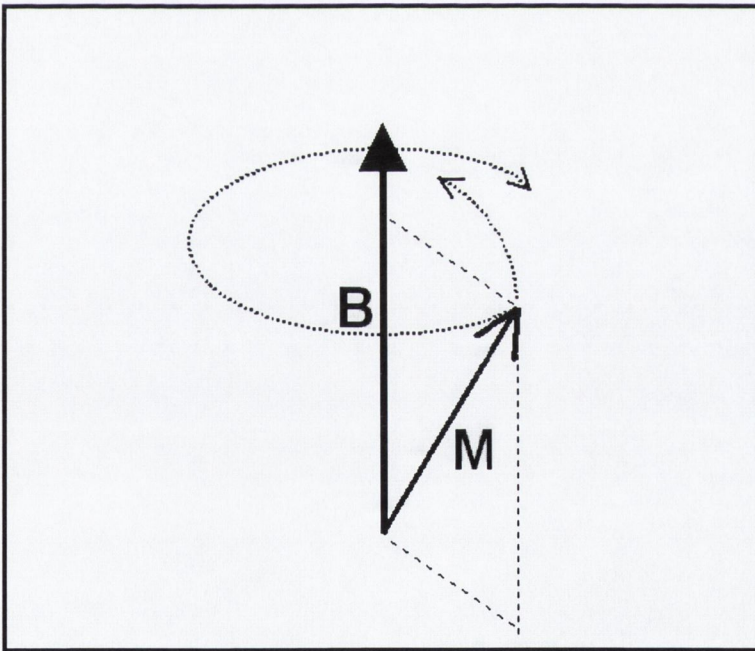


Figure 1.2: When the magnetisation $\mathbf{M}$ is not aligned with the external field $\mathbf{B}$, the forces on $\mathbf{M}$ cause it to precess around the direction of $\mathbf{B}$. Thus, the magnetisation vector traces a spiral during relaxation.

After the magnetisation of the water protons is allowed to relax (i.e., line up with the external field), microwave radiofrequency (RF) radiation can disturb the alignment of the magnetisation vector. If the frequency of the RF pulse is equal to the Larmor

frequency, it will cause $\mathbf{M}$ to rotate away from the direction of $\mathbf{B}$. This effect is called *resonance*. After the RF pulse is switched off, $\mathbf{M}$ will relax back to the direction of $\mathbf{B}$ (as described above) and will emit RF at the Larmor frequency as it does so. This radiofrequency radiation is the basis for the MR signal. The fundamental procedures of *magnetic resonance imaging* (MRI) are to place a subject in a magnetic field $\mathbf{B}$, transmit an RF pulse to excite $\mathbf{M}$ away from the direction of $\mathbf{B}$ and turn on a receiver to detect the small RF signal that is re-transmitted by the subject's protons.

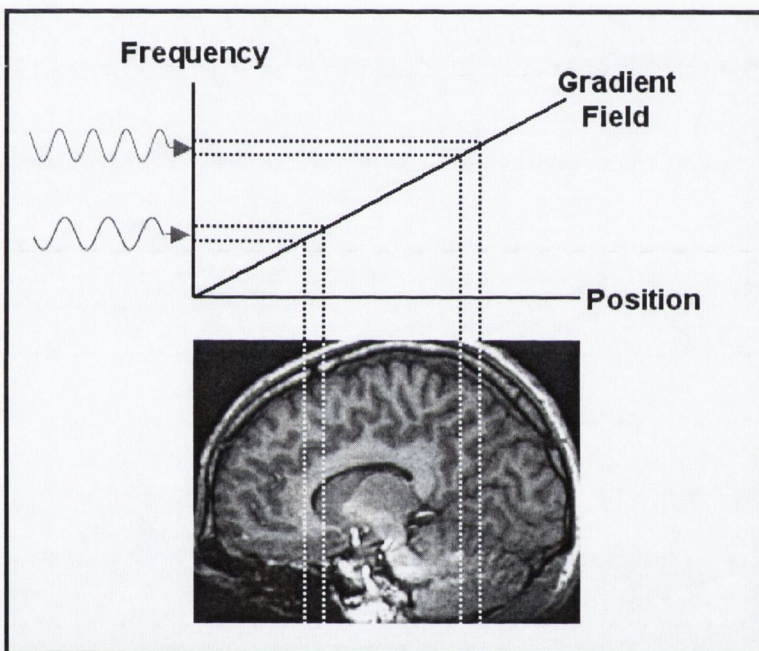


Figure 1.3: When the gradient field is applied, an RF pulse of a certain frequency will excite only those protons that have the same resonant frequency. In this example, a high frequency pulse will only excite a slice at the back of the head, whereas a low frequency pulse will only excite a slice at the front.

The RF signal detected is a summation of the RF radiation from all the excited protons. If different locations in the subject's brain are to be distinguished, these locations must be selectively excited. This is achieved by using linear magnetic field gradients. If the magnetic field is uniform, the resonant frequency of all the protons is the same because this frequency depends on the strength of the magnetic field at that point. However, if a magnetic gradient is applied when the RF pulse is transmitted,

only those protons that have the same resonant frequency as the RF radiation will be excited (see Figure 1.3). Thus, after excitation the re-emitted RF signal emanates from only one slice through the subject's brain. This is called *slice selective excitation*. It is possible to excite only a single voxel (cube) in the brain by applying similar magnetic gradients in the three different spatial directions. However, due to the length of time it would take to measure signal from all voxels this method is not suitable for constructing a rapid picture of a brain.

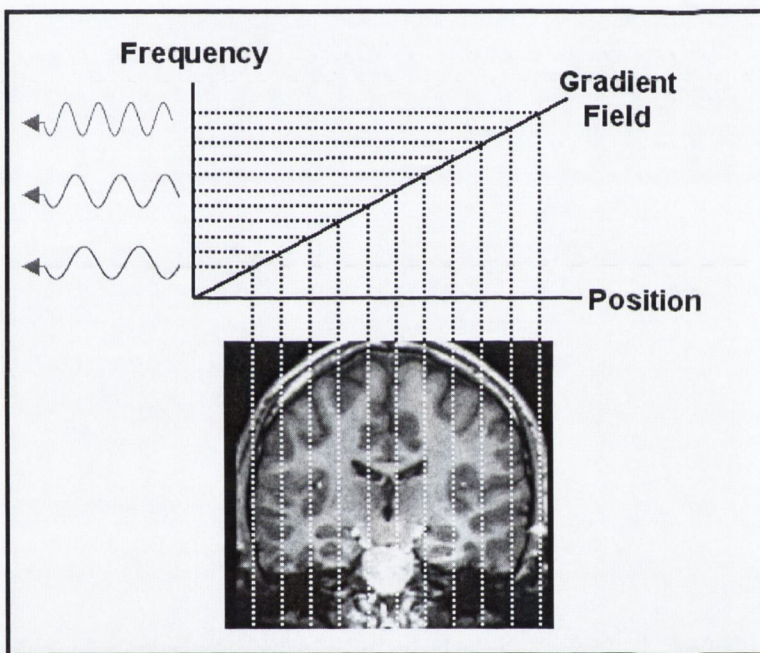


Figure 1.4: When a gradient is applied after excitation in the frequency encoding direction (in this example from left-to-right), the protons on the left side of the brain will precess at a lower frequency than those on the right and so emit RF at a lower frequency.

A more efficient way to form an image is to use *frequency encoding*. After using a gradient to choose a slice, the measured RF signal is transmitted from the entire slice. If we apply a gradient along the slice after excitation and during the readout, for example from left-to-right, the left side of the slice will precess at a lower frequency than the right side because the strength of the magnetic field is lower on the left (see Figure 1.4). It is then possible to distinguish from which position the signals were

transmitted due to the difference in their frequencies. In this way, we can determine the signal from each column of voxels in the slice.

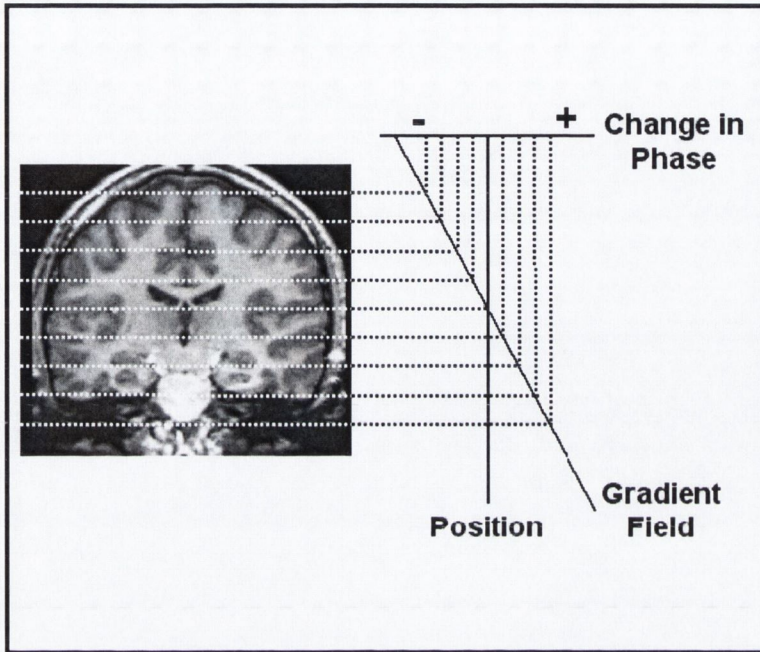


Figure 1.5: In the phase encoding direction, a gradient is switched on for a short period. This changes the phase of precession of the protons. Those protons at the top of the brain receive a negative phase shift while those at the bottom receive a positive one.

The localisation of the signal in the third direction (i.e., perpendicular to the slice selection direction and the frequency encoding direction) is provided by another gradient. The purpose of this gradient is to make the phase of the RF signal depend on the location along this direction. This is called *phase encoding* (see Figure 1.5). When the phase encoding gradient is switched on, say from top-to-bottom, the protons will precess slower at the top of the column of voxels than those at the bottom. When this gradient is switched off, the magnetic field strength returns to its original value and the precessional frequency of the protons in the column become equal again. Even though all the protons in this column are now rotating at the same frequency, there will be a phase difference from top to bottom. By repeatedly applying the phase encoding gradient (keeping the slice selection and the frequency encoding gradients

constant) whilst reading out the signal, it is possible to determine position along this direction by measuring the fluctuation in phase across all the readouts.

In this way, the signal from every voxel in the brain can be calculated. The amplitude of the RF signal from each voxel is directly related to the number of protons in that voxel and hence, the density of tissue. Different manipulations of the applied RF and the magnetic field allow emphasis of different tissue properties. In this way, images that highlight various tissue types can be acquired.

1.2 Physiology of blood flow and fMRI

The purpose of *functional magnetic resonance imaging*, fMRI, is to measure signal changes caused by neuronal activity. At present, it is not possible to detect the electrical activity (or the associated magnetic activity) due to neurons firing using this method. However, it is possible to measure the haemodynamic effects, such as changes in blood oxygenation, which occur within a few seconds of neuronal activity. This indirect measurement of neuronal activity is enabled by the differences in magnetic properties of oxygenated and deoxygenated haemoglobin. Oxyhaemoglobin is diamagnetic whereas deoxyhaemoglobin is paramagnetic. A paramagnetic material introduces inhomogeneities into the magnetic field and thus detracts from the MR signal. When neurons fire, oxygen is metabolised, so there is a slight increase in deoxyhaemoglobin concentration. The arteries open to supply blood to the active area resulting in an oversupply of oxyhaemoglobin. Therefore, there is a decrease in deoxyhaemoglobin concentration and thus an increase in the MR signal. This is called the *BOLD effect* (blood oxygen level dependence) and is the basis for the vast majority of fMRI experiments.

This BOLD response is very slow compared to the related neuronal activity to stimulation (see Figure 1.6). It takes approximately 1 to 2 seconds for a change in signal to occur after stimulation. It then takes 4 to 6 seconds for the signal to reach its maximum and another 4 to 6 seconds for it to return to baseline, that is, the level it was at before excitation. This return to baseline is further complicated by a phenomenon known as the *post-undershoot*. Often the BOLD signal does not stop at the pre-stimulus value but travels below the baseline and takes some extra time to rebound back to its original value. This does not necessarily occur in all voxels.

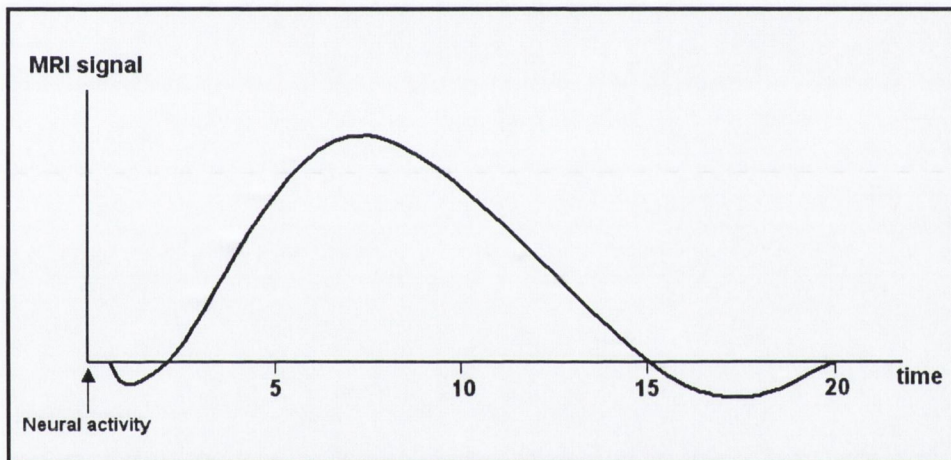


Figure 1.6: The shape of the haemodynamic response is displayed. It takes 1 to 2 seconds for the signal to change after neural activity. It reaches its peak between 6 and 10 seconds and returns to baseline between 14 and 18 seconds. The post-undershoot is displayed.

With current MRI technology, it is possible to measure enough slices to construct a whole brain image in less than 2 seconds. By using techniques that concentrate on detecting the BOLD signal, a time course of this signal can be constructed for every voxel in the brain on a 1 to 2 sec time resolution. Thus, signal detection techniques can determine whether specific voxels are close to stimulus driven neuronal activity.

1.3 Task Designs in fMRI

The first successful detections of human brain activation using the BOLD contrast mechanism described above were performed in the early '90s (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992). Initially, most investigators borrowed techniques from the more established field of positron emission tomography (PET) to design experiments. These studies were performed using block designs, that is, extended periods of “on” and “off” activations. Blocks of stimuli are repeatedly presented to the subject interspersed with blocks of a rest condition. In this way, it is possible to compare those time points acquired during the “on” condition with those of the “off” condition in each voxel. If the “on” condition is significantly greater than the “off” condition, then one can assume that the voxel is supplying blood to neurons that are involved in processing of the task. Although this approach works reasonably well, the types of possible experiments are limited due to the need for the stimuli to be presented in a block. This is clearly not required since activity in fMRI experiments is detectable within seconds of the stimulus onset.

It was first shown in 1992 that a haemodynamic response is measurable using BOLD contrast fMRI when brief visual stimuli lasting 2 seconds were presented to a subject (Blamire et al., 1992). Bandettini and colleagues demonstrated signal changes to events as short as $\frac{1}{2}$ second (cited in Rosen et al., 1998). This opened the prospect of more interesting paradigms utilising fMRI that were not possible with block designs. Experiments that required stimuli to occur randomly or infrequently were now achievable. However, there was one limitation: the length of the haemodynamic response. The haemodynamic response to the events of interest could not overlap if one wanted to borrow established techniques such as those in the field of EEG.

Investigations into the properties of overlapping haemodynamic responses showed that not only could activations to brief stimuli be detected but those haemodynamic responses summated in a roughly linear fashion (Boynton et al., 1996; Dale and Buckner, 1997). This linearity breaks down when the events occur extremely close together (i.e., <1sec apart) (Friston et al., 1998b). However, this still leaves the option of separating and analysing events that occur in close succession (e.g., 1 event per 2 sec). Indeed, Dale and Buckner showed that when randomly presenting lateralised visual stimuli every 2 secs within an intermixed trial paradigm, it was possible to extract the lateralisation pattern (Dale and Buckner, 1997). The BOLD responses for primary sensory systems were generally considered to be greater and more robust than those of higher cognitive paradigms. The question of whether it was possible to detect individual cognitive trial events was unresolved until Buckner and colleagues demonstrated the validity of these methods in a cognitive task (Buckner et al., 1996). This led to the proposal of trial-based fMRI experimental designs (Zarahn et al., 1997). These are now better known as event-related fMRI tasks.

The advantages of event-related fMRI designs over block designs are numerous. The ability to randomise the presentation of the trials allows more possibilities in designing cognitive paradigms such as GO/NOGO tasks (Fassbender et al., 2004; Garavan et al., 1999; Kaufman et al., 2003), oddball tasks (Linden et al., 1999; Stevens et al., 2000; Yoshiura et al., 1999) and priming tasks (Buckner et al., 1998a; Buckner et al., 2000; James et al., 1999) that would not be possible using block design techniques. Event-related designs also enable the performance of post-hoc classification of the trials based on behavioural measures, for example, correct and incorrect responses. This is an important step when analysing data because the

inclusion of incorrect trials in the signal has been shown to have detrimental effects on the resulting activation maps (see Chapter 5 and Murphy and Garavan, 2004a). The shape characteristics of a haemodynamic response, such as amplitude, latency and width, can be examined in event-related designs to allow for more quantitative measures of brain function such as mental chronometry (Bellgowan et al., 2003). A more subtle advantage of event-related designs is that there is no need for cognitive subtraction. In a block design, it is only possible to compare the task blocks with a resting condition or some other task blocks. In this way, any activation measures are the difference between the two block types and not the activation measure of the task block alone. In an event-related design, the haemodynamic response over the tonic baseline of the task is calculated, so there is no comparison drawn with another mental state.

1.4 Analysis of fMRI data

The earliest method for detecting signal changes in fMRI block design time series data made use of the repeated measures Student t-test. All time points in the task block were compared to those in the rest block using a t-test and thus statistically significant differences identified. Those voxels with a t-statistic above a certain threshold were considered to be active. The main problem with this technique is that it does not account for those time points that are in the transition between active and inactive states.

A more general method for analysing block designs was developed that utilised correlation coefficients (Bandettini et al., 1992). A reference time series (or ideal time series) that has the shape of the expected fMRI response is determined. The

correlation of the time series in each voxel to the ideal is calculated. The correlation coefficient can then be converted into a t-statistic and those voxels passing a certain threshold are considered to be active. In the case where the reference time series consists of zeros for the rest condition and ones for the task condition, this method would yield exactly the same results as the previous t-test method. However, with this technique it is possible to account for the rise and fall times in the haemodynamic response when determining the reference waveform. The simplest approach is to apply several reference functions to the data with a range of different delays. The value of delay that produces the highest correlation coefficient is chosen for each voxel.

A more advanced choice of reference time series can be produced by convolving a stimulus timing function (which is equal to 0 when no stimulus is applied and 1 when a stimulus is presented) with a haemodynamic response function (HRF) that accurately represents the shape of the BOLD response after a single event. In a block design this amounts to convolving the reference waveform above (without delays) with a HRF. The gamma variate function, $\gamma(t) = t^r e^{-t/b}$, has been shown to effectively model the haemodynamic response to brief stimuli (Cohen, 1997), with parameters $r = 8.6$ and $b = 0.51$, and is a popular choice for modelling the haemodynamic shape. Thus, convolving the stimulus timing function with the gamma variate function to give a reference waveform will yield more reliable results when using the correlation method. However, it is important that the HRF model chosen accurately reflects the true shape of the response. If, for some reason, the haemodynamic shape of a subject's response is compromised (e.g. they take a vascular-affecting drug), then the results of the analysis could be confounded by this difference in shape.

The correlation method is equivalent to the linear least squares regression of the equation: $y(t) = \beta \cdot x(t) + \alpha + \varepsilon(t)$, where $y(t)$ is the voxel time series data, $x(t)$ is the reference function with β its scaling factor, α is a constant and $\varepsilon(t)$ is a random Gaussian error term. Both β and α are unknown parameters that are fit by the linear regression method. The advantage of using this formulation over the correlation method is that it is easy to extend the equation to include extra regressors to remove unwanted trends in the data, such as baseline drift, whilst simultaneously computing the correlation. The scaling factor, β , can then be used as the activation measure in each voxel.

Multiple reference waveforms can now be easily included in this type of analysis, denoted by the term *multiple linear regression*. In an experiment with two active conditions (1 and 2) the equation $y(t) = \beta_1 \cdot x_1(t) + \beta_2 \cdot x_2(t) + \alpha + \delta \cdot t + \varepsilon(t)$ is fitted to the data. For this model, four parameters are estimated, the two scaling factors β_1 and β_2 and the baseline α and also a baseline drift rate δ which accommodates for linear changes in the baseline over time. Now it is possible to investigate whether β_1 or β_2 are nonzero and whether β_1 is different from β_2 ? The statistical significance of answers to these questions can be calculated using F-tests. Not only does this method allow one to determine active areas in the brain but also one can determine if an area is more active in one condition than another (useful in parametrically varied tasks).

This equation can be further generalised to $\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{Y}$ is a column vector of the voxels time series data, $\mathbf{X}$ is the design matrix, $\boldsymbol{\beta}$ is a column vector of scaling factors and $\boldsymbol{\varepsilon}$ is a column vector of Gaussian error terms (Friston et al., 1995; Worsley and Friston, 1995). This equation is called the *General Linear Model* (GLM) and is

the basis for most fMRI analytic techniques. Many software packages use this approach, for examples see SPM (Statistical Parametric Mapping - <http://www.fil.ion.ucl.ac.uk/spm/>) or the 3dDeconvolve program in AFNI (Analysis of Functional NeuroImages - <http://afni.nimh.nih.gov/afni/>). The columns of the design matrix $\mathbf{X}$ model the effects of interest and also confounding variables. They are in essence the reference waveforms mentioned above.

It should be obvious that the columns of the design matrix need not be reference waveforms for block designs only. It is possible to make reference waveforms for event-related designs, in a similar way to above, by convolving the stimulus timing function for the events with an appropriate HRF. Indeed, the columns of the design matrix need not be of exclusively one type of design: by mixing both block and event-related reference waveforms, tonic levels of activation and event-related activations above and beyond can be calculated simultaneously (Hester et al., 2004).

The reference waveforms model the expected haemodynamic response to a task. The majority of event-related analyses base their activation measure on the scaling factors from the multiple linear regression. Throughout the studies in this thesis, a technique is employed, first used by Garavan and colleagues (Garavan et al., 1999), in which the scaling factors are not the basis of the activation measure. This method has the advantage that it can accommodate the differences in haemodynamic shape, both within- and between-subject, that have been reported in the literature (Aguirre et al., 1998; Handwerker et al., 2004; Neumann et al., 2003).

Using deconvolution, this technique performs an estimation of the impulse response function (IRF) for each voxel. Employing multiple regression techniques, this procedure calculates the IRF (i.e., the average best-fitting “shape”) that follows each event. To achieve this, a regressor that specifies the temporal position of each event is defined. This regressor is then time-shifted by 1 TR to create a new one. This is repeated up to 8 TRs (i.e. 16 seconds). The deconvolution procedure then calculates a scaling value for each regressor using a least-squares multiple regression algorithm. These scalar values define the IRF for each voxel. The deconvolution procedure also calculates intercept and slope parameters, which can be used to calculate a baseline (i.e., the ongoing mean activity level after excluding the variance due to the events) for the IRF. The best-fitting haemodynamic shape (as defined by the gamma-variate function above) is determined for each voxel’s IRF by using a non-linear regression algorithm (Ward et al., 1998). This estimated haemodynamic shape for each voxel is converted into a percentage area under the curve score (%AUC) by expressing the area under the haemodynamic curve as a percentage of the area under the baseline. This %AUC represents the activation measure for the voxel (see Figure 1.7). Any number of regressors representing different event types can be inserted into the deconvolution procedure in this way and activation measures can be calculated for each. The gamma variate function has been shown to effectively model the haemodynamic response (Cohen, 1997), however, another model could be used to find the best-fitting shape, from which a %AUC measure could be calculated.

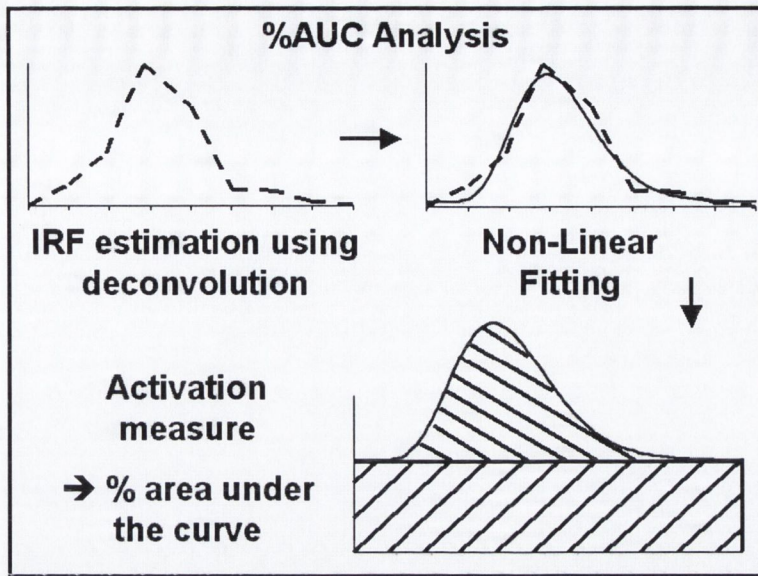


Figure 1.7: The main steps for the %AUC analysis are displayed. First, an IRF is estimated using deconvolution. Second, the best-fitting gamma variate shape is derived using a non-linear fitting technique. Finally, this is converted into the %AUC activation measure.

1.5 Optimising the experimental design

There are many variables in an fMRI design on which a researcher needs to decide. The timing of the task, the number of events and the number of subjects are some of the decisions a researcher must make. Efforts have been made in the literature to optimise these variables for both block and event-related fMRI to give the most reliable and robust results. Much of the emphasis has been placed on the issue of experimental timing and very little is known about the optimal number of events and the required number of subjects (especially in event-related tasks).

The issue of experimental timing is very important in fMRI tasks due to the relatively poor temporal resolution of the technique. As mentioned above, most techniques in event-related fMRI analysis assume that the haemodynamic shape of the BOLD signal is linearly additive. However, it has been shown that the haemodynamic shape displays a 17-25% reduction in amplitude when trial onsets are spaced (on average)

5secs apart compared to those spaced 20secs apart (Miezin et al., 2000). Therefore, it is very important to determine the optimal interstimulus interval (ISI). Dale demonstrated that if the ISI is jittered, the statistical efficiency improves monotonically with decreasing mean ISI and that the efficiency can be up to 10 times greater than that of a fixed ISI design (Dale, 1999). However, there are two fundamentally different goals when analysing event-related fMRI tasks: estimation of the impulse response function (IRF) and detection of signal change. Birn and colleagues showed that the estimation of the IRF is optimised when stimuli are frequently alternated between task and control states, whereas detection of activated areas is optimised by block designs (Birn et al., 2002). To this end, Liu and colleagues have developed a method that can simultaneously achieve the estimation efficiency of randomised designs and the detection power of block designs at a cost of increasing the length of the experiment by less than a factor of two (Liu et al., 2001). Genetic algorithms that can produce designs that outperform random designs on estimation efficiency, detection efficiency and design counterbalancing have been developed (Wager and Nichols, 2003).

Little research has gone into the question of the optimal number of events. Saad and colleagues demonstrated in a block design experiment that the spatial extent of activation increases monotonically as the number of blocks included in the analysis increase (Saad et al., 2003). Using a similar measure, Huettel and McCarthy found that spatial extent of activation failed to asymptote even after 150 events in an event-related design (Huettel and McCarthy, 2001). This suggests that in both block and event-related designs, the required number of events to provide sufficient power is unattainable due to external constraints such as scanning duration and costs.

