

SOME PROBLEMS IN THE STATISTICAL ANALYSIS  
OF NEUROIMAGES

by

Alan Graham McCormack

Thesis submitted for the degree  
of Doctor of Philosophy  
in the University of Glasgow

September 1990

ProQuest Number: 13834283

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13834283

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

Thesis  
8741  
Copy 2



## ACKNOWLEDGEMENTS

I am indebted to my supervisors, Dr. I. Ford, at the department of Statistics and Professor J. McCulloch at the Wellcome Surgical Institute, University of Glasgow for introducing me to the statistical problems in the analysis of neuroimages and for their help and encouragement throughout this period of research.

I am also grateful to Dr. E. T. MacKenzie at Lers - Synthelabo for providing useful discussion and data for the study.

I would also like to thank Elaine for her continual support and encouragement.

This work was supported by a SERC CASE research grant with the cooperation of Lers Synthelabo.

TABLE OF CONTENTS

	Page
SUMMARY	i
INTRODUCTION	1
1. TRACER KINETIC MODELLING OF MK - 801	3
1.0 Background	3
1.1 Introduction	4
1.2 Theory	11
1.3 Experimental methods and considerations	16
1.4 Results	23
1.5 Discussion	48
2. RANKING REGIONAL RESPONSES	51
2.0 Introduction	51
2.1 Ranking the components of a multivariate mean vector	53
2.2 Investigating the pattern of response	67
2.3 Summary of results	94
3. THE BLOOD FLOW - METABOLISM COUPLING PROBLEM	96
3.1 Introduction	96
3.2 Notation and the general structure of the problem	98
3.3 Repeated measures analysis of variance	100
3.4 An example	111
3.5 Simulation study	113
3.6 Results	115
3.7 The relative power of the procedures	132
3.8 Multiple comparison procedures in repeated measures designs	135
3.9 Conclusions from the simulation studies	157

4.	STATISTICAL ANALYSIS OF IMAGE DIFFERENCES	160
4.0	Introduction	160
4.1	Data and Methods	161
4.2	Results	169
4.3	Discussion	182
5.	CONCLUSIONS AND FURTHER WORK	185
	APPENDIX A:	189
	APPENDIX B:	197
	REFERENCES	207

## SUMMARY

This thesis discusses some of the problems which arise in the statistical analysis of neuroimages. In particular, we develop and evaluate statistical methods which, we hope, will provide greater and more reliable insight into the biological processes which are illustrated in images generated by positron emission tomography (PET), single photon emission computerized tomography (SPECT) and quantitative autoradiography.

In chapter one, a mathematical model is developed to characterise the kinetics of the drug MK - 801 in normal and ischaemic tissue and to explore the use of radiolabelled MK - 801 as an *in vivo* ligand for studying glutamenergic mechanisms. Using a three compartment model and assuming negligible dissociation from the specific receptor site, kinetic constants are found to be numerically identifiable in four of the nine brain regions in ischaemic tissue (frontal parietal cortex, frontal cortex, occipital cortex and striatum). Convergence to a unique set of parameters is not obtained for normal central nervous system tissue using this model. However, in all ischaemic and normal tissue a two compartment model can be fitted to the data. Thus, in pathological states in which extracellular concentrations of glutamate are elevated and levels of cerebral blood flow are reduced (e.g. ishaemia) during the period of measurement, it would appear that MK-801 has some potential as an *in vivo* ligand for imaging glutamate release.

In chapter two, we address the problems of ranking the response to a drug over a set of brain regions and comparing the patterns

of response between drugs or treatments. In the first instance a theoretical approach is taken for the case of ranking three brain regions. We aim to identify the covariance structure for maximising and minimising the probability of a correct ranking, assuming multivariate normality. For higher dimensions, the probability of correctly ranking the observation vector is investigated using the Bonferroni inequality. Due to the complex nature of the response vector, these theoretical approaches are seen to have severe limitations.

As an alternative approach, we have investigated, empirically, the performance of a simple measure to characterise the response to drug treatment over a large number of brain regions. In a simulation study, we establish that, within the set of covariance matrices studied, fairly reliable measures of association can be computed. Moreover, doubling the between animal variance component or increasing the within animal variability appeared to have little effect on the reliability of the derived rankings.

In chapter three, four univariate repeated measures ANOVA techniques (the traditional F-test, the Huynh-Feldt and Greenhouse-Geisser adjusted tests, and a conservative test based on the lower bound of Box's correction factor) are studied. This form of analysis is important in studying, for instance, the relationship between local cerebral blood flow and local cerebral glucose utilisation. A common feature of the data from these experiments is the high dimensionality of the observation vector on a, relatively speaking, small number of experimental units. Ten multiple comparison procedures are also studied within the

same framework. Four of these methods (the Tukey, Scheffe, Bonferroni and Sidak procedures) are constructed under the assumption that the covariance matrix displayed sphericity. Of the remaining procedures, two Scheffe-type pairwise intervals, based on the Greenhouse-Geisser and Huynh-Feldt correction factors, take account of departures from sphericity by adjusting the degrees of freedom of the ANOVA F distribution. Four other methods (Bonferroni, Sidak, Greenhouse-Geisser and Hunyh-Feldt adjusted intervals) are based on the specific estimated variance for each contrast rather than on a pooled estimate based on the assumption of sphericity.

In the hypothesis testing situation, the test based on the Huynh-Feldt correction factor has a significance level close to the nominal 5% level over a wide range of covariance matrices. However, if a conservative test is necessary, the Greenhouse-Geisser approach is preferable. Of the multiple comparison procedures, the Bonferroni and Sidak approaches, based on a specific error term for each contrast perform consistently well, giving joint confidence levels close to the nominal 95% level.

In chapter four, we examine the spatial patterns in 'control - minus - control' subtractions in the case of six phantom images using a NOVOSPECT scanner. The results indicate that the clusters of 'high noise' which are present in the differenced images may lead to difficulties in the interpretation of any apparent effects observed in images obtained in activation type studies.

In chapter five further work in the area of neuroimaging is

discussed. In particular, we emphasise the need for interval estimation in non-linear regression and the determination of the signal to noise ratio, in activation-type studies, which is required to have some assurance that any apparent effects in the difference image are not artefacts of the measurement process.

## INTRODUCTION

Quantitative autoradiography has found widespread use in neuroscience research most notably for localising radioisotopically labelled tracers in histological sections of the central nervous system. The sections are exposed to X-ray film for days to months, depending on the tracer employed, to yield images of a variety of neurobiological processes. Densitometric analysis of the images with reference to precalibrated standards can then be used to generate rigorous measurements of the dynamic biochemical parameter of interest within specified brain regions via an appropriate mathematical model.

There are two broad strategies for isotopic labelling of cerebral tissue for autoradiography; *in vitro* labelling (where the brain is removed after death and histological sections of the central nervous system incubated with the radioisotopic material) and *in vivo* labelling (where the tracer is introduced intravenously in life and taken up by the central nervous system in life; the animal is then sacrificed and histological sections prepared). Autoradiography with *in vitro* labelling has been widely employed to characterise the sites of chemical transmission between brain cells in normal and disease states in man and animals (Whitehouse, 1985), whilst autoradiography with *in vivo* labelling has been used for mapping dynamic events such as glucose utilisation (Sokoloff et al., 1977), cerebral blood flow (Sakurada et al., 1978), protein synthesis (Smith et al., 1984), blood brain barrier permeability (Blasberg et al., 1983) and

tissue pH (Kobatake et al., 1984).

The counterparts in man of in vivo autoradiography are positron emission tomography (PET) and single photon emission computerised tomography (SPECT). Briefly, with these technologies, subjects are injected with positron or photon emitting substances and placed in a scanner. The number of emissions, recorded by a ring of detectors surrounding the head, are then converted to yield an image of the dynamic process of interest, in a planar section of the brain, via a reconstruction algorithm.

One consequence of the development of PET, SPECT and quantitative autoradiography is that it has become increasingly easy for experimenters to observe complex data structures on each experimental unit. Due to the basic cost of experiments, a common feature of all these data structures is the high dimensionality of the observation vector measured on a relatively speaking, small number of experimental units. In these circumstances, the sample correlation matrix will be singular. This special nature of the data presents challenging problems in many aspects of multivariate analyses, particularly in the areas of selecting and ordering populations. The principle aim of this thesis is to develop and evaluate statistical methods which will provide greater and more reliable insight into the biological process which are illustrated in neuroimages.

## CHAPTER 1    TRACER KINETIC MODELLING OF MK-801

### 1.0    BACKGROUND

MK-801 is a noncompetitive glutamate N-methyl-D-aspartate (NMDA) receptor antagonist (Wong et al, 1986) with potent anti-ischaemic effects (Oyzurt et al, 1989). Prior to its development, a reduction in ischaemic damage had only been achieved with intracerebral administration of NMDA antagonists, due to the low blood-brain barrier (BBB) penetration of these agents. However, MK-801 is highly lipophilic and unlike competitive antagonists of the NMDA receptors, which act at the transmitter recognition site, MK-801 is a use dependent blockader which acts at a site solely related to the ion channel (at higher concentrations of MK-801 the ion channels will be open for a longer period of time). Furthermore, the onset of NMDA receptor blockade is more rapid in the presence of glutamate (a major excitatory neurotransmitter whose extracellular concentrations are markedly elevated in cerebral ischaemia).

In this chapter we aim to develop a mathematical model describing the uptake and retention of MK-801 in the rat brain to:

- 1) characterise the kinetics of the drug in normal and ischaemic tissue and to gain insight into the anti-ischaemic effect of the compound and
- 2) to explore the use of radiolabelled MK-801 as an in vivo ligand for studying glutamenergic mechanisms.

However before developing a kinetic model to describe the uptake and retention of MK-801 in the central nervous system certain biochemical properties of the drug should be known. In particular, knowledge of the metabolic degradation of the compound in blood plasma and brain tissue is essential. At 2 hours after the administration of [ $^3\text{H}$ ]MK-801 in the rat, Hucker et al (1983) reported that relatively high concentrations of labelled metabolites are present in the blood; however it would appear that these metabolites do not cross the BBB, and moreover that there is minimal metabolism of [ $^3\text{H}$ ]MK-801 by the brain itself, implying that virtually all radioactivity in the brain can be attributed to MK-801. Other concerns are that the tracer should act specifically with the receptor of interest and must either be structurally related to the natural substance or have similar transport properties.

## 1.1 INTRODUCTION

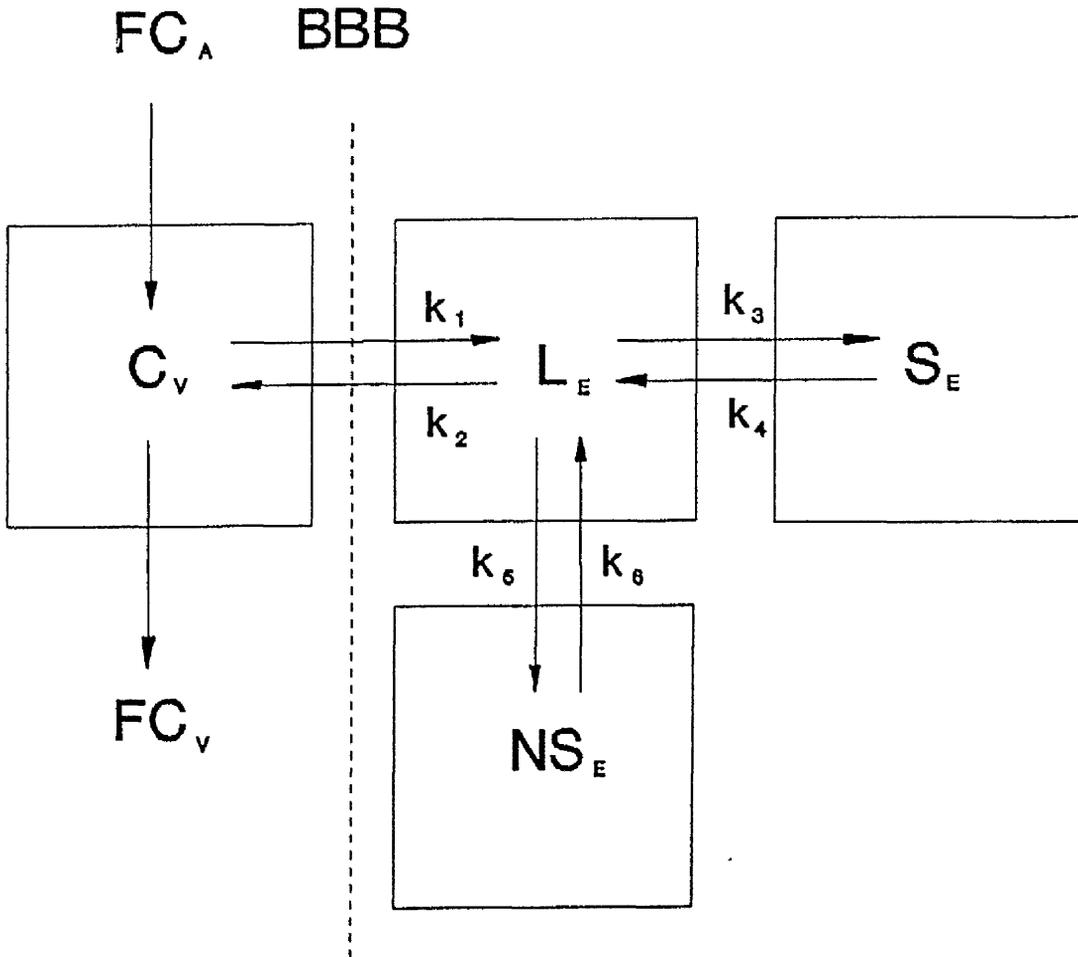
Recently many radiolabelled ligands have been synthesised which allow for in vivo studies of neurotransmitter systems using Positron Emission Tomography (PET), Single Photon Emission Computerised Tomography (SPECT) and Autoradiography. These studies involve imaging living human or baboon brains, as well as tissue dissection of animal brains at specific times after a bolus injection of radioligand.

To obtain a quantitative interpretation of the radioligand-neurotransmitter interactions, a variety of models have been proposed, including models for the dopaminergic (Mintun et al 1984; Wong et al, 1986a,b; Perlmutter et al, 1986; Farde et al,

1986; Logan et al, 1987), muscarinic cholinergic (Gibson et al, 1984; Frey et al, 1985) and opiate (Frost and Wagner, 1984) systems. The purpose of all these models is to determine the first order rate constants which are expected to reflect the binding characteristics of the radioligand. In addition, several techniques also estimate neuroreceptor densities and binding affinities which are more indicative of the regional amount of specific neuroreceptors and the rate of association of free ligand with the receptor.

All models fall into one of two categories (dynamic and equilibrium) and are based on an appropriate mathematical model (Figure 1.1) similar to the four rate constant model (Phelps et al, 1979; Huang et al, 1980) developed as an extension to the [<sup>14</sup>C]-2-deoxyglucose model for the measurement of local cerebral glucose utilisation (Sokoloff et al, 1977).

**Figure 1.1.** Compartmental model used in kinetic analysis of<sup>6</sup> ligand distribution in brain.



In the model,  $C_A$  represents the systematic arterial concentration of radioligand,  $C_v$  the capillary plasma concentration of ligand,  $L_E$ ,  $S_E$  and  $NS_E$  are tissue concentrations of tracer in the free, specifically bound and nonspecifically bound compartments,  $F$  represents the rate of local cerebral blood flow and  $k_1$ ,  $k_2$ ,  $k_3$ ,  $k_4$ ,  $k_5$  and  $k_6$  are transfer coefficients between compartments.

In this basic comprehensive model for analysing kinetic data for in vivo receptor ligand studies, the free ligand is delivered via the arterial blood and transported into an extravascular compartment by passive diffusion. Once extracted, the ligand can either be nonspecifically bound in the extracellular compartment or specifically bound to the receptor of interest or back-diffuse into the blood. In most models the nonspecific binding sites and the free ligand in tissue are considered to be in a common compartment, if it is assumed that nonspecific binding is rapidly reversible. Furthermore, the transfer coefficients between compartments are assumed to be first-order constants; first order input and elimination implies that the input and elimination rates are proportional respectively, to the amount of drug at the absorption site and the amount of drug in the compartment through which the drug is eliminated. Notice that  $k_3$ , the transfer coefficient describing the movement of free ligand into the specifically bound compartment (defined as the product of the rate constant for association of free ligand with the receptor,  $k_1$ , and the concentration of free receptors,  $R$ ) will only become a pseudo first-order rate constant under conditions where the tracer only occupies a negligible fraction of the available receptors.