Similarly, little is known about the number of subjects required to yield stable activation maps. The first paper dealing with this issue showed that conjunction analyses with a fixed-effect model is sufficient to make inferences about population characteristics thus reducing the number of subjects required to infer differences between populations (Friston et al., 1999a). However, it did not give a clear indication of the number of subjects required. By estimating the mean differences and variability between two block conditions, Desmond and Glover were able to recommend that at least 24 subjects be scanned to achieve reasonable power (Desmond and Glover, 2002). Again, the recommended number of subjects is very high with few fMRI studies every reaching that goal. Whether event-related studies will require a smaller sample is unknown.

1.6 Topics addressed in this thesis

Addressing some of the methodological concerns in event-related fMRI task design and analysis is the focus of this thesis. Four practical questions have been chosen that are of particular importance but have been poorly investigated in the literature. These questions span the entire field of fMRI, from physiology through experimental design to analytic techniques. It is hoped that the answers obtained in these studies will equip the average fMRI researcher with useful guidelines to the design and analysis of fMRI experiments; in particular, those that investigate group comparisons.

A study is presented in Chapter 2 that addresses the concern of group difference confounds due to deviations in the haemodynamic properties of each group. Given that fMRI is such an indirect measure of neuronal activity, to investigate with confidence differences between groups of interest one must be sure that no

confounding factors in haemodynamic properties exist, that are independent of neuronal differences. Since many drugs have direct vascular effects (e.g. cocaine: Gollub et al., 1998; Kaufman et al., 1998b; Lee et al., 2003; Luo et al., 2003), one would expect that comparisons between drug groups and controls would be extremely compromised if this were the case (maybe more so than any other group comparison). To this end, the first study in this thesis compares the haemodynamic properties of three drug groups (cocaine, nicotine and THC) with those of controls, to assess the likelihood of a confound when performing group comparisons with event-related fMRI. The results show that although these drugs have direct effects on vasculature, the BOLD response as measured by fMRI is not compromised. Thus, comparisons between these drug groups and controls are not affected by the administration of the drugs. This result may extend to comparisons of other non-drug groups, since it is likely that the integrity of the haemodynamic shapes is greater than those of a vascular-affecting drug using population.

The number of events required to yield a stable activation measure in an event-related design appears to be very extreme if Huettel and McCarthy are to be believed (see above). Chapter 3 presents a study that tries to determine if this result is correct. Using the correlation method presented by Huettel and McCarthy and two other analytic techniques (GLM and %AUC), data from a finger-tapping task containing 150 non-overlapping events (i.e. events were sufficiently spread so that haemodynamic shapes did not overlap) were analysed. The Huettel and McCarthy result is replicated very closely. However, the other two analytic methods display asymptotic behaviour at radically fewer events. Since most event-related designs have events that overlap, the analyses were repeated on a finger-tapping task with 150

overlapping events. Again the GLM and %AUC analysis techniques display asymptotic behaviour at drastically fewer events. The recommendation for an fMRI investigator would be to choose their analytic methods carefully since only 25 events are required to yield stable activation maps if a GLM or %AUC analysis is employed.

To yield a reliable and stable group activation map using event-related fMRI, the optimal number of subjects must be identified. Chapter 4 presents an empirical study that attempts to derive this number using a cohort of 58 subjects. The behaviour of the activation maps was investigated as a function of subject numbers using a variety of measures. The results demonstrate that in the range of most conventional studies ($n=10$ to 20), the power of activation maps is very poor when using voxelwise overlap as the measure. Although this is the case, the majority of voxels in these activation maps are true positives. Reproducibility of activation maps was investigated by splitting the sample into two groups. Voxelwise overlap seems to be a very poor measure of reproducibility so an alternative measure was created that quantifies the distances between the centres-of-mass of activated areas into one similarity measure. The results reveal that although the power is poor, 80% of the activated areas fall within 25mm of each other when $n=20$. The main recommendation is that at least 20 subjects are required to yield stable activation maps in an event-related study.

The final study in Chapter 5 investigates what effect performance differences between groups have on the results of an fMRI group study. FMRI is often employed to determine distinctions between groups (e.g., schizophrenia and controls: Braus et al., 2000; Menon et al., 2001; Mueller et al., 2002) that often, by design, differ in performance measure. If a researcher does not use the advantage of event-related

analysis techniques to exclude incorrect trials, artifactual group differences in activation measure may result, solely due to this discrepancy in performance measure. Two studies were reanalysed: one that compared an old and young group and another that compared a hard and easy condition, both with significant differences in performance. The data were analysed twice using event-related techniques, one treating correct and error responses separately, the other treating error responses as corrects. The activation maps differ considerably between the two types of analyses, with the inclusion of errors leading to many false positive and false negative differences between the groups/conditions. Further data simulations that varied the number of errors included in the analyses find that surprisingly few errors can significantly alter activation maps. This study renders suspect those papers that use block design techniques when comparing groups that differ in performance. An analysis technique that allows one to selectively remove errors, such as an event-related method, should be employed.

Armed with the results of these studies, an fMRI researcher will be able to design group comparison experiments with more confidence. By advocating the use of event-related fMRI designs, removing the fear of confounding factors and making recommendations about fundamental experimental variables, this thesis makes a significant contribution to the knowledge a researcher requires to produce reliable and robust results in the field of fMRI.

Chapter 2

A validation of event-related fMRI comparisons between drug groups and controls

Abstract

Non-invasive brain imaging techniques are powerful tools for researching the effects of drug abuse on brain activation measures. However, as many drugs have direct vascular effects, the validity of techniques dependent on blood flow measures, which purport to reflect neuronal activity, could be called into question. This may be of particular concern in event-related fMRI where current analytic techniques search for a specific shape in the haemodynamic response to neuronal activity. To investigate possible alteration in task-related activation, four groups performed a simple event-related fingertapping task: controls, cocaine, nicotine and THC (marijuana) users. It was found that activation measures, determined by two different analytic methods, did not differ between the groups. A comparison between an IV-saline and an IV-cocaine condition in cocaine addicts found a similar null result. Further in-depth analyses into the shape of the haemodynamic responses in each group also showed no differences. This demonstrates that it is possible to compare drug groups with controls using event-related fMRI without the need for any post-hoc correction procedures to correct for possible drug-induced cardiovascular alterations. Thus, fMRI activation differences reported between these drug groups can be more confidently interpreted as reflecting neuronal differences.

2.1 Introduction

The investigation of clinical populations to find inherent differences in brain activity is one of the functions of non-invasive brain imaging that has great potential. Research on the effects of drugs of abuse is one such application whereby cognitive, affective as well as direct drug differences are placed under investigation. However, given that many drugs may have direct vascular effects, and that fMRI is an indirect measure of neuronal activity based on blood flow, it is important that researchers should determine the extent to which these non-neuronal effects might affect fMRI activation measures. If drugs of abuse do affect cardiovascular properties, this renders suspect the interpretation of fMRI differences as indications of between-group neuronal differences.

2.1.1 Cocaine

Cocaine increases mean arterial blood pressure (MABP), reduces cerebral blood volume (CBV) and causes cerebral vasoconstriction (Gollub et al., 1998; Kaufman et al., 1998a; Kaufman et al., 1998b; Luo et al., 2003). How complex global and regional vascular changes interact with the fMRI BOLD signal is still poorly understood (Buxton et al., 1998) and in the literature, there are conflicting conclusions on what effect these might have. It has been shown that the BOLD signal is affected even when no cocaine has been administered in that chronic cocaine abusers displayed a significantly enhanced positive BOLD response to photic stimulation when compared to control subjects (Lee et al., 2003). The length of prior cocaine usage has been demonstrated to predict the relationship between cocaine administration and vasoconstriction (Kaufman et al., 1998b), suggesting that cocaine

may have a cumulative residual effect that could be present both with the drug in the bloodstream and without. These studies imply that task-induced BOLD fMRI signal may be confounded by chronic vasoconstrictive effects in a cocaine using population. However, Luo and colleagues, (Luo et al., 2003), showed that the potential confounding factor of the MABP changes had little effect on the interpretation of cocaine-induced BOLD changes and thus concluded that BOLD fMRI could be extended to map drug-induced neuronal activity. Arriving at a similar conclusion, Gollub and colleagues were able to observe an expected 2-3% increase in visual stimulus-induced BOLD signal on top of a 14% decrease in baseline blood flow following acute cocaine administration (Gollub et al., 1998). Breiter and colleagues have also demonstrated that fMRI can map localised dynamic changes in brain activation following cocaine infusion (Breiter et al., 1997). Combined, these studies lead us to conclude that although there may be chronic or baseline effects of cocaine on the vasculature that underlies the BOLD signal, it is nonetheless possible to observe robust task-induced activation changes in cocaine users using fMRI.

2.1.2 Nicotine

Nicotine is a prominent and more socially acceptable drug of abuse in modern society and as a consequence, a number of studies have examined the cognitive and cardiovascular effects of the drug. It has been shown that nicotine affects cognitive strategies on working memory tasks (Ernst et al., 2001), improves attention (Lawrence et al., 2002) and improves performance on working memory tasks (Kumari et al., 2003). These improvements in performance have been associated with changes in activation in task-relevant areas. There are conflicting opinions on whether administration of nicotine increases (Domino et al., 2000; Grunwald et al., 1991) or

decreases (Ghatan et al., 1998; Zubieta et al., 2001) cerebral blood flow (CBF) and most studies have found a complex pattern of decreases and increases across the brain. Terborg and colleagues stated that although they discovered significant increases in cerebral blood flow velocity, oxyhaemoglobin and total haemoglobin, a substantial increase in regional CBF was not found (Terborg et al., 2002). However, in contrast, nicotine has been found to cause vasodilation (Iida et al., 1998) and other evidence suggests that chronic users display a global reduction in CBF (Mathew and Wilson, 1991). Little is known about how these increases and decreases would affect the BOLD signal in nicotine users. It has been shown, however, that there is no evidence for an effect of nicotine on a task-activated BOLD signal response to photic stimulation (Jacobsen et al., 2002). On an optimistic note, this suggests that whatever direct effect it may have on the vasculature, nicotine may not alter the coupling between BOLD signal and neuronal activity in the cortex.

2.1.3 THC

The effects of marijuana and its active ingredient THC on cardiovascular function have been the subject of several reviews (Dewey, 1986; Hollister, 1986; Mendelson, 1987; Randall et al., 2002). The most commonly observed cardiovascular effect of these drugs in humans is a dose-related tachycardia (Benowitz and Jones, 1981; Heishman et al., 1990; Lemberger et al., 1974; Lex et al., 1984) that follows a time course very similar to that seen for their psychoactive effects. The overall effect of marijuana and THC is to increase myocardial workload and oxygen demand. Consequently, there is some evidence that high doses of cannabinoids can depress the respiratory system in animals (Bloom et al., 1997; Dewey, 1986). At the same time, effects on blood pressure are minimal (Heishman et al., 1990; Hollister, 1986) and

moderate doses of IV THC (Mathew et al., 1997) or smoked marijuana (Mathew et al., 1992) in humans have not been associated with increases in carbon dioxide. Functional CB1 receptors are also found on cerebral blood vessels (Liu et al., 2000; Randall et al., 2002) although the functional significance of this is still not clear. It has been shown using fMRI that there are individual differences in blood flow reactions to THC (Mathew and Wilson, 1991; Mathew et al., 1998). Mathew and colleagues found that a majority of subjects showed an increase in CBF after THC, but those who showed a decrease also had a significant alteration in time sense (Mathew et al., 1998). With evidence for both decreases and increases in CBF, even in the same experiment, the effect of THC administration on task-induced BOLD fMRI signal change is not well known.

To summarise, there is evidence of direct cardiovascular effects with the administration of cocaine, nicotine or THC, but what consequences these have on brain vasculature and, more importantly for present purposes, on neuronally mediated vascular changes is quite unclear. In chronic abusers of such drugs, changes in cardiovascular properties may be altered permanently and thus may exist when the drug of abuse is not present in the system. As a result, it is possible that this alteration in cerebrovascular dynamics could confound the interpretation of results from fMRI studies in substance abusing populations. This would be of particular concern when trying to determine differences in brain function between drug abusers and a control group and may be especially problematic in event-related fMRI designs in which haemodynamic responses may overlap, thereby requiring deconvolution analyses that assume robust and linearly additive haemodynamic responses.

To investigate possible alteration in task-related activation, four groups performed a simple event-related fingertapping task: controls, cocaine, nicotine and THC (marijuana) users. The same task was run on cocaine users after an injection of saline (the IV-saline condition) and an injection of cocaine (the IV-cocaine condition). An event-related task was chosen as a particularly appropriate test for assessing haemodynamic differences. Two types of analyses were performed on the functional data. First, to ensure the robustness of any findings and second, given that different researchers employ different analytic techniques, it is important to show that the results are unbiased by the technique used. The first technique determines an impulse response function (IRF) for each voxel using a deconvolution procedure. A separate haemodynamic shape is then fitted to each voxel's IRF and an activation measure is obtained from this shape. One advantage of this method is the ability to quantitatively describe the IRF, both voxelwise and subjectwise. This method should be particularly valuable if the different groups have differing haemodynamic responses. The second technique uses the standard SPM haemodynamic shape convolved with the event stimulus stream as a regressor in a multiple regression procedure. Z-score activation maps are found for each subject based on the scaling factor of this regressor.

2.2 Methods

2.2.1 Participants and Task Design

Seventy-six participants (30 female, mean age: 29, range: 19-46), all right-handed, completed the experiment. The participants could be categorised into 4 groups: 25 controls (14 female, mean age: 25, range: 19-36), 18 cocaine users (5 female, mean age: 38, range: 22-45), 13 nicotine users (6 female, mean age: 34, range: 22-46) and

20 THC users (5 female, mean age: 23, range: 19-45). The cocaine group had been using cocaine for an average of 12.5 years and spent, on average, \$350 a week on cocaine. The nicotine group had been smoking, on average, for 18 years and smoked a mean of 23 cigarettes a day (average Fagerstrom score: 5.3). The THC group had been using the drug for an average of 6.5 years and did so on average 6 days a week. Some of the cocaine and THC groups also smoked tobacco. Nine of the cocaine group smoked cigarettes (average nicotine use of 9 years and 6 cigarettes a day) while eight of the THC group smoked (average nicotine use 5 years and 7.5 cigarettes a day). However, the nicotine usage in the nicotine group was significantly greater in both years of usage and number of cigarettes a week than either the cocaine or the THC groups. Controls were naïve to all drugs. Each of the cocaine and THC users was abstinent from their drug of use for 24 hours prior to testing. Nicotine users had smoked within hours of commencement of the task, thus avoiding concerns of acute withdrawal in this group. After complete description of the study to the subjects, written informed consent was obtained.

The event-related finger-tapping task consisted of a black and white flashing checkerboard (4Hz frequency) lasting 1 second. This was presented 30 times over 6 minutes with interstimulus intervals of between 8 and 16 seconds. A countdown clock was presented prior to each checkerboard presentation to prepare the participants. While the checkerboard appeared, the participants were required to finger-tap bilaterally for the duration of the stimulus. The participants were asked to finger-tap as quickly as possible in the hope that contamination of results due to performance differences (i.e., lower activation measures due to slower finger-tapping) would be minimised.

A further thirteen cocaine addicts (2 female, mean age: 37, range: 21-45) completed the same experiment, after an injection of saline and after an injection of cocaine. This group had been using cocaine for an average of 12 years and spent, on average, \$220 a week on cocaine. Eleven of this group smoked cigarettes (average nicotine use of 15 years and 9 cigarettes a day). Written informed consent was also obtained from this group. The finger-tapping task was run in conjunction with a GO/NOGO task (Kaufman and Garavan, Under Review). The cocaine users participated in two scan sessions, approximately three hours apart, during the same day: one scan session in which saline was infused twice and the other in which cocaine was infused twice (the order of infusions was counterbalanced across subjects). In each session, subjects completed the finger-tapping twice but we limited the analysis to just the first performance of the task so as to be comparable to the data of the other groups in which the task was performed just once. Subjects were manually injected over 120 seconds through a catheter port with either cocaine at 40mg/70kg body weight or normal saline. The finger-tapping task was performed approximately 4.5 minutes after the infusion.

2.2.2 Scanning Parameters

Scanning was conducted in the Medical College of Wisconsin where ethical approval was obtained. Seven mm contiguous sagittal slices covering the entire brain were acquired from a 1.5T GE Signa scanner using a blipped gradient-echo, echo-planar pulse sequence (TE = 40 msec; TR = 2000 msec; FOV = 24 cm; 64 x 64 matrix; 3.75 mm x 3.75 mm in-plane resolution). High resolution spoiled GRASS anatomic images (TR = 24 msec, TE = 5msec, flip angle = 45°, FOV = 24cm, thickness = 1.0

mm with no gap, matrix size = 256x256x124) were acquired prior to functional imaging.

2.2.3 Image Analysis

All data processing was conducted using the AFNI software package (Cox, 1996) (<http://afni.nimh.nih.gov/afni>). Initially, each voxel's time series was shifted so that the separate slices acquired at varying times were aligned to the same temporal origin. Each 3D image corresponding to a particular time point was then volume registered to a "base" image using a Fourier interpolation algorithm to align the voxels. Voxels lying outside the brain were then removed.

The estimation of the impulse response function (IRF) for each voxel was performed using a deconvolution technique (for a more detailed explanation, see Section 1.4). This technique uses a multiple regression analysis to estimate the signal contributed by the events to the overall time-series. A regressor indicating the positions of the fingertapping events along with motion regressors derived from the volume registration were included. The deconvolution model includes both intercept and slope parameters, which can be used to calculate a baseline for the IRF. The best fitting haemodynamic shape (a gamma-variate function, $y = k t^r e^{-t/b}$) was determined for each voxel's fingertapping IRF using a non-linear regression algorithm (Ward et al., 1998). The estimated haemodynamic shape for each voxel was converted into a percentage area under the curve score (%AUC) by expressing the area under the haemodynamic curve as a percentage of the area under the baseline.

Another analysis was performed which emulated that used by Statistical Parametric Mapping (SPM; <http://www.fil.ion.ucl.ac.uk/spm/spm2.html>). Two regressors, one obtained by convolving the event stimulus train with the packaged SPM haemodynamic response function and the other with its time derivative, were used in the regression procedure (the inclusion of the time derivative in the multiple regression analysis should accommodate for differences in temporal delay between voxels (Friston et al., 1998a)). The scaling factor for the first regressor was converted into a z-score map for each participant.

The %AUC and z-score maps were then resampled at a higher 1 μ l resolution and converted to the standard stereotaxic coordinate system of Talairach and Tournoux (Talairach and Tournoux, 1988). The images were then spatially smoothed using a Gaussian kernel with 3mm r.m.s. isotropic deviation.

To create regions-of-interest (ROIs), for in-depth between group comparisons, independent sample t-tests against the null hypothesis of no effect (zero activation) were performed on a voxelwise basis using the %AUC activation maps. To negate any bias in combining these maps due to uneven numbers of participants in each group, 50 t-test maps were made by randomly selecting 13 activation maps from each of the four groups (13 being the minimum number of participants in any group) and t-testing each collection of 52 subjects. These t-test maps were then thresholded at a strict p-value of $p=0.000001$ with a 100 μ l cluster criterion and combined by ANDing the maps together (i.e., a voxel was deemed to show a significant effect only if it showed a significant effect in all 50 t-test maps after thresholding). A similar procedure was followed using the z-score activation maps.

The mean activations, both for the %AUC maps and the z-score maps, were calculated in each ROI for each participant. The mean %AUCs were inserted into an ANOVA to check for differences between each of the four groups for each ROI. The difference was considered significant if it passed the standard thresholded of $p=0.05$ (to enable the detection of any differences between the groups, no corrections for multiple comparisons were made). A similar procedure was applied to the mean z-score values. Paired-sample t-tests were performed for each ROI on the %AUC scores and the z-scores comparing the IV-cocaine and IV-saline conditions in the cocaine group.

A more in-depth analysis was performed testing for differences in the shape of the IRFs for each group. The average IRF shape in each ROI was calculated and subjected to the non-linear fitting procedure for each participant. This yielded descriptive values for each participant's haemodynamic responses, such as the time to depart from baseline (t_0), the variables r and b in the equation $y = k t^r e^{-t/b}$, the time to peak (T_{max}) and the standard deviation between the fitted model and the IRF (σ_{res}). Similar to above, the four groups were compared on each parameter for each ROI with one-way ANOVAs. Paired-sample t-tests were also performed for cocaine users between the IV-cocaine and IV-saline conditions.

2.3 Results

There were four areas identified in the %AUC AND-map: Left Precentral Gyrus ($x = -31, y = -21, z = 49$, volume = $710\mu l$), Left Medial Frontal Gyrus ($-1, -1, 54$, volume = $382\mu l$), Right Precentral Gyrus ($38, -15, 50$, volume = $320\mu l$) and Left Cingulate Gyrus ($-4, -1, 46$, volume = $182\mu l$) (see Figure 2.1). The z-score AND-map found similar

clusters in the left and right precentral gyrii but failed to detect both of the midline clusters (this is true even if the scaling factors of both the event regressor and its temporal derivative are included in the activation measure).

	Group	%AUC	z-score	t_0	r	b	$T_{\max}$	σ_{res}
LpreG	Controls	0.036 ± 0.024	0.29 ± 0.22	1.36 ± 0.44	8.39 ± 0.16	0.41 ± 0.05	4.80 ± 0.66	9.59 ± 4.07
	Cocaine	0.035 ± 0.031	0.31 ± 0.46	1.19 ± 0.56	8.40 ± 0.16	0.42 ± 0.07	4.69 ± 0.88	11.75 ± 4.72
	Nicotine	0.032 ± 0.024	0.37 ± 0.29	1.52 ± 0.55	8.29 ± 0.17	0.40 ± 0.06	4.96 ± 0.75	8.34 ± 4.56
	THC	0.035 ± 0.021	0.40 ± 0.31	1.39 ± 0.51	8.31 ± 0.21	0.43 ± 0.05	5.00 ± 0.78	10.80 ± 3.38
	IV-saline	0.032 ± 0.024	0.32 ± 0.36	1.41 ± 0.62	8.34 ± 0.14	0.41 ± 0.06	4.92 ± 0.91	9.99 ± 3.85
	IV-cocaine	0.034 ± 0.016	0.26 ± 0.23	1.67 ± 0.23	8.26 ± 0.19	0.43 ± 0.03	5.27 ± 0.33	7.98 ± 1.83
RpreG	Controls	0.045 ± 0.037	0.31 ± 0.35	1.27 ± 0.53	8.39 ± 0.23	0.40 ± 0.05	4.66 ± 0.91	9.62 ± 3.67
	Cocaine	0.038 ± 0.035	0.26 ± 0.19	1.19 ± 0.64	8.42 ± 0.17	0.40 ± 0.06	4.61 ± 0.92	10.87 ± 2.90
	Nicotine	0.034 ± 0.020	0.34 ± 0.31	1.57 ± 0.27	8.36 ± 0.20	0.42 ± 0.06	5.08 ± 0.57	8.42 ± 3.28
	THC	0.049 ± 0.032	0.41 ± 0.47	1.27 ± 0.58	8.36 ± 0.20	0.43 ± 0.05	4.90 ± 0.93	11.78 ± 4.27
	IV-saline	0.035 ± 0.030	0.15 ± 0.30	1.40 ± 0.65	8.32 ± 0.24	0.40 ± 0.06	4.81 ± 1.07	8.95 ± 2.68
	IV-cocaine	0.051 ± 0.023	0.27 ± 0.30	1.52 ± 0.46	8.33 ± 0.18	0.41 ± 0.04	4.96 ± 0.63	8.24 ± 2.23
LmedFG	Controls	0.036 ± 0.025	0.25 ± 0.33	1.45 ± 0.41	8.33 ± 0.15	0.41 ± 0.06	4.74 ± 0.79	10.51 ± 4.33
	Cocaine	0.041 ± 0.031	0.15 ± 0.28	1.48 ± 0.49	8.38 ± 0.19	0.39 ± 0.06	4.81 ± 0.81	11.01 ± 3.55
	Nicotine	0.033 ± 0.033	0.30 ± 0.29	1.26 ± 0.68	8.44 ± 0.26	0.39 ± 0.06	4.54 ± 1.03	9.80 ± 3.95
	THC	0.038 ± 0.030	0.25 ± 0.31	1.23 ± 0.61	8.32 ± 0.17	0.41 ± 0.06	4.58 ± 0.96	12.01 ± 3.39
	IV-saline	0.032 ± 0.021	0.19 ± 0.26	1.55 ± 0.58	8.35 ± 0.20	0.40 ± 0.08	5.00 ± 1.00	9.79 ± 1.96
	IV-cocaine	0.033 ± 0.014	0.19 ± 0.36	1.35 ± 0.59	8.38 ± 0.22	0.39 ± 0.07	4.50 ± 0.87	9.17 ± 1.59
LCingG	Controls	0.037 ± 0.034	0.22 ± 0.47	1.27 ± 0.62	8.35 ± 0.19	0.40 ± 0.06	4.56 ± 0.91	12.34 ± 5.44
	Cocaine	0.036 ± 0.036	0.16 ± 0.32	1.12 ± 0.61	8.39 ± 0.26	0.41 ± 0.07	4.67 ± 1.01	13.23 ± 5.13
	Nicotine	0.041 ± 0.024	0.22 ± 0.25	1.21 ± 0.61	8.39 ± 0.14	0.41 ± 0.06	4.62 ± 0.98	9.79 ± 3.12
	THC	0.036 ± 0.029	0.25 ± 0.46	1.23 ± 0.64	8.37 ± 0.22	0.39 ± 0.07	4.58 ± 1.09	12.51 ± 3.88
	IV-saline	0.039 ± 0.034	0.17 ± 0.34	1.28 ± 0.60	8.40 ± 0.19	0.40 ± 0.05	4.69 ± 0.95	12.31 ± 2.70
	IV-cocaine	0.035 ± 0.013	0.12 ± 0.31	1.36 ± 0.49	8.38 ± 0.15	0.41 ± 0.06	4.85 ± 0.72	10.30 ± 2.20

Table 2.1: The mean ± the standard deviation for each of the variables tested are displayed. The values are split by region-of-interest for each of the six groups. The Left Precentral Gyrus is denoted by LPreG, the Right Precentral Gyrus by RPreG, the Left Medial Frontal Gyrus by LMedFG and the Left Cingulate Gyrus by LCingG. Both t_0 and $T_{\max}$ are in seconds.

No significant differences in %AUC were found when testing between the four groups using a one-way ANOVA in any of the four ROI areas ($F(3,72)=0.07$, $p=0.98$; $F(3,72)=0.71$, $p=0.55$; $F(3,72)=0.2$, $p=0.89$; $F(3,72)=0.08$, $p=0.97$ are the statistics for the Left and Right Precentral Gyrii, the Left Medial Frontal Gyrus and the Left

Cingulate Gyrus respectively). No significant differences in z-score measure were found ($F(3,72)=0.49$, $p=0.69$; $F(3,72)=0.61$, $p=0.61$; $F(3,72)=0.68$, $p=0.57$; $F(3,72)=0.18$, $p=0.9$). Similarly, no differences in both %AUC ($p=0.82$; $p=0.12$; $p=0.91$; $p=0.65$) and z-score ($p=0.68$; $p=0.33$; $p=0.97$; $p=0.78$) were found between the IV-saline and IV-cocaine activation maps in each of the ROI areas.

The in-depth analysis on the shape of the IRFs revealed no significant differences in any ROI when comparing the four groups on each of the parameters t_0 , r , b , $T_{\max}$ and σ_{res} (see Figure 2.1). The relevant statistics for each of the parameters are: t_0 – $F(3,72)=0.02$, $p=0.99$; $F(3,72)=1.38$, $p=0.26$; $F(3,72)=1.04$, $p=0.38$; $F(3,72)=0.23$, $p=0.88$; r – $F(3,72)=1.77$, $p=0.16$; $F(3,72)=0.4$, $p=0.75$; $F(3,72)=1.37$, $p=0.26$; $F(3,72)=0.19$, $p=0.9$; b – $F(3,72)=0.94$, $p=0.42$; $F(3,72)=0.97$, $p=0.41$; $F(3,72)=0.27$, $p=0.85$; $F(3,72)=0.33$, $p=0.8$; $T_{\max}$ – $F(3,72)=0.63$, $p=0.59$; $F(3,72)=1.02$, $p=0.39$; $F(3,72)=0.36$, $p=0.78$; $F(3,72)=0.05$, $p=0.98$; σ_{res} – $F(3,72)=2.027$, $p=0.12$; $F(3,72)=2.72$, $p=0.06$; $F(3,72)=0.99$, $p=0.4$; $F(3,72)=1.5$, $p=0.22$;

Similarly, there were no significant differences found between the IV-saline and IV-cocaine activation maps in any of the parameters: t_0 – $p=0.21$; $p=0.35$; $p=0.45$; $p=0.73$; r – $p=0.26$; $p=0.9$; $p=0.66$; $p=0.76$; b – $p=0.32$; $p=0.35$; $p=0.59$; $p=0.85$; $T_{\max}$ – $p=0.27$; $p=0.5$; $p=0.24$; $p=0.6$; σ_{res} – $p=0.08$; $p=0.25$; $p=0.42$; $p=0.75$.

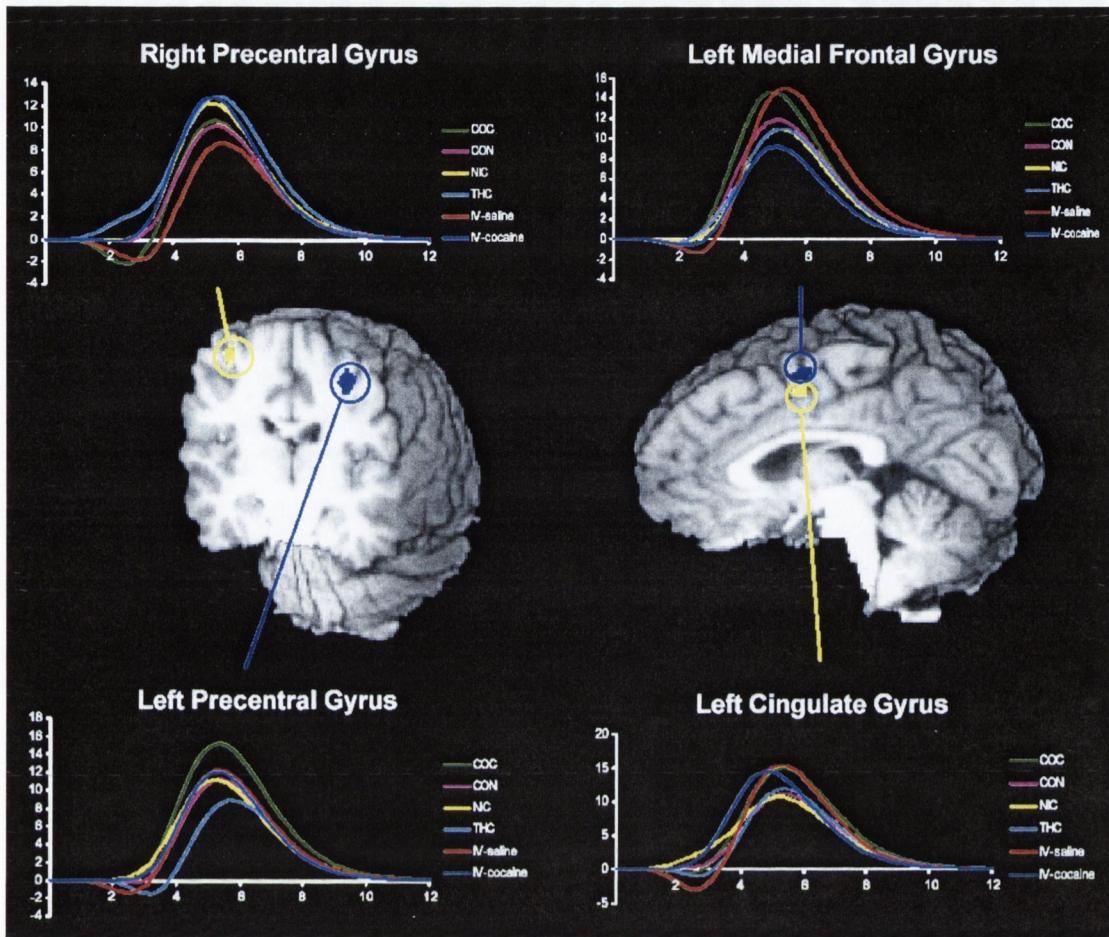


Figure 2.1: The four regions-of-interest used in the analysis are shown (Left and Right Precentral Gyrii, Left Medial Frontal Gyrus and Left Cingulate Gyrus). Also displayed are the averaged fitted haemodynamic shapes for each group (cocaine, controls, nicotine, THC, IV-saline and IV-cocaine) in each area. These shapes are averaged across voxels in each area and then across subjects in each group.

2.4 Discussion

The lack of significant differences between the groups in %AUC and z-score activation measures would suggest that the BOLD signal intensity of drug users is not compromised and might be indicative of intact vascular/neuronal coupling in these groups. Similarly, no significant differences between the IV-saline and IV-cocaine conditions were found. This result is particularly interesting since it is known that cocaine causes vasoconstriction. The in-depth analysis results reveal that despite the existing literature suggesting drug-related alterations in the properties of blood

vessels, the haemodynamic shape is not significantly changed. Thus, the results of analysis techniques that depend on haemodynamic shape may not be distorted when investigating groups that abuse drugs. The fact that no differences were found with the z-score analysis suggests that the standardised haemodynamic shape it uses is sufficiently sensitive to yield reliable results even when cocaine has been administered. Combined, these results imply that these drugs of abuse have either no impact on the vascular processes that underlie the BOLD signal or that they affect the tonic baseline but not the dynamic task-induced event-related BOLD changes (Gollub et al., 1998).

There were differences between the final activation maps of the %AUC and z-score analyses. One possible reason for this discrepancy is that the standard haemodynamic shape used in the z-score analysis was different to the actual haemodynamic response measured in the midline areas. In the majority of the groups the measured IRF is noisier in these areas (see σ_{res} column in Table 2.1), thus it is possible that the regressor used in the analysis did not correlate as well with the shape of the timeseries in those voxels and thus did not identify those voxels as significantly active (these areas do appear at a lower statistical threshold in the z-score analysis). However, the more robust %AUC analysis can accommodate voxelwise differences in haemodynamic shape (by being capable of fitting a different gamma-variate function to each voxel) was able to detect activations in these areas. For this reason, the regions-of-interest for the analyses were taken from the %AUC AND-map.

A possible confounding factor in this study is age since both the cocaine and nicotine groups differ significantly from the control group in this measure ($p < 0.001$ and

p 0.002, respectively). It has been shown that some BOLD fMRI haemodynamic properties change with normal aging (Huettel et al., 2001; Richter and Richter, 2003). To verify whether this situation biased the results, the groups were collapsed into one and age was correlated with all experimental measures (i.e., %AUC, z-score, t_0 , r , b , $T_{\max}$ and σ_{res}). No significant correlations were found, suggesting that age related differences in the BOLD response between the groups are negligible. Furthermore, if age has an effect on non-neuronally-driven vasculature, then we would expect this to have led to group differences, which were not found. A possible gender related confound could also exist given that the groups were not perfectly balanced on this demographic factor. Male and female differences were investigated by using independent samples t-tests on all experimental measures after collapsing across all groups. Again, no significant differences were discovered. Furthermore, multiple regression analyses in which age, gender and group were independent variables found no significant prediction of any of the experimental measures.

Although the sample sizes of the present study were reasonably large, one might propose that the failure to find group differences was due to a lack of statistical power. It has been shown, however, that a group of 20 participants will yield stable activation maps and that the activation measures of groups with lower sample number correlate well with those of higher power (see Chapter 4 and Murphy and Garavan, 2004b). Three recent studies in this lab have used the %AUC technique to detect differences between a cocaine user group and controls in areas associated with executive functions with group sizes of 13, 13 and 15 respectively ((Kaufman et al., 2003); Kaufman and Garavan, 2004, Under Review; Hester and Garavan, 2004, Under Review). Detection of age-related activation differences with groups of 10 or

fewer employing this technique is also possible (Nielson et al., 2002). This suggests that the null result discovered by this study is not due to a lack of power. To demonstrate this further, the drug users were collapsed into one group and analyses comparing all users to controls were repeated. Again, no significant differences were found, thus strengthening the claim that no group differences exist in haemodynamic properties.

2.5 Conclusions

Although the present results rely on accepting the null hypothesis of no difference, it is clear from Figure 2.1 that robust haemodynamic responses were present in each group. The in-depth analyses show that these haemodynamic responses are not altered by the drug of choice and, in the case of cocaine, do not differ even after an acute administration of the drug. This evidence demonstrates that it is possible to compare drug groups with controls on an equal basis, using the analyses techniques described, without the need for post-hoc corrections due to drug-related vascular changes. Thus, neuroimaging studies that identify activation differences between controls and drug-abusing groups may conclude with increased confidence that the observed differences reflect true neuronal differences.

Chapter 3

Deriving the optimal number of events for an event-related fMRI study based on the spatial extent of activation

Abstract

Event-related fMRI is a powerful tool for localising psychological functions to specific brain areas. However, the number of events required to produce stable activation maps is a poorly investigated and understood problem. Huettel and McCarthy (2001: *Neuroreport* 12, 2411-2416) have shown that the spatial extent of activation increases monotonically with the number of events in an analysis. In the present paper, this result is replicated and shown to be a consequence of the cross-correlation technique used to determine active voxels and does not hold, for example, for a GLM analysis. Another analysis technique, that does not depend on goodness-of-fit to the data, is also proposed. This technique calculates an impulse response function (IRF) for each voxel, finds the best fitting haemodynamic shape to the IRF and returns an area-under-the-curve (%AUC) activation measure. Using spatial extent as a measure, both the GLM and %AUC analysis techniques display asymptotic behaviour after as few as 25 events in two fingertapping tasks: one with overlapping haemodynamic responses and one without. The experimental validity of the %AUC technique to identify active brain regions while minimising false positive levels is demonstrated in a group study with 25 participants.

3.1 Introduction

Functional magnetic resonance imaging (fMRI) has become a widely employed and useful technique in the field of brain imaging. Its superior spatial resolution allows a researcher to localise functions to specific areas of the brain. However, as it is an expensive method, efforts must be made to optimise the design of studies to reduce costs without compromising results. Given these considerations, how many events to include in an event-related experimental design is a pertinent issue and is the focus of the present chapter.

The related issue of experimental timing is also of importance in fMRI tasks due to the relatively poor temporal resolution of the technique. Therefore, it is very important to determine the optimal interstimulus interval (ISI). As discussed in Section 1.5, it has been shown that a jittered ISI improves statistical efficiency (Dale, 1999). Birn and colleagues showed that the estimation of the IRF is optimised when stimuli are frequently alternated between task and control states (Birn et al., 2002). Consequently, an important consideration in designing and analysing event-related studies is that the analysis technique is able to accommodate overlapping haemodynamic responses and that any estimation of the required number of events should be applicable to both overlapping and non-overlapping designs.

The question of the optimal number of events has been poorly addressed in the literature. One possible reason for this could be that there is no standard metric by which to measure the required number of events. If one examined how spatial extent of activation varies as the number of events increases, one would expect that as the optimal number of events is reached, the spatial extent would cease to grow.

However, it has been shown, in a block design experiment, that when progressively averaging over scans (one scan = 200secs: 20secs ON, 20secs OFF,...) the spatial extent of activation increased monotonically and failed to asymptote even after 22 scans (Saad et al., 2003). Huettel and McCarthy have reported a similar result in an event-related task (Huettel and McCarthy, 2001). It was found that at a number of events typical of fMRI experiments (i.e., 50 trials), only 50 % of the eventually active voxels were found to be significant. Indeed, the spatial extent of activation failed to asymptote even after 150 events. This would suggest that the number of events required in an event-related study far exceeds practical capabilities and existing conventions.

In this chapter, it is proposed that the Huettel and McCarthy result is a consequence of the correlation method used. It is demonstrated that utilising a different technique, first employed by Garavan and colleagues (Garavan et al., 1999), asymptotic values are reached in the range typical of fMRI experiments. Furthermore, in keeping with the advantages of jittering events such that their haemodynamic responses overlap, we demonstrate an early asymptote with both overlapping and non-overlapping haemodynamics.

3.2 Materials and Methods

3.2.1 Participant and Task Design

One female right-handed participant (age: 38) completed two event-related finger-tapping tasks. Each task consisted of a black and white flashing checkerboard (4Hz frequency) lasting 1 second. In the first experiment, this was presented 30 times over a 10-minute period with an interval of 20 secs between each trial, thus avoiding

overlapping haemodynamics (Non-Overlap task). In the second experiment, this was presented 30 times over 6 minutes at intervals of between 8 and 16 seconds (Overlap task). A countdown clock was presented prior to each checkerboard presentation to prepare the participant. While the checkerboard appeared on screen, the participant was required to finger-tap bilaterally. Five successive blocks of the first experiment were run giving 150 finger-tapping events. On a separate day, five blocks of the second experiment were run, also yielding 150 events.

To verify the sensitivity of the different analyses methods used, 25 participants (14 female, mean age: 25, range: 19-36) completed one 6-minute block of the Overlap task.

3.2.2 Scanning Parameters

Scanning was conducted in the Medical College of Wisconsin where ethical approval was obtained. Images were acquired using contiguous 7 mm sagittal slices covering the entire brain from a 1.5T GE Signa scanner using a blipped gradient-echo, echo-planar pulse sequence (TE = 40 msec; TR = 2000 msec; FOV = 24 cm; 64 x 64 matrix; 3.75 mm x 3.75 mm in-plane resolution). High resolution spoiled GRASS anatomic images (TR = 24 msec, TE = 5msec, flip angle = 45°, FOV = 24cm, thickness = 1.0 mm with no gap, matrix size = 256x256x124) were acquired prior to functional imaging.

3.2.3 Image Processing

All data processing was conducted using the AFNI software package (Cox, 1996) (<http://afni.nimh.nih.gov/afni>). Initially, each voxel's time series was shifted so that

the separate slices acquired at varying times were aligned to the same temporal origin. For each task, the five experimental blocks were concatenated into one dataset, after removing linear trends and normalising to a common baseline. Each 3D image corresponding to a particular time point was then volume registered to a "base" image using a Fourier interpolation algorithm to align the voxels. Voxels lying outside the brain were removed.

3.2.4 Analysis Techniques

Three types of analyses were performed on the Non-Overlap data (see Figure 3.1). First, following Huettel and McCarthy's technique (2001), a ***correlation analysis*** was performed in which the ten time points (corresponding to 20 secs) after each event were averaged across all events for each voxel in the brain. This 20-second long timeseries was then correlated with a standard haemodynamic shape taken from the Statistical Parametric Mapping package (SPM). The correlation coefficient, a goodness-of-fit measure, was used as the activation measure. Second, a ***GLM analysis*** was performed in which a regressor was obtained by convolving the event stimulus train with the standard SPM haemodynamic response. A scaling factor for this regressor was found using a multiple regression procedure in each voxel. This procedure calculates how much it has to "scale" the regressor so that the best fit to the data is found. These scaling factors were then converted into a z-score map. Third, a method that can accommodate differences in haemodynamic shape, both within- and between-subject was used (see Section 1.4 for a more detailed explanation). This involved an estimation of the impulse response function (IRF) for each voxel using a deconvolution technique. Using multiple regression techniques, this procedure calculates the IRF (i.e., the average best-fitting "shape") that follows each event. The

deconvolution procedure also calculates intercept and slope parameters, which can be used to calculate a baseline (i.e., the ongoing mean activity level after excluding the variance due to the events) for the IRF. The best fitting haemodynamic shape (a gamma-variate function, $y = k t^r e^{-t/b}$) was determined for each voxel's IRF using a non-linear regression algorithm (Ward et al., 1998). The estimated haemodynamic shape for each voxel was converted into a percentage area under the curve score (%AUC) by expressing the area under the haemodynamic curve as a percentage of the area under the baseline. This method was termed the *%AUC analysis*.

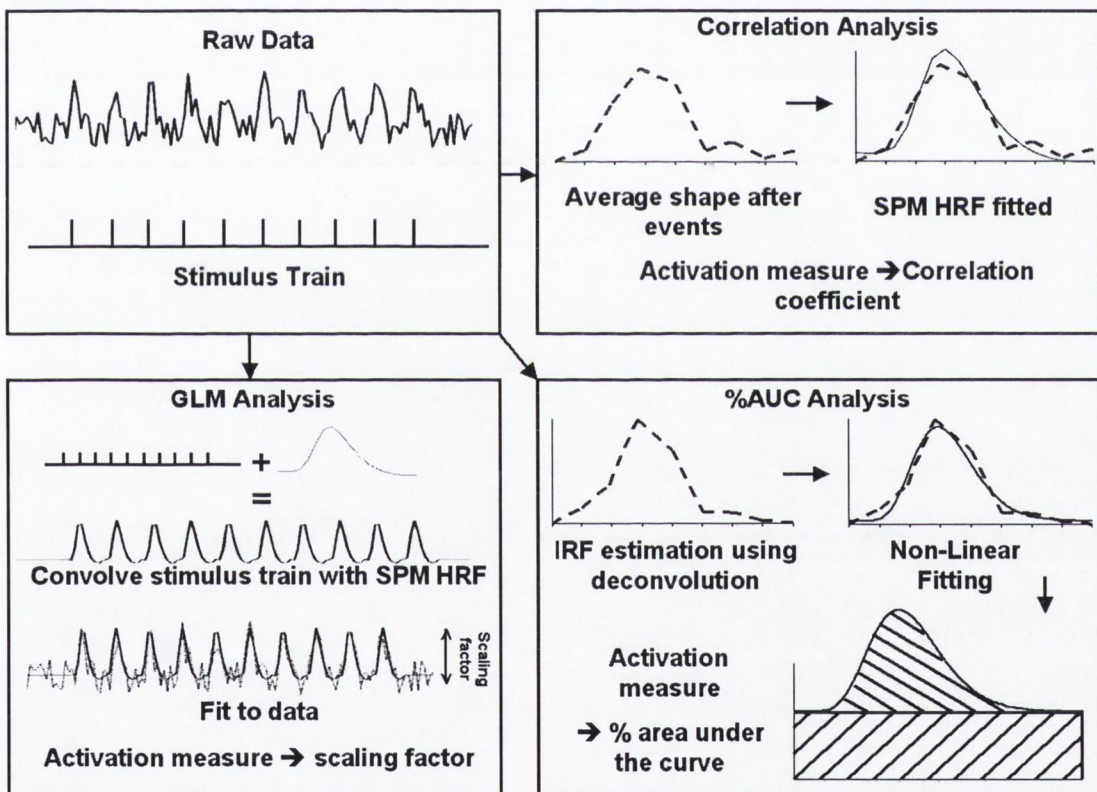


Figure 3.1: The three different analysis techniques are depicted: the correlation analysis (top right), the GLM analysis (bottom left) and the %AUC analysis (bottom right).

To create regions-of-interest (ROIs), the final activation maps from the three analyses were combined by finding the top 10% of activation scores in each map and ANDing them together (i.e., a voxel was considered to be in the ROI if it was in the top 10% of

all three activation maps). This resulted in a ROI map that had 115 voxels in the primary motor and visual areas. This liberal method for calculating the ROI was chosen so that further calculations were not biased towards any particular analysis technique.

3.2.5 Permutation Analyses

Further analyses were carried out to determine the behaviour of activation maps as a function of the number of events. For the Non-Overlap data, twenty random samplings (with replacement) of 5, 10, ..., 150 events were chosen. The correlation analysis was repeated by averaging over just those events chosen for each sample size. Similarly, the GLM analysis was repeated with a regressor created by convolving the standard haemodynamic shape with the stimulus train consisting only of the events in the sample. All other events were excluded from the analysis by including another regressor in the multiple regression procedure consisting of the convolution of the standard haemodynamic shape with the stimulus train of those events not in the sample. The %AUC analysis used a regressor containing only the events in the sample. Again all other events were excluded from the analysis with a second regressor. This procedure produced 600 activation maps for each analysis type (i.e., 20 maps for each number of events 5, 10, ..., 150). These activation maps were then thresholded and the number of active voxels in the ROIs derived above was counted. The correlation analysis maps were thresholded at $p < 0.001$ ($t > 3.89$) as in Huettel and McCarthy (2001). The GLM analysis activation maps were also thresholded at $p < 0.001$ ($z = 3.1$). The %AUC activation maps were converted into z-scores by calculating the mean and standard deviation of %AUC across all voxels in the brain, then using the formula $(\%AUC - \text{mean})/(\text{standard deviation})$ to change each

voxels %AUC into a z-score. These maps were then thresholded with the same z-value as used for the GLM analysis.

Two types of analyses were performed on the Overlap data, the GLM and the %AUC analyses (the correlation analysis was not carried out, as it was impossible to average haemodynamic shapes due to their overlapping characteristics). Activation maps were found that included all 150 events and ROIs (124 voxels) were calculated using the same procedure as above (i.e., voxels in the ROI must be present in the top 10% of both analysis types). Twenty random samples were chosen of 5, 10, ..., 150 events. Due to the overlapping nature of the haemodynamics, two regressors were made for each analysis type. First, a regressor including all the events in the sample and second, a regressor including all other events so that the contribution of these events to the timeseries would be removed. The same procedure was followed to determine the behaviour of the activation maps as a function of number of events.

3.2.6 Group Analysis

To determine if the %AUC analysis yields similar results to the GLM analysis, the data from the 25 participants were analysed using both approaches. Each dataset was then resampled to a higher 1 μ l resolution (thus 1 voxel = 1 μ l) and converted to the standard stereotaxic coordinate system of Talairach and Tournoux (Talairach and Tournoux, 1988). The images were then spatially smoothed using a Gaussian kernel with 3mm r.m.s. isotropic deviation. A voxelwise t-test was performed on the 25 maps for each analysis type and a threshold of $p < 0.00005$ with a 100 μ l cluster criterion was imposed.

3.3 Results and Discussion

3.3.1 Non-Overlap Analysis

The ROIs derived from the three analysis types of the Non-Overlap data comprised an activation map with 115 voxels, mainly situated in left and right primary motor cortex and in the visual cortex. It is important to note that when the activation maps were thresholded at $p < 0.001$, none of the analyses found all 115 voxels to be statistically significant. This is due to the method used to select the voxels for the ROI (i.e., each included voxel was required only to be in the top 10 % of each activation map), which did not depend on significance levels. This procedure resulted in a liberal selection of voxels and guaranteed that all those voxels passing the threshold criterion would be included within the ROIs.

Figure 3.2a shows the number of voxels that are found to be significantly active in the ROIs as a function of the number of events using the cross-correlation analysis. The mean of the 20 samples is a monotonically increasing line that appears to be slightly concave. This graph corresponds closely to that found by Huettel and McCarthy, with their derived equation, $V = V_{\max}(1 - e^{(-0.016 \times N)})$, fitting well over the mean data points, thus replicating their result.

Figure 3.2b shows the same graph for the GLM analysis. This graph monotonically increases as a function of the number of events. However, after 40 events, the slope of the mean line decreases (slope of line before 40 events = 1.0526, slope of line after 40 events = 0.1777). Thus between 40 and 150 events the number of significant voxels increased by only 19 (a 27% increase over the 68 voxels that were significant at 40 events).

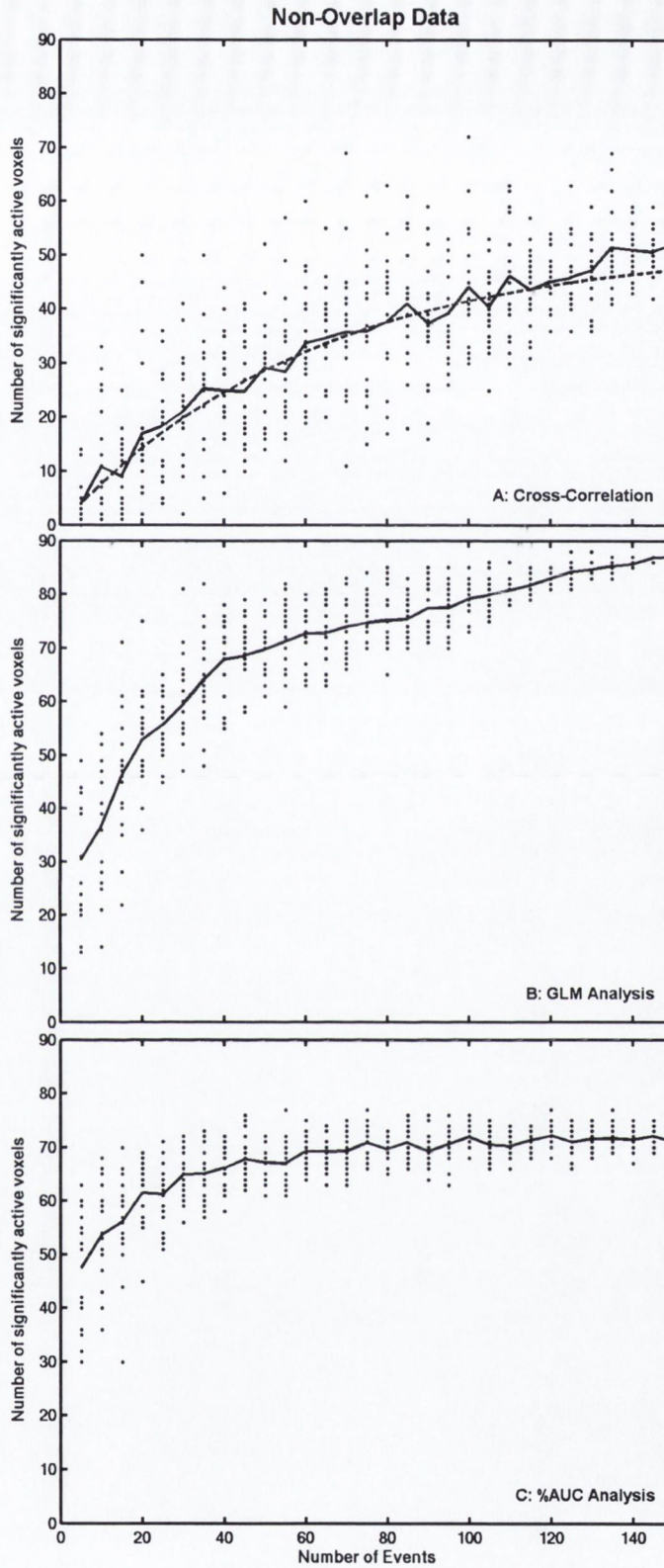


Figure 3.2: The number of significantly active voxels as a function of the number of events is plotted for (A): the cross-correlation analysis, (B): the GLM analysis and (C): the %AUC analysis on the Non-Overlap data. The solid line displays the mean number of voxels over the 20 samples at each event size. The dotted line in (A) shows the equation derived by Huettel and McCarthy (2001): $V = V_{\max}(1 - e^{(-0.016 \times N)})$, where $V_{\max}$ = number of voxels at 150 events, and N =number of events.