The advantages of both the dynamic and equilibrium approaches have been discussed by Huang, Barrio and Phelps (1986). Briefly, the dynamic approach requires measurement of the time course of the radioactivity in both brain tissue and blood. Using an appropriate model these measurements can then be used to derive estimates of the rate constants for transport, binding and release of ligand using the normalised graphical method

(Patlak et al, 1983; Gjedde et al, 1985) or nonlinear regression techniques. The former method involves a transformation of the tissue-activity curve to estimate a number of the model parameters in Figure 1.1. With the dynamic approach, it is assumed that the radioligand is injected in tracer amounts and that the ligand occupies only a negligible fraction of the available receptors for  $k_3$ , to approximate a first order rate constant. One drawback from this approach is that  $k_3$  and  $R$  cannot be estimated separately.

With the equilibrium approach, multiple determinations of the radioactivity present in the tissue at different specific activities of the ligand are obtained to measure the fraction of bound ligand under various levels of unbound ligand concentrations. A main assumption for this particular method is that the binding and release of ligand at specific binding sites are in equilibrium during the time of measurement and the concentration of unbound ligand can be estimated. Consequently, this approach is suitable for ligands with fast release rates from the receptor of interest. Its main advantage over the dynamic approach is that estimates of the individual values of the association and dissociation constants of the receptor and receptor density can be established using such techniques as Scatchard and Hills plots. In this chapter we will only consider dynamic approaches.

Any model taking into account all the compartments in Figure 1.1 will certainly result in parameters which are not numerically identifiable from the analysis of typical "receptor binding" data. Therefore, to extract useful physiological information

from such data a number of simplifying assumptions must be made to arrive at a workable model. Consequently, a variety of models of differing complexity have been proposed for tomographic and dissection type studies. However, no matter how simple the model, it is worth noting that any estimated physiological parameters should not be interpreted without careful consideration to the many assumptions/approximations which have been made.

Mintun et al (1984) developed a three compartment model for use with PET that results in the quantification of physiological parameters using [ $^{18}\text{F}$ ] spiperone. The model involves a total of nine parameters; reduced to eight by assuming that the specific activity of the radioligand is sufficiently high that the injected ligand is present only in tracer quantities. Of the eight or nine model parameters, four are assigned values prior to data analysis. The remaining parameters are estimated in a two step analysis: two parameters being estimated from cerebellum tissue data, using the assumption that the cerebellum is an area of the brain devoid of specific spiperone binding sites; the others being estimated from the tissue data in the area of interest.

Wong, Gjedde and Wagner (1986a) presented an alternative dynamic model to describe the binding of  $^{14}\text{C}$ -labelled N-methylspiperone (NMSP) to the  $\text{D}_2$  dopamine receptor in the caudate nucleus. Based on a four compartment model a system of differential equations were derived employing the simplifying assumption that nonspecific binding is rapidly reversible in comparison with the rates of other processes in the model. Further requirements for

derivations of the operational equation were that the unimolecular dissociation rate was negligibly small and that the tracer was only present in tracer quantities throughout the period of study. Except for specific binding, an additional assumption was that the kinetic behaviour in the caudate nucleus and cerebellum were identical. After correcting the plasma for metabolites the parameters of interest were estimated using Patlak plots.

Logan et al (1987) derived a further set of ordinary differential equations to describe a kinetic model for neuroleptic tracers for applications with PET, assuming the quantity of radioligand is only present in trace amounts, nonspecific binding is rapidly reversible and the ratio of free ligand to nonspecifically bound ligand remains constant for the duration of the experiment. Estimates of the model parameters were obtained using nonlinear regression with the plasma tracer activities being described by a sum of four exponentials.

Frey et al (1985), to our knowledge, have provided the only kinetic model which has been successfully applied to data from the dissection of rat brains at different times after the injection of a radiotracer. Assuming nonspecific binding to brain is nonsaturable and occurs instantaneously compared to the rates of BBB transport and receptor binding, kinetic rate constants were obtained from six brain regions using a nonlinear least squares curve fitting procedure; estimates of the parameters were obtained after the rate of dissociation of the tracer from the specific binding site was assumed negligible.

Other authors have investigated a number of the simplifying assumptions which have been made to derive kinetic models of receptor-ligand binding. In a simulation study, based on exact analytical solutions of the relevant system of differential equations, Zeeberg and Wagner (1987) indicate that numerical errors can arise if the assumption of rapidly reversible binding to nonspecific receptors is invalid. Bahn et al (1989) have considered the suitability of different model configurations to describe the kinetics in caudate and cerebellum using linear and nonlinear models. These authors also assessed the effect of assuming a negligible dissociation rate from the specific binding sites: the parameters from the linear and nonlinear models were in better agreement if reversible binding is assumed.

## 1.2 THEORY

A general compartmental model defining tracer movement among compartments of the central nervous system is described in Figure 1.1. Tracer is considered to be present in three compartments in brain tissue corresponding to free, nonspecifically bound and specifically bound ligand. Denoting the concentrations of tracer in these compartments by LE, NSE and SE respectively, the total tissue concentration of tracer, CT, is related to its constituents by the equation

$$CT = LE + NSE + SE \quad 1.1$$

Before defining the system of differential equations relating MK-801 movement between compartments in Figure 1.1 it is first necessary to derive an equation relating the capillary blood

concentration or venous outflow, CV, to the systematic arterial concentration, CA. According to the Fick principle (Fick, 1870) the rate of accumulation of an inert diffusible tracer in tissue over a defined time interval of length T can be expressed mathematically as

$$\frac{dC_i}{dt} = F(CA - CV) \quad 1.2$$

where  $C_i$  is the concentration of the diffusible tracer in tissue and F represents blood flow.

Furthermore, Kety (1951) defined the relationship between the concentration of tracer in a homogeneous tissue and its venous outflow by the equation

$$CA - CV = m(CA - C_i/\lambda) \quad 1.3$$

where  $\lambda$  is the tissue:blood partition coefficient for the tracer and m is a constant between 0 and 1 and represents the extent to which diffusion equilibrium between blood and tissue is achieved during passage from the arterial to the venous end of the capillary.

By combining equations 1.2 and 1.3, the rate of accumulation of tracer can be rewritten as

$$\frac{dC_i}{dt} = K(\lambda CA - C_i) \quad 1.4$$

where  $K = mF/\lambda$ .

Since MK-801 is highly lipophilic it seems reasonable that the ratio of permeability-surface area product to flow is high enough to assume  $m$  approaches 1. Equation 1.3 therefore implies that the concentration of tracer in venous outflow is proportional to the tracer in the tissue in the absence of any binding to receptors.

If tracer binds to receptors in the tissue and it is subsequently trapped, at any instant in time the concentration of tracer in the tissue free to partition is equal to  $LE$ . Equation 1.4 can thus be rewritten for unmetabolised freely diffusible tracers in the presence of binding to be

$$\frac{dC_i}{dt} = K(\lambda CA - LE) \quad 1.5$$

Assuming that passive diffusion across the blood-brain barrier is bidirectionally symmetrical (i.e.  $k_1 = k_2 = K$ ), the kinetics describing tracer movement between the compartments can be defined by the following system of differential equations:

$$\frac{dLE}{dt} = K\lambda CA + k_4SE + k_6NSE - (K + k_3 + k_5)LE \quad 1.6$$

$$\frac{dSE}{dt} = k_3LE - k_4SE \quad 1.7$$

$$\frac{dNSE}{dt} = k_5LE - k_6NSE \quad 1.8$$

where  $K$ ,  $k_3$ ,  $k_4$ ,  $k_5$  and  $k_6$  are transfer coefficients between compartments in the model.

Assuming rapid equilibrium between the concentration of nonspecifically bound and free tracer, equation 1.8 can be

approximated by

$$\frac{dNSE}{dt} = k_5LE - k_6NSE = \alpha \frac{dLE}{dt} \quad 1.9$$

where  $\alpha = k_5/k_6$ .

Substitution of the above relationship into equation 1.6 implies that

$$(1+\alpha) \frac{dLE}{dt} = K\lambda CA - KLE - k_3LE + k_4SE \quad 1.10$$

and

$$CT = (1+\alpha)LE + SE \quad 1.11$$

The resulting system of equations 1.7, 1.10 and 1.11 can be solved for CT using Laplace transform techniques (Godfrey, 1983). Assuming there is no initial concentration of tracer in the brain, the Laplace transforms of these equations after a unit impulse function are given by:

$$SE(s) = k_3LE(s)/(s + k_4) \quad 1.12$$

$$LE(s) = [K\lambda + k_4SE(s)]/((1 + \alpha)s + k_3 + K) \quad 1.13$$

and 
$$CT(s) = (1 + \alpha)LE(s) + SE(s) \quad 1.14$$

After a little algebra, the elimination of LE(s) and SE(s) in equation 1.14 yields

$$CT(s) = [(s+k_4)+k_3/(1+\alpha)]K\lambda / [s^2+s((K+k_3)/(1+\alpha)+k_4)+Kk_4/(1+\alpha)] \quad 1.15$$

Taking the inverse Laplace transform of equation 1.15 and expanding by partial fractions, the time course of tracer after

a unit impulse in arterial plasma is given by

$$CT = K\lambda [(k_4+k_3/(1+\alpha)-x_1)e^{-x_1 t} + (x_2-k_4-k_3/(1+\alpha))e^{-x_2 t}] / (x_2-x_1) \quad 1.16$$

where  $t$  represents time and

$$x_1 = \{ [(K+k_3)/(1+\alpha)+k_4] + [((K+k_3)/(1+\alpha)+k_4)^2 - 4Kk_4/(1+\alpha)]^{1/2} \} / 2$$

$$x_2 = \{ [(K+k_3)/(1+\alpha)+k_4] - [((K+k_3)/(1+\alpha)+k_4)^2 - 4Kk_4/(1+\alpha)]^{1/2} \} / 2$$

Now under a general input function the tissue response at time  $T$  is defined by

$$CT = \frac{k\lambda}{(x_2-x_1)} [(k_4 + \frac{k_3}{(1+\alpha)} - x_1)e^{-x_1 t} + (x_2 - k_4 - \frac{k_3}{(1+\alpha)})e^{-x_2 t}] * CA \quad 1.17$$

where  $*$  denotes the operation of convolution.

Finally, approximating the time course of plasma after a bolus injection of MK-801 by a sum of  $n$  decaying exponentials, the parameters of interest are identifiable from the fitted response curve of the form:

$$\begin{aligned} CT &= \sum_{m=1}^2 (A_{2m-1}) e^{-(A_{2m})t} * \sum_n (B_{2n-1}) e^{-(B_{2n})t} \\ &= \int_0^t \sum_{m=1}^2 \sum_n (A_{2m-1}) (B_{2n-1}) e^{-(A_{2m}(t-u) + B_{2n}u)} du \\ &= \sum_{m=1}^2 \sum_n \frac{(A_{2m-1}) (B_{2n-1})}{((A_{2m}) - (B_{2n}))} [e^{-(B_{2n})t} - e^{-(A_{2m})t}] \end{aligned} \quad 1.18$$

where  $A_1 = K\lambda (k_4 + k_3/(1 + \alpha) - x_1)/(x_2 - x_1)$ ,

$A_2 = x_1$ ,

$A_3 = K\lambda (x_2 - k_4 - k_3/(1 + \alpha))/(x_2 - x_1)$ ,

$A_4 = x_2$

and  $A_1$ ,  $A_2$ ,  $A_3$  and  $A_4$  are constrained to be greater than or equal to 0.

### 1.3 EXPERIMENTAL METHODS AND CONSIDERATIONS

#### 1.3.1 BLOOD: BRAIN PARTITION COEFFICIENT

To permit the deterministic identifiability of the kinetic constants in the above model knowledge of the blood:brain partition coefficient ( $\lambda$ ) of MK-801 in the rat is essential. For the purpose of this investigation,  $\lambda$  was estimated experimentally in the rat using in vitro techniques and was considered to be constant over all brain regions.

In the experiments to determine the blood:brain partition coefficient for MK-801, 4ml of blood was withdrawn from a Sprague-Dawley rat and mixed with heparin, 10  $\mu$ Ci of [ $^3$ H]MK-801 and 0.2ml of unlabelled MK-801 (nonradioactive MK-801). After vortexing the sample every 15 minutes for 1 hour, samples of whole blood and plasma (removed after centrifugation) were counted in a liquid scintillation counter to assess the amount of radioactivity present. To assess radioactivity in the tissue, sections of the cerebellum (a brain region assumed to contain no NMDA receptors) were incubated with plasma at 4°C for 24 hours. After washing in distilled water

and drying in a stream of cold air, the sections were opposed to Amersham 3H Hyperfilm and left for 14 days. The developed film was then analyzed on a Quantimet 970 image analyser: optical densities being converted to nCi/g using a reference set of prepared standards.

An estimate of the partition coefficient of 4.0 was then derived by dividing the quantified concentration of radioactivity in the cerebellum by the concentration of radioactivity in the whole blood.

### 1.3.2 FREELY DIFFUSIBLE PROPERTY OF MK-801

To derive equation 1.5 in section 1.2 we have assumed that [<sup>3</sup>H]MK-801 is a freely diffusible tracer. To investigate this property, "local cerebral blood flow (LCBF) values" were assessed for [<sup>3</sup>H]MK-801 and compared to these elicited in the same rat by [<sup>14</sup>C] iodoantipyrine (a known freely diffusible compound). For each tracer, LCBF values were determined using a previously published blood flow model (Sakurada et al, 1978) based on the uptake of a freely diffusible tracer and relating the concentration of the radiotracer in tissue to the time course history of the tracer in plasma (blood:brain partition coefficients of 0.79 and 4.0 were used in the operational equations for [<sup>14</sup>C] iodoantipyrine and [<sup>3</sup>H]MK-801 respectively).

In this particular study, each of 6 anaesthetised Sprague-Dawley rats was infused intravenously with 40 $\mu$ Ci of [<sup>3</sup>H]MK-801 and 20 $\mu$ Ci of [<sup>14</sup>C] iodoantipyrine and over the succeeding 30 seconds 14 timed samples of arterial blood were collected in preweighed filter discs.

On collection of the last sample, the discs were immediately reweighed and the concentration of both [<sup>3</sup>H]MK-801 and [<sup>14</sup>C] iodoantipyrine determined by liquid scintillation counting.

At 30 seconds after the start of the infusion, the animal was decapitated and the brain rapidly removed. Samples of tissue corresponding to the cerebellum (left and right), medulla, pons, cortex (left and right) and spinal cord were then dissected from the brain, weighed and then left to dissolve overnight in ecosint. Tissue concentrations of both [<sup>3</sup>H]MK-801 and [<sup>14</sup>C] iodoantipyrine were determined the following day by liquid scintillation counting.

The results of the investigation are displayed in Table 1.1. Logged LCBF values have been used since the variance of the LCBF value tends to increase with its expectation (McCulloch et al, 1982). Although mean values of LCBF generated using the MK-801 data tended to be lower than those of iodoantipyrine (IAP), in all regions investigated the difference was not statistically significant using a one sample t-test with a nominal significance level of 0.05, implying that MK-801 enters the central nervous system with no or negligible diffusion limitations.

To determine whether the blood flow model was robust with respect to the blood:brain partition coefficient ( $\lambda$ ), several values of  $\lambda$  were selected for use with the MK-801 data. With values of  $\lambda$  greater than 1, in no circumstance were there any appreciable increases in the observed rate of LCBF.