The graph for the %AUC analysis is quite different (see Figure 3.2c). The starting point is greater and the mean number of voxels does not increase quite as quickly as it does for the previous two analyses. Again there is a dramatic change in slope from 0.5087 before 40 events to almost a flat line slope of 0.0393 thereafter. This demonstrates that after 40 events the number of significant voxels increases very little as a function of the number of events using the %AUC method. Indeed, after 40 events only five fewer voxels are deemed to be significantly active than after 150 events. The mean number of significant voxels after only 5 events is 48 compared with 71 at 150 events.

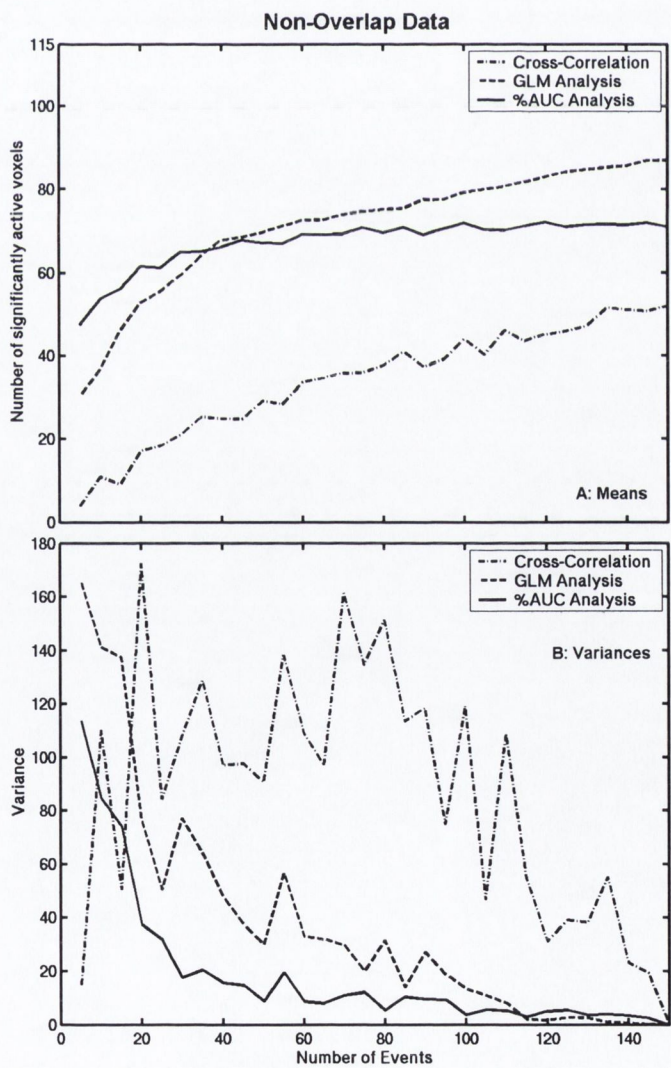


Figure 3.3: The mean number of significantly active voxels as a function of the number of events is replotted in (A) for all three analyses on the Non-Overlap data. The variance across the 20 samples for each of the analysis types is plotted in (B).

Figure 3.3a plots the means of the three analyses together to show how many of the 124 ROI voxels are deemed to be significantly active by each. The %AUC measure performs the best up to 40 events while the GLM analysis performs better thereafter. However, the %AUC analysis has a lower variance across the 20 samples, especially at lower numbers of events (see Figure 3.3b).

The reason the %AUC technique, compared to the cross-correlation technique and, to a lesser extent, the GLM analysis, is more powerful at low numbers of events is that it is almost “blind” to noise in the data since it does not depend on a goodness-of-fit measure. To demonstrate the rationale behind this, take one haemodynamic response that is exactly the same as the SPM HRF, but with random Gaussian noise added at each time-point. The cross-correlation technique compares this event with the SPM HRF and will find a poor correlation value because of the included noise. As more and more of these events are averaged together, the random noise will start to cancel out and the average will begin to look like the standard HRF. Thus, the correlation value will increase and the voxel will pass a statistical threshold and be deemed active. The GLM analysis performs better because it fits its regressor to the timeseries but then derives an activation measure from the scaling factor that does not depend on the goodness-of-fit. However, if the voxel’s IRF is quite different from the standard HRF, the GLM analysis will have difficulty fitting the regressor to the data, thus compromising the scaling factor. Of course, it should be acknowledged that this situation could be improved by such techniques as adding the first derivative of the regressor in order to accommodate phase-shifting (Calhoun et al., 2004). The %AUC analysis, on the other hand, does not assume a shape for the haemodynamic response. Rather, it fits a gamma variate shape to the central tendency of the data. In a situation

in which the central tendency remains unchanged but the amount of noise surrounding it varies, the %AUC measure will be unchanged. As this procedure does not depend on a goodness-of-fit measure of the fitted curve to the data, it should produce relatively stable activation measures with very few events (assuming that the number of events provide an adequate estimate of the central tendency). This is borne out by the variability in the number of voxels deemed to be significant across the 20 samples. The variation in the cross-correlation data (see Figure 3.3b) actually increases as the number of events increases to roughly 75 and then starts to decrease. The spread across the 20 samples is a lot less for the GLM analysis and decreases even further with the %AUC analysis. Thus, the %AUC technique is more robust and replicable at smaller numbers of events than the other two analysis methods.

This explanation might also have implications for the results found by Saad et al. (2003) mentioned earlier. It was found that the spatial extent of activation in a block design increased monotonically as the number of scans increased. The activation measure in this study was based on the correlation of the timeseries with a reference block function. If a measure that did not depend on the goodness-of-fit was used instead (e.g. ON-OFF percentage activation change), the spatial extent might asymptote more quickly, thus requiring fewer scans.

3.3.2 Overlap Analysis

It can be seen from Figure 3.4a that the GLM analysis performs better on the Overlap data than on the Non-Overlap data (compare with Figure 3.2b). There is a sharper rise in the mean number of voxels deemed significantly active at smaller numbers of events after which a shallower asymptote is reached. With 25 events in the analysis,

only 13 fewer voxels are deemed active than at 150 events. The overlapping haemodynamic shapes seem to be helping the multiple-regression procedure when it is trying to fit the convolved regressor to the data. The %AUC analysis also displays an improved performance (see Figure 3.4b), possibly because of the greater efficiency afforded by the overlapping design. The asymptote occurs around 25 events where there are 9 fewer voxels deemed active than at 150 events. The mean number of voxels is almost identical between the two analyses methods (see Figure 3.5a).

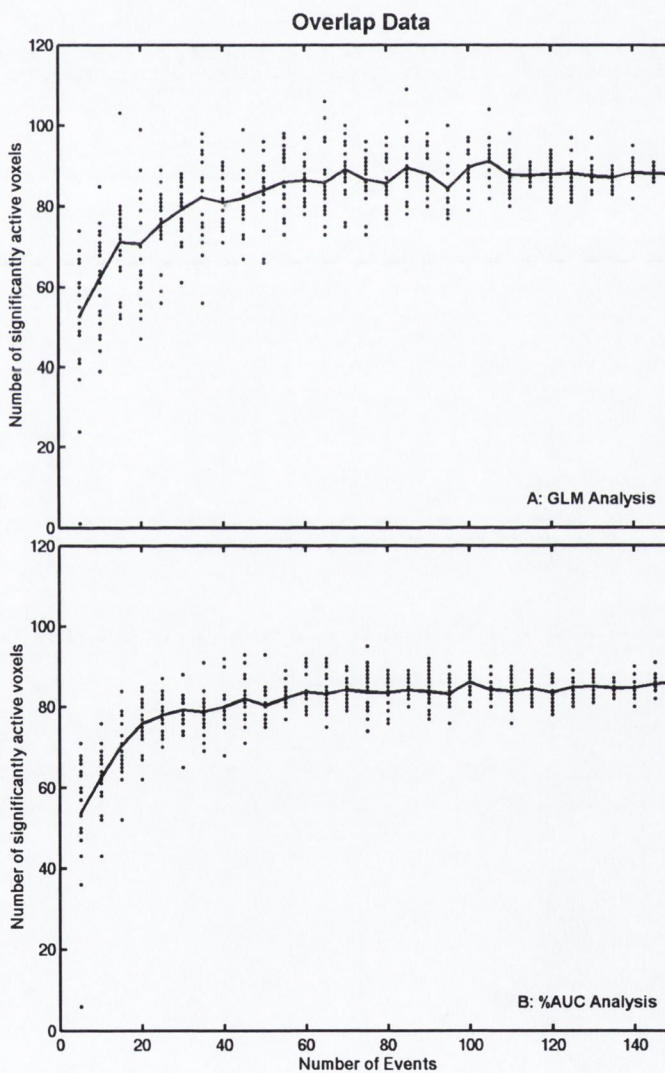


Figure 3.4: The number of significantly active voxels as a function of the number of events is plotted for (A): the GLM analysis and (B): the %AUC analysis on the Overlap data. The solid line displays the mean number of voxels over the 20 samples at each event size.

However, the variability across the 20 samples using the %AUC method is less than in the GLM analysis (see Figure 3.5b). This demonstrates the robustness of the %AUC technique when employed on data with overlapping haemodynamic shapes and suggests that as few as 25 events are sufficient to provide stable activation maps.

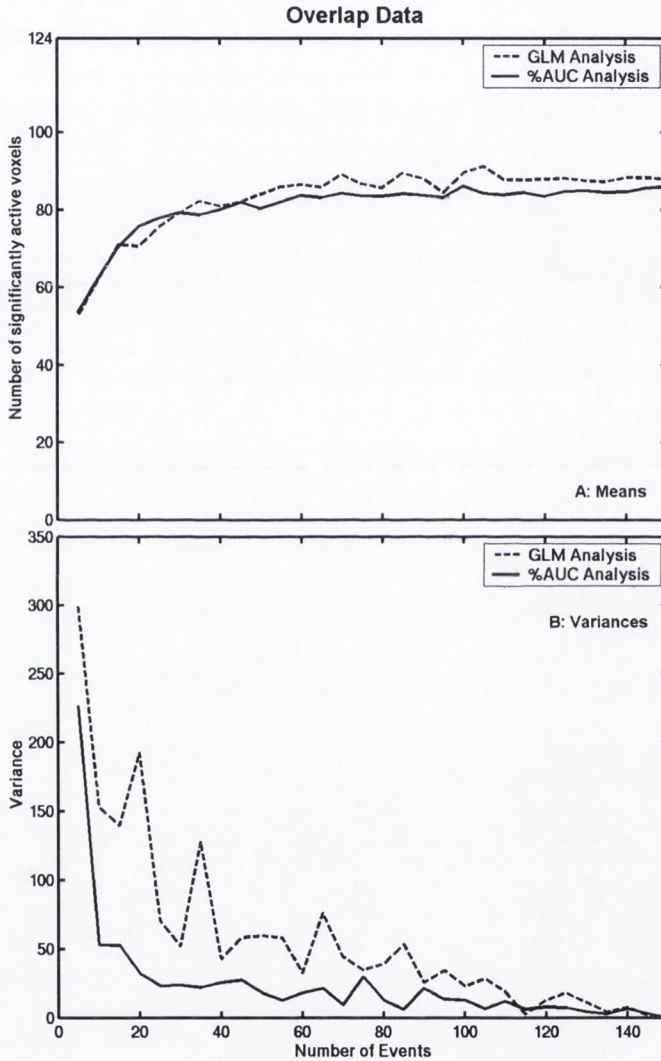


Figure 3.5: The mean number of significantly active voxels as a function of the number of events is replotted in (A) for all three analyses on the Overlap data. The variance across the 20 samples for each of the analysis types is plotted in (B).

The generalisability of the results might be queried since fMRI experiments routinely use more rapid event presentation rates of, say, 2 to 3 seconds. However, the %AUC technique depends on an accurate estimation of the impulse response function and it has been shown that estimation accuracy of the haemodynamic IRF actually increases

with shorter interstimulus intervals (ISIs) and reaches a maximum at an average ISI of roughly 2 secs (Birn et al., 2002). This would suggest that the performance of the %AUC might improve with more rapid presentation rates and that the results may be generalisable to other event-related fMRI experimental designs.

3.3.3 Group Comparison

The %AUC technique, which appears to asymptote after as few as 25 events, derives a best-fitting haemodynamic curve for every voxel in the brain and converts it into an activation measure (%AUC). As the curve fitting can be either positive or negative going, returning a positive or negative %AUC value, non-responsive voxels should return a mean of 0 when voxelwise group statistics are performed. Nonetheless, the flexibility of this method may give rise to concerns about high levels of false positives. Therefore, it is important to demonstrate that in a group study the technique will determine relevant areas to be active whilst minimising Type I errors. Twenty-five subjects (a number established to give stable activation maps in Chapter 4 and Murphy and Garavan, 2004b) performed one block of the Overlap task (30 events) and were analysed using both the GLM and %AUC analysis techniques. At a threshold of $p < 0.00005$ with a cluster criterion of 100 μ l, the GLM analysis identified an area in the Left Precentral Gyrus (432 μ l in size, centred on (-34,-19,49)) but none in the Right Precentral Gyrus. The %AUC analysis discovered a similar area in the Left Precentral Gyrus (1891 μ l in size, centred on (-30,-24,51)), two in the Right Precentral Gyrus (496 μ l in size, centred on (36,-20,40), 419 μ l in size, centred on (37,-12,51)) and one in the Right Postcentral Gyrus (484 μ l in size, centred on (31,-32,52)). If the cluster criterion is relaxed for the GLM analysis, an area in the Right Precentral Gyrus is found to be active (92 μ l in size, centred on (37,-20,47)). The pattern of

activation identified by these techniques, being restricted to motor areas, demonstrates the ability of the %AUC measure to identify areas of activation while minimising false positives.

3.4 Conclusions

The monotonic increase of the spatial extent of activation as a function of number of events, demonstrated by Huettel and McCarthy (2001), would appear to be a consequence of the cross-correlation method used to determine active voxels. A technique that does not depend on the goodness-of-fit to the data has been proposed that displays the desired effect of reaching an asymptote after a more manageable number of events. This %AUC technique, and, perhaps to a lesser extent, the GLM technique, are robust, replicable and can locate functionally relevant areas in a group study. The number of events required to yield stable activation maps was determined to be approximately 25, which is well within the practical limits of event-related fMRI studies.

Chapter 4

An empirical investigation into the number of subjects required for an event-related fMRI study

Abstract

Optimising the number of subjects required for an event-related functional imaging study is critical for ensuring sufficient statistical power. We report an empirical investigation of this issue by employing a resampling approach to the data of 58 subjects drawn from 4 previous GO/NOGO studies. Using voxelwise measures and setting the activation map from the complete sample to be a "gold standard," analyses revealed the statistical power to be surprisingly low at typical sample sizes ($n = 20$). However, voxels that were significantly active from smaller samples tended to be true positives, that is, they were typically active in the gold standard map and correlated well with the gold standard activation measure. The numerous false negatives that resulted from the lower SNR of the smaller samples drove the poor statistical power of those samples. Splitting the sample into two groups provided a test of the reproducibility of activation maps that was assessed using an alternative measure that quantified the distances between centres-of-mass of activated areas. These analyses revealed that, although the voxelwise overlap may be poor, the locations of activated areas provide some optimism for studies with typical sample sizes. With $n = 20$ in each of two groups, it was found that the centres-of-mass for 80% of activated areas fell within 25 mm of each other. The reported analyses, by quantifying the spatial reproducibility for various sample sizes performing a typical event-related cognitive task, thus provide an empirical measure of the disparity to be expected in comparing activation maps.

4.1 Introduction

The number of subjects scanned in an fMRI study is very often dictated by practical constraints such as access to scanning time and costs. Under these conditions, an investigator must make a trade-off between the number of subjects to scan and the length of the experiment. Even though these decisions are made frequently, little is known about how many trials/scans/subjects are needed to yield reliable results.

Previous research addressing these issues has shown that the spatial extent of BOLD signal activation maps increases as the number of single-trials averaged increases (Huettel and McCarthy, 2001). These authors have demonstrated that at an average of 50 trials (a typical number of trials in an fMRI study), even though the haemodynamic shape was stable, only 50 % of the eventually activated voxels were deemed significant. The volume of the activation maps only reached asymptotic values after 150 trials were averaged (although Chapter 3 derives a reduced value of 25 events). Similarly, for block design studies, it has been shown that when averaging across progressively increasing numbers of scans (where a scan, in this case, is defined as a time series of 100 volumes obtained during one 200-s stimulus presentation period: 20s ON, 20s OFF, etc.), the spatial extent of the activated voxels increased monotonically and failed to asymptote with as many as 22 scans (Saad et al., 2003).

Practically, it could be very difficult to obtain the required number of trials and scans as dictated by the above studies for each subject. This could also be highly dependent on the type of study involved. For example, a GO/NO-GO study needs to develop a pre-potency to respond and thus the trials of interest (NOGOs), by design, must be infrequent. Under these circumstances, the number of trials will be dictated by the

length of time the subject can comfortably remain in the scanner whilst maintaining their ability to perform the task. In this case, to increase the power and thus the reliability of the study, one viable option is to increase the number of subjects scanned. This, in turn, leads one to ask how many subjects are necessary to obtain a reliable group activation map.

To our knowledge, very few published studies have addressed this question. The first such paper (Friston et al., 1999a) showed that conjunction analysis with a fixed-effect model is sufficient to make inferences about characteristics that are typical of populations. Using this method can reduce the number of subjects needed to infer differences between populations that are normally required using a standard random-effects model. Although this method is very useful, it doesn't give a clear indication of how many subjects are necessary to perform an event-related fMRI study. Desmond and Glover (2002) estimated mean differences and variability between two block conditions with fMRI data. These values were used to generate power curves and an estimation of the number of subjects needed to yield reliable results. For a threshold of $p=0.05$, 12 subjects were required to achieve 80% power. At more realistic fMRI thresholds (i.e., after correcting for multiple comparisons) approximately twice as many subjects were required to yield similar power. However, this study addressed statistical power in block design experiments and may not extend to event-related designs.

This paper reports an empirical approach to the question of sample size and statistical reliability. Fifty-eight subjects performing similar event-related GO-NOGO tasks were tested. By varying the number of subjects included in the group activation maps,

we were able to derive empirically the stability of these maps for different sample sizes.

4.2 Materials and Methods

4.2.1 Subjects and Task Design

Fifty-eight right-handed subjects (35 female, mean age: 30, range: 18 - 46) completed a GO/NOGO task after providing written informed consent. The GO/NOGO task required frequent responses and occasional response inhibitions. Subjects were presented with a serial stream of letters. A response was required for every occurrence of the alternating target letters, X and Y, unless the alternation order was broken. Minor variations in the task were presented to four different groups. Fourteen subjects (6 female, mean age: 30, range 19 - 44) completed the task whereby the letters were presented serially every 500msec with 0-sec interstimulus interval (Garavan et al., 1999). A response inhibition was required on average every 20 secs with GO trials occurring on average every 3.5secs. Another fourteen subjects (10 female, mean age: 31, range: 19 - 45) completed an experiment in which only the target letters, X and Y, were presented (Garavan et al., 2002). The stimuli were presented at a frequency of 1Hz. The interstimulus interval, ranging from 100 to 400ms in 100ms steps, was tailored for each subject based on their performance in a pilot task. Fifteen subjects (9 female, mean age: 29, range: 18-46) completed the task with two conditions (Garavan et al., 2003), a high conflict condition in which the stimuli were presented for 600msec followed by a 400msec blank screen and a low conflict condition with the stimuli lasting 900msec followed by a 100msec blank. Only the low conflict condition was included in the analysis for this paper. The final variation, completed by 15 subjects (10 female, mean age: 30, range: 23 - 40), introduced a visual cue to this task,

2 to 7 secs prior to half of the NOGO events (Hester et al., 2004) but this analysis only included activation maps based on the non-cued NOGO events.

4.2.2 Scanning Parameters

Scanning for three of the studies (Garavan et al., 2003; Garavan et al., 2002; Garavan et al., 1999) was conducted in the Medical College of Wisconsin using contiguous 7 mm sagittal slices covering the entire brain from a 1.5T GE Signa scanner using a blipped gradient-echo, echo-planar pulse sequence (TE = 40 msec; TR = 2000 msec; FOV = 24 cm; 64 x 64 matrix; 3.75 mm x 3.75 mm in-plane resolution). High resolution spoiled GRASS anatomic images (TR = 24 msec, TE = 5msec, flip angle = 45°, FOV = 24cm, thickness = 1.0 mm with no gap, matrix size = 256x256x124) were acquired prior to functional imaging. Scanning for the fourth study (Hester et al., 2004) was conducted in the Nathan Kline Institute using a 1.5T Siemens VISION scanner. Contiguous 5 mm sagittal slices covering the entire brain were collected using a single-shot, T2* weighted echo planar imaging sequence (TE = 50 msec; TR = 2000 msec; FOV = 256 mm; 64 x 64 mm matrix; 4 mm x 4 mm in-plane resolution). High-resolution T1-weighted structural MPRAGE images (FOV = 256mm, isotropic 1mm voxels) were acquired following functional imaging. Ethical approval was obtained for all studies before scanning.

4.2.3 Image Analysis

All data processing was conducted using the AFNI software package (Cox, 1996). Initially, each voxel's time series was shifted so that the separate slices acquired at varying times were aligned to the same temporal origin. Each 3D image corresponding to a particular time point was then volume registered to a "base" image

using a Fourier interpolation algorithm to align the voxels. Voxels lying outside the brain were then removed.

The estimation of the impulse response function (IRF) for each voxel was performed using a deconvolution technique (for a more detailed explanation, see Section 1.4). This technique uses a multiple regression analysis to estimate the signal contributed by each individual event type to the overall time-series. A correct inhibition (STOP) regressor and a commission error (ERROR) regressor were included (it has been shown that the inclusion of an ERROR regressor leads to a more accurate activation map (see Chapter 5 and Murphy and Garavan, 2004a). The deconvolution model includes both intercept and slope parameters, which can be used to calculate a baseline for the IRF. The best fitting haemodynamic shape (a gamma-variate function) was determined for each voxel's STOP IRF using a non-linear regression algorithm (Ward et al., 1998). The estimated haemodynamic shape for each voxel was converted into a percentage area under the curve score (%AUC) by expressing the area under the haemodynamic curve as a percentage of the area under the baseline. The %AUC maps, which describe the activity due to a successful inhibition of a NOGO stimulus, were then resampled at a higher 1 μ l resolution and converted to the standard stereotaxic coordinate system of Talairach and Tournoux (Talairach and Tournoux, 1988). The images were then spatially smoothed using a Gaussian kernel with 3mm r.m.s. isotropic deviation.

T-test maps were made from various numbers of subjects in the following way. One subject from each experiment was chosen at random and a voxelwise t-test of the four activation maps against zero was performed, i.e., for each voxel in the brain, the four

activation measures, one from each subject, were t-tested against zero. This yields an average activation measure and a t-value for each voxel. Another subject from each group was chosen at random and an 8-subject t-test against zero was performed. This was repeated up to 56 subjects and then the remaining two subjects were added for a final sample size of 58. Sample sizes were incremented by four through the addition of one subject from each study in order to minimise variance due to differences between the studies. Fifty iterations of this sequence were completed producing 750 t-test maps. Each of the t-test activation maps was thresholded, using a 100 μ l cluster size criterion, with five differing t-values corresponding to uncorrected p-values of 0.01, 0.001, 0.0001, 0.00001 and 0.000001.

4.2.4 Comparison of Activation Maps

The power of an activation map was defined as the percentage of significantly active voxels in the map that were also significantly active in the “gold standard” 58-subject map (Desmond and Glover, 2002). The power for N subjects (where N = 4, 8, 12, ... 56, 58) at each p-value was determined by measuring the overlap of the N-subject map with the 58-subject map (both thresholded at the same p-value) and averaging over the 50 iterations.

It is a common practice to let a preliminary group analyses with a certain number of subjects determine if more subjects should be run to increase statistical power. This typically occurs if the preliminary results are promising but not sufficiently robust to survive statistical thresholding. To investigate the effect of this practice, the percentage of voxels in the N-subject activation map that were present in the N+4-

subject activation map (as a percentage of the number of voxels in the N-subject map) was calculated and averaged over the 50 iterations for each p-value.

The previous analyses depend on activation maps thresholded at a certain p-value. As N increases, the number of voxels in an activation map also increases. To compare maps without this sample-size dependent characteristic, we examined the unthresholded mean voxelwise activation maps. Regions-of-interest (ROIs) were defined by the 58-subject map, thresholded at the strictest p-value: $p=0.000001$ (7829 voxels). Within these ROIs, a correlation coefficient was calculated between the mean activation in the N-subject t-test map and the 58-subject t-test map.

All the above analyses compare activation maps with one map composed of a subset of the participants in the other map. As a consequence, these analyses are constrained to reach a power of unity, as the N-subject map approaches the gold standard. The result is to inflate the effect of increasing sample size. To facilitate the comparison of maps with different participants, split-half analyses were performed. One subject from each experiment was chosen and placed into the first group; another was chosen and placed into the second group. Four-subject t-test maps were then calculated for each group. In a similar fashion to above, t-test maps for 4, 8, 12, 16, 20, 24, 28 and 29 subjects were produced for each group. These maps were thresholded, using a 100 μ l cluster size criterion, at four different p-values: 0.01, 0.001, 0.0001 and 0.00001. This process was repeated for 50 iterations.

Comparisons were made between the N-subject maps in each group at each p-value level for each of the 50 iterations. A percentage overlap score, similar to a power

score, was calculated by counting the number of voxels that overlapped between the N-subject maps and expressing them as a percentage of the mean number of voxels deemed active in both maps.

The final comparison compared the centres-of-mass (CMs) of the split-half activation maps thresholded at $p = 0.05$ corrected ($p=0.001$ uncorrected). The distance between the CMs was calculated and those clusters that were less than 10mm apart were deemed to be similar. This was achieved by taking each CM in map A, comparing it with *every* CM in map B and if the distance between any two clusters was less than 10mm then both of these clusters were deemed similar. The percentage of clusters across the two maps that were found to be similar yielded a similarity measure for each sample size. This was repeated for distances of 15, 20, 25, 30, 35, 40, 45 and 50mm.

4.3 Results and Discussion

4.3.1 “Gold Standard” Analyses

The results of the power analyses are shown in Figure 4.1. It was expected that we would find a “shoulder” in the graph after a certain number of subjects, which would then asymptote to a straight line up to 58 subjects. As can clearly be seen, this did not happen. The best-case scenario was at $p=0.01$ where the power only reaches 0.5 after 32 subjects. As the p-value became stricter ($p=0.000001$), this deteriorated to 0.5 at 50 subjects. It is obvious that these activation maps are severely underpowered. As previously mentioned, it should be borne in mind that these results are confounded by the fact that they must converge to 1 after 58 subjects. It is expected that the results

would be worse if a gold standard activation map from a different set of 58 subjects was used to define the ROIs.