Table 1.1 Comparison of blood flow values

Region	n	raw data		logged data difference (s.e.)	p value
		MK-801 mean	IAP mean		
cerebellum left	6	79.2	86.0	-0.126 (0.093)	0.23
cerebellum right	6	79.5	86.3	-0.133 (0.095)	0.22
medulla	6	81.0	91.3	-0.144 (0.102)	0.22
pons	6	86.5	96.3	-0.139 (0.092)	0.19
cortex left	6	129.7	136.7	-0.075 (0.088)	0.43
cortex right	6	127.7	140.2	-0.088 (0.068)	0.25
spinal cord	6	76.3	68.8	0.040 (0.132)	0.78

### 1.3.3 DETERMINATION OF [<sup>3</sup>H]MK-801 IN TISSUE AND PLASMA

To assess the uptake and retention of MK-801 in the normal and ischaemic rat brain Sprague-Dawley rats were anaesthetised with halothane/nitrous oxide anaesthesia and mechanically ventilated via a tracheostomy. All animals were monitored for temperature, arterial blood gases and mean arterial pressure. Fifteen animals further underwent left middle arterial artery occlusion to induce ischaemia. Following an intravenous bolus injection of [<sup>3</sup>H]MK-801 (20 $\mu$ Ci per rat) timed arterial blood samples were obtained 0.25, 0.5, 0.75, 1, 2, 3, 5, 7, 5, 10, 15, 30, 45, 60, 75, 90, 105 and 120 minutes after injection, centrifuged immediately and samples of plasma removed. The ligand concentration in each of the plasma samples was then assessed by liquid scintillation counting.

To generate time-activity curves for individual brain regions, 3 normal and ischaemic rats were decapitated at 5, 15, 30, 60 and 120

minutes after [<sup>3</sup>H]MK-801 injection. Brains were removed rapidly and dissected into subdivisions corresponding to the frontal parietal cortex, frontal cortex, hippocampus, naccumbens, occipital cortex, olfactory bulb, parietal cortex, striatum and cerebellum. With the exception of the cerebellum, duplicate samples of tissue were removed from normal animals. Only tissue in the left hand side of the brain was removed from ischaemic animals. After dissection, the tissue samples were weighed and then left to solubilise overnight. Samples were analysed the following day for <sup>3</sup>H content by liquid scintillation counting.

#### 1.3.4 CORRECTION OF PLASMA SAMPLES FOR METABOLITES

As mentioned previously, Hucker et al (1983) reported that relatively high concentrations of labelled metabolites of [<sup>3</sup>H]MK-801 will be present in the blood after 2 hours. Arnett et al (1985) have suggested that the percentage of activity due to unmetabolised drug can be adequately explained by a sum of two exponentials. Although Hucker et al (1983) do not provide enough data to apply this procedure for [<sup>3</sup>H]MK-801, in the first six hours after administration of the tracer the authors estimated that the half-life of unmetabolised drug in plasma was approximately 86 minutes.

To correct the plasma samples in the experiments at the Wellcome Surgical Institute, we have assumed that the percentage of activity due to the unmetabolised drug at time  $t$  ( $\%Ca(t)$ ) can be represented by a function of the form  $\%Ca(t) = \theta^t$  where  $\theta$  is a parameter which has to be estimated from the data.

By plotting the logged mean plasma activity against time and assuming that the half-life of unmetabolised tracer was 86 minutes, it was estimated, using only the time points after 10 minutes, that 50% of the radioactivity detected in plasma was due to metabolites at 90 minutes, implying a value of  $\theta$  of 0.9923. All plasma samples were then corrected by taking the product of the plasma tracer activity and the percentage of this activity due to unmetabolised drug.

### 1.3.5 NORMALISATION OF PLASMA SAMPLES

Although great care is taken to administer the same dose of radiotracer to each animal, inevitably slight differences in dosage will occur. Depending on the specific activity of the radiotracer, these differences could have a great effect on the noise in the data and subsequently on the number of exponentials which can realistically be fitted to the data. To reduce the amount of noise in the system, due to slight differences in dosage, the activity in the tissue and plasma - measured in disintegrations per minute (dpm)/mg and dpm/ml respectively - were normalised by dividing the number of counts in each tissue/plasma sample by the area under the first two minutes of the time-plasma activity curve (two minute integral). The two minute integral was chosen since data was available for all animals up to this time point.

### 1.3.6 FITTING EXPONENTIALS

Fitting multiexponential models to experimental data has been used often enough in biomedical research that a number of practical

problems are well recognised. One such problem which often arises is the determination of the number of exponentials which can realistically be fitted to the experimental data. Consider the case where the arterial input function at time  $t_i$ ,  $C(t_i)$ , can be represented by the relationship

$$Ca(t_i) = \sum_{j=1}^n A_j e^{-\lambda_j t_i} + \varepsilon_i$$

where  $A_j$  ( $j=1, \dots, n$ ) are positive,  $\lambda_j > \lambda_{j+1} \geq 0$  and  $\varepsilon_i \sim N(0, \sigma^2)$  and denotes the noise in the system. Godfrey (1983) suggests that some guidance on the order of the model can be obtained by using a statistical test based on the F-ratio, which is given by

$$F = \frac{(SSQ_{n-1} - SSQ_n) / (df_n - df_{n-1})}{SSQ_n / df_n} \quad (1.19)$$

where  $SSQ_n$  and  $SSQ_{n-1}$  are the residual sum of squares of the models of order  $n$  and  $n-1$ ; and  $df_n$  and  $df_{n-1}$  are the corresponding numbers of degrees of freedom. If the error variance is not constant for all  $t_i$ , ie  $\text{var}(\varepsilon_i) = \sigma_i^2 = v_i \sigma^2$ , then weights proportional to  $1/v_i$  are calculated and the weighted sum of squared differences between the observed  $Ca(t_i)$  and the model predictions is used to replace  $SSQ$  in (1.19) above. If multiple observations are available at each  $t_i$ , then typically data points will be weighted according to the reciprocal of the variance at that time point.

However, there are statistical problems with the use of the F-ratio, a special case of the likelihood ratio test, in such situations. Suppose, for example, we want to test  $H_0: A_j=0$  against  $H_1: A_j \neq 0$ . The difficulty of using the F-ratio arises since the regularity

conditions used to obtain the asymptotic properties of the least squares estimator and test statistic are violated (Gallant, 1987). An additional problem arises, at least in this application, since the A's will be constrained to be positive.

Glass and Garreta (1967, 1971) have used simulation techniques to assess the ability of the Marquart (1964) and the Newton-Raphson (Jennrich and Sampson, 1968) techniques to fit exponentials to biological data. Factors which were found to affect the accuracy with which amplitudes and exponents could be estimated included the measurement accuracy, the magnitude of the noise in the system, the sampling frequency and the time range over which samples are taken (measurement should begin early enough and at short intervals for a fast transient and should continue long enough to identify a slow transient).

Other methods for fitting data to multiexponentials include the Prony and modified Prony algorithms (Osborne, 1975). However, these methods are appropriate only for data equally spaced in time and are thus seldom used for physiological data.

## 1.4 RESULTS

### 1.4.1 PLASMA CLEARANCE CURVE

The estimated plasma decay curve obtained after a bolus intravenous injection of [<sup>3</sup>H]MK-801 is given by the equation

$$Ca(t) = 5410e^{-4.04t} + 1090e^{-0.670t} + 176e^{-0.00818t} \quad (1.20)$$

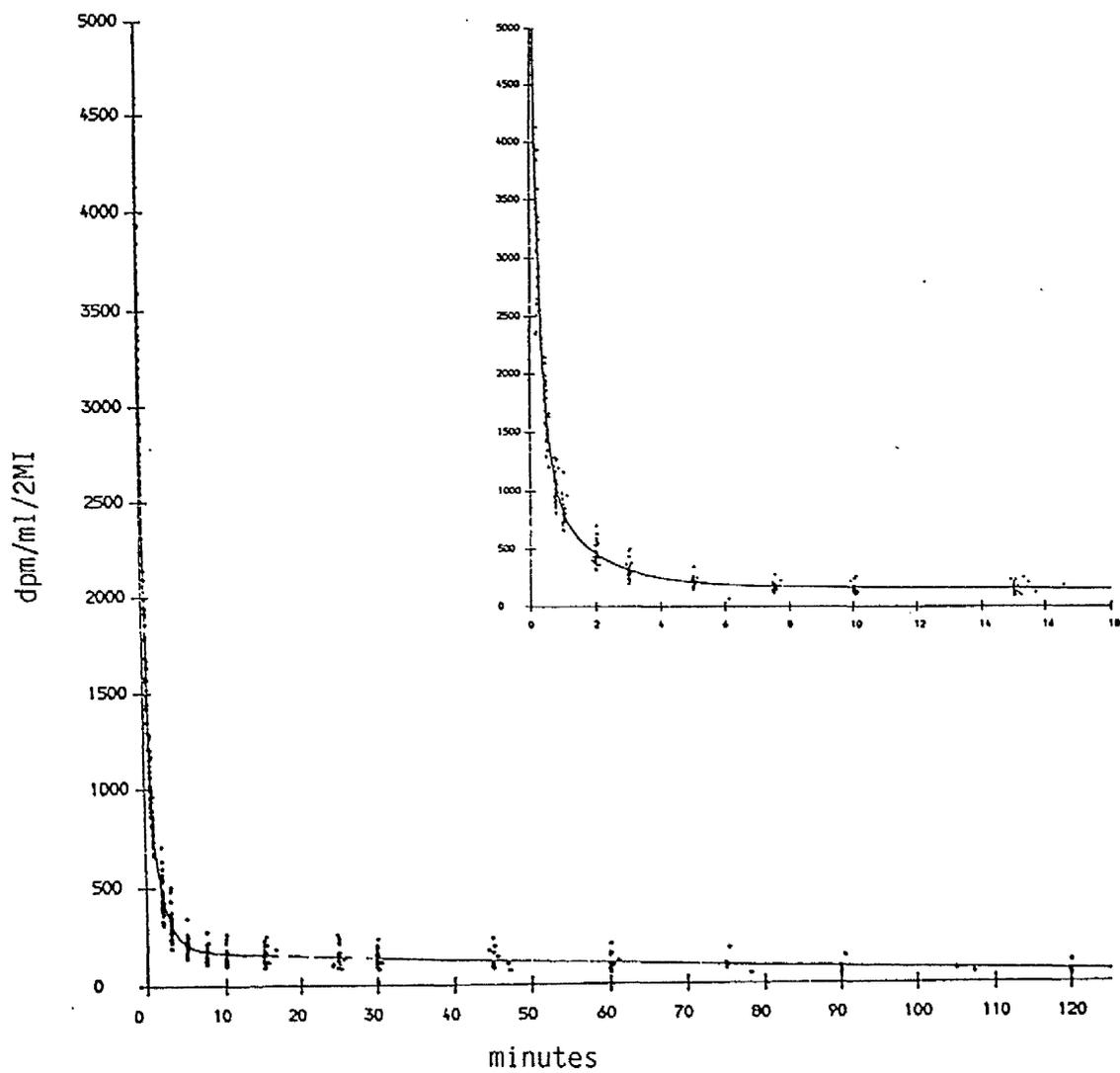
Data were corrected for metabolites and normalised using the two

minute integral. The six parameters were identified using a standard BMDP (1987) nonlinear, weighted least squares, curve fitting algorithm (weights were set to the reciprocal of the estimated variance at each time point). A plot of the data with the 'best fit' curve is given in Figure 1.2.

The inclusion of a fourth component in the plasma decay curve did not produce a significant improvement in the fit to the data (the parameter estimate of the fourth exponential was very small and would require experimental data over a longer time period to be determined accurately). Also the slowest component in (1.20) represents a half-life of approximately 86 minutes. This is in good agreement with the estimate of 85 minutes from the analysis of data by Hucker et al (1983).

To determine the effect of weighting and normalisation on the limitations and the quality of the fit to the data, plasma clearance curves were estimated using unnormalised data and/or equal, weighting in the least squares curve-fitting procedure. The estimates of the parameters  $B_1, B_2, \dots, B_6$  in equation (1.18) and the corresponding coefficients of variation are given in Table 1.2a for the unnormalised data and Table 1.2b for the normalised data. In Table 1.2b the estimates of the parameters  $B_1, B_2$  and  $B_3$  are far less than those in Table 1.2a since the data have been normalised. However, it would be expected that they will be roughly proportional to the corresponding parameter estimates using the unnormalised data.

Figure 1.2 Plasma clearance curve for MK-801



2MI = two minute integral

**Table 1.2a** Estimates of the arterial plasma curve (unnormalised data)

Parameter	unweighted LS		weighted LS	
	Estimate	Coefficient of variation	Estimate	Coefficient of variation
$B_1$	333000	0.074	347000	0.176
$B_2$	2.96	0.200	3.78	0.210
$B_3$	49600	0.611	77900	0.249
$B_4$	0.450	0.684	0.642	0.150
$B_5$	11000	0.443	12000	0.038
$B_6$	0.00651	1.913	0.00826	0.106

**Table 1.2b** Estimates of the arterial plasma curve (normalised data)

Parameter	unweighted LS		weighted LS	
	Estimate	Coefficient of variation	Estimate	Coefficient of variation
$B_1$	5060	0.036	5410	0.095
$B_2$	3.20	0.092	4.04	0.099
$B_3$	709	0.285	1090	0.118
$B_4$	0.484	0.326	0.670	0.087
$B_5$	164	0.193	176	0.036
$B_6$	0.00608	0.896	0.00818	0.108

Tables 1.2a,b suggest that normalising the data by dividing through by the two minute integral tends to produce more reliable estimates of the parameters of interest. Secondly, although there is a slight increase in the magnitude of the coefficient of variation of  $B_1$  when using weighted least squares estimation, this is more than compensated for by the increase in precision of  $B_3$ ,  $B_4$ ,  $B_5$  and  $B_6$ . Furthermore, it is also worth noting that the decay constants  $B_2$ ,

$B_4$  and  $B_6$  are increased by weighting the data, whilst the amplitudes  $B_1$ ,  $B_3$  and  $B_5$  decrease. This may be expected since more weight is being given to the later time points. It must be stressed, however, that these results have been based on only one data set. Estimates of the standard errors of the parameters can really only be assessed in a simulation study.

#### 1.4.2 REGIONAL ESTIMATION OF MODEL PARAMETERS

Although, by using a dynamic approach to the quantification of neuroreceptors, the receptor density,  $R$ , in tissue, cannot be separated from the associated constant of receptor binding,  $k_4$ , relative receptor densities can be assessed in different parts of the brain as the ratio of total binding (specific and nonspecific) to nonspecific binding. In the case of NMDA receptors, the radioactivity in the region of interest can be compared with the activity in the cerebellum, a region assumed to have no NMDA receptors. Figures 1.3(i)-1.10(ii) show the ratio of radioactivities in each of the regions with respect to the cerebellum as a function of time. Data from normal and ischaemic animals have been plotted separately. The solid line in each plot is the estimated tissue: cerebellum ratio one would expect if the dynamic process can be explained by a two compartment blood flow model.

Within the time course of the experiments, the results indicate that in the normal central nervous system, isotope distribution can be adequately explained by a two compartment model. In all regions, the tissue:cerebellum ratio remained constant throughout

Figure 1.3 Ratio of frontal parietal cortex:cerebellum

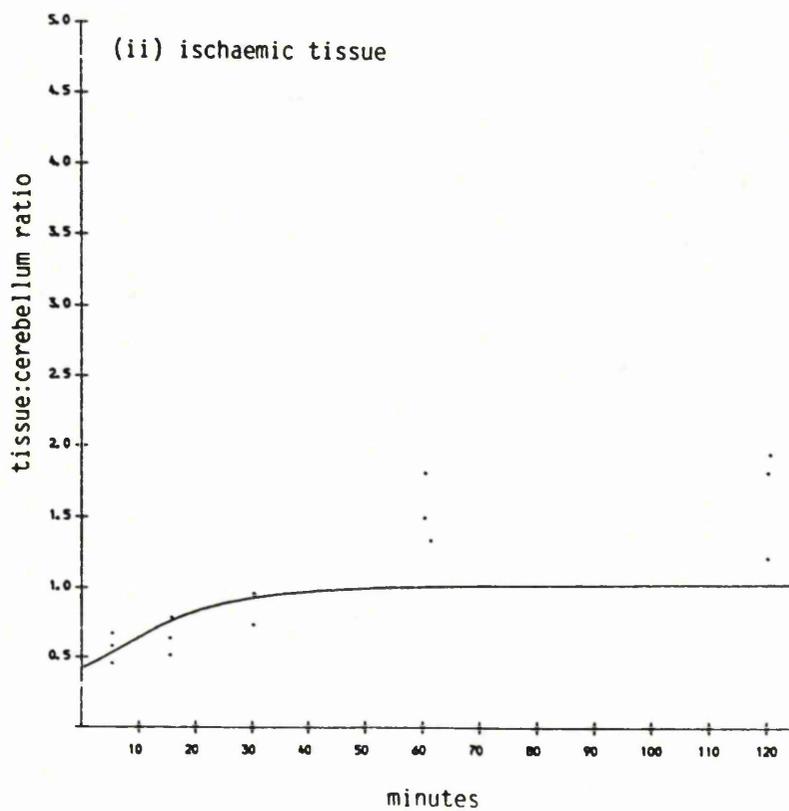
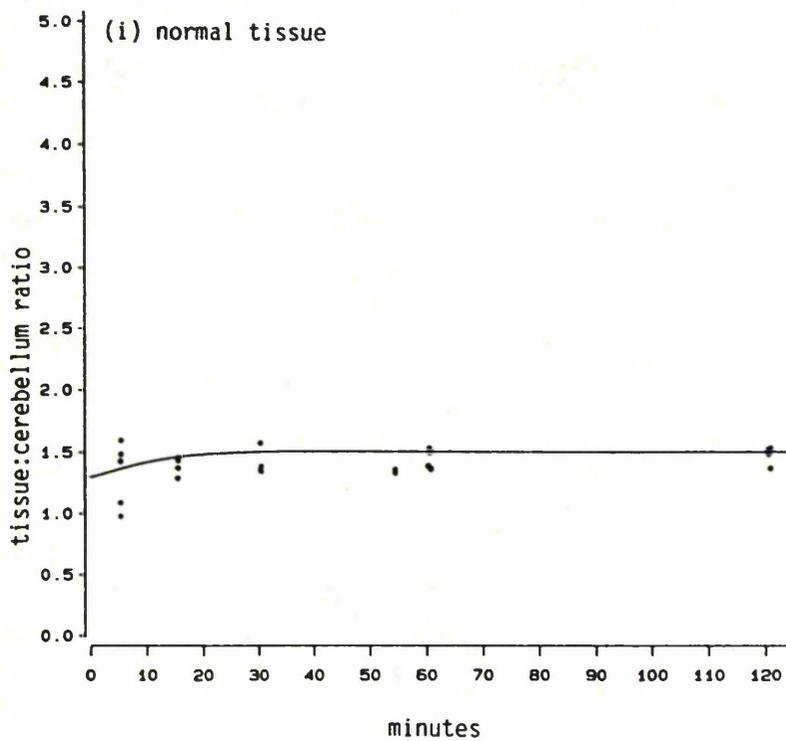


Figure 1.4 Ratio of frontal cortex:cerebellum

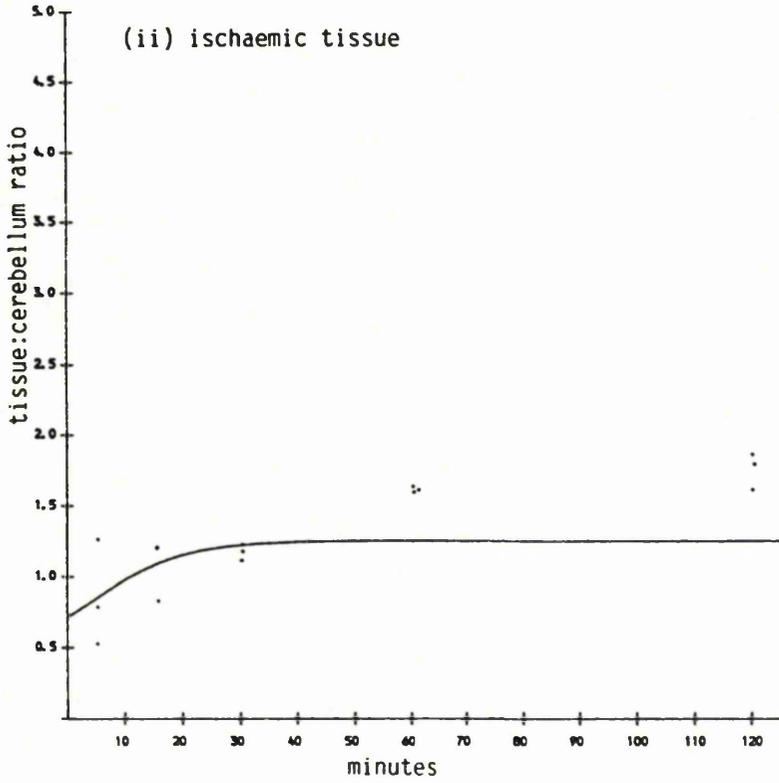
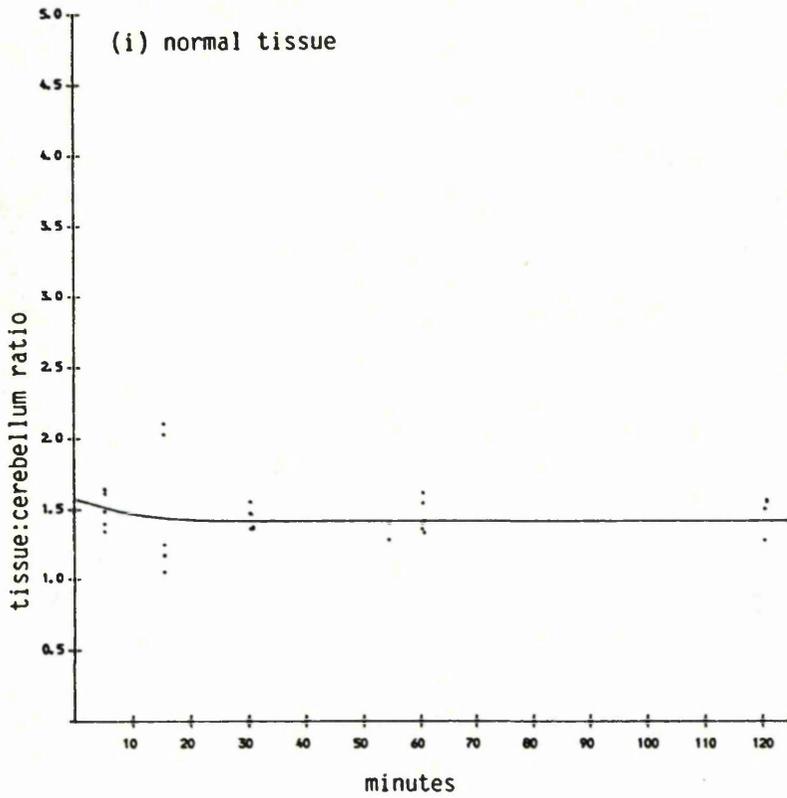


Figure 1.5 Ratio of hippocampus:cerebellum

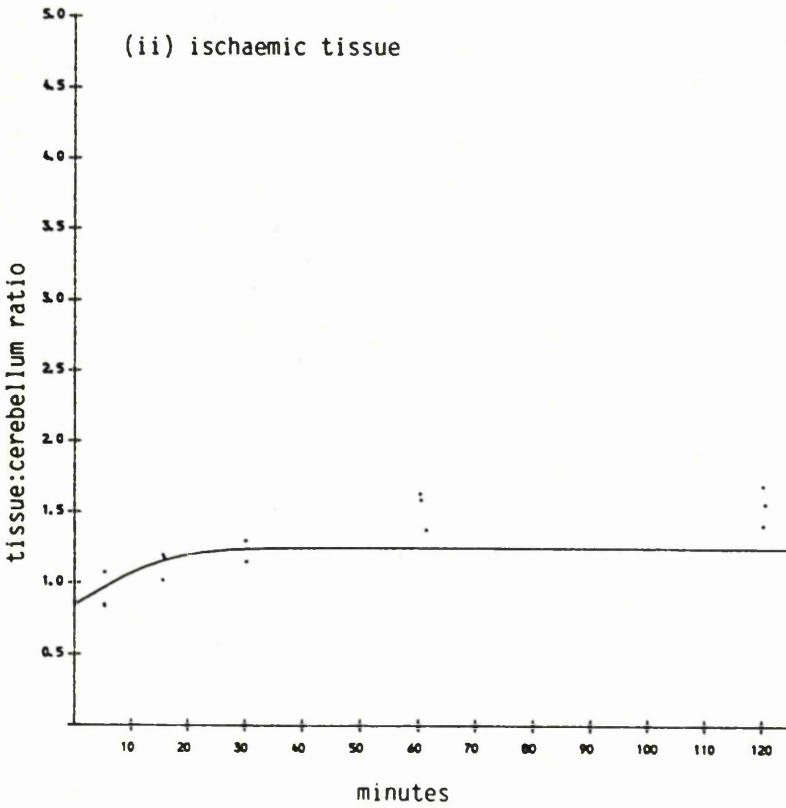
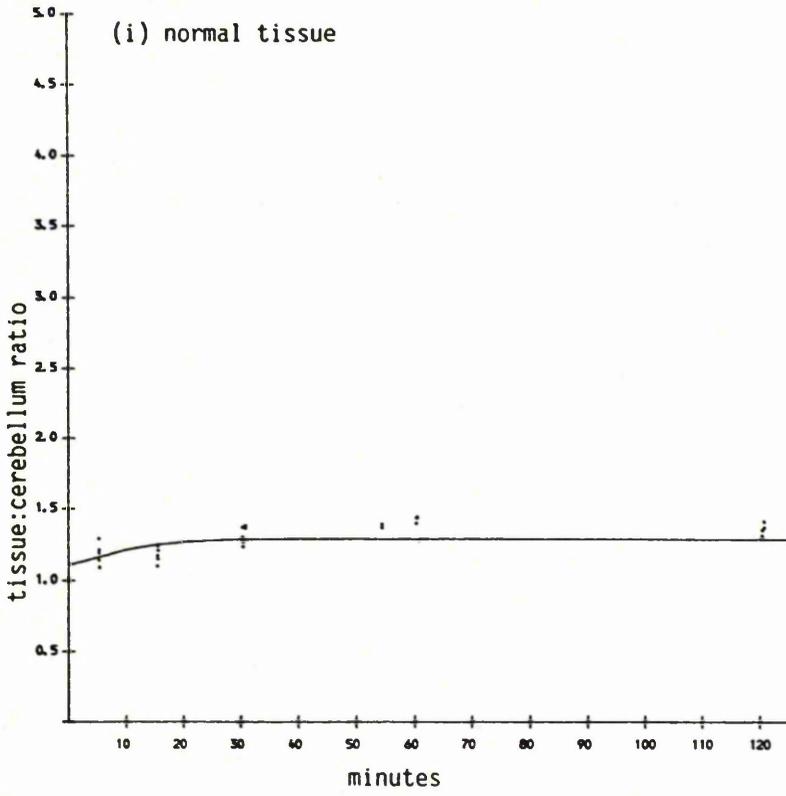


Figure 1.6 Ratio of naccumbens:cerebellum

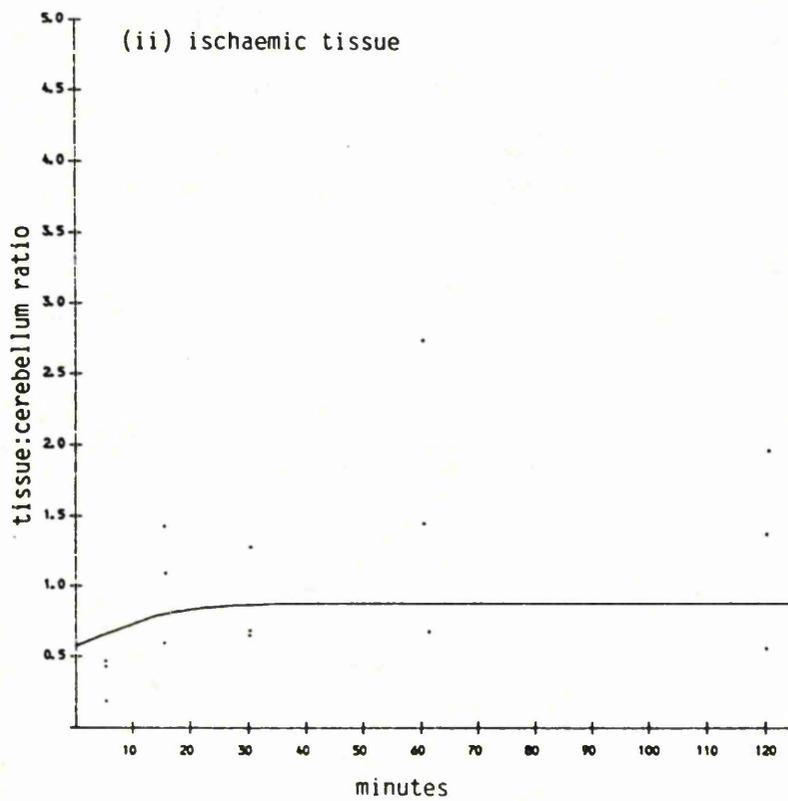
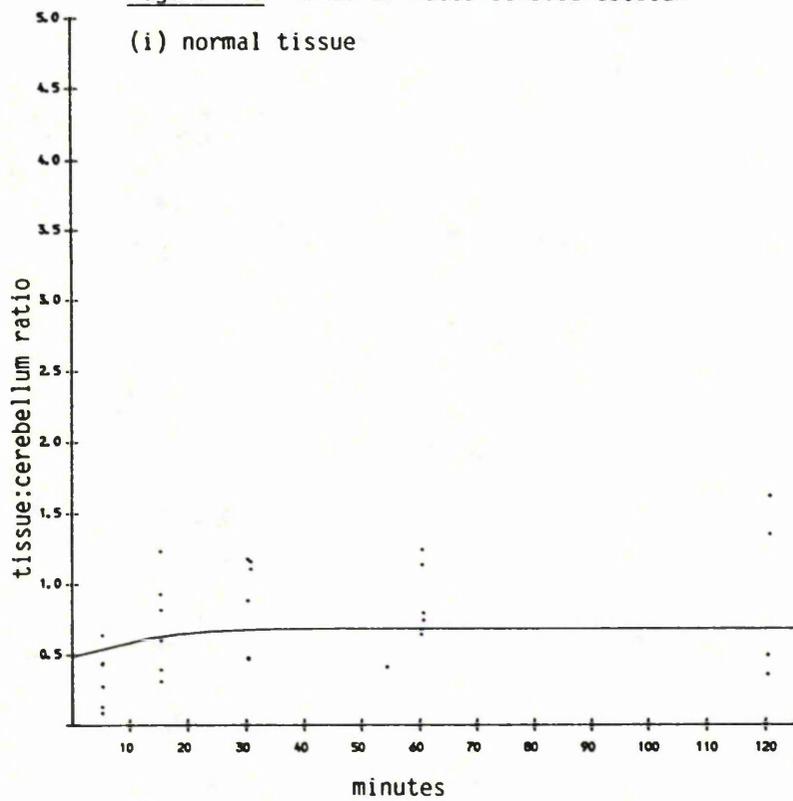