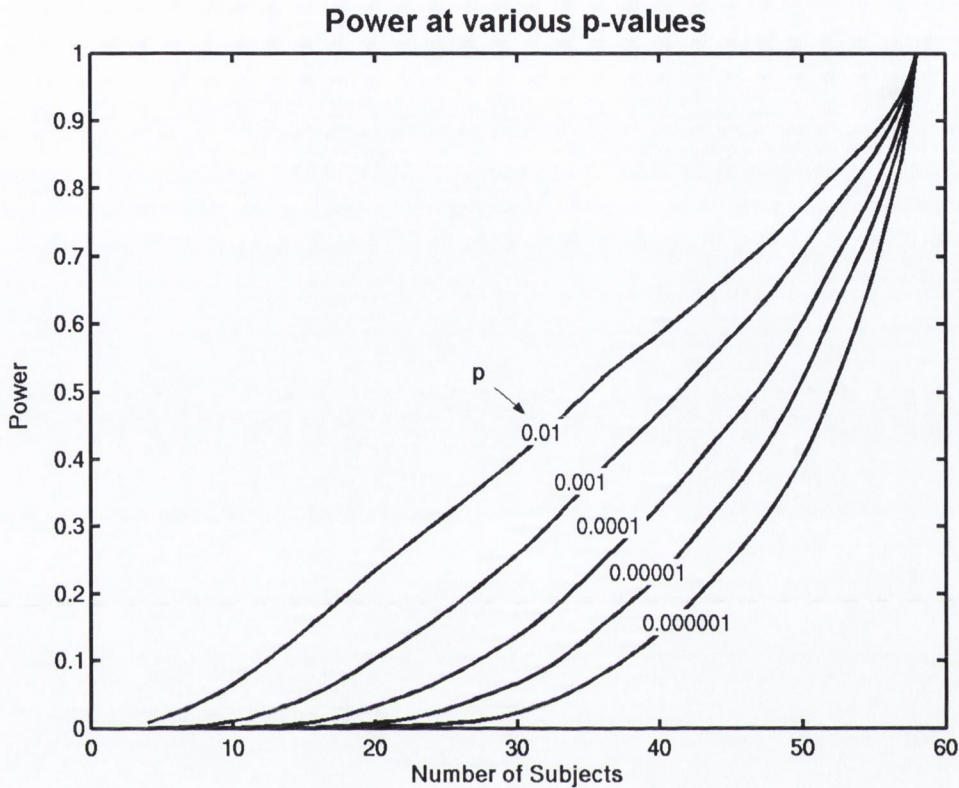


Figure 4.1: The mean power at different numbers of subjects for various p-values is depicted. Power is defined as the percentage of voxels that are in common with the 58-subject activation map when thresholded at the specified p-value

Figure 4.2 depicts the percentage of the N-subject map that was present in the 58-subject map. This graph differs from the preceding one in that the overlap was calculated as a percentage of the number of voxels in the N-subject map. That is, whereas the preceding analysis investigated the percentage of the gold standard map that was detected at each level if N, this analysis investigated the percentage of voxels dictated in an N-sample map that can be considered true positives by virtue of also being present in the gold standard map. We can see that although the power is poor (see Figure 4.1), the majority of significant voxels were also significant in the gold standard map. For example, at $p=0.01$, 0.001 and 0.0001, roughly 80% of the map

overlapped with the 58-subject map at 20 subjects. At the stricter p-values, this was not the case, due to the low-N activation maps containing few or no significantly active voxels. From these first two analyses, one might conclude that although the low-N activation maps found only a small proportion of voxels in the gold standard map to be active, the bulk of the voxels found were true positives, hence the majority of the activation map was correct. However, due to the low power of the low-N samples, substantial numbers of false negatives were recorded.

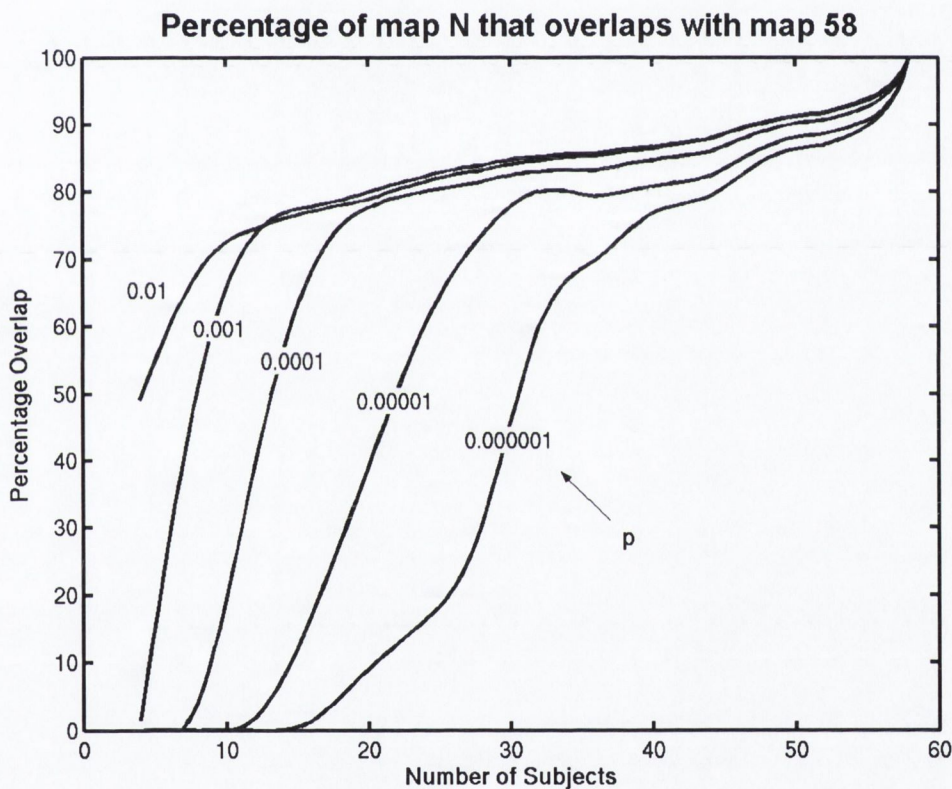


Figure 4.2: The mean percentage of map N that overlaps with map 58 is plotted for various p-values.

Figure 4.3 displays the results when map N was compared with map N+4. For $p = 0.01$, 0.001 and 0.0001 , at least three quarters of the N-subject activation map was found to be significantly active in the N+4-subject activation map after $N=22$. This might lead an experimenter to think that they are approaching an asymptote with regards to the activation maps, but we have seen from Figure 4.1 that the power is

very poor with this number of subjects. An explanation into this apparent contradiction is that the perceived asymptote deteriorates if one increases the number of additional subjects (data not shown), that is, these activation maps are deceptively labile: although the N -subject map may be similar to the $N+4$ -subject map and the $N+4$ -subject map may be similar to the $N+8$ -subject map, the N -subject map is quite different to the $N+8$ -subject map.

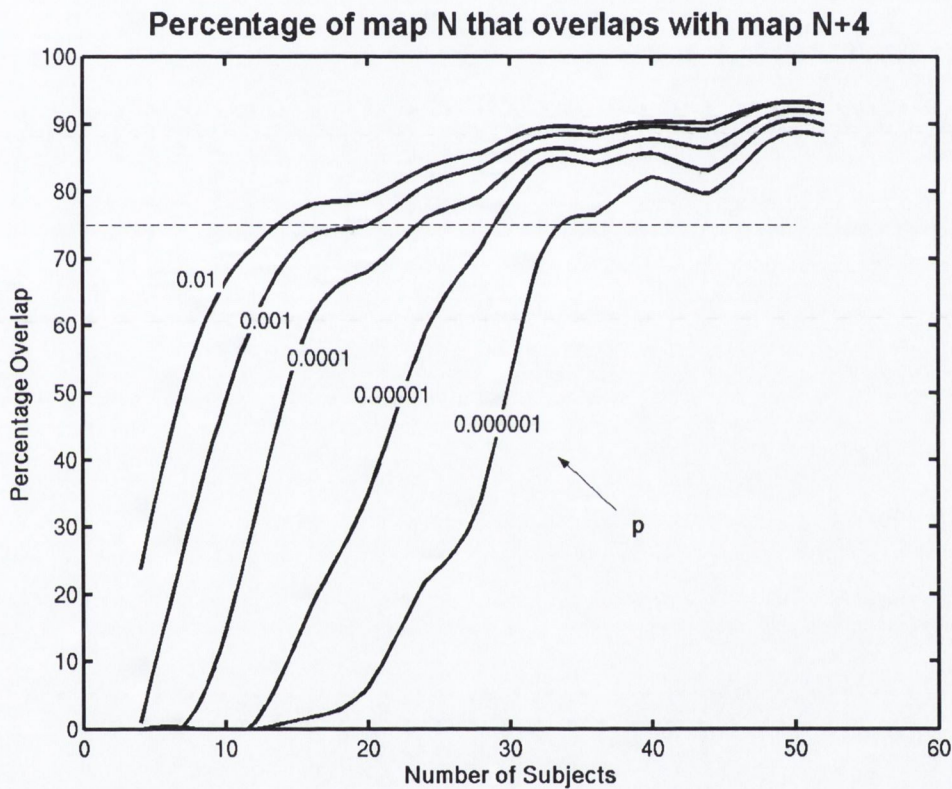


Figure 4.3: The mean percentage of map N that overlaps with map $N+4$ is plotted for various p -values. The dashed line indicates the 75% overlap level.

The results for the correlation analysis are shown in Figure 4.4. It can be seen that although the power was poor, there was a 0.8 correlation when 24 subjects were included. The conclusion is that the underlying activation maps are a lot more similar than the thresholded activation maps would lead us to believe. This result is similar to that reported in Figure 4.2, that is, the activation measure at an N of roughly 20 is reasonably accurate (i.e., similar to the gold standard), only not sufficiently robust to

survive thresholding. However, the apparent “solution” to lower the statistical threshold, of course, runs the risk of admitting many false positives into the group activation maps of low-N studies.

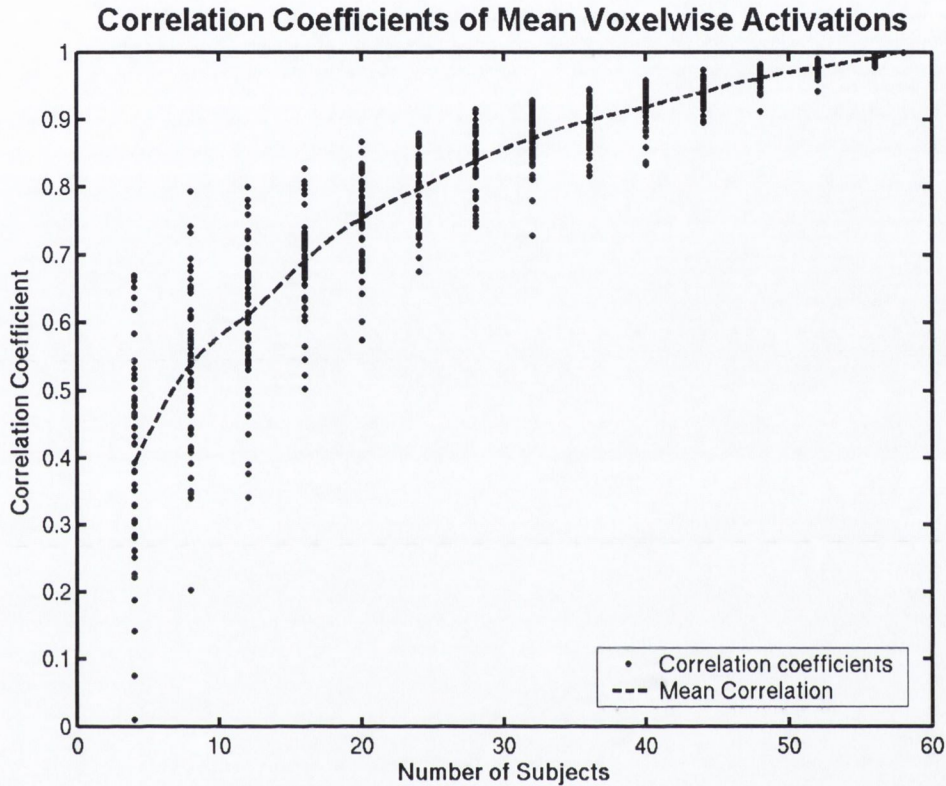


Figure 4.4: The correlation coefficients between the voxelwise activations of map N and those of map 58 are plotted for all 50 iterations. The correlation coefficients were calculated within regions-of-interest defined by the 58-subject map thresholded at the strictest p-value of 0.000001.

4.3.2 Independent Groups Analyses

Split-half analyses were performed to enable comparisons between different groups of participants. Figure 4.5 shows the percentage overlap between N-subject activation maps consisting of different participants at different p-values. This is analogous to the power calculation above and represents the reproducibility of activation maps. Unfortunately, the percentage overlap was very poor with the highest overlap only reaching 30% after 29 subjects at $p=0.01$. This result calls into question the reproducibility on a voxelwise basis of the majority of event-related fMRI studies.

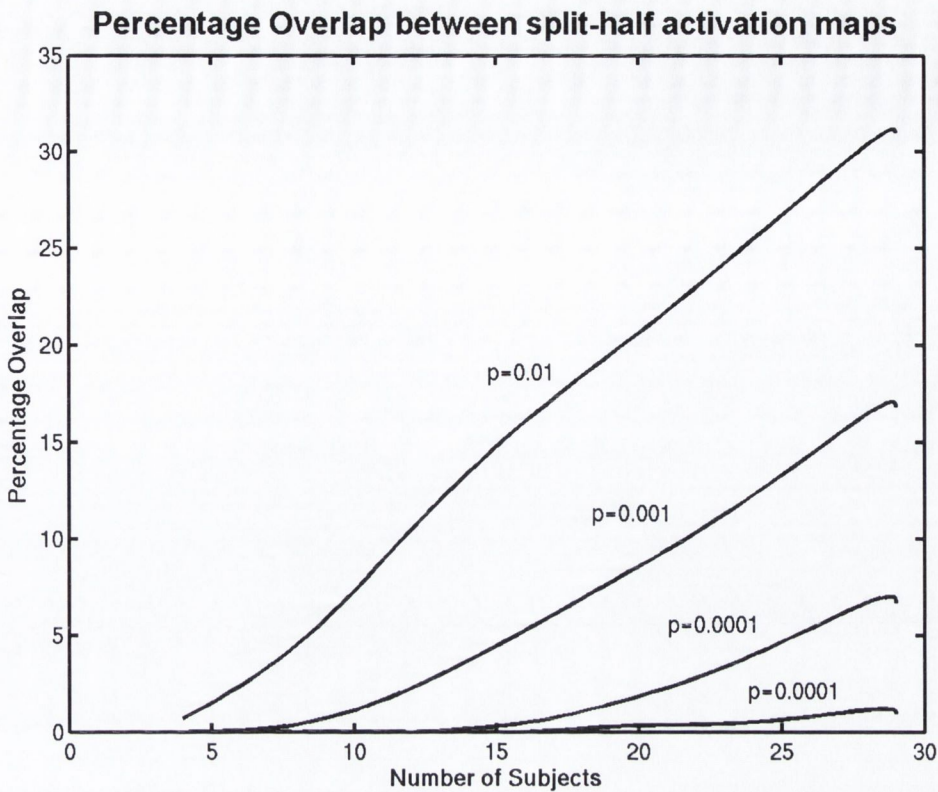


Figure 4.5: The mean number of voxels over the 50 iterations that are present in both of the split-half activation maps are plotted as a percentage of the mean size of the two maps.

The poor reproducibility of these results does, however, run counter to the experience of many fMRI researchers. For example, fingertapping studies invariably find activation in sensori-motor cortex, even with relatively few subjects. One possible criticism of the previous analyses is that the voxelwise overlap (or power) measure is too strict a criterion for measuring similarity between activation maps. Qualitatively, one 20-subject map looks very much like another despite the dismal voxelwise overlap. In most published articles, the locations of activations (the spatial location of activation being the most critical information for an fMRI researcher) are conveyed to the reader using the centre-of-masses or peak activations within clusters of significant activation. Therefore, a more intuitive measure of similarity between two activation maps might be the distance between CMs.

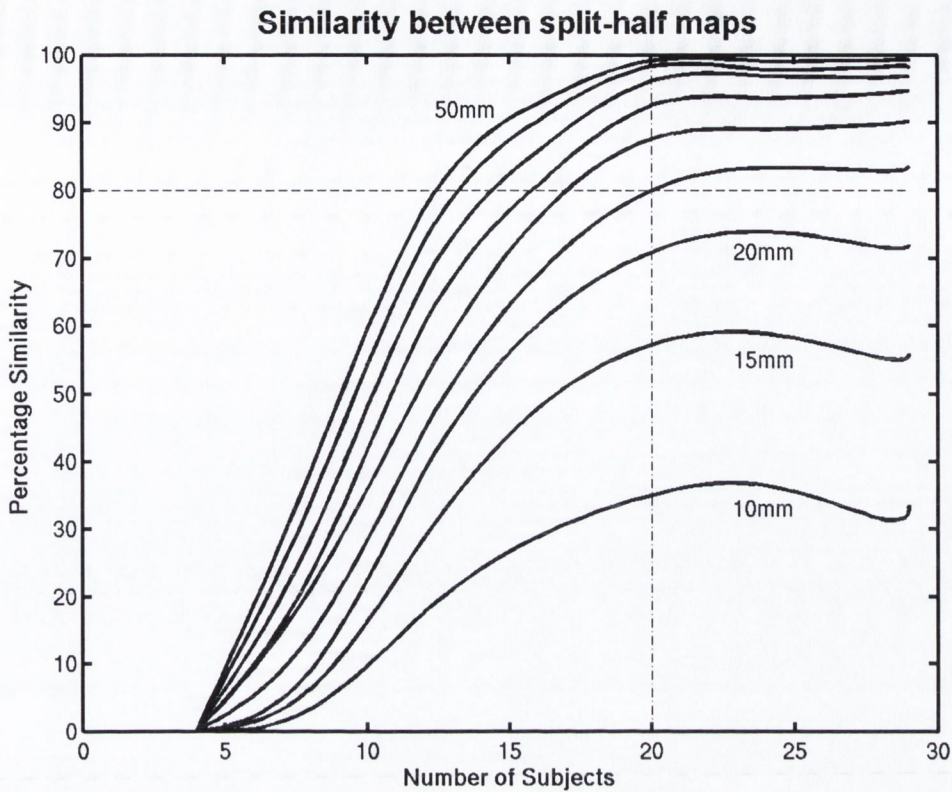


Figure 4.6: The mean similarity between split-half activation maps for various distances is depicted. Similarity is defined as the percentage of clusters in the activation maps under comparison whose centres-of-mass are less than the specified distances apart (5mm increments). The horizontal dashed line indicates the 80% similarity level. The vertical dashed line indicates the similarity values when 20 subjects are included in the activation maps

The final comparison compared the CMs of the split-half activation maps. The results are shown in Figure 4.6. After 20 subjects, all lines asymptoted to approximately a straight line, revealing that within the range of subjects tested, the addition of more subjects beyond this point would not increase the similarity between the activation maps. It can be seen that with 20 subjects in each of the two groups, the CMs of only 30% of the activated clusters in the two maps will fall within 10mm if each other (dashed vertical line). Approximately another 20% of clusters fall within 10-15mm of each other. Finally, it can be seen that 80% of clusters in the two maps fall within 25mm of each other with $N = 20$. Figure 4.6 also allows us to quantify the degree of similarity between maps as a function of sample size. If we set 80% as an acceptable similarity criterion, then we can plot the distance between similar clusters as a

function of sample size (dotted horizontal line). It can be seen that with just 10 subjects, distances between clusters of 50mm will have to be tolerated to achieve 80% similarity. With 15 subjects this drops to less than 40mm and with 20 subjects this reduces to 25mm. It is evident that the reliability of an activation map that a researcher wishes to achieve and the number of subjects required to achieve it will depend on the degree of spatial imprecision with which the researcher can contend.

The centre-of-mass comparison was repeated using a stricter thresholding value ($p = 0.0001$ uncorrected). We found that the similarity measures took longer to plateau than those found in Figure 4.6 (data not shown). For example, at $N=20$ subjects, only 60% of voxels fell within 25mm of each other. To obtain an 80% similarity between the maps, while allowing a spatial disparity of 25mm, another 6 subjects were needed. Thus, with a stricter p -value, more subjects were required to achieve an acceptable similarity measure. This was caused by the fact that one small cluster in a map that was spatially close to a cluster in the corresponding map, might disappear under a stricter thresholding criterion while its counterpart in the other map remained, thus reducing the similarity measure. However, at the stricter threshold the larger activation clusters remained in the *same* position but were reduced in volumetric size.

Wide ranges of smoothing kernels are used in fMRI studies and that employed for the present analyses (3mm r.m.s. or 4.078mm f.w.h.m.) might be smaller than is typical. To investigate the impact of a bigger smoothing kernel, the split-half percentage overlap analysis was repeated at a typical fMRI p -value of $p=0.001$ using smoothing kernels of 6mm, 9mm and 12mm r.m.s. The results are shown in Figure 4.7. We see that by increasing the size of the smoothing kernel, the percentage overlap increased

as expected. However, it must be noted that increasing the kernel also increased the number of voxels in the activation map which, presumably, attenuates the spatial precision of the map: the average largest cluster in the 20-subject activation map for each blurring kernel 3mm, 6mm, 9mm and 12mm is 3532, 15019, 36783 and 62822 voxels (μl) in size, respectively. Thus, even the 6mm r.m.s. smoothing might be too large to yield an informative activation map, that is, an activation map that can localise the cognitive function under scrutiny to a specific area of the brain. If we use the 6mm r.m.s smoothing kernel on the centre-of-mass comparison however, we find that the results are relatively unchanged. Even though the larger smoothing kernel increased the percentage overlap measure, the centres-of-mass of each of the clusters are in approximately the same position. Hence, the centre-of-mass similarity measure is not affected as greatly as the percentage overlap measure.

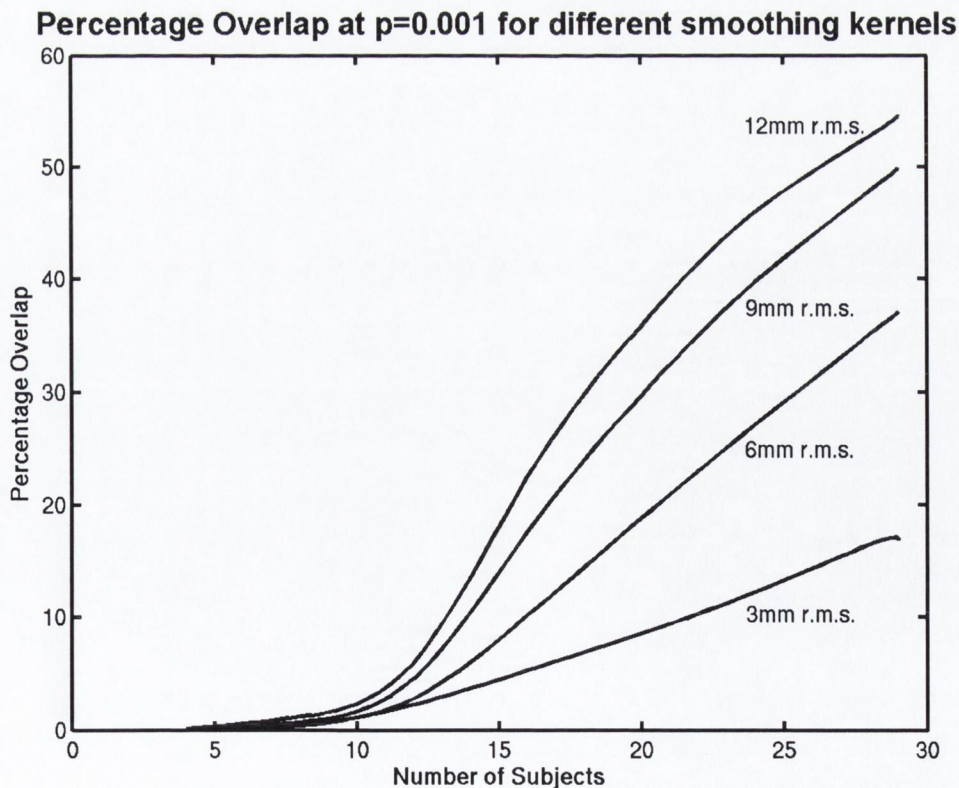


Figure 4.7: The mean number of voxels over the 50 iterations that are present in both of the split-half activation maps thresholded at $p=0.001$ are plotted as a percentage of the mean size of the two maps for different size smoothing kernels: 3mm, 6mm, 9mm and 12mm r.m.s.

A potential criticism of this study might be that we have combined the data of four different experiments. Our sampling technique, whereby we composed each N-subject map with equal numbers of subjects from each of the studies, helps minimise the effect of differences between the activation maps of the constituent studies. After splitting the subjects into two 29-subject groups using this technique, we found no significant behavioural differences between the groups in reaction time, number of correct inhibitions and scores on the Cognitive Failures Questionnaire (Broadbent et al., 1982). Furthermore, the voxelwise overlap measures between the component studies fell within the range of the values found by the 50 iterations of the resampling procedure at the relevant subject level. Hence, the inter-study differences were not greater than the inter-subject differences that would be expected given the sample sizes of those studies.

4.4 Conclusions

When planning an event-related fMRI study, it is important to know how many subjects are required to yield reliable results. This paper attempted to answer that question empirically. Although these results might be applicable to the majority of fMRI researchers investigating cognitive processes (such as inhibition), it is important to note that these results may not translate to studies with a higher signal-to-noise ratio or that suffer smaller inter-subject neuroanatomical variability. The results in this paper, however, show that when comparing N-subject activation maps to a gold standard (consisting of 58 subjects), even though the power was very poor in the range of most conventional fMRI studies (10 to 20 subjects), the majority of activated areas were true positives and voxelwise correlations in the gold standard areas were reasonably robust (0.6 – 0.75). On a cautionary note, assessing the reliability of an

activation map by determining how stable it is with the addition of a few new subjects can be misleading. Split-half comparisons showed that activation maps with the same number of different subjects overlapped poorly, suggesting that the reproducibility of studies as measured on a voxelwise basis is inadequate. However, using a more intuitive measure of similarity involving centres-of-mass demonstrated that after 20 subjects the similarity between activation maps would not be improved by adding more subjects (within the range of subjects presented in this analysis). If one could tolerate distances of 25mm between centres-of-mass, we have demonstrated that 80% of clusters proved reproducible with a sample size of 20 subjects. This analysis also allows one to determine the spatial disparity that must be tolerated to achieve a certain percentage similarity using a specified number of subjects. Desmond and Glover (2002) determined that a similar number of subjects ($n=24$) were required to give an accurate activation map with a sufficient level of power (i.e., an 80% true positive rate). Both this result and the current study would suggest that the majority of fMRI studies with typical subject numbers ($n=10$ to 20) are underpowered. However, we have shown that an activation map determined using this subject range consists of mainly true positive voxels. This leads us to the conclusion that a study with 15 subjects, say, is not necessarily inaccurate but incomplete: activated areas are likely to be true positives but there will be a sizeable number of false positives. Other brain areas might be included in the functional neuroanatomy of a cognitive process if a greater sample size was used, however, this does not mean that we must disregard results that have been found with smaller sample sizes. Due to the spatial imprecision and large number of false negatives found using smaller sample sizes, it is important that meta-analyses be carried out to give an accurate reflection of the brain structures involved in a specific cognitive process.

Chapter 5

Artifactual fMRI group and condition differences driven by performance confounds

Abstract

Analysis techniques comparing groups or conditions that vary in performance are open to a possible confound driven by those performance differences, if these errors are ignored. Disproportionate numbers of errors may either introduce noise into the signal of interest or may confound the signal of interest with additional signal associated with specific error-related processes. Two inhibitory task datasets were reanalysed, one comparing young and elderly groups, the other comparing high and low conflict conditions within the same group of subjects. The data were analysed twice using event-related techniques, one treating correct and error responses separately, the other treating error responses as though they were correct. It was found that the activation maps differed considerably, with the inclusion of errors leading to many false positive and false negative activation clusters. Using performance as a covariate, ANCOVA analyses were used to try to correct these differences without success. Data simulations that varied the number of errors included in the analyses found that surprisingly few errors could significantly alter activation maps. Consequently, brain imaging investigations that do not accommodate error contributions to functional signals are at risk of misinterpreting activation patterns.

5.1 Introduction

With the advent of non-invasive brain imaging techniques, there has followed a surge of interest in determining the functional differences that might characterise certain populations of interest. As pointed out in Chapter 1, there is a multitude of studies comparing neurologically impaired groups with clinical syndromes, such as schizophrenia (Braus et al., 2000; Menon et al., 2001; Mueller et al., 2002; Surguladze et al., 2001), ADHD (Rubia et al., 1999; Smith et al., 2002) and addiction (Pfefferbaum et al., 2001; Wexler et al., 2001) to controls. Related to this, many investigations also compare demographic differences, such as sex differences (Cowan et al., 2000) and age differences (Grossman et al., 2002). Finally, many cognitive tasks can be varied parametrically, allowing within-subject contrasts to be made. Comparisons between differing levels of n-back (Jansma et al., 2000), working and long-term memory (Braver et al., 2001) and varying difficulty in inhibition GO/NOGO tasks (Durstun et al., 2002) are examples. However, performance very often differs between the groups/conditions of interest. This chapter will focus on the detrimental effects these differences can have on an analysis when they are ignored.

Performance differences frequently occur by design as the investigation focuses on a psychological process that *a priori* is expected to be aberrant in the population or altered by the condition of interest. Consequently, group activation differences may be attributable to inherent group differences *or* to contamination of the poorer performers' activation by increased errors. Two adverse effects on a time series may result from the inclusion of errors. First, the impulse response functions of the task-related activity will be adversely affected by the inclusion of events (i.e., errors) that don't exhibit a haemodynamic response, thus affecting any signal processing

strategies. Second, the introduction of post-error processes (which can produce substantial activation in their own right (Garavan et al., 2002; Kiehl et al., 2000)) into the impulse response functions may contaminate the activation maps by introducing a haemodynamic response not related to the task.

The purpose of this study was to investigate the degree to which performance differences might confound brain activation differences. The extent to which performance can be factored out using covariate analysis as a post-hoc procedure was also addressed. With event related analyses, each individual response can be determined to be correct or incorrect and thus included or excluded from the analysis. This enables us to establish what effect the inclusion of errors might have on the final activation maps of a study. Two data sets were reanalysed for this investigation, both GO/NOGO tasks, one comparing young and elderly groups (Nielson et al., 2002) and the other comparing high conflict and low conflict conditions in a young group (Garavan et al., 2003). For each of these datasets, two types of analysis were performed: the tailored analysis – error responses were excluded from the analysis, and the untailored analysis – error responses were treated as correct responses.

The question then arises: how does the inaccuracy of an activation map vary with the number of errors included? To investigate this, the high conflict condition above was reanalysed using simulation techniques in which the correct responses were contaminated systematically with error events to establish the number of errors needed to significantly affect final activation maps.

5.2 Methods

5.2.1 *Between-group comparisons*

14 young participants (9 male, 5 female, range: 19-44) and 14 elderly participants (6 male, 8 female, range: 62-77) completed the experiment. A response inhibition task was used to compare the functional neuroanatomy of inhibitory control in each group. The task consisted of a stream of letters presented serially at 500 ms intervals with no interstimulus interval (ISI). Participants were required to press a button whenever alternating target letters (*X* or *Y*) were presented (GO). A response inhibition (NOGO) was required if the alternation order of the target letters was broken. Participants completed four runs (250 letters each) with a valid target occurring on average every 3.5s and an inhibition required on average every 20s. In total 150 correct targets (GOs) and 25 inhibition targets (NOGOs) were presented. Performance differed significantly ($t(27)=3.4$, $p<0.002$), with the young group making on average 1.9 commission errors compared to the elderly group's 6.3 commission errors. For more details see Nielson et al (Nielson et al., 2002).

5.2.2 *Within-group comparisons*

16 young participants (6 male, 10 female, range: 18-46) completed the experiment. This paradigm differed from the previous in two ways. First, there were two conflict conditions, HIGH and LOW, and second, only the letters *X* and *Y* were displayed (448 GO stimuli and 52 NOGO stimuli per conflict condition). The LOW CONFLICT condition presented stimuli for 900 ms duration (followed by a 100 ms blank ISI) and the HIGH CONFLICT condition presented stimuli for 600 ms duration (followed by a 400 ms blank screen ISI). The presentation order was counterbalanced across subjects. Performance differed significantly between conditions ($t(15)=7.2$,

$p < 0.0001$), with 9.3 commission errors made on average in the low conflict condition and 24.2 made on average in the high conflict condition. For more details see Garavan et al (Garavan et al., 2003).

5.2.3 Simulations

The HIGH CONFLICT condition of the within-group comparison was used for the simulation analysis.

5.2.4 Image Acquisition

Functional and anatomical brain imaging for both between- and within-group studies was conducted in the Medical College of Wisconsin using a 1.5T GE Signa scanner. Contiguous 7-mm sagittal slices that covered the entire brain were collected using a blipped gradient echo, echo-planar pulse sequence (TE = 40ms; TR = 2000; FOV = 24cm; 64 x 64 matrix). High resolution spoiled GRASS anatomic images were acquired prior to functional imaging to allow subsequent activation localisation. Ethical approval for both studies was obtained before scanning.

5.2.5 Image Analysis

All data processing was conducted using the AFNI software package (Cox, 1996). The following analyses were performed on each subject for each of the analysis types – tailored, untailored and simulated. Each 3D image corresponding to a particular time point was then volume registered to a "base" image using a Fourier interpolation algorithm to align the voxels. Voxels lying outside the brain were then removed.

The estimation of the impulse response function (IRF) for each voxel was performed using a deconvolution technique (see Section 1.4 for a more detailed explanation). This technique uses a multiple regression analysis to estimate the signal contributed by each individual event type to the overall time-series. For the tailored analysis, a correct inhibition (STOP) regressor and a commission error (ERROR) regressor were included. For the untailored analysis, both of the above regressors were collapsed into one. The simulations entailed sequentially relocating 1 to 12 errors from the ERROR regressor to the STOP regressor.

The deconvolution model includes both intercept and slope parameters, which can be used to calculate a baseline for the IRF. The best fitting haemodynamic shape (a gamma-variate function) was determined for each voxel's IRF using a non-linear regression algorithm (Ward et al., 1998). The estimated haemodynamic shape for each voxel was converted into a percentage area under the curve score (%AUC) by expressing the area under the haemodynamic curve as a percentage of the area under the baseline. The %AUC maps were converted to the standard stereotaxic coordinate system of Talairach and Tournoux (Talairach and Tournoux, 1988). The images were then spatially blurred using a Gaussian blurring algorithm with 3mm r.m.s. isotropic deviation.

5.2.6 Old/Young Comparison

One sample t-tests against the null hypothesis of no effect (zero activation) were performed for each group, young and old, with respect to each analysis type, tailored and untailored, on a voxelwise basis using the STOP %AUC activation maps. As in the original report, each of the four t-test maps were then thresholded with a t-value of

4.6 ($p < 0.0005$) combined with a minimum volumetric size of 100mm^3 (Nielson et al., 2002). The tailored young map and the tailored old map were then combined into an old OR young map such that a voxel was included if significant in either. The mean activations of the resulting functionally defined regions-of-interest (ROIs) were used for between-group comparisons (independent samples t-test). Similar steps were performed for the untailored analysis.

To compare tailored and untailored maps, a strategy to determine if significant regions in each map constituted the same area was devised. An area in the untailored map that overlapped with more than 50% of the volume of an area in the tailored map was considered to be a corresponding functionally active area (however, a more conservative criterion could have been used, as the minimum overlap between clusters judged to be the same was 76%). Areas could be categorised into four groups: true positives, false positives, true negatives and false negatives (see Figure 5.1). True positives are areas in which both the tailored and untailored analyses found significant group differences. False positives are areas found to show significant group differences in the untailored analysis but not in the tailored. We define two types of false positives: Type A - the area displaying a significant difference in the untailored analysis overlaps with an area displaying no difference in the tailored analysis, and Type B – the significant area in the untailored analysis does not overlap with any area in the tailored map. True negatives are areas that were present both in the tailored and untailored OR maps but which did not display a significant difference between old and young activation in either analysis. Finally, false negatives, areas found to be significantly different in the tailored analysis but not in the untailored analysis, can also be classified into two types. Type A: the significantly different area in the

tailored map overlaps with an area in the untailored map displaying no difference, and Type B: the significantly different area does not overlap with any area in the untailored maps.

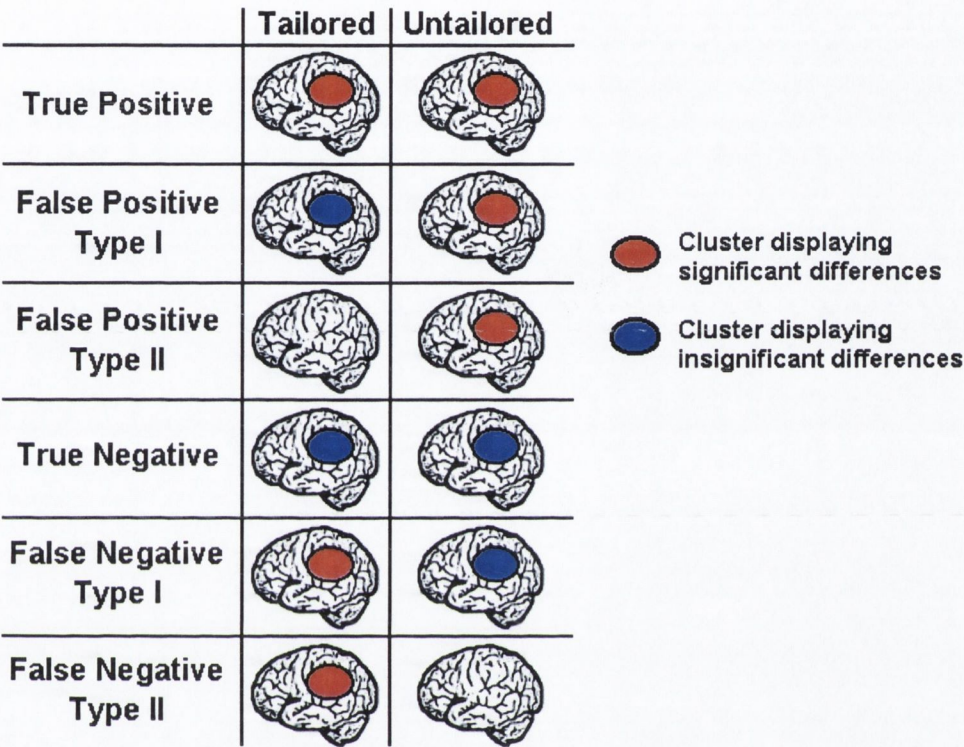


Figure 5.1: Schematic explaining the definitions for classification of untailored areas.

Often, to correct for performance differences between groups, an analysis of covariance (ANCOVA) is performed. To test the effectiveness of this post-hoc correction procedure, commission errors were used as a covariate for the old/young analysis.

At times, predefined ROIs are used in studies to perform analyses on group datasets. To emulate this type of analysis, the tailored ROIs were used on the untailored datasets to determine whether the false negative areas would display significant differences between old and young groups (as desired).

5.2.7 High/Low Conflict Comparison

A similar sequence of analyses to that used in the young/old comparison, but a higher voxelwise threshold ($t=5$; $p \leq 0.0001$) (Garavan et al., 2003), was applied to the high/low conflict within-group comparison. Thus, within-group comparisons of the high/low conflict condition effect were performed on functional ROIs defined by the tailored and untailored analyses. The same strategy as outlined above was employed to compare the tailored and untailored maps. A similar covariance correction was performed using the difference in commission errors between the two conditions as the covariate. A similar analysis using the tailored ROIs with the untailored datasets was performed as above.

5.2.8 Simulations

The simulations, performed on the high conflict condition, involved redesignating 1 to 12 errors as correct (the minimum number of errors committed) and repeating all the above analyses. A threshold of $t=5.0$ ($p < 0.0001$) was used with no volumetric cluster size criterion imposed. To determine a measure of inaccuracy, each significant voxel in a simulation was contrasted with its corresponding voxel in the “true” (i.e., tailored) map. The tailored analysis map (consisting of 1838 significant voxels when no cluster size criterion was applied) contains no contamination from errors so can be set as a gold standard against which maps contaminated by varying numbers of errors can be contrasted. If a voxel passed threshold in both the simulation and tailored map, it was considered a true positive, but if it was present in the simulation only, it was deemed a false positive. Each of the twelve resulting NOGO activation maps was compared with the tailored map and the number of false positive and true positive

voxels was calculated. This process was repeated 20 times per number of errors (i.e., 20 different random samples of error were performed for 1 to 12 errors).

5.3 Results

5.3.1 Old/Young

	Result	Effect	Area	(x, y, z)	L/R	Volume	t-values	ANCOVA		
Tailored	True Pos	O > Y	Cingulate Gyrus	(-22, -22, 44)	L	106	2.86			
			Medial Frontal Gyrus	(-12, 15, 47)	L	164	3.89			
			Supramarginal Gyrus	(-46, -51, 26)	L	361	2.84			
			Supramarginal Gyrus	(-48, -51, 38)	L	318	2.40			
			Insula	(34.5, 12.8, -1)	R	257	2.70			
		Y > O	Culmen	(36, -52, -28)	R	633	5.68			
			Inferior Occipital Gyrus*	(36, -80, -4)	R	107	2.26			
			Inferior Temporal Gyrus*	(42, -68, 1)	R	169	2.17			
			Medial Frontal Gyrus*	(8, 25, 43)	R	115	2.43			
			Middle Frontal Gyrus	(40, 22, 40)	R	233	2.14			
Untailored	True Pos	O > Y	Supramarginal Gyrus**	(-40, -54, 38)	L	3678	2.49	successful		
			Medial Frontal Gyrus	(-12, 15, 47)	L	274	3.30	successful		
		Y > O	Culmen	(35, -50, -26)	R	441	3.29	successful		
			Middle Frontal Gyrus	(40, 22, 41)	R	354	2.91	successful		
	True Neg		Middle Frontal Gyrus	(-30, -1, 49)	L	801	0.12	successful		
			Precentral Gyrus	(-45, 4, 34)	L	352	0.16	successful		
	False Pos	O > Y	-----	-----	----	-----	-----	-----		
			Type A	Y > O	Inferior Parietal Lobule	(40, -59, 47)	R	406	2.22	successful
		Type B	O > Y	Precentral Gyrus	(-48, 6, 8)	L	210	2.20	successful	
				Precentral Gyrus	(-36, -16, 44)	L	139	2.20	successful	
				Superior Temporal Gyrus	(-55, 13, 2)	L	409	2.12	successful	
				Precentral Gyrus	(29, 0, -15)	R	190	4.08	unsuccessful	
			Y > O	Angular Gyrus	(32, -55, 35)	R	105	2.49	successful	
				Middle Temporal Gyrus	(46, -40, -2)	R	135	3.47	unsuccessful	
		False Neg	Type A	O > Y	Insula	(35, 13, -1)	R	257	0.14	unsuccessful
				Y > O	Inferior Temporal Gyrus	(42, -68, 1)	R	169	1.88	unsuccessful
					Medial Frontal Gyrus	(8, 25, 43)	R	115	1.27	unsuccessful
		False Neg	Type B	O > Y	Cingulate Gyrus	(-22, -22, 44)	L	106	-----	unsuccessful
Y > O				Inferior Occipital Gyrus	(36, -80, -4)	R	107	-----	successful	

* These areas display the same young greater than old significant difference in the untailored datasets as they do in the tailored.

** This centre-of-mass of this area is in the Left Inferior Parietal Lobule, but it overlaps with the two Supramarginal Gyrus areas in the tailored analysis.

Table 5.1: Areas revealed by the Old/Young comparison

Clusters of activation for the tailored and untailored analyses of the Old/Young comparison. The second column indicates the direction in which a significant difference was observed. For the False Positive areas this indicates the direction

of significance revealed by the untailored analysis. For the False Negative areas, the direction in which the tailored analysis revealed a significant difference is indicated. The t-value obtained from the t-tests of the old versus young group is denoted in the t-values column for each of the areas (note there are no values for the False Negative Type B areas since no area was present in the untailored analysis). The ANCOVA column specifies whether the ANCOVA correction procedure was successful or unsuccessful (i.e., returned the same result as the tailored analysis or not).

The results of the original Old/Young comparison, which employed a tailored analysis, can be read elsewhere (Nielson et al., 2002). The untailored analysis revealed a substantially altered set of between-group results. Table 5.1 lists the 10 significantly different areas (5 old > young) found in the tailored analysis. Also listed are 4 true positive, 2 true negative, 7 false positive (1 Type A) and 5 false negative (3 Type A) areas found in the untailored analysis. Table 5.1 shows that the ANCOVA analysis did not correct for the inclusion of errors in the activation maps. Although the ANCOVA did not alter the true positive and true negative results (as desired), it had mixed success correcting false positives and was very poor at correcting false negatives. Using the tailored ROIs on the untailored datasets, we find that three of the five false negative areas display the same difference as shown in the tailored analysis

5.3.2 High/Low Conflict

The full results of the original High/Low Conflict comparison can be read elsewhere (Garavan et al., 2003). Two areas revealing significant differences (high > low) in the tailored analysis are listed in Table 5.2. No true positives, 3 true negatives, 7 false positives (1 Type A) and 2 false negatives (both Type B) were found in the untailored analysis (see Table 5.2). The ANCOVA procedure, as above, did not affect true negatives (there were no true positives) but failed to correct for the false positives and false negatives. One of the two false negative areas showed a high greater than low conflict difference when used in an ROI analysis on the untailored datasets.

	Result	Effect	Area	(x, y, z)	L/R	Volume	t-values	ANCOVA
Tailored	True Pos	H > L	Declive	(5, -63, -16)	R	150	4.31	
			Superior Frontal Gyrus*	(25, 44, 32)	R	109	2.77	
		L > H	-----	-----	----	-----	-----	
Untailored	True Pos	H > L	-----	-----	----	-----	-----	-----
		L > H	-----	-----	----	-----	-----	-----
	True Neg		Inferior Parietal Lobule	(37, -50, 48)	R	110	0.88	successful
			Middle Frontal Gyrus	(41, 27, 27)	R	118	1.31	successful
			Superior Parietal Lobule	(26, -64, 51)	R	411	0.53	successful
	False Pos Type A	H > L	-----	-----	----	-----	-----	-----
		L > H	Middle Temporal Gyrus	(54, -36, 1)	R	225	2.60	successful
	False Pos Type B	H > L	Middle Frontal Gyrus	(-39, 35, 17)	L	135	2.56	successful
			Anterior Cingulate	(7, 27, 20)	R	115	4.13	successful
			Medial Frontal Gyrus	(4, 11, 45)	R	359	3.36	unsuccessful
			Middle Frontal Gyrus	(24, -8, 47)	R	124	2.38	successful
			Superior Temporal Gyrus	(56, -25, 4)	R	121	2.56	unsuccessful
		L > H	Inferior Temporal Gyrus	(46, -74, -1)	R	115	2.14	successful
	False Neg Type A	H > L	-----	-----	----	-----	-----	-----
		L > H	-----	-----	----	-----	-----	-----
	False Neg Type B	H > L	Declive	(5, -63, -16)	R	150	-----	successful
			Superior Frontal Gyrus	(25, 44, 32)	R	109	-----	unsuccessful
		L > H	-----	-----	----	-----	-----	-----

* This area displays a high greater than low difference in the untailored datasets

Table 5.2: Areas revealed by the High/Low Conflict comparison

This table shows the position of the clusters revealed by the tailored and untailored analysis. Similar information is supplied as in Table 5.1.

5.3.3 Simulations

The numbers of voxels in the simulation maps increased exponentially as errors were included in the “correct” NOGO activation map ($y = 1892.11 \exp(0.0459x)$, which is almost linear, see Figure 5.2). The regression analysis revealed 1980.98 voxels after one error, increasing to 3282.13 voxels after 12 errors, an addition of approximately 125 voxels per error. The number of false positives also increased exponentially with the number of redesignated errors ($y = 1213.34 \exp(0.0765x)$, which again is almost linear; after one error: 1309.8, after 12 errors: 3038.53, see Figure 5.3). Furthermore, the number of true positives decreased logarithmically ($y = 781.63 - 157.21 \ln(x)$). After one error only 781.63 true positives (compared with 1838 in the tailored map) were present. This decreased to 390.97 voxels after 12 errors (see Figure 5.4).

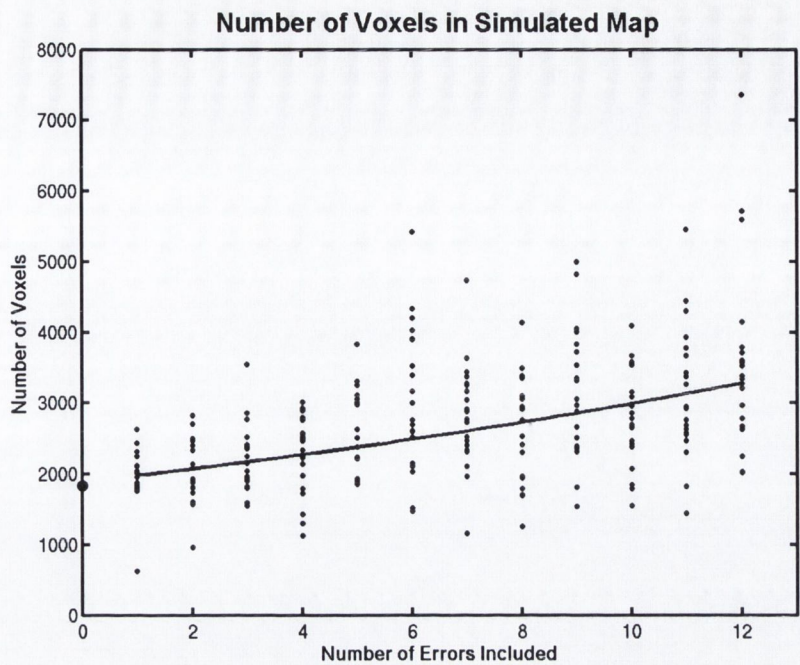


Figure 5.2: The number of significant voxels is plotted against the number of errors included in the simulation analysis. Points for all 20 simulations are shown. The solid line indicates the best-fitting exponential regression line described in the Section 5.3.3. The point on the Y-axis indicates the number of significant voxels in the tailored (zero errors) activation map.

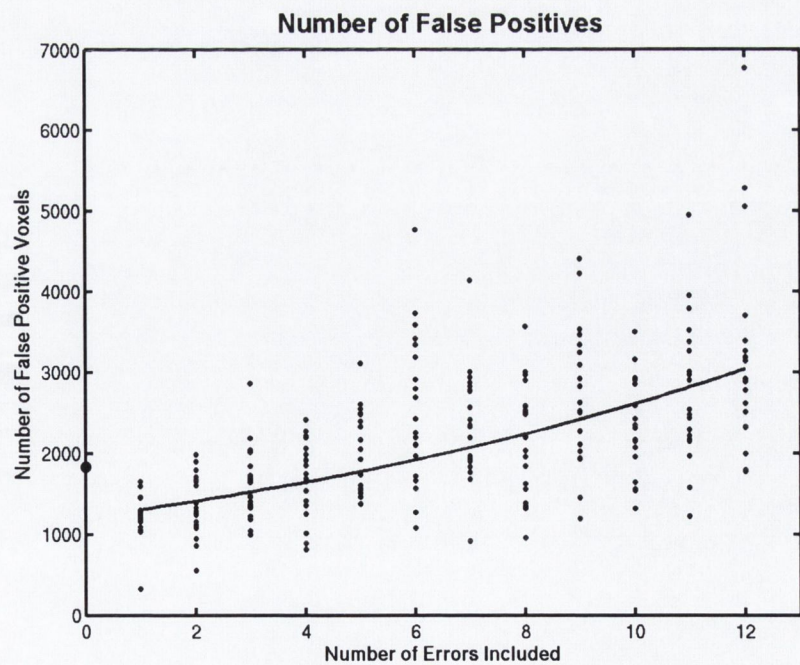


Figure 5.3: The number of false positive voxels is plotted against the number of errors included in the simulation analysis. All features of the graph are the same as in Figure 5.2.

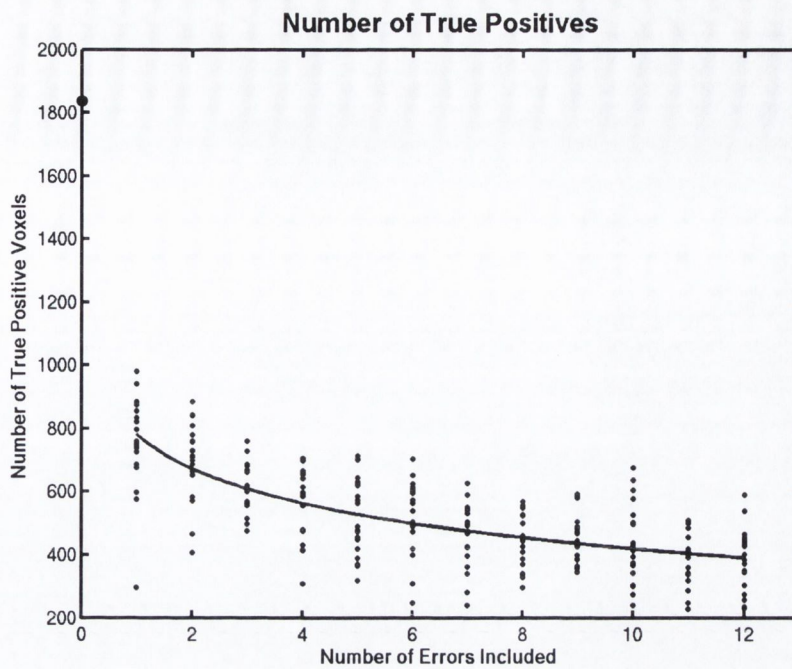


Figure 5.4: The number of true positive voxels is plotted against the number of errors included in the simulation analysis. All features of the graph are the same as in Figure 5.2.

Figure 5.5 shows the effect the inclusion of errors has on a specific cluster of significant voxels. The original tailored cluster (centred in the Left Superior Frontal Gyrus (2, -14, 48)) is shown in the “Truth” panel. Each voxel in the remaining panels is colour-coded to illustrate the number of the simulated maps that deemed the voxel significant.

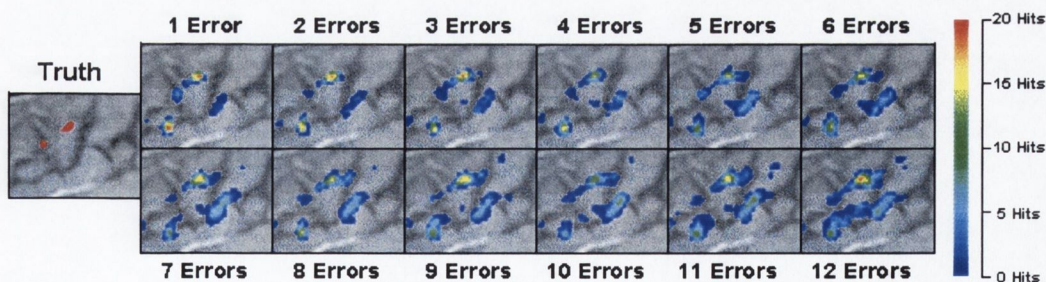


Figure 5.5: Shown is a small area of the brain in the Left Superior Frontal Gyrus (centre 2, -14, 48). Each panel shows the effect of the addition of errors on the activation map. The count of simulation maps that deemed the voxel significant is colour coded according to the scale on the right-hand side. As the number of errors included increases, the spatial extent of the simulated activation maps increases. Each voxel is deemed significant by fewer simulated activation maps as indicated by the increase of voxels on the blue end of the scale.

5.4 Discussion

5.4.1 Error-Related Activation Confounding

The analyses have shown that the inclusion of errors creates many discrepancies in brain maps when comparing groups/conditions with varying performance.