Figure 1.7 Ratio of occipital cortex:cerebellum

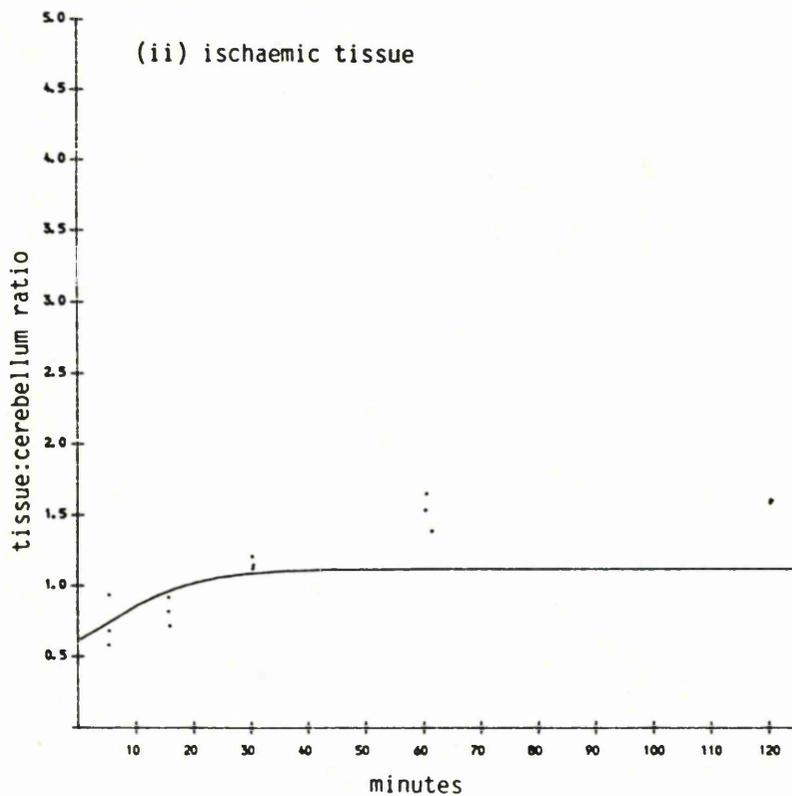
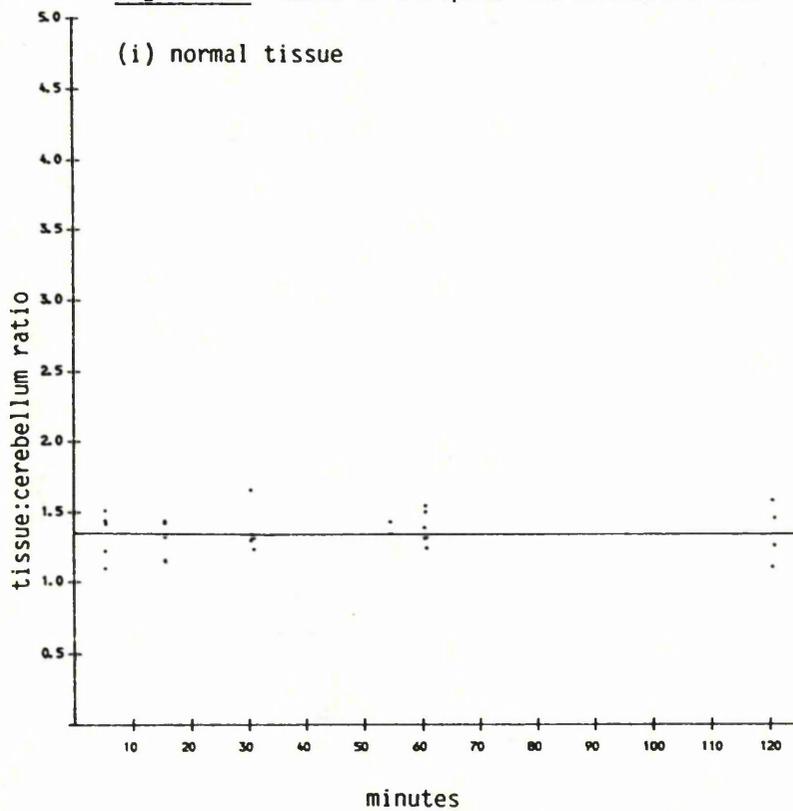


Figure 1.8 Ratio of olfactory bulb:cerebellum

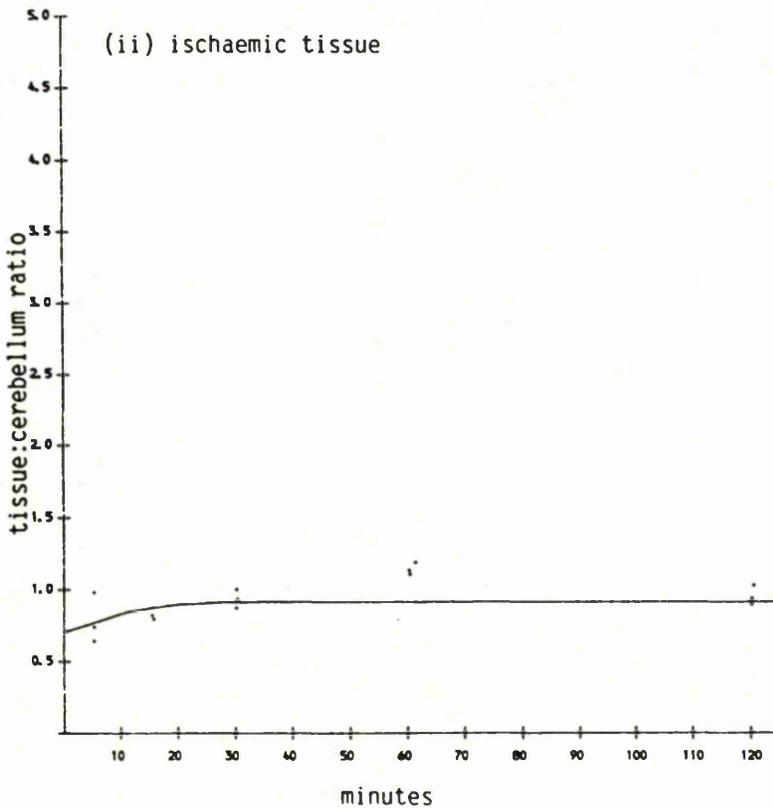
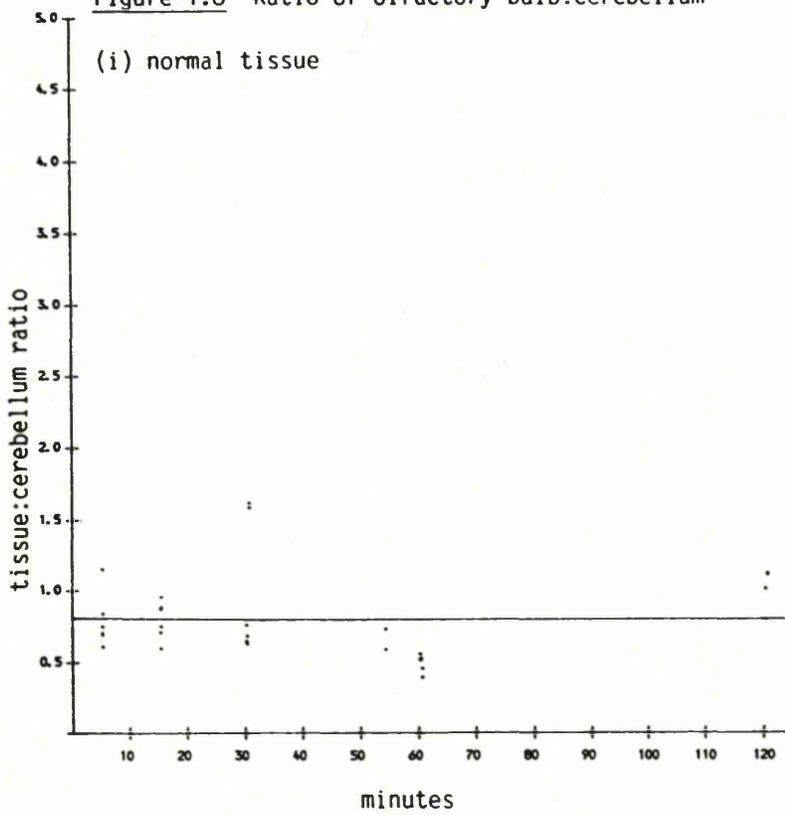


Figure 1.9 Ratio of parietal cortex:cerebellum

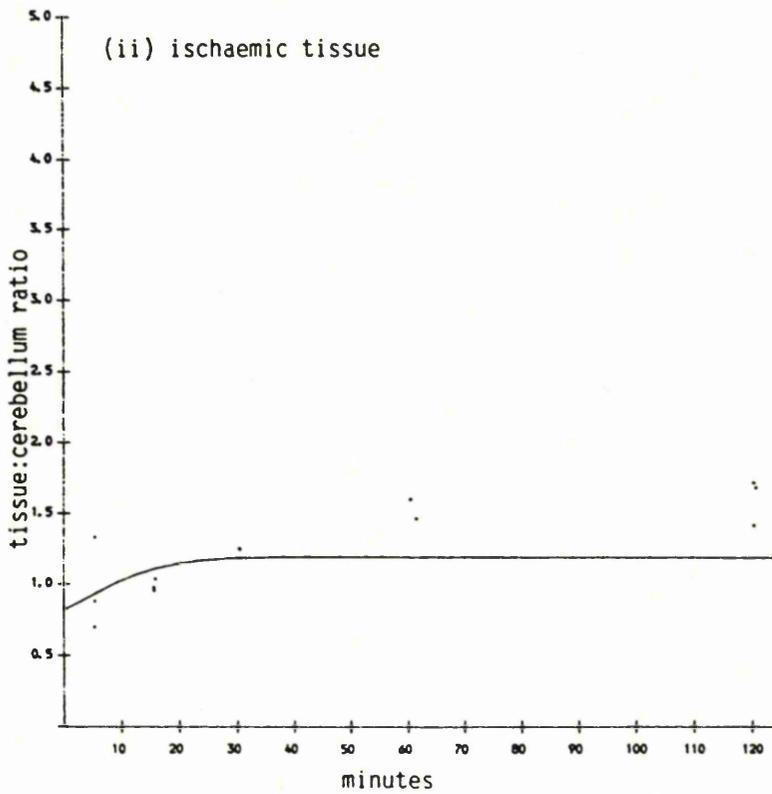
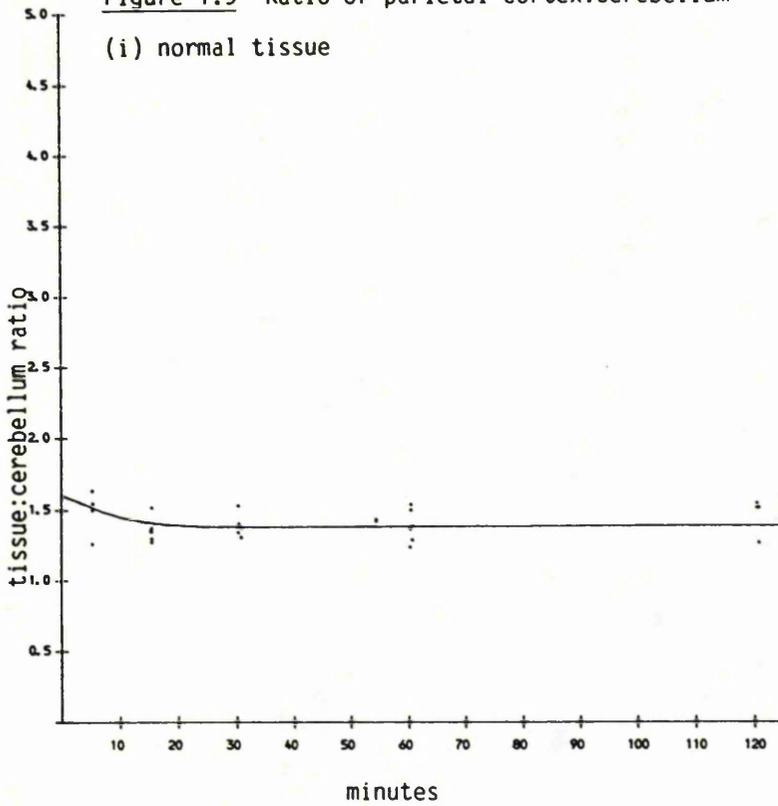
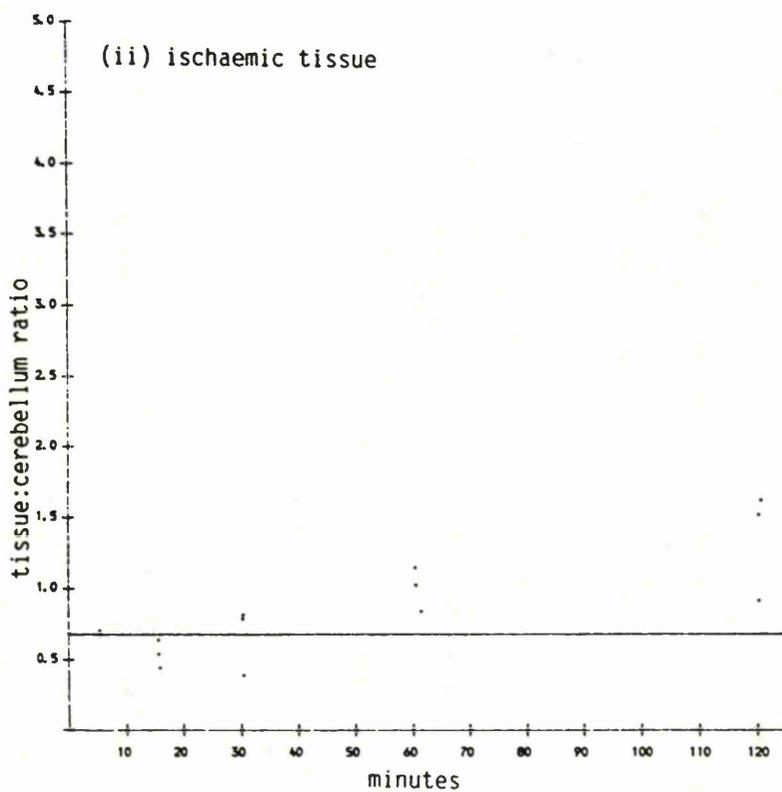
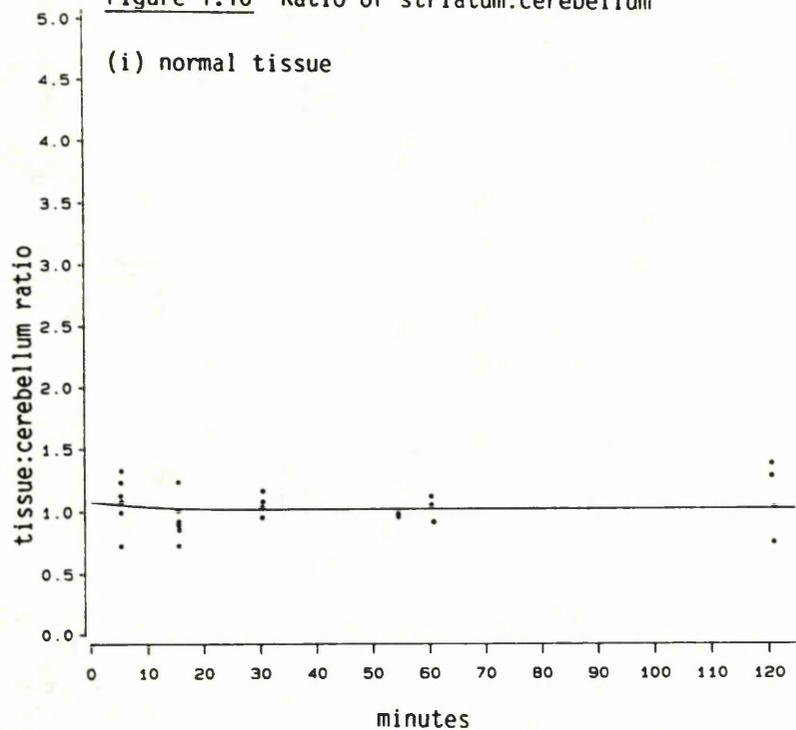


Figure 1.10 Ratio of striatum:cerebellum



the duration of the experiment. In contrast with normal tissue, the ratio of tissue: cerebellum activity in ischaemic animals increased progressively over the first 30-60 minutes after isotope administration and then levelled out for the remainder of the experiment. Although the two compartment model predicts an increase in the tissue: cerebellum activity ratio in ischaemic animals similar to that observed over the first 30-60 minutes, in six of the regions of interest (frontal parietal cortex, frontal cortex, hippocampus, occipital cortex, parietal cortex, striatum), the ratio for the later time points is underestimated.

Thus by directly comparing the relative receptor densities it would appear that the uptake of MK-801 into a number of ischaemic areas cannot be attributed solely to the reduced level of cerebral blood flow.

Using the arterial input function in equation (1.20), regional estimates of the model parameters describing the kinetics of MK-801 distribution in normal and ischaemic tissue were derived by analysing the time course of tracer in each brain region. The adequacy of a two and three compartment model was examined after fitting the kinetic data to the convolution of the input with the sum of one or two exponential functions, using a BMDP (1987) nonlinear least-square curve fitting procedure. A model was defined as being adequate if the parameters converged to a nonnegative set of parameters. Figures 1.11(i)-1.19(ii) show the time-activity curves for the regions of interest with the corresponding 'best fit' curve.

Figure 1.11 Time-activity curve for the cerebellum

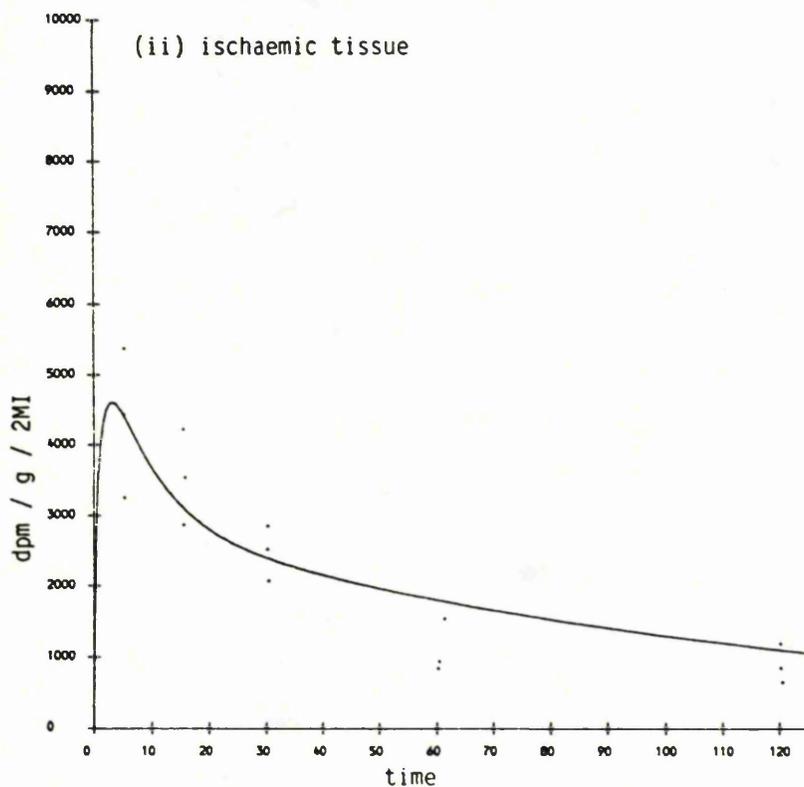
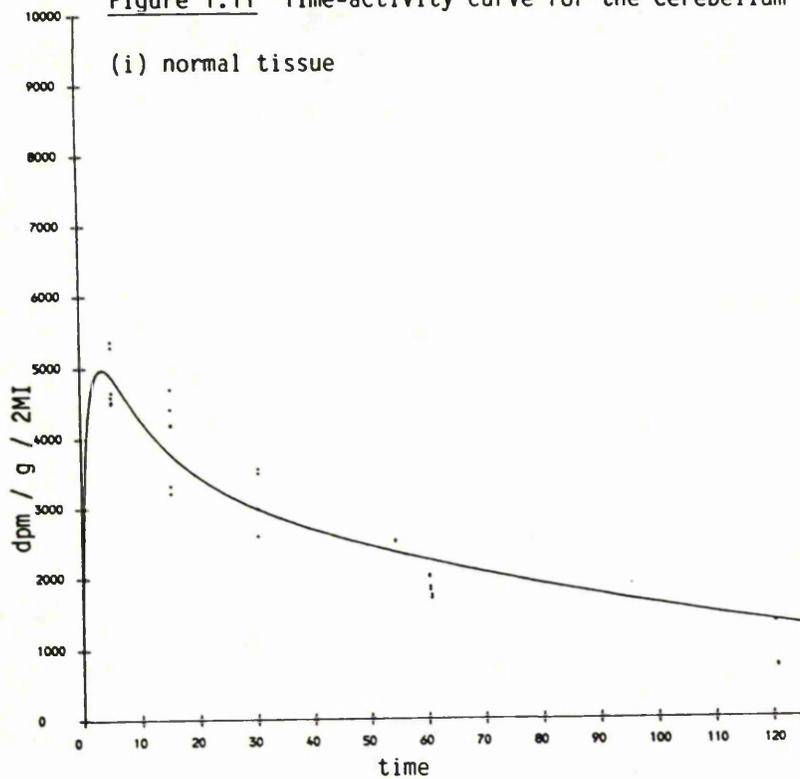


Figure 1.12 Time-activity curve for the frontal parietal cortex <sup>38</sup>.

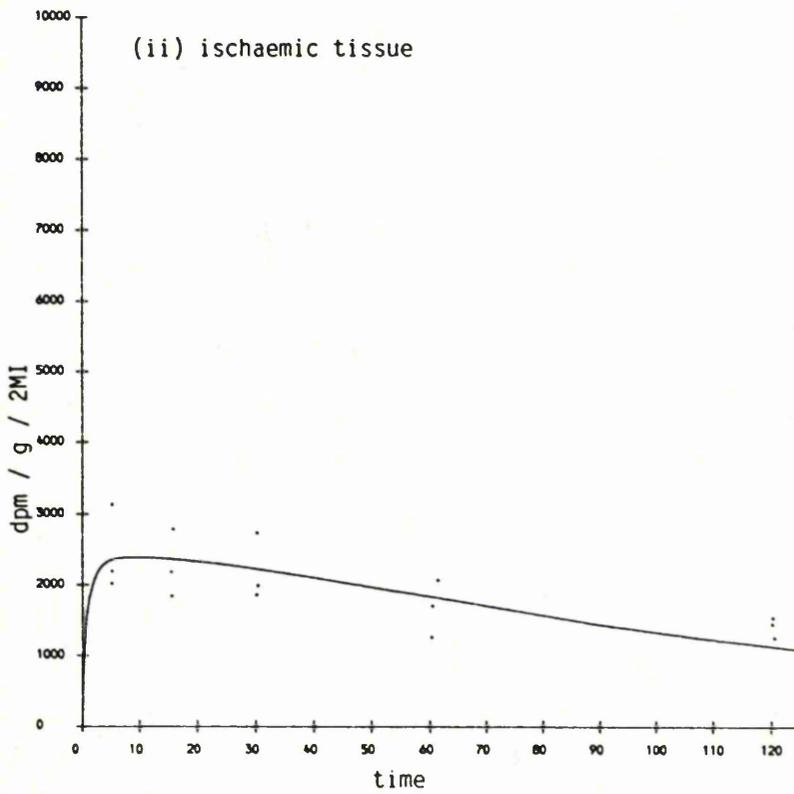
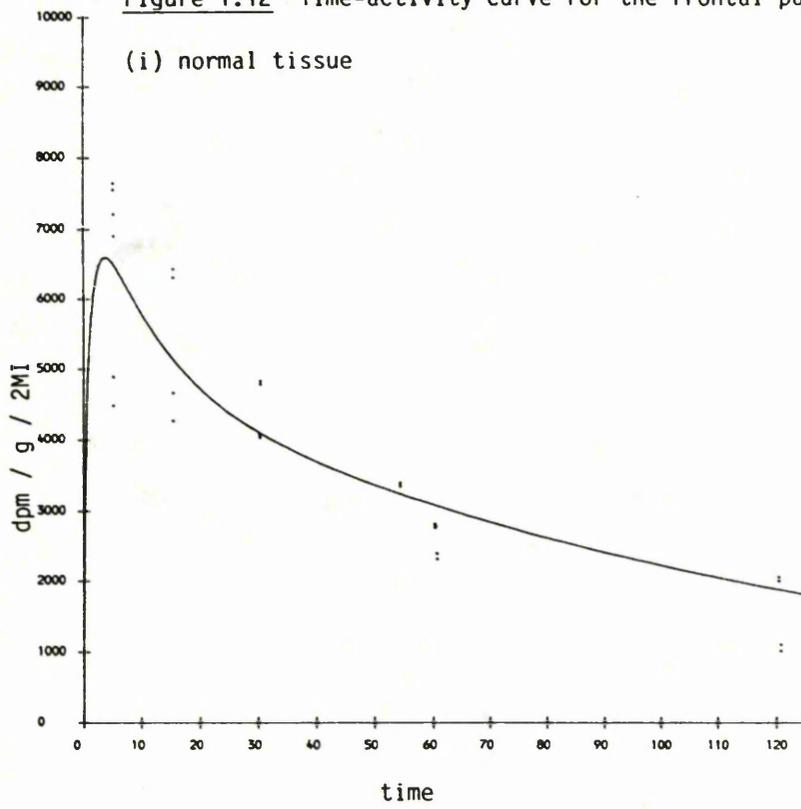


Figure 1.13 Time-activity curve for the frontal cortex

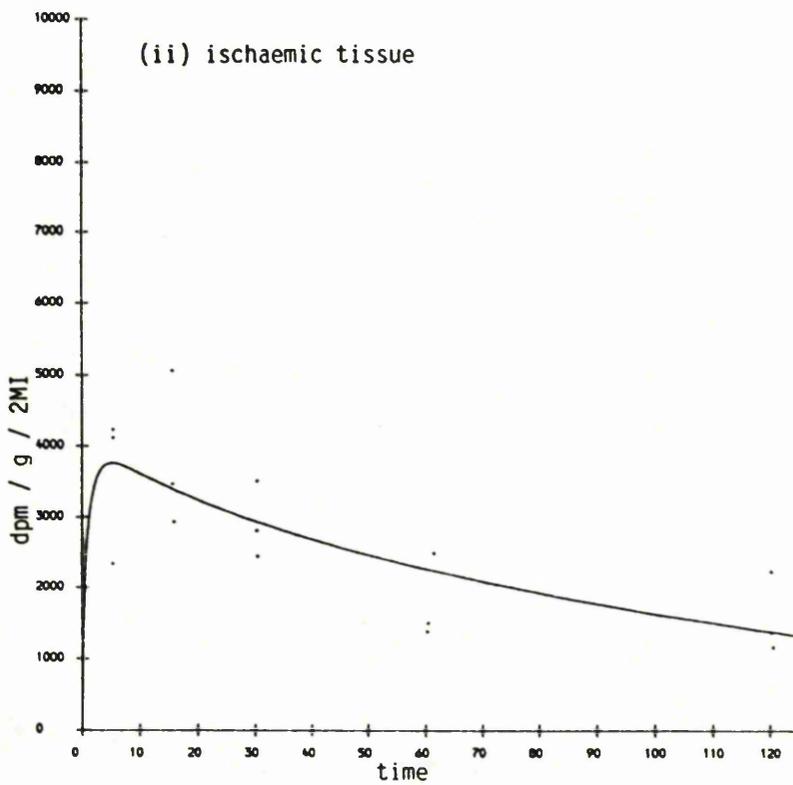
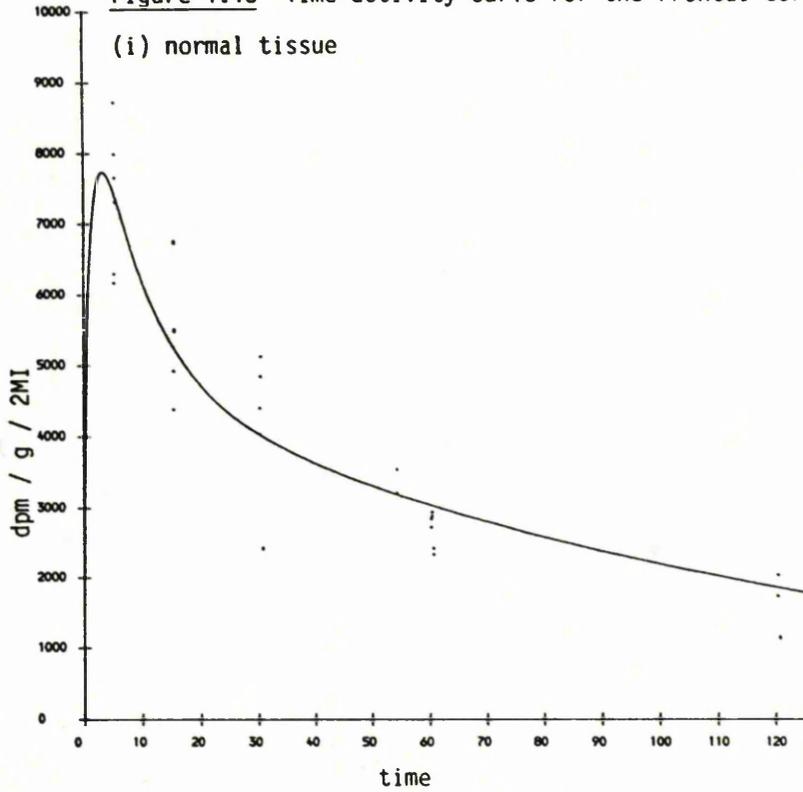


Figure 1.14 Time-activity curve for the hippocampus

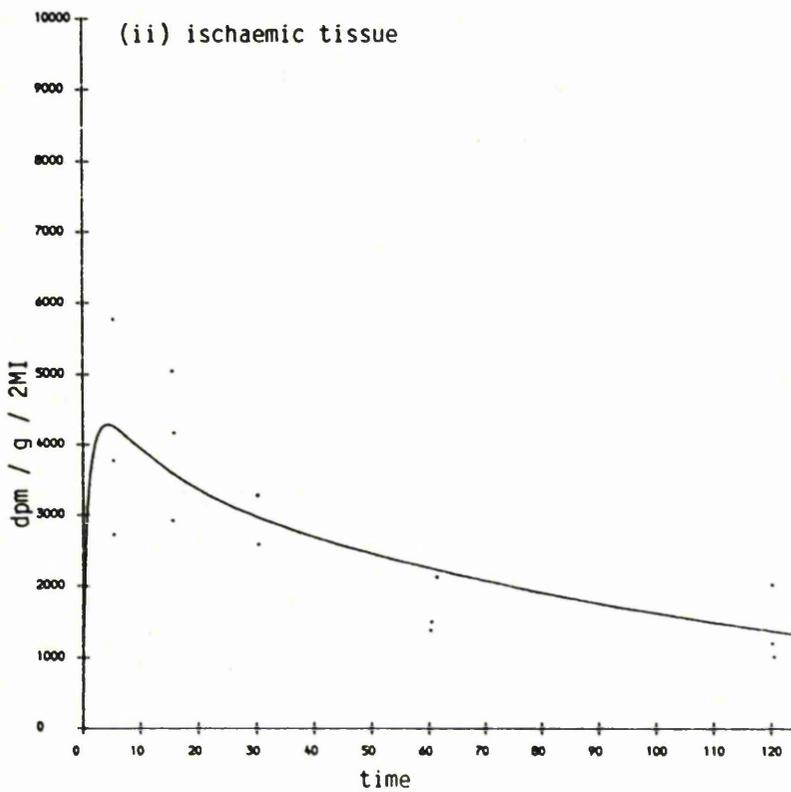
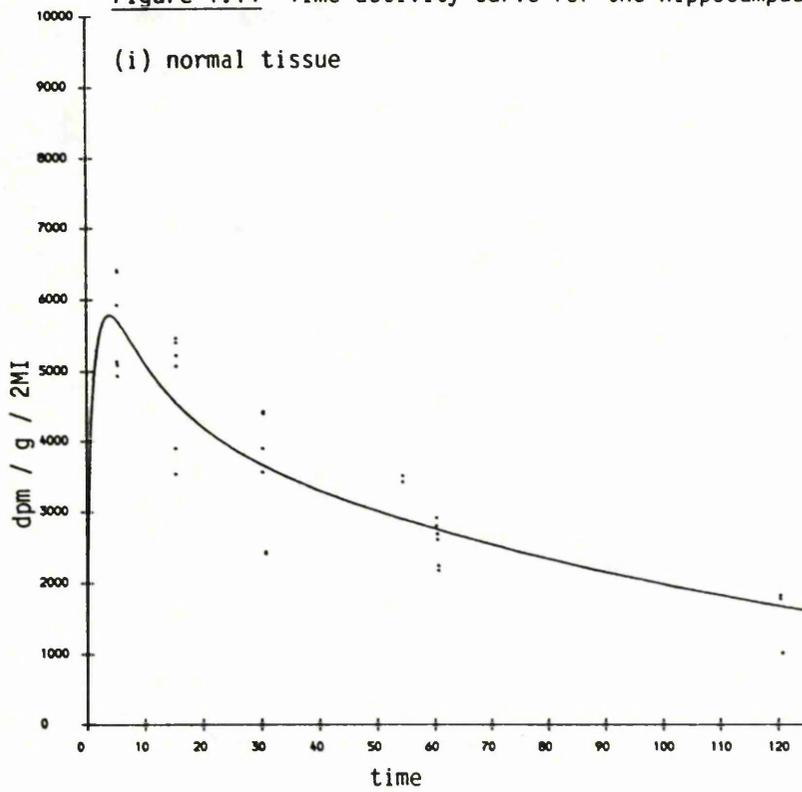


Figure 1.15 Time-activity curve for the naccumbens

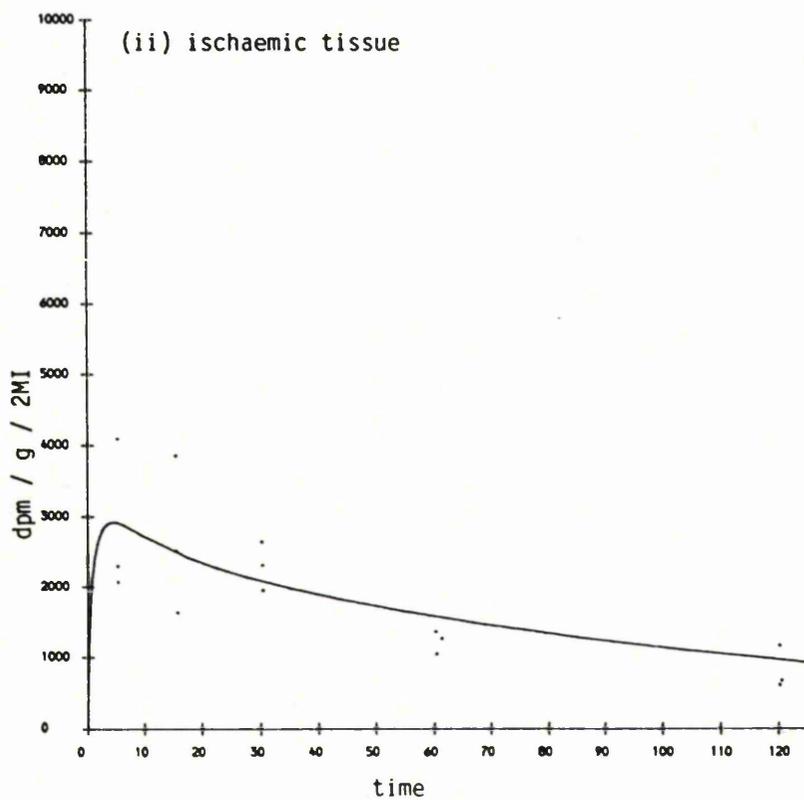
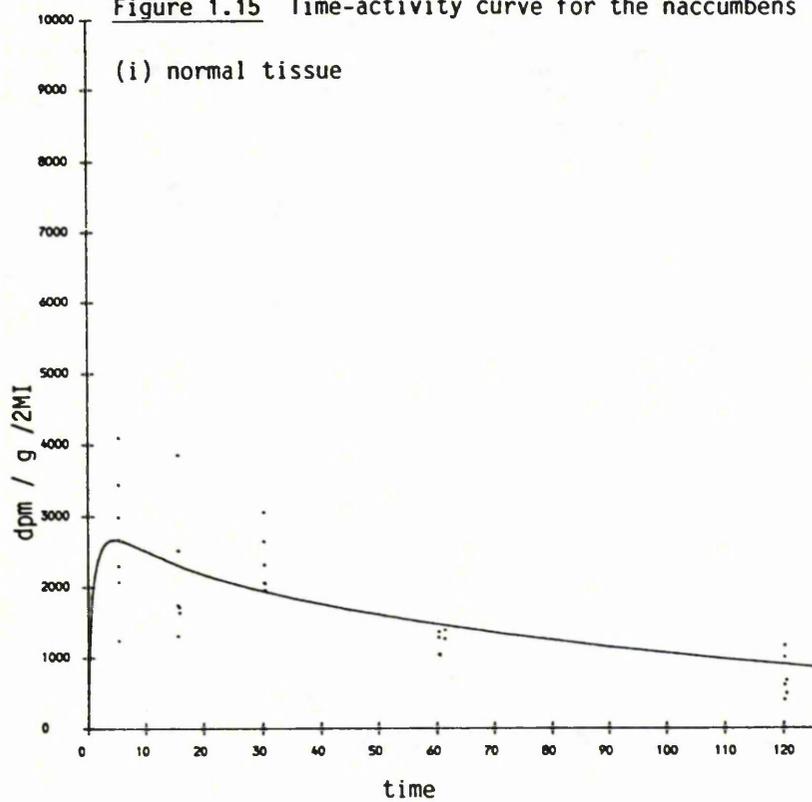


Figure 1.16 Time-activity curve for the occipital cortex

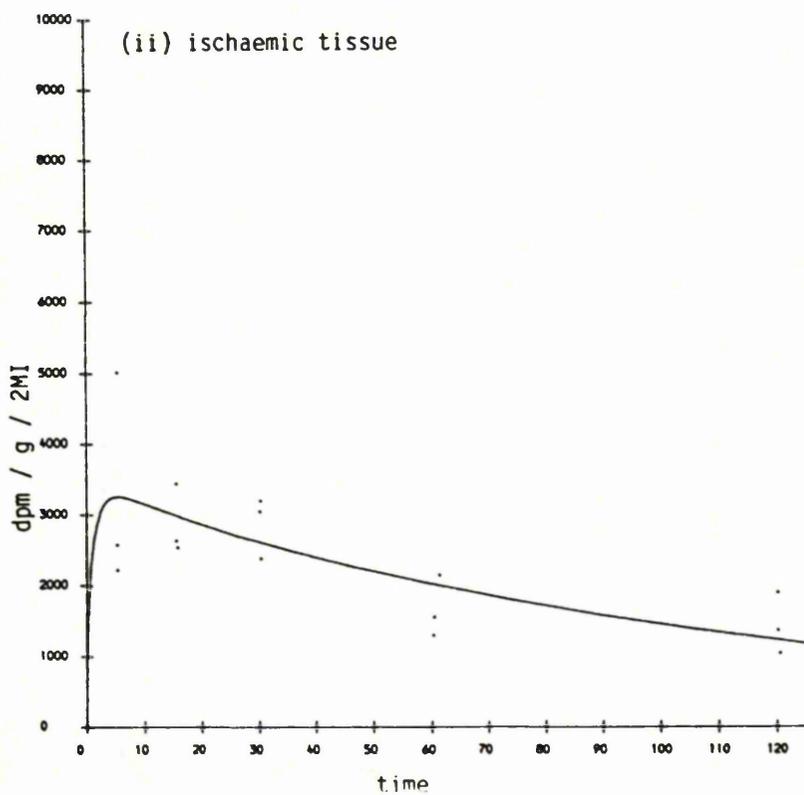
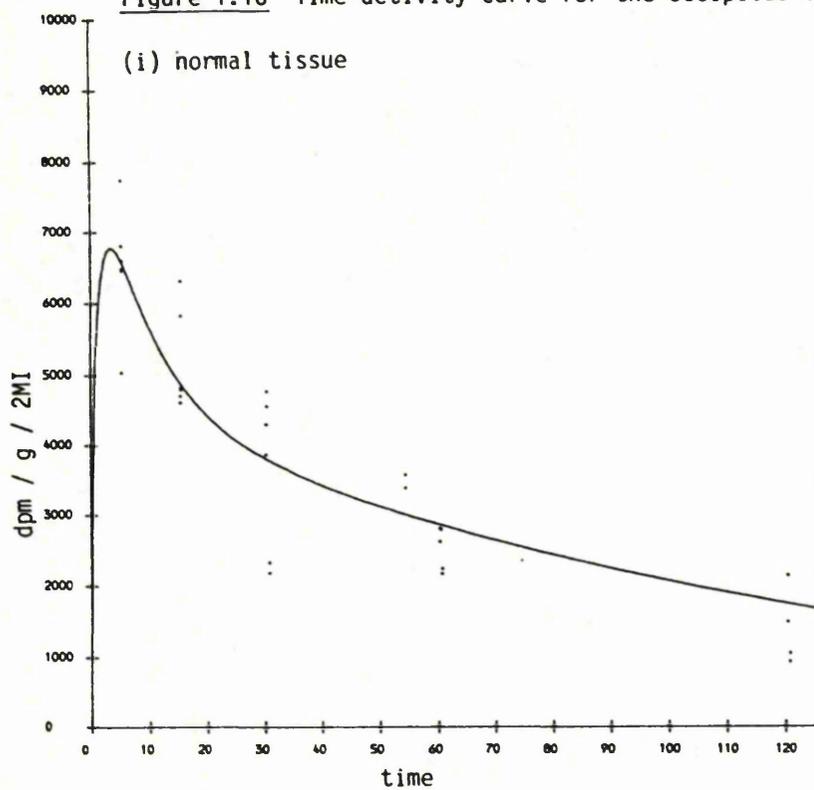


Figure 1.17 Time-activity curve for the olfactory bulb

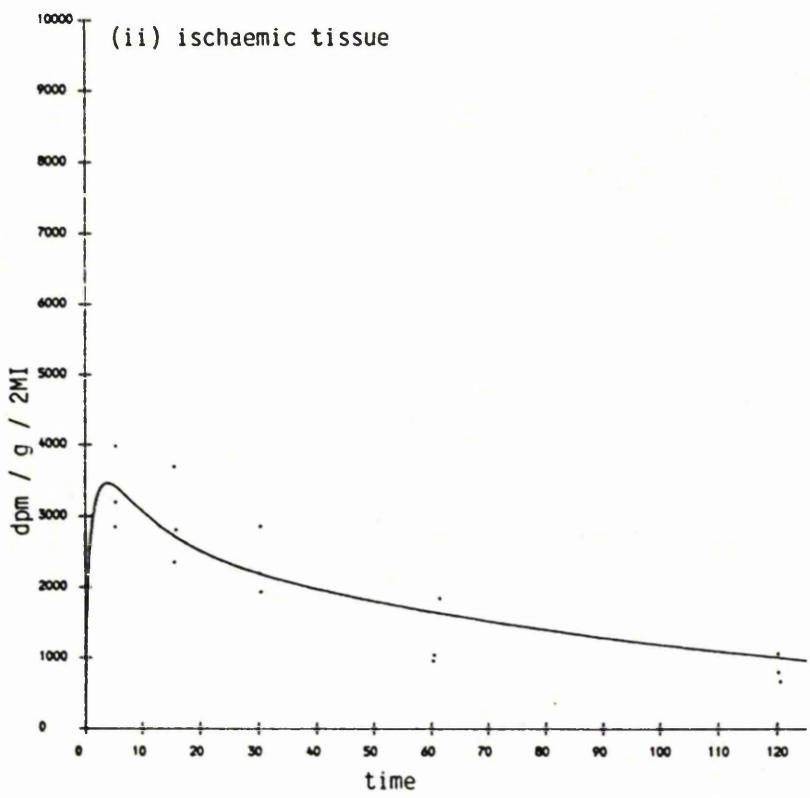
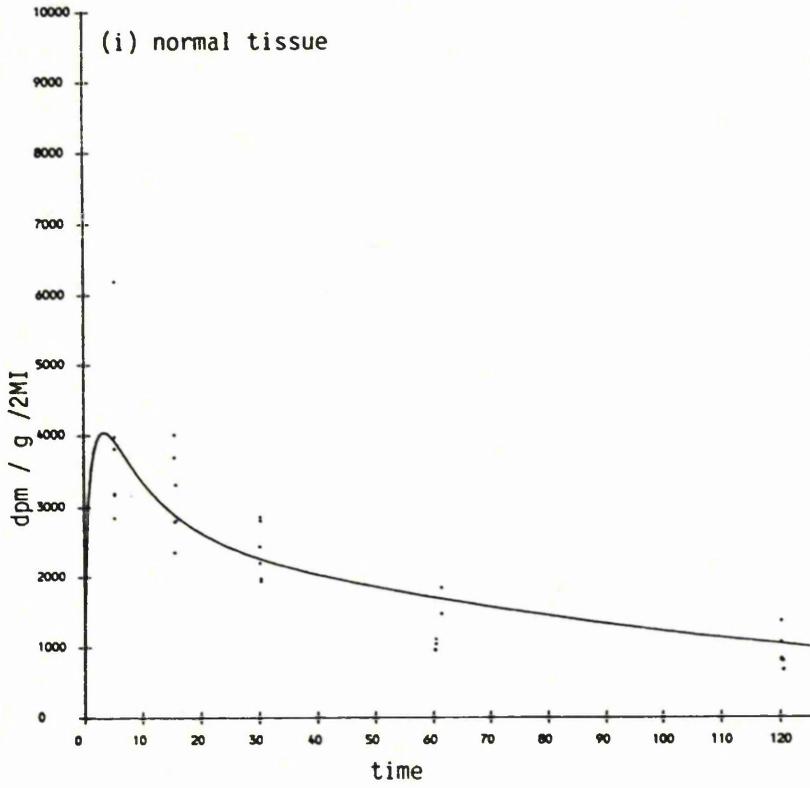