For the Old/Young between-group comparison, four true positive areas were discovered in the untailored analysis, but it must be noted that each of these was volumetrically larger than its tailored analysis counterpart. The greatest disparity can be found in the untailored region in the Supramarginal Gyrus (Left Inferior Parietal Lobule), which overlapped with the two tailored areas in the same vicinity. This region overestimates the size of the tailored area by 541%. The majority of false positive areas (6 out of 7) were found to be Type B. This indicates that the untailored analysis erroneously identified many areas to be of between-group interest. After an analysis that included errors, we were left with a brain map that showed 11 significantly different areas, only four of which were accurate.

A similar confounding of activation maps was observed for the within-group high and low conflict comparison. Here, the untailored analysis failed to detect the two areas displaying a significant difference in the tailored analysis. Seven false positive areas were found, one of which (the only Type A false positive area), the Right Middle Temporal Gyrus, encompassed the Superior Frontal Gyrus which was associated with errors in the original study (Garavan et al., 2003). The most striking finding is that the tailored and untailored activation maps only overlap in one small volume (101mm³). Including errors in the analysis produced a brain map that included 7 significantly

different areas, none of which were significantly different when errors were excluded from the activation measure

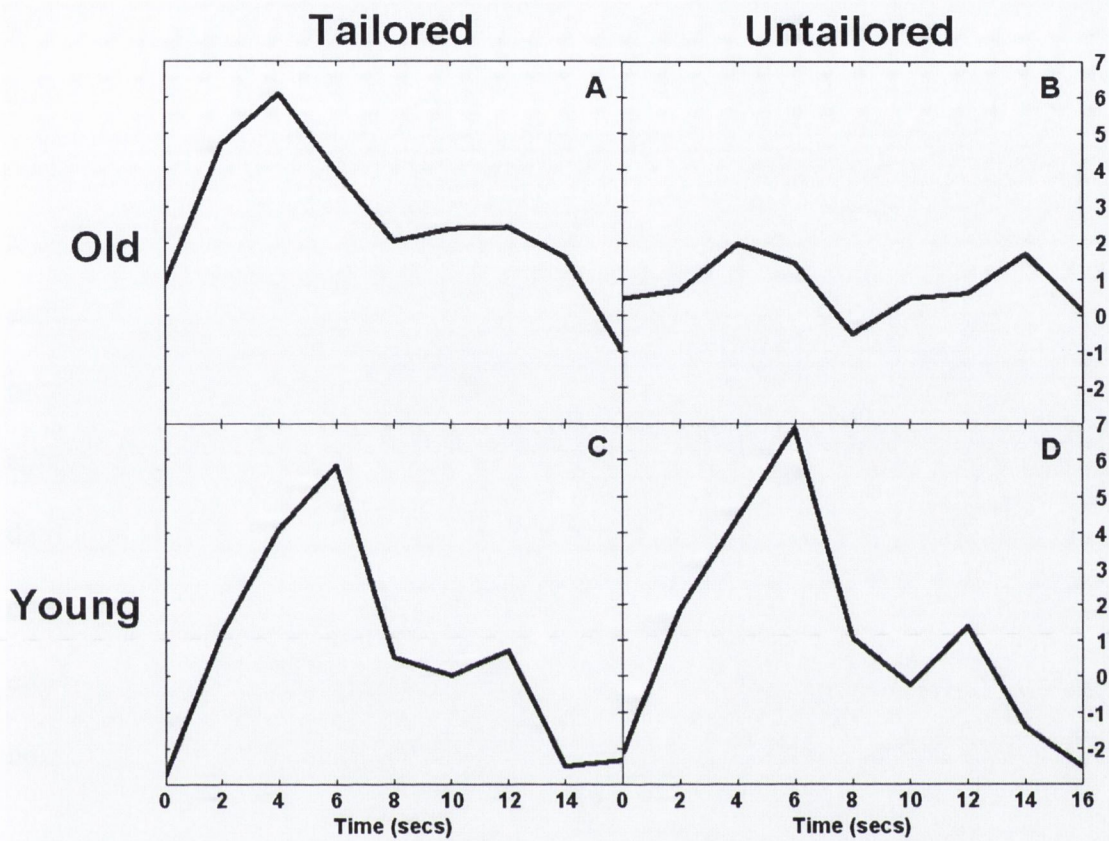


Figure 5.6: This figure shows the average impulse response function (IRF) in a False Positive Type A voxel (Right Inferior Parietal Lobule in the Old/Young analysis). Panel A depicts the average IRF for the old group in the tailored analysis, B the old group in the untailored analysis, C the young group in the tailored analysis and D the young group in the untailored analysis. Since this is a false positive voxel, there is no significant difference between the old and young group in the tailored analysis (compare A and C). The addition of errors in the untailored analysis changes the shape of the IRF. In the young group, where few errors are added, the change is negligible (compare D with C). However, the change is more noticeable in the older group, where many errors have been added - the IRF loses its haemodynamic shape (see panel B). Therefore, the untailored analysis reveals a significant young greater than old difference.

To explain how the untailored analysis can reveal false positive areas, the average IRFs in a false positive voxel are shown in Figure 5.6. This area revealed no significant differences in the tailored analysis. The inclusion of errors in the correct regressor changes the shape of the IRF in the untailored analysis. When few errors are added, as in the young group, the change is negligible. However, it can be clearly seen

in the older group that the addition of many errors into the deconvolution procedure affects the shape of the IRF greatly. Since events that display no haemodynamic shape are now averaged with those that do, the IRF tends towards a flatter line. This in turn drives the activation measure down in this voxel. Hence, the younger activation score now seems to be significantly greater than the older in the untailored analysis and thus a false positive voxel results. Similarly, the addition of this type of noise into the IRF estimation is responsible for false negative areas. It is possible that in areas related to error specific cognitive processes, the addition of errors could artificial inflate the activation score of one group more than another, thus leading to significant differences. While there was some evidence of this in the high/low conflict comparison, overall, the topography of the false positive activations make this explanation unlikely as it is not consistent with what is known of the functional neuroanatomy of error-related processes (Garavan et al., 2003; Kiehl et al., 2000).

It is possible that differences between the tailored and untailored analyses may result not only from error-related confounds, but also from differences in efficiency of response estimation between the two analyses (Friston et al., 1999b). It was found that the efficiency of response estimation did not differ significantly between the tailored and untailored analyses in the old ($p \leq 0.693$) and young ($p \leq 0.247$) groups. However, efficiency did differ significantly in the high/low conflict comparison ($p \leq 0.006$). The efficiency of response estimation for the untailored analysis was on average 10.54% greater than that for the tailored analysis across the subjects. Could this difference be responsible for the dissimilarity in activation maps in the high/low conflict comparisons? To investigate this, the efficiency of response estimation was calculated for each number of errors in the simulation analysis. This analysis uses the

same regressors as the high/low conflict condition with the exception of the transfer of events from the high conflict incorrect to the high conflict correct regressor. Thus, the “zero error” simulation is exactly the same as the analysis performed in the high/low conflict comparison. It was found that there was no significant difference in efficiency between the 0- and 1-error analysis ($p \leq 0.875$) and the 0- and 12-error analysis ($p \leq 0.153$). The simulation analysis shows that even without a change in response estimation efficiency, we find a sizeable reduction in true positives and a sizeable increase in false positives. This demonstrates that the difference in efficiency between the tailored and untailored analyses in the high/low conflict comparison is not the driving force for the disparity between activation maps.

5.4.2 Covariate Analysis

For the Old/Young between-group comparison, using performance as a covariate made it possible to correct the spurious activation maps to some extent. The untailored analysis with the ANCOVA correction procedure produced six areas displaying a significant difference between young and old activation. Only four of these areas were accurate with the untailored analysis failing to identify a further six areas that should have shown significant old/young differences.

Using performance difference as a covariate, the ANCOVA carried out on the high/low conflict within-group comparison produced even poorer results. The untailored analysis, after the ANCOVA, identified two false positives, neither of which had any correspondence to areas in the tailored analysis.

5.4.3 *Predefined ROIs*

As mentioned above, it is often the case that predefined ROIs are used to perform analyses between groups/conditions. When the tailored ROIs were used on the untailored datasets, we find that only four of the seven false negative areas displayed a significant difference. This implies that even an analysis that uses predefined ROIs, rather than depending on data-defined ROIs, will still give erroneous results due to the contamination of the signal by the inclusion of errors.

5.4.3 *Simulations*

We found that after the inclusion of just one error in the NOGO activation map, the number of false positive voxels was 71% of the original true NOGO activation map. This rose to 165% after the inclusion of 12 errors. Additionally, the number of true positives decreased logarithmically which means that after relabelling just one error as correct, only 42% of the true NOGO map remained statistically significant. After 12 errors this decreased to just 21%. The effect of this can be seen clearly in Figure 5.5. Each voxel is colour-coded to reveal the number of simulation maps that returned it as significant. As the number of errors included in the analysis increased, it can be seen that the spatial extent of activation increased. As the spatial extent increased, each voxel was found to be significant by fewer simulated maps (as indicated by the increase in blue areas in Figure 5.5). After twelve errors were included, very few of the “true” voxels were revealed by all of the simulated maps.

These results show that with the addition of errors, the false positive rate increases faster than the true positive rate decreases and demonstrate that the inclusion of errors can substantially affect the outcome of an analysis.

5.5 Conclusion

The main point of this chapter is to demonstrate that when comparing groups/conditions with differing performance levels, the impact of errors on activation measures, which have been shown to be appreciable, will be greater for the group/condition with the higher number of errors. If significant group/condition activation differences are found, this may not necessarily be due to different activation levels, but to the differential contamination by error events. Similarly, error detection areas may find their activation score increased by the inclusion of more errors. This could be problematic in block analysis techniques in which errors are often averaged with successful task performance (Buckner et al., 1998b; Kwon et al., 2001; Moses et al., 2002; Phillips et al., 1999; Thomas et al., 1999; Vaidya et al., 1998). However, recent developments in mixed design analysis have shown that it is possible to exclude errors from block design experiments using both block and event-related regressors (Burgund et al., 2003; Donaldson et al., 2001; Otten et al., 2002). Therefore, an analysis in which errors can be selectively removed, either event-related or mixed, should be the preferred option, allowing one to compare groups/conditions on an equal basis.

Chapter 6

Discussion

The four studies presented in this thesis provide valuable information to an fMRI researcher. By using an empirical approach, important, practical questions that span the entire field of fMRI, from physiology through experimental design to analytic techniques have been addressed. The fear of confounds due to haemodynamic differences in drug groups has been removed, suggestions for important experimental variables such as number of events and number of subjects have been given and the use of event-related fMRI to compare groups with performance differences has been advocated. These results are not only useful to the subset of fMRI researchers studying group differences, but to fMRI users as a whole. However, caution must be used when interpreting these findings since variations in magnetic field strength, experimental paradigms and analytic techniques could produce different conclusions than those found in this thesis. For example, by scanning at a higher field strength, the increased signal-to-noise ratio might reduce the number of events required, whereas an experimental paradigm with a smaller effect size might increase the number of subjects needed to produce a stable activation map.

It is demonstrated in Chapter 2 that, by comparing the haemodynamic responses of drug groups with those of controls, the integrity of the BOLD response may not be affected due to vascular confounds in drug-addicted populations. Comparisons

between these groups and controls are therefore reliable, even after acute administration of the drugs. Not only is this information valuable to people investigating the effects of drug addiction on brain function, it is useful to anybody performing fMRI group comparisons. Since the groups studied in Chapter 2 use vasoactive drugs, it is unlikely that the integrity of the haemodynamic shape is less in any other group. Therefore, this possible confound may not have implications for studies that, for example, examine functional deficits in clinical syndromes or explore how demographic differences affect brain activations. This result could even benefit studies that derive activation maps from a single group; for example, there may be no need for concern if any members of the group use a substance that might affect vasculature, such as smokers or coffee drinkers.

After addressing this issue with regard to physiology, the focus now shifts to single-subject analysis. The question discussed in Chapter 3 concerns such a topic: how many events does it take to produce a stable activation map for one subject? This is a very relevant issue for anybody who wishes to use event-related tasks with fMRI. The previous research by Huettel and McCarthy showed that a point of stability was not reached even after 150 events, thus rendering nearly all event-related fMRI studies to date somewhat unreliable (Huettel and McCarthy, 2001). However, by using analytic techniques that do not depend on a goodness-of-fit measure, it is possible to show that asymptotic behaviour in activation maps is reached using a much lower number. In Chapter 3, the more reasonable result of 25 events is shown to provide sufficient power to produce a stable activation measure. The lesson, for the average fMRI researcher, is to choose their analytic method carefully. Twenty-five is not an unreasonable number of events and should be within the reaches of most experimental

tasks. Apart from the obvious improvement in reduction of cost, an advantage of employing a small number of events is that subjects are less likely to fatigue and so will perform more competently throughout the duration of the task. Similarly, it is easier to investigate experiments that require events to be spaced far apart, such as GO/NOGO and oddball tasks that are quite long even with few events. The importance of achieving accurate single subject maps cannot be understated since these are the building blocks for a group analysis.

If one can assume the accuracy of an individual's activation map, then a reasonable question is: how many of these activation maps are needed to produce a reliable group map? The results from Chapter 4 show that this is not an easy question to answer. Power in the group maps with subject numbers in the standard range for fMRI studies is very poor. Reproducibility, something that is key to validating fMRI studies, is almost non-existent at the same number of subjects. However, this conclusion may hinge on the metric used, since the metric of voxelwise overlap appears to be very restrictive. A more intuitive measure, based on the centres-of-mass of activated areas, shows that the researcher needs to tolerate some spatial imprecision when approaching the question of reproducibility. The results demonstrate that after 20 subjects, the similarity between group maps does not increase with the addition of more subjects. It has been demonstrated that a similar number of subjects are required to produce stable activation maps in a block-design (Desmond and Glover, 2002). The fact that these two results are comparable is not a coincidence. The analyses in this chapter performed comparisons between different groups of single-subject activation maps. How these activation maps were derived is inconsequential: they could be from either a block or an event-related design. Thus, performing similar analyses on block-

design activation maps should produce a comparable result. Again, the fMRI researcher is provided with a simple message: at least 20 subjects are required to have confidence in a group activation map.

If a researcher has followed the last two recommendations, a more reliable and accurate group activation map should be produced. People often want to compare two different groups that might have significant performance differences between them and so an fMRI investigator might ask the question: are there any other methodological concerns? By comparing groups with performance differences and treating error events as correct responses, it was shown in Chapter 5 that the inclusion of these errors in an analysis can create many discrepancies in between-group activation differences. Not only are functionally defined activation maps skewed, but activation measures in those areas are drastically altered. Since block design techniques cannot exclude error events, they are particularly susceptible to erroneous results. The final recommendations for an fMRI researcher, who is interested in determining group differences, are to use an event-related design and to remove the influence of error trials from the analysis.

A further positive consequence of this thesis is the justification of the %AUC method. This technique is an interesting approach to fMRI analysis that has many advantages but is not widely employed, compared to the popular GLM analysis. Its ability to take account of the variability in the haemodynamic shape makes it a robust method. For example, in Chapter 2, the GLM analysis failed to recognise the activation in midline areas uncovered by the %AUC method. Similarly, the Right Precentral Gyrus was deemed significant by only the %AUC technique in Chapter 3. The lower variability

in number of voxels when using the %AUC points to a more reliable activation measure, especially at small numbers of events (typical in the majority of fMRI studies). An additional advantage is that a lot of information is gained about the haemodynamic response in each voxel, such as, time to leave baseline, time to peak, signal maximum and time to reach maximum that is not available with other techniques. Further analyses can be performed on these variables to investigate subtle within-subject, between-subject and between-group differences in haemodynamic shape.

fMRI is a widely used technique. Often important basic variables are chosen arbitrarily and unsuitable analytic methods are utilised. This thesis attempts to approach basic user-driven questions with techniques that are easily reproduced and simple to understand. These results provide the researcher with valuable information that will help in the design of more accurate and reliable event-related fMRI studies.

References

- Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *Neuroimage* 8, 360-369.
- Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., and Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magn Reson Med* 25, 390-397.
- Bellgowan, P. S., Saad, Z. S., and Bandettini, P. A. (2003). Understanding neural system dynamics through task modulation and measurement of functional MRI amplitude, latency, and width. *Proc Natl Acad Sci U S A* 100, 1415-1419.
- Benowitz, N. L., and Jones, R. T. (1981). Cardiovascular and metabolic considerations in prolonged cannabinoid administration in man. *J Clin Pharmacol* 21, 214S-223S.
- Birn, R. M., Cox, R. W., and Bandettini, P. A. (2002). Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. *Neuroimage* 15, 252-264.
- Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 34, 537-541.
- Blamire, A. M., Ogawa, S., Ugurbil, K., Rothman, D., McCarthy, G., Ellermann, J. M., Hyder, F., Rattner, Z., and Shulman, R. G. (1992). Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. *Proc Natl Acad Sci U S A* 89, 11069-11073.
- Bloom, A. S., Tershner, S., Fuller, S. A., and Stein, E. A. (1997). Cannabinoid-induced alterations in regional cerebral blood flow in the rat. *Pharmacol Biochem Behav* 57, 625-631.

- Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16, 4207-4221.
- Braus, D. F., Ende, G., Hubrich-Ungureanu, P., and Henn, F. A. (2000). Cortical response to motor stimulation in neuroleptic-naïve first episode schizophrenics. *Psychiatry Research: Neuroimaging Section* 98, 145-154.
- Braver, T. S., Barch, D. M., Kelley, W. M., Buckner, R. L., Cohen, N. J., Miezin, F. M., Snyder, A. Z., Ollinger, J. M., Akbudak, E., Conturo, T. E., and Peterson, S. E. (2001). Direct Comparison of Prefrontal Cortex Regions Engaged by Working and Long-Term Memory Tasks. *NeuroImage* 14, 48-59.
- Breiter, H. C., Gollub, R. L., Weisskoff, R. M., Kennedy, D. N., Makris, N., Berke, J. D., Goodman, J. M., Kantor, H. L., Gastfriend, D. R., Riorden, J. P., *et al.* (1997). Acute effects of cocaine on human brain activity and emotion. *Neuron* 19, 591-611.
- Broadbent, D. E., Cooper, P. F., FitzGerald, P., and Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *Br J Clin Psychol* 21 (Pt 1), 1-16.
- Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc Natl Acad Sci U S A* 93, 14878-14883.
- Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., and Dale, A. M. (1998a). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron* 20, 285-296.

- Buckner, R. L., Koutstaal, W., Schacter, D. L., Wagner, A. D., and Rosen, B. R. (1998b). Functional-Anatomic Study of Episodic Retrieval Using fMRI I. Retrieval Effort versus Retrieval Success. *NeuroImage* 7, 151-162.
- Buckner, R. L., Koutstaal, W., Schacter, D. L., and Rosen, B. R. (2000). Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain* 123 Pt 3, 620-640.
- Burgund, E. D., Lugar, H. M., Miezin, F. M., and Petersen, S. E. (2003). Sustained and transient activity during an object-naming task: a mixed blocked and event-related fMRI study. *Neuroimage* 19, 29-41.
- Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med* 39, 855-864.
- Calhoun, V. D., Adali, T., McGinty, V. B., Pekar, J. J., Watson, T. D., and Pearlson, G. D. (2001). fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. *Neuroimage* 14, 1080-1088.
- Calhoun, V. D., Stevens, M. C., Pearlson, G. D., and Kiehl, K. A. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *Neuroimage* 22, 252-257.
- Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1999). Time course of fMRI-activation in language and spatial networks during sentence comprehension. *Neuroimage* 10, 216-224.
- Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *Neuroimage* 6, 93-103.

- Cowan, R. L., Frederick, B. d. B., Rainey, M., Levin, J. M., Maas, L. C., Bang, J., Hennen, J., Lukas, S. E., and Renshaw, P. F. (2000). Sex differences in response to red and blue light in human primary visual cortex: a bold fMRI study. *Psychiatry Research: Neuroimaging Section* 100, 129-138.
- Cox, R. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers in Biomedical Research* 29, 162-173.
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8, 109-114.
- Dale, A. M., and Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping* 5, 329-340.
- de Zubicaray, G. I., Andrew, C., Zelaya, F. O., Williams, S. C., and Dumanoir, C. (2000). Motor response suppression and the prepotent tendency to respond: a parametric fMRI study. *Neuropsychologia* 38, 1280-1291.
- Desmond, J. E., and Glover, G. H. (2002). Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J Neurosci Methods* 118, 115-128.
- D'Esposito, M., Postle, B. R., Ballard, D., and Lease, J. (1999). Maintenance versus manipulation of information held in working memory: an event-related fMRI study. *Brain Cogn* 41, 66-86.
- Dewey, W. L. (1986). Cannabinoid pharmacology. *Pharmacol Rev* 38, 151-178.
- Domino, E. F., Minoshima, S., Guthrie, S., Ohl, L., Ni, L., Koeppe, R. A., and Zubieta, J. K. (2000). Nicotine effects on regional cerebral blood flow in awake, resting tobacco smokers. *Synapse* 38, 313-321.

- Donaldson, D. I., Petersen, S. E., Ollinger, J. M., and Buckner, R. L. (2001). Dissociating state and item components of recognition memory using fMRI. *Neuroimage 13*, 129-142.
- Durston, S., Thomas, K. M., Worden, M. S., Yang, Y., and Casey, B. J. (2002). The Effect of Preceding Context on Inhibition: An Event-Related fMRI Study. *NeuroImage 16*, 449-453.
- Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., and Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature 369*, 525.
- Ernst, M., Matochik, J. A., Heishman, S. J., Van Horn, J. D., Jons, P. H., Henningfield, J. E., and London, E. D. (2001). Effect of nicotine on brain activation during performance of a working memory task. *Proc Natl Acad Sci U S A 98*, 4728-4733.
- Fassbender, C., Murphy, K., Foxe, J. J., Wylie, G. R., Javitt, D. C., Robertson, I. H., and Garavan, H. (2004). A topography of executive functions and their interactions revealed by functional magnetic resonance imaging. *Brain Res Cogn Brain Res 20*, 132-143.
- Friederici, A. D., Meyer, M., and von Cramon, D. Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain Lang 75*, 289-300.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998a). Event-related fMRI: characterizing differential responses. *Neuroimage 7*, 30-40.

- Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., and Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage* 2, 45-53.
- Friston, K. J., Holmes, A. P., and Worsley, K. J. (1999a). How many subjects constitute a study? *Neuroimage* 10, 1-5.
- Friston, K. J., Josephs, O., Rees, G., and Turner, R. (1998b). Nonlinear event-related responses in fMRI. *Magn Reson Med* 39, 41-52.
- Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., and Dale, A. M. (1999b). Stochastic designs in event-related fMRI. *Neuroimage* 10, 607-619.
- Gaillard, W. D., Pugliese, M., Grandin, C. B., Braniecki, S. H., Kondapaneni, P., Hunter, K., Xu, B., Petrella, J. R., Balsamo, L., and Basso, G. (2001). Cortical localization of reading in normal children: an fMRI language study. *Neurology* 57, 47-54.
- Garavan, H., Ross, T. J., Kaufman, J., and Stein, E. A. (2003). A midline dissociation between error-processing and response-conflict monitoring. *Neuroimage* 20, 1132-1139.
- Garavan, H., Ross, T. J., Murphy, K., Roche, R. A., and Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *Neuroimage* 17, 1820-1829.
- Garavan, H., Ross, T. J., and Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proc Natl Acad Sci U S A* 96, 8301-8306.
- Ghatan, P. H., Ingvar, M., Eriksson, L., Stone-Elander, S., Serrander, M., Ekberg, K., and Wahren, J. (1998). Cerebral effects of nicotine during cognition in smokers and non-smokers. *Psychopharmacology (Berl)* 136, 179-189.

- Gollub, R. L., Breiter, H. C., Kantor, H., Kennedy, D., Gastfriend, D., Mathew, R. T., Makris, N., Guimaraes, A., Riorden, J., Campbell, T., *et al.* (1998). Cocaine decreases cortical cerebral blood flow but does not obscure regional activation in functional magnetic resonance imaging in human subjects. *J Cereb Blood Flow Metab* 18, 724-734.
- Goodyear, B. G., and Menon, R. S. (1998). Effect of luminance contrast on BOLD fMRI response in human primary visual areas. *J Neurophysiol* 79, 2204-2207.
- Grossman, M., Cooke, A., DeVita, D., Alsop, D., Detre, J., Chen, W., and Gee, J. (2002). Age-Related Changes in Working Memory during Sentence Comprehension: An fMRI Study. *NeuroImage* 15, 302-317.
- Grunwald, F., Schrock, H., and Kuschinsky, W. (1991). The influence of nicotine on local cerebral blood flow in rats. *Neurosci Lett* 124, 108-110.
- Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639-1651.
- Heishman, S. J., Huestis, M. A., Henningfield, J. E., and Cone, E. J. (1990). Acute and residual effects of marijuana: profiles of plasma THC levels, physiological, subjective, and performance measures. *Pharmacol Biochem Behav* 37, 561-565.
- Hester, R. L., Murphy, K., Foxe, J. J., Foxe, D. M., Javitt, D. C., and Garavan, H. (2004). Predicting success: patterns of cortical activation and deactivation prior to response inhibition. *J Cogn Neurosci* 16, 776-785.
- Hollister, L. E. (1986). Health aspects of cannabis. *Pharmacol Rev* 38, 1-20.
- Huettel, S. A., and McCarthy, G. (2001). The effects of single-trial averaging upon the spatial extent of fMRI activation. *Neuroreport* 12, 2411-2416.

- Huettel, S. A., Singerman, J. D., and McCarthy, G. (2001). The effects of aging upon the hemodynamic response measured by functional MRI. *Neuroimage 13*, 161-175.
- Iida, M., Iida, H., Dohi, S., Takenaka, M., and Fujiwara, H. (1998). Mechanisms underlying cerebrovascular effects of cigarette smoking in rats in vivo. *Stroke 29*, 1656-1665.
- Jacobsen, L. K., Gore, J. C., Skudlarski, P., Lacadie, C. M., Jatlow, P., and Krystal, J. H. (2002). Impact of intravenous nicotine on BOLD signal response to photic stimulation. *Magn Reson Imaging 20*, 141-145.
- James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., and Goodale, M. A. (1999). Repetition priming and the time course of object recognition: an fMRI study. *Neuroreport 10*, 1019-1023.
- Jansma, J. M., Ramsey, N. F., Coppola, R., and Kahn, R. S. (2000). Specific versus Nonspecific Brain Activity in a Parametric N-back Task. *NeuroImage 12*, 699-697.
- Kaufman, J. N., Ross, T. J., Stein, E. A., and Garavan, H. (2003). Cingulate hypoactivity in cocaine users during a GO-NOGO task as revealed by event-related functional magnetic resonance imaging. *J Neurosci 23*, 7839-7843.
- Kaufman, M. J., Levin, J. M., Maas, L. C., Rose, S. L., Lukas, S. E., Mendelson, J. H., Cohen, B. M., and Renshaw, P. F. (1998a). Cocaine decreases relative cerebral blood volume in humans: a dynamic susceptibility contrast magnetic resonance imaging study. *Psychopharmacology (Berl) 138*, 76-81.

- Kaufman, M. J., Levin, J. M., Ross, M. H., Lange, N., Rose, S. L., Kukes, T. J., Mendelson, J. H., Lukas, S. E., Cohen, B. M., and Renshaw, P. F. (1998b). Cocaine-induced cerebral vasoconstriction detected in humans with magnetic resonance angiography. *JAMA* 279, 376-380.
- Kiehl, K. A., Liddle, P. F., and Hopfinger, J. B. (2000). Error processing and the rostral anterior cingulate: An event-related fMRI study. *Psychophysiology* 37, 216-223.
- Kumari, V., Gray, J. A., ffytche, D. H., Mitterschiffthaler, M. T., Das, M., Zachariah, E., Vythelingum, G. N., Williams, S. C., Simmons, A., and Sharma, T. (2003). Cognitive effects of nicotine in humans: an fMRI study. *Neuroimage* 19, 1002-1013.
- Kwon, H., Menon, V., Eliez, S., Warsofsky, I. S., White, C. D., Dyer-Friedman, J., Taylor, A. K., Glover, G. H., and Reiss, A. L. (2001). Functional neuroanatomy of visuospatial working memory in fragile X syndrome: relation to behavioural and molecular measures. *American Journal of Psychiatry* 158, 1040-1051.
- Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., and et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci U S A* 89, 5675-5679.
- Lahaye, P. J., Poline, J. B., Flandin, G., Dodel, S., and Garnero, L. (2003). Functional connectivity: studying nonlinear, delayed interactions between BOLD signals. *Neuroimage* 20, 962-974.
- Lawrence, N. S., Ross, T. J., and Stein, E. A. (2002). Cognitive mechanisms of nicotine on visual attention. *Neuron* 36, 539-548.