Figure 1.18 Time-activity curve for the parietal cortex

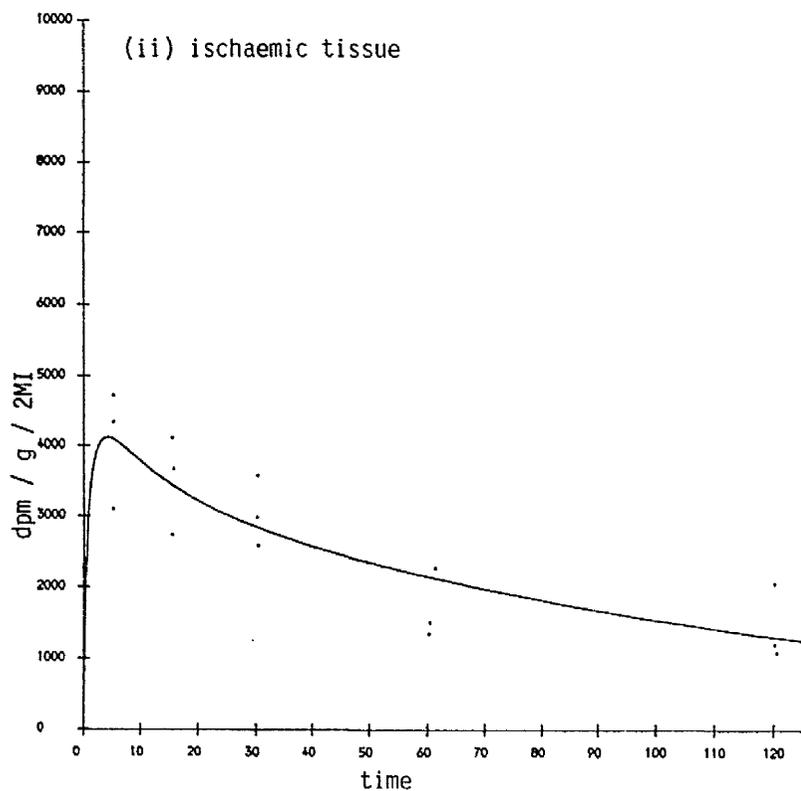
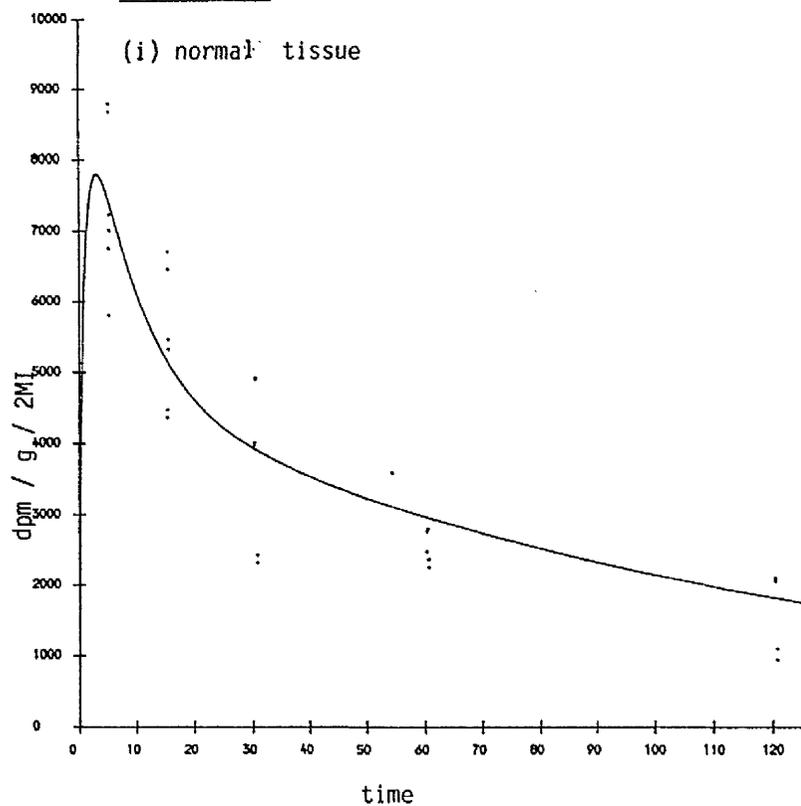
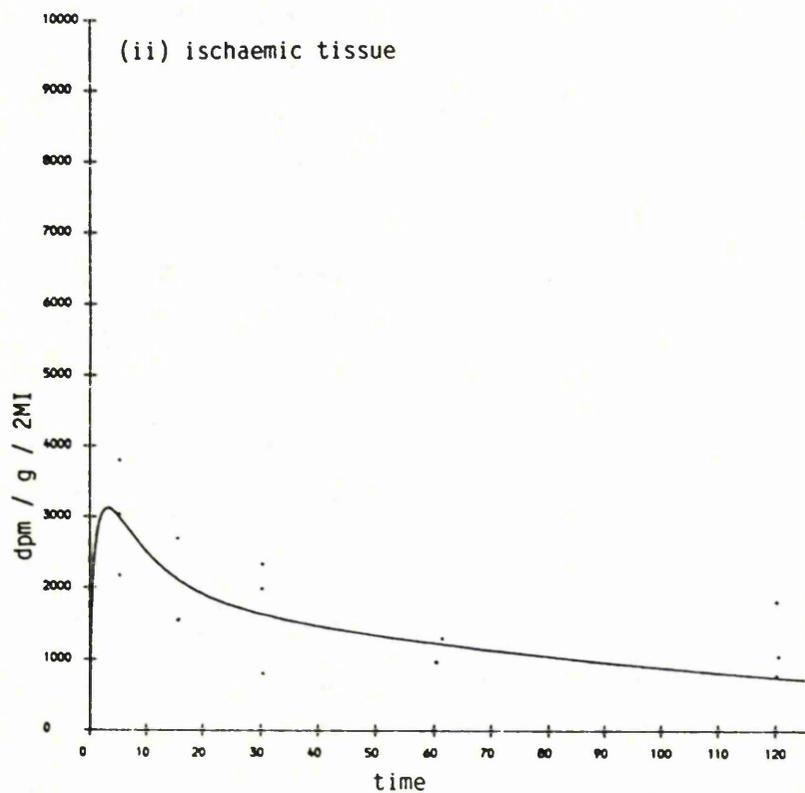
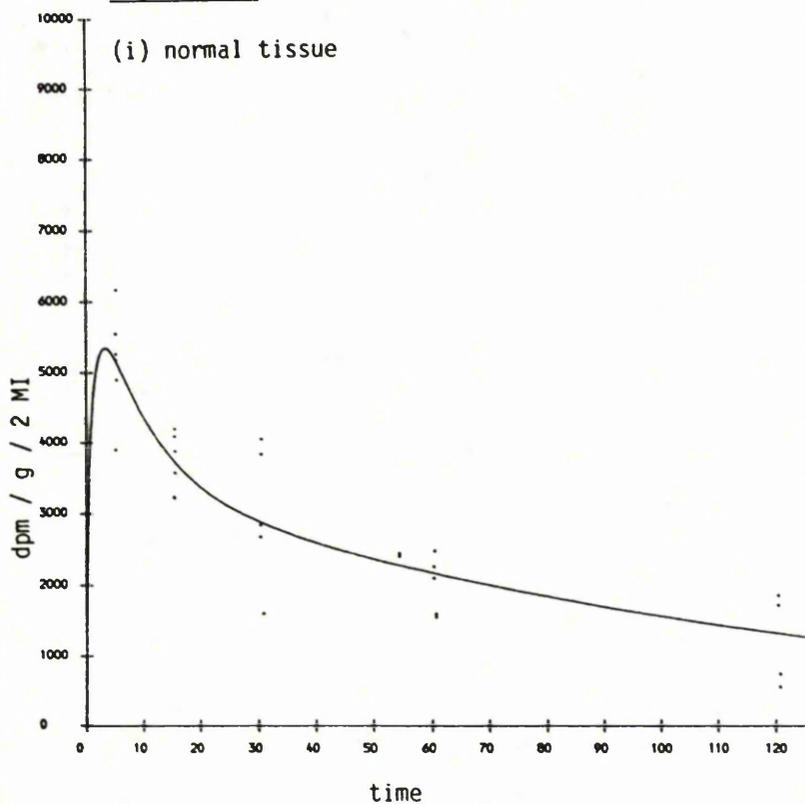


Figure 1.19 Time-activity curve for the striatum



On no occasion did a fit of the full model in equation (1.18) converge using regional data. However, recalculation of the parameters with  $A_4$  in equation (1.18) set to zero (equivalent to the condition where dissociation from the receptor is negligible) resulted in the identification of a unique set of estimates in four of the nine regions in ischaemic tissue. No convergence to a unique nonnegative set of parameters was obtained using this model for the normal central nervous system tissue data, implying that the model was overparameterised. In all ischaemic and normal tissue data, the fitting converged with both  $A_3$  and  $A_4$  set to zero. The regional model parameter values derived from the two and three compartment fits are given in Table 1.3. Based on the F-test, the fit of the three compartment model in the frontal parietal cortex, frontal cortex, occipital cortex and striatum was not significantly better than the two compartment model.

The parameter estimates of  $\alpha$  and the capillary permeability-surface area product,  $k$ , in normal central nervous system tissue are rather heterogeneous. Both parameters are markedly higher in cortical regions; areas corresponding to high levels of cerebral blood flow.

In general, the regional parameter estimates of  $k$  and  $\alpha$  using the two compartment model tended to be lower in ischaemic tissue (only the naccumbens showed an increase in both parameters). Furthermore, with the exception of the cerebellum and the frontal parietal cortex, regional estimates of  $k$  appear to be homogenous; the estimate of  $k$  in the cerebellum exceeded all others whilst the estimated  $k$  in the frontal parietal cortex appeared to be lower. Indeed, with respect to the cerebellum, the estimates of  $k$  and  $\alpha$

are only slightly reduced in ischaemic tissue suggesting that this<sup>47</sup> region is largely unaffected after left middle artery occlusion. The estimates of the parameter  $k_3$  using the three compartment fit were all small, emphasising that specific binding represents only a small contribution to the total activity in these regions. However, by including an extra parameter in the model estimates of  $k$  increased whilst those of  $\alpha$  decreased.

**Table 1.3a Estimates of kinetic constants from the two compartment model normal tissue**

Region	$k$	$\alpha$
Cerebellum	0.483	3.57
Frontal parietal cortex	0.628	5.76
Frontal cortex	0.762	5.54
Hippocampus	0.536	4.81
Naccumbens	0.237	2.04
Occipital cortex	0.652	5.11
Olfactory bulb	0.389	2.63
Parietal cortex	0.776	5.42
Striatum	0.521	3.66

**Table 1.3b Estimates at kinetic constants from the two and three compartment model (ischaemic tissue)**

Region	2 compartments		3 compartments		
	$k$	$\alpha$	$k$	$k_3$	$\alpha$
Cerebellum	0.454	2.88	-	-	-
Frontal parietal cortex	0.190	2.63	0.207	0.0079	2.18
Frontal cortex	0.324	3.62	0.324	0.0001	3.65
Hippocampus	0.385	3.68	-	-	-
Naccumbens	0.261	2.28	-	-	-
Occipital cortex	0.279	3.11	0.283	0.0021	3.01
Olfactory bulb	0.321	2.47	-	-	-
Parietal cortex	0.371	3.48	-	-	-
Striatum	0.308	1.65	0.343	0.0072	1.37

## 1.5 Discussion

In this section we have introduced a mathematical model to describe the uptake and retention of MK-801 in the rat brain. In addition, we have focused our investigation on the model configurations that are identifiable from modelling MK-801 in normal and ischaemic central nervous tissue. In the nine regions of interest studied, our results indicate that the kinetic data, pooled in a series of rats from dissection experiments, support only two distinguishable compartments in the model for normal and ischaemic tissue. In four of the nine regions in ischaemic tissue (frontal parietal cortex, frontal cortex, occipital cortex and striatum) a three compartment model could be fitted to the data although the F-test, which is possibly unreliable, indicated that the fit, with the inclusion of the extra parameter, was not significantly better. However, the modelling of the tissue: cerebellum ratio indicated that the kinetics of MK-801 uptake into these ischaemic areas at the later time points could not be attributed solely to the rate of cerebral blood flow.

The inability to fit a three compartment model to the data in circumstances under which the release of glutamate is not enhanced can be explained by the diffusibility and lipophilic properties of MK-801 coupled with the high level of cerebral blood flow. These properties dominate the kinetic model and greatly outweigh any binding of the drug to specific binding sites. However, in conditions with low cerebral blood flow and enhanced glutamate release (e.g. ischaemia) the possibility of fitting a three compartment model assuming no dissociation from the specific binding

site is increased. Possible explanations as to why the three compartment model does not improve significantly the fit or cannot be fitted at all could be measurement noise in the blood radioactivity concentration, difficulty in estimating accurately the variance of data values (leading to suboptimal weighting in the nonlinear least squares curve fitting algorithm), the limited time range of data (especially the lack of data at very early experimental times) and small sample size.

A further explanation for the apparent lack of fit could be the result of tissue dissection errors. Speculation regarding this cause could be removed by conducting an autoradiographic study. Moreover, greater anatomical resolution would be obtainable allowing the investigator to study *in vivo* MK-801 binding in more brain structures. Although such studies would reduce the tissue dissection variability in the kinetic data, this source of variation is likely to be small compared with the variation in the data caused by pooling data from different animals. Bearing in mind that PET and SPECT studies would permit the investigator to observe data in the form of time versus regional activities for each individual, the results of the present study are of significance with respect to the eventual feasibility of studying MK-801 receptor binding in experiments with enhanced glutamate release, employing these technologies. Not only would the signal have reduced noise over a wide range of time, in addition the investigator has the advantage of obtaining a pictorial display of regional parameter values.

The validity of the simplifying assumptions which were necessary to derive the working model is another important issue which has to be

raised. Although tracer kinetic studies have shown consistency between the specified model and the data, biochemical measurements are required to verify the approximating assumptions and to confirm model predictions of compartmental concentrations of tracer at several time points. Moreover, the approximating assumptions should always be examined under a wide range of physiological conditions (e.g. receptor densities, rate of blood flows, specific activities, partition coefficients). To illustrate this point, due to the similarities of the experimental procedure we attempted to fit the data in this study to the model by Frey et al (1985). The parameters  $A_1$ ,  $A_2$  and  $A_3$  in equation (1.18) can be derived from the data in a similar fashion. However these did not satisfy the constraints necessary for the transformation of these parameters into the kinetic parameters of interest ( $K_1$ ,  $K_3$  and  $\alpha$ ). The inability to identify the set of parameters can be explained by the freely diffusibility property and the high brain: blood partition coefficient of MK-801. Thus, the selection of a particular approach should depend very much on the characteristics of the radiolabelled ligand and the appropriateness of the simplifying assumptions.

In conclusion, by introducing a mathematical model which incorporates the freely diffusibility property of MK-801 we have been able to derive parameters which characterise the kinetics behaviour of MK-801. Furthermore, in conditions with low cerebral blood flow and enhanced glutamate release, there is enough evidence to suggest that in vivo MK-801 binding can be employed to examine glutamergic mechanisms which cannot be assessed in vitro, if the variability in the experimental data can be controlled.

## CHAPTER 2 RANKING REGIONAL RESPONSES

### 2.0 INTRODUCTION

Neuropharmacologists are often interested in investigating the hierarchy or the pattern of response of some dynamic biochemical process of interest in a large number of brain regions after drug administration or other treatment, and possibly in comparing and contrasting the observed response with different treatments or drugs. Consider, as an illustration, the investigation of the level of local cerebral glucose utilisation (LCGU) in discrete regions of interest (ROI's), as measured by the [<sup>14</sup>C]-2-deoxyglucose quantitative autoradiographic technique. Typical applications in this area would involve measuring LCGU in up to 100 ROI's at each of 3 to 5 doses of a drug (including a control group), with experimental animals nested within individual doses of the drug.

In the literature, a variety of methods to deal with the quantification of the regional response to drug treatment have evolved. For example, in situations where a treatment or a single dose group is to be compared with a control group, regional mean differences or percentage changes relative to the control group, are calculated and interpreted (Eckardt et al, 1986; Sokoloff et al, 1977), whereas in multidose situations a one-way analysis of variance between treatment groups may be carried out within regions and either the resulting F statistics or p values used to rank responses (Dunn et al, 1980).

The main feature of all these methods is that no attempt is made to assess the reliability of the rankings derived. Furthermore the use of F statistics to measure the response is potentially misleading since the F statistic just measures a general variability in the group means and does not necessarily rank the response to drug treatment sensibly. For instance, a region affected only at the highest dose, at which maximal response is reached, would be equally ranked with a region in which the maximal response is reached at the lowest dose. Another problem with the F statistic is that division by the regional residual mean square error may introduce spurious variability to the ranking problem.

In this chapter we investigate the reliability of methods of ranking regions with respect to the strength of response. In section 2.1 we address, mathematically, some of the problems associated with ranking the components of a multivariate normal mean vector and identify extreme forms of the covariance structure which lead to the least and most favourable forms of analysis. In section 2.2 we investigate, empirically, the performance of methods to characterise the response to drug treatment over the whole brain, or at least the set of regions selected for study. The performance of these methods are assessed over a variety of covariance structures using simulation and we identify extreme forms of the covariance structure which should lead to conservative forms of analysis.

## 2.1 RANKING THE COMPONENTS OF A MULTIVARIATE MEAN VECTOR

In this section we are concerned with the multiple-decision problem of evaluating the probability associated with the complete ranking of the components of a multivariate mean vector.

The results, of course, can be extended in an obvious manner to the case where we are interested in the differences between two multivariate normal vectors.

The indifference-zone philosophy of multiple decision procedures was first formulated by Bechhofer in a pioneering paper in 1954, in which the author addressed several possible ranking and selection goals as alternatives to conventional tests of homogeneity. Gupta (1956) introduced the related subset selection approach. The mathematical problems associated with both these approaches are equivalent in the present case of interest, and so as a matter of convenience only the indifference-zone philosophy will be considered. Although many authors have investigated various goals using these approaches (see Gibbons, Olkin and Sobel (1977) and Gupta and Panchapakesan (1979) for a review of the general statistical problems) very little work in this area has focused on selection and ranking procedures relating to a single  $k$ -variate normal distribution. Frischtak (1973), in a doctoral dissertation, considered the problems of selecting the sample size which, for a guaranteed probability will (a) select the component with the largest population mean (b) select the component with the smallest variance and (c) select a subclass of components with the smallest population generalised variance.

### 2.1.1 FORMULATION OF THE PROBLEM

Suppose the random vector  $\underline{X}^T = (X_1, \dots, X_k)$  has a multivariate normal distribution with mean vector  $\underline{\mu}$  and covariance matrix  $\Sigma$ . Assume  $\underline{\mu}^T = (\mu_1, \dots, \mu_k)$  is unknown and  $\Sigma = \sigma^2 R$ , where  $\sigma^2$  is known and  $R = \{\rho_{ij}\}$  denotes the correlation matrix.

Let  $\mu_{(1)} \leq \mu_{(2)} \leq \dots \leq \mu_{(k)}$  be the ranked values of the vector  $\underline{\mu}$ . It is assumed that it is not known beforehand which population is paired with  $\mu_{(i)}$  ( $i=1, \dots, k$ ). Our goal, in this problem, will be to find the correct (but unknown) ordering

$$\mu_{(1)} \leq \mu_{(2)} \leq \dots \leq \mu_{(k)}$$

To formulate the probability associated with this goal, when using the indifference zone approach, we introduce a "distance" measure between each successive