- Lee, J. H., Telang, F. W., Springer, C. S., Jr., and Volkow, N. D. (2003). Abnormal brain activation to visual stimulation in cocaine abusers. *Life Sci* 73, 1953-1961.
- Lemberger, L., McMahon, R., Archer, R., Matsumoto, K., and Rowe, H. (1974). Pharmacologic effects and physiologic disposition of delta 6a,10a dimethyl heptyl tetrahydrocannabinol (DMHP) in man. *Clin Pharmacol Ther* 15, 380-386.
- Lex, B. W., Mendelson, J. H., Bavli, S., Harvey, K., and Mello, N. K. (1984). Effects of acute marijuana smoking on pulse rate and mood states in women. *Psychopharmacology (Berl)* 84, 178-187.
- Linden, D. E., Prvulovic, D., Formisano, E., Vollinger, M., Zanella, F. E., Goebel, R., and Dierks, T. (1999). The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. *Cereb Cortex* 9, 815-823.
- Liu, J., Gao, B., Mirshahi, F., Sanyal, A. J., Khanolkar, A. D., Makriyannis, A., and Kunos, G. (2000). Functional CB1 cannabinoid receptors in human vascular endothelial cells. *Biochem J* 346 Pt 3, 835-840.
- Liu, T. T., Frank, L. R., Wong, E. C., and Buxton, R. B. (2001). Detection power, estimation efficiency, and predictability in event-related fMRI. *Neuroimage* 13, 759-773.
- Luo, F., Wu, G., Li, Z., and Li, S. J. (2003). Characterization of effects of mean arterial blood pressure induced by cocaine and cocaine methiodide on BOLD signals in rat brain. *Magn Reson Med* 49, 264-270.
- Mathew, R. J., and Wilson, W. H. (1991). Substance abuse and cerebral blood flow. *Am J Psychiatry* 148, 292-305.

- Mathew, R. J., Wilson, W. H., Coleman, R. E., Turkington, T. G., and DeGrado, T. R. (1997). Marijuana intoxication and brain activation in marijuana smokers. *Life Sci* 60, 2075-2089.
- Mathew, R. J., Wilson, W. H., Humphreys, D. F., Lowe, J. V., and Wiethe, K. E. (1992). Regional cerebral blood flow after marijuana smoking. *J Cereb Blood Flow Metab* 12, 750-758.
- Mathew, R. J., Wilson, W. H., Turkington, T. G., and Coleman, R. E. (1998). Cerebellar activity and disturbed time sense after THC. *Brain Res* 797, 183-189.
- Mendelson, J. H. (1987). Marijuana. In *Psychopharmacology: The Third Generation of Progress*, H. Y. Meltzer, ed. (New York, Raven Press), pp. 1565-1571.
- Menon, V., Anagnosan, R. T., Glover, G. H., and Pfefferbaum, A. (2001). Functional Magnetic Resonance Imaging Evidence for Disrupted Basal Ganglia Function in Schizophrenia. *American Journal of Psychiatry* 158, 646-649.
- Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., and Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage* 11, 735-759.
- Moses, P., Roe, K., Buxton, R. B., Wong, E. C., Frank, L. R., and Stiles, J. (2002). Functional MRI of Global and Local Processing in Children. *NeuroImage* 16, 415-424.
- Mueller, J. L., Roeder, C., Schuier, G., and Klein, H. E. (2002). Subcortical overactivation in untreated schizophrenic patients: A functional magnetic resonance image finger-tapping study. *Psychiatry and Clinical Neurosciences* 56, 77.

- Murphy, K., and Garavan, H. (2004a). Artifactual fMRI group and condition differences driven by performance confounds. *Neuroimage* 21, 219-228.
- Murphy, K., and Garavan, H. (2004b). An empirical investigation into the number of subjects required for an event-related fMRI study. *Neuroimage* 22, 879-885.
- Neumann, J., Lohmann, G., Zysset, S., and von Cramon, D. Y. (2003). Within-subject variability of BOLD response dynamics. *Neuroimage* 19, 784-796.
- Nielson, K. A., Langenecker, S. A., and Garavan, H. (2002). Differences in the functional neuroanatomy of inhibitory control across the adult life span. *Psychol Aging* 17, 56-71.
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A* 89, 5951-5955.
- Otten, L. J., Henson, R. N., and Rugg, M. D. (2002). State-related and item-related neural correlates of successful memory encoding. *Nat Neurosci* 5, 1339-1344.
- Pfefferbaum, A., Desmond, J. E., Galloway, C., Menon, V., H., G. G., and Sullivan, E. V. (2001). Reorganization of Frontal Systems Used by Alcoholics for Spatial Working Memory: An fMRI Study. *NeuroImage* 14, 7-20.
- Phillips, M. L., Williams, L., Senior, C., Bullmore, E. T., Brammer, M. J., Andrew, C., Williams, S. C. R., and David, A. S. (1999). A differential neural response to threatening and non-threatening negative facial expressions in paranoid and non-paranoid schizophrenics. *Psychiatry Research: Neuroimaging Section* 92, 11-31.

- Pollmann, S., and von Cramon, D. Y. (2000). Object working memory and visuospatial processing: functional neuroanatomy analyzed by event-related fMRI. *Exp Brain Res* 133, 12-22.
- Postle, B. R., and D'Esposito, M. (1999). Dissociation of human caudate nucleus activity in spatial and nonspatial working memory: an event-related fMRI study. *Brain Res Cogn Brain Res* 8, 107-115.
- Randall, M. D., Harris, D., Kendall, D. A., and Ralevic, V. (2002). Cardiovascular effects of cannabinoids. *Pharmacol Ther* 95, 191-202.
- Richter, W., and Richter, M. (2003). The shape of the fMRI BOLD response in children and adults changes systematically with age. *Neuroimage* 20, 1122-1131.
- Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998). Event-related functional MRI: past, present, and future. *Proc Natl Acad Sci U S A* 95, 773-780.
- Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S. C. R., Simmons, A., and Bullmore, E. T. (1999). Hypofrontality in Attention Deficit Hyperactivity Disorder During Higher-Order Motor Control: A Study With Functional MRI. *American Journal of Psychiatry* 156, 891-896.
- Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., Simmons, A., Williams, S. C., Giampietro, V., Andrew, C. M., and Taylor, E. (2001). Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage* 13, 250-261.
- Saad, Z. S., Ropella, K. M., DeYoe, E. A., and Bandettini, P. A. (2003). The spatial extent of the BOLD response. *Neuroimage* 19, 132-144.

- Smith, A., Taylor, E., Rogers, J. W., Newman, S., and Rubia, K. (2002). Evidence for a pure time perception deficit in children with ADHD. *Journal of Child Psychology and Psychiatry* 43, 529.
- Stevens, A. A., Skudlarski, P., Gatenby, J. C., and Gore, J. C. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magn Reson Imaging* 18, 495-502.
- Surguladze, S. A., Calvert, G. A., Brammer, M. J., Campbell, R., Bullmore, E. T., Giampietro, V., and David, A. S. (2001). Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiatry Research: Neuroimaging Section* 106, 1-14.
- Talairach, J., and Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain* (New York, Theime Medical).
- Terborg, C., Birkner, T., Schack, B., and Witte, O. W. (2002). Acute effects of cigarette smoking on cerebral oxygenation and hemodynamics: a combined study with near-infrared spectroscopy and transcranial Doppler sonography. *J Neurol Sci* 205, 71-75.
- Thomas, K. M., King, S. W., Franzen, P. L., Welsh, T. F., Berkowitz, A. L., Noll, D. C., Birmaher, V., and Casey, B. J. (1999). A Developmental Functional MRI Study of Spatial Working Memory. *NeuroImage* 10, 327-338.
- Vaidya, C. J., Austin, G., Kirkorian, G., Ridlehuber, H. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. E. (1998). Selective effects of methylphenidate in attention deficit hyperactivity disorder: A functional magnetic resonance study. *Proceedings of the National Academy of Sciences* 95, 14494-14499.
- Wager, T. D., and Nichols, T. E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage* 18, 293-309.

- Ward, B., Garavan, H., Ross, T. J., Bloom, A., Cox, R., and Stein, E. A. (1998). Nonlinear regression for fMRI time series analysis. *NeuroImage* 7, S767.
- Wexler, B. E., Gottschalk, C. H., Fulbright, R. K., Prohovnik, I., Lacadie, C. M., Rounsaville, B. J., and Gore, J. C. (2001). Functional Magnetic Resonance Imaging of Cocaine Craving. *American Journal of Psychiatry* 158, 86-95.
- Worsley, K. J., Cao, J., Paus, T., Petrides, M., and Evans, A. C. (1998). Applications of random field theory to functional connectivity. *Hum Brain Mapp* 6, 364-367.
- Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited--again. *Neuroimage* 2, 173-181.
- Yoshiura, T., Zhong, J., Shibata, D. K., Kwok, W. E., Shrier, D. A., and Numaguchi, Y. (1999). Functional MRI study of auditory and visual oddball tasks. *Neuroreport* 10, 1683-1688.
- Zarahn, E., Aguirre, G., and D'Esposito, M. (1997). A trial-based experimental design for fMRI. *Neuroimage* 6, 122-138.
- Zubieta, J., Lombardi, U., Minoshima, S., Guthrie, S., Ni, L., Ohl, L. E., Koeppe, R. A., and Domino, E. F. (2001). Regional cerebral blood flow effects of nicotine in overnight abstinent smokers. *Biol Psychiatry* 49, 906-913.

Appendices

Appendix A: Abstracts of Publications

Chapter 2 is derived from the following article:

Murphy, K., Dixon, V., LaGrave, K., Kaufman, J., Risinger, R., Bloom, A., & Garavan, H. *A validation of event-related fMRI comparisons between drug groups and controls.* American Journal of Psychiatry (Under Review).

Non-invasive brain imaging techniques are powerful tools for researching the effects of drug abuse on brain activation measures. However, as many drugs have direct vascular effects, the validity of techniques dependent on blood flow measures, which purport to reflect neuronal activity, could be called into question. This may be of particular concern in event-related fMRI where current analytic techniques search for a specific shape in the haemodynamic response to neuronal activity. To investigate possible alteration in task-related activation, four groups performed a simple event-related fingertapping task: controls, cocaine, nicotine and THC (marijuana) users. It was found that activation measures, determined by two different analytic methods, did not differ between the groups. A comparison between an IV-saline and an IV-cocaine condition in cocaine addicts found a similar null result. Further in-depth analyses into the shape of the haemodynamic responses in each group also showed no differences. This demonstrates that it is possible to compare drug groups with controls using event-related fMRI without the need for any post-hoc correction procedures to correct for possible drug-induced cardiovascular alterations. Thus, fMRI activation differences reported between these drug groups can be more confidently interpreted as reflecting neuronal differences.

Chapter 3 is derived from the following article:

Murphy, K., & Garavan, H. *Deriving the optimal number of events for an event-related fMRI study based on the spatial extent of activation.* NeuroImage (Under Review).

Event-related fMRI is a powerful tool for localising psychological functions to specific brain areas. However, the number of events required to produce stable activation maps is a poorly investigated and understood problem. Huettel and McCarthy (2001: *Neuroreport* 12, 2411-2416) have shown that the spatial extent of activation increases monotonically with the number of events in an analysis. In the present paper, this result is replicated and shown to be a consequence of the cross-correlation technique used to determine active voxels and does not hold, for example, for a GLM analysis. Another analysis technique, that does not depend on goodness-of-fit to the data, is also proposed. This technique calculates an impulse response function (IRF) for each voxel, finds the best fitting haemodynamic shape to the IRF and returns an area-under-the-curve (%AUC) activation measure. Using spatial extent as a measure, both the GLM and %AUC analysis techniques display asymptotic behaviour after as few as 25 events in two fingertapping tasks: one with overlapping haemodynamic responses and one without. The experimental validity of the %AUC technique to identify active brain regions while minimising false positive levels is demonstrated in a group study with 25 participants.

Chapter 4 is derived from the following article:

Murphy, K. & Garavan, H. (2004). *An empirical investigation into the number of subject required for an event-related fMRI study.* NeuroImage 22 (2), 879-885

Optimising the number of subjects required for an event-related functional imaging study is critical for ensuring sufficient statistical power. We report an empirical investigation of this issue by employing a resampling approach to the data of 58 subjects drawn from 4 previous GO/NOGO studies. Using voxelwise measures and setting the activation map from the complete sample to be a "gold standard," analyses revealed the statistical power to be surprisingly low at typical sample sizes ($n = 20$). However, voxels that were significantly active from smaller samples tended to be true positives, that is, they were typically active in the gold standard map and correlated well with the gold standard activation measure. The numerous false negatives that resulted from the lower SNR of the smaller samples drove the poor statistical power of those samples. Splitting the sample into two groups provided a test of the reproducibility of activation maps that was assessed using an alternative measure that quantified the distances between centres-of-mass of activated areas. These analyses revealed that, although the voxelwise overlap may be poor, the locations of activated areas provide some optimism for studies with typical sample sizes. With $n = 20$ in each of two groups, it was found that the centres-of-mass for 80% of activated areas fell within 25 mm of each other. The reported analyses, by quantifying the spatial reproducibility for various sample sizes performing a typical event-related cognitive task, thus provide an empirical measure of the disparity to be expected in comparing activation maps.

Chapter 5 is derived from the following article:

Murphy, K. & Garavan, H. (2004). *Artifactual fMRI group and condition differences driven by performance confounds.* *NeuroImage* 21, 219-228.

Analysis techniques comparing groups or conditions that vary in performance are open to a possible confound driven by those performance differences, if these errors are ignored. Disproportionate numbers of errors may either introduce noise into the signal of interest or may confound the signal of interest with additional signal associated with specific error-related processes. Two inhibitory task datasets were reanalysed, one comparing young and elderly groups, the other comparing high and low conflict conditions within the same group of subjects. The data were analysed twice using event-related techniques, one treating correct and error responses separately, the other treating error responses as if they were correct. It was found that the activation maps differed considerably, with the inclusion of errors leading to many false positive and false negative activation clusters. Using performance as a covariate, ANCOVA analyses were used to try to correct these differences without success. Data simulations that varied the number of errors included in the analyses found that surprisingly few errors could significantly alter activation maps. Consequently, brain imaging investigations that do not accommodate error contributions to functional signals are at risk of misinterpreting activation patterns.

Burke, D., **Murphy, K.**, Garavan, H., & Reilly, R. (2004). *A Pattern Recognition Approach to the Detection of Single-Trial Event-Related fMRI*. Medical & Biological Engineering and Computing (In Press).

Functional Magnetic Resonance Imaging (fMRI) is an imaging technique for determining which regions of the brain are activated in response to a stimulus or event. Early fMRI experiment paradigms were based upon those used in Positron Emission Tomography (PET), i.e. employing a block design consisting of extended periods of “on” versus “off” activations. More recent experiments are based on event-related fMRI, harnessing the fact that very short stimuli trains or single events can generate robust responses. Functional MRI data suffers from low signal-to-noise ratios and typical event-related experiment paradigms employ selective averaging over many trials prior to using statistical methods for determining active brain regions. In this paper we report on a pattern recognition approach to the detection of single-trial fMRI responses without recourse to averaging and at modest field strengths (1.5 tesla). Linear Discriminant Analysis (LDA) is applied in conjunction with different feature extraction techniques. Employing the unprocessed data samples as features results in single-trial events being classified with an accuracy of $61.0 \pm 9.5\%$ over 5 subjects. To improve classification accuracy, knowledge of the ideal template haemodynamic response is used in the feature extraction stage. A novel application of parametric modelling yields an accuracy of $69.8 \pm 6.3\%$, while a matched filtering approach yields an accuracy of $71.9 \pm 5.4\%$. Single-trial detection of event-related fMRI may yield new ways of examining the brain by facilitating new adaptive experiment designs and enabling tight integration with other single-trial electrophysiological methods.

Fassbender, C., **Murphy, K.**, Foxe, J.J., Wylie, G., Javitt, D.C., Robertson, I.H., & Garavan, H. (2004). *A Topography of Executive Functions revealed by functional Magnetic Resonance Imaging*. Cognitive Brain Research 20(2), 132-143.

We used fMRI to study the brain processes involved in the executive control of behavior. The Sustained Attention to Response Task (SART), which allows unpredictable and predictable NOGO events to be contrasted, was imaged using a mixed (block and event-related) fMRI design to examine tonic and phasic processes involved in response inhibition, error detection, conflict monitoring and sustained attention. A network of regions, including right ventral prefrontal cortex (PFC), left dorsolateral PFC (DLPFC) and right inferior parietal cortex, was activated for successful unpredictable inhibitions, while rostral anterior cingulate was implicated in error processing and the pre-SMA in conflict monitoring. Furthermore, the pattern of correlations between left dorsolateral PFC, implicated in task-set maintenance, and the pre-SMA were indicative of a tight coupling between prefrontally mediated control and conflict levels monitored more posteriorly. The results reveal that the executive control of behavior can be separated into distinct functions performed by discrete cortical regions.

Hester, R. **Murphy, K.** & Garavan, H. (2004). *Beyond Common Resources: The Cortical Basis for Resolving Task Interference*. *NeuroImage* 23(1), 202-212.

Recent studies have suggested that declining inhibitory control observed during simultaneous increases in working memory (WM) demands may be due to sharing common neural resources, although it is relatively unclear how these processes are successfully combined at a neural level. Event-related functional MRI was used to examine task performance that required inhibition of varying numbers of items held in WM. Common activation regions for WM and inhibition were observed and this functional overlap may constitute the cortical basis for task interference. However, maintaining successful inhibitory control under increasing WM demands tended not to increase activation in these overlapping regions as might be expected if these common areas reflect common resources essential for task performance. Instead, increased activation was observed predominantly in unique, inhibition-specific regions including dorsolateral prefrontal cortex. The finding that successfully maintaining weaker stimulus-response relationships in the face of competition from stronger, prepotent responses requires greater activity in these regions reveals the means by which the brain resolves task interference and supports theories of how top-down attentional control is implemented.

Hester, R., **Murphy, K.**, Foxe, D.M., Foxe, J., & Garavan, H. (2003). *Predicting Success: The effect of pre-target cueing on inhibition performance*. Journal of Cognitive Neuroscience 16(5), 776-785.

The present study investigated the relationships between attention and other preparatory processes prior to a response inhibition task and the processes involved in the inhibition itself. To achieve this, a mixed fMRI design was employed to identify the functional areas activated during both inhibition decision events and the block of trials following a visual cue introduced 2 to 7 sec prior (cue period). Preparing for successful performance produced increases in activation for both the cue period and the inhibition itself in the frontoparietal cortical network. Furthermore, preparation produced activation decreases in midline areas (insula and medial prefrontal) argued to be responsible for monitoring internal emotional states, and these cue period deactivations alone predicted subsequent success or failure. The results suggest that when cues are provided to signify the imminent requirement for behavioral control, successful performance results from a coordinated pattern of preparatory activation in task-relevant areas and deactivation of task-irrelevant ones.

Kübler, A., **Murphy, K.**, Kaufman, J., Stein, E.A., & Garavan, H. (2003). *Coordination within and between verbal and visuospatial working memory: Network modulation and anterior frontal recruitment*. *NeuroImage* 20, 1298-1308.

Attention switching between items being stored and manipulated in working memory (WM) is proposed to be an elementary executive function. Experiment 1 reveals a similar attentional limitation within and between verbal and visuospatial WM and identifies a supramodal switching process required for switching between WM items. By using functional magnetic resonance imaging, Experiment 2 investigated brain activation correlates of parametrically varied attention switching within and between these two WM modalities. Attention switching activation was broadly distributed, was quite similar across the three conditions, and, in almost all areas, increased with increasing switching demand, indicating that attention switching recruits and modulates the entire WM network. Dorsolateral prefrontal cortex was implicated in both within- and between-modality attention switching, but no significant activation was found in ventrolateral areas, supporting dorsal-ventral process models of prefrontal organization. A functional dissociation between anterior frontal and dorsolateral prefrontal cortex was found with the former being more activated when switching attention between modalities was required. The data challenge the notion of an anatomically separate attention switching executive function, but suggest that anterior frontal areas are recruited for the additional demand of coordinating the verbal and visuospatial WM slave systems.

Garavan, H., Ross, T. J., **Murphy, K.**, Roche, R. A. P., & Stein, E. A. (2002). *Dissociable executive functions in the dynamic control of behaviour: Inhibition, error detection and correction*. *NeuroImage* 17, 1820-1829.

The present study employed event-related fMRI and EEG to investigate the biological basis of the cognitive control of behavior. Using a GO/NOGO task optimized to produce response inhibitions, frequent commission errors, and the opportunity for subsequent behavioral correction, we identified distinct cortical areas associated with each of these specific executive processes. Two cortical systems, one involving right prefrontal and parietal areas and the second regions of the cingulate, underlay inhibitory control. The involvement of these two systems was predicated upon the difficulty or urgency of the inhibition and each was employed to different extents by high- and low-absent-minded subjects. Errors were associated with medial activation incorporating the anterior cingulate and pre-SMA while behavioral alteration subsequent to errors was associated with both the anterior cingulate and the left prefrontal cortex. Furthermore, the EEG data demonstrated that successful response inhibition depended upon the timely activation of cortical areas as predicted by race models of response selection. The results highlight how higher cognitive functions responsible for behavioral control can result from the dynamic interplay of distinct cortical systems.

Kelly, A.M.C., Hester, R., **Murphy, K.**, Javitt, D.C., Foxe, J.J. & Garavan, H. *Prefrontal-Subcortical Dissociations Underlying Inhibitory Control Revealed by Event-Related fMRI*. *European Journal of Neuroscience* 19, 3105-3112.

Using event-related fMRI, this study investigated the neural dynamics of response inhibition under fluctuating task demands. Fourteen participants performed a GO/NOGO task requiring inhibition of a prepotent motor response to NOGO events that occurred as part of either a Fast or Slow presentation stream of GO stimuli. We compared functional activations associated with correct withholds (Stops) required during the Fast presentation stream of stimuli to Stops required during the Slow presentation stream. A predominantly right hemispheric network was activated across conditions, consistent with previous studies. Furthermore, a functional dissociation of activations between conditions was observed. Slow Stops elicited additional activation in anterior dorsal and polar prefrontal cortex and left inferior parietal cortex. Fast Stops showed additional activation in a network that included right dorsolateral prefrontal cortex, insula and dorsal striatum. These results are discussed in terms of our understanding of the impact of preparation on the distributed network underlying response inhibition and the contribution of subcortical areas, such as the basal ganglia, to executive control processes.

Fassbender, C., **Murphy, K**, Hester, R., Foxe, J. J., Foxe, D. M., Javitt, D. C., & Garavan, H. *Prefrontal and midline interactions mediating behavioural control.*

The control of human behaviour is thought to involve interactions between top-down, attentional processes and stimulus-driven, bottom-up demands. To investigate this interaction, we utilised a GO/NO-GO task with cued and uncued inhibitory events and assessed the effect of cue-induced levels of top-down control on stimulus-driven response conflict. We report that on a within-subjects, trial-for-trial basis, high levels of top-down control, as indexed by left dorsolateral prefrontal activation prior to the NO-GO, resulted in lower levels of activation on the NO-GO trial in midline areas. Furthermore, the midline effects were particular to the pre-supplementary motor area and statistically dissociable from the anterior cingulate. These results are consistent with a prefrontally-mediated reduction in response conflict and suggest a tight coupling between prefrontal and midline regions in the implementation of cognitive control.

Fassbender, C., **Murphy, K.**, Hester, R., Meaney, J., Robertson, I. H., & Garavan, H. *The role of right fronto-parietal cortex in cognitive control: Common activations for "cues-to-action" and response inhibition.*

Seemingly diverse cognitive tasks often activate similar anatomical networks. For example, right fronto-parietal cortex is active across a wide variety of paradigms suggesting that these regions may subserve a general cognitive function. We utilized fMRI and a GO/NOGO task with intermittent unpredictable "cues-to-action" in order to investigate areas involved in inhibition of a prepotent response and top-down attentional control. A similar network of right dorsolateral prefrontal and inferior parietal regions were active for cues and inhibitions. Inhibitions in the un-cued condition also activated additional right fronto-parietal regions. These results suggest that this network is involved in general cognitive control processes through the allocation of top-down attentional resources.

Kübler, A., **Murphy, K.**, & Garavan, H. *Effects of cocaine use on attentional control within and between verbal and visuospatial working memory.*

Many studies have shown the negative effects of cocaine on neuropsychological and cognitive performance in drug dependent individuals but little is known about the underlying neuroanatomy of these dysfunctions. The present study addressed attention switching between items held in working memory with a task in which subjects were required to store and update two items held in verbal or visuospatial working memory. Attention switching frequency varied between trials, thereby allowing us to isolate the switching component of task performance. Behavioural data revealed that cocaine addicts were specifically impaired in visuospatial working memory. With functional magnetic resonance imaging, we identified attenuated responses in prefrontal and cingulate cortices and in the dorsal striatum while other areas such as dorsolateral prefrontal cortex did not differ between healthy controls and users. Group differences in activation were exacerbated when performance differences were not taken into account underlying the importance of addressing performance confounds in interpreting brain activation measures. The results reveal that addiction may be accompanied by specific rather than ubiquitous hypoactivation in prefrontal and subcortical areas and suggest a compromised ability in users to control their attention to their thoughts as might be particularly relevant when required to switch away from drug-related thoughts, and thus the dysfunction in attention switching may contribute to the maintenance of addiction.

Garavan, H., Fassbender, C., **Murphy, K.**, Sweet, L. H., Rao, S., & Stein, E. A. *Quantitative Age-Related Differences In Functional Recruitment In A Working Memory Task.*

The greater activation patterns typically observed in older participants in performing cognitive tasks, which have been interpreted to reflect compensatory recruitment, may result in fewer additional resources being available to these participants as a task increases in difficulty. Participants in the young adult age range (<65, n=28) completed a verbal working memory N-back task, in which memory load can be parametrically manipulated, during imaging. Results revealed that greater activity amongst the older participants held for only the easiest level of task difficulty. Furthermore, age was negatively correlated with the extent of functional recruitment in the majority of the brain regions that were identified as showing an N-back load dependency. The results challenge simple accounts of age-related activation increases and suggest that age-related functional recruitment may limit subsequent recruitment in response to additional task challenges.

Garavan, H., Fassbender, C., **Murphy, K**, Sweet, L. H., Rao, S., & Stein, E. A. *Quantitative Age-Related Differences In Functional Recruitment In A Working Memory Task.*

The greater activation patterns typically observed in older participants in performing cognitive tasks, which have been interpreted to reflect compensatory recruitment, may result in fewer additional resources being available to these participants as a task increases in difficulty. Participants in the young adult age range (<65, n=28) completed a verbal working memory N-back task, in which memory load can be parametrically manipulated, during imaging. Results revealed that greater activity amongst the older participants held for only the easiest level of task difficulty. Furthermore, age was negatively correlated with the extent of functional recruitment in the majority of the brain regions that were identified as showing an N-back load dependency. The results challenge simple accounts of age-related activation increases and suggest that age-related functional recruitment may limit subsequent recruitment in response to additional task challenges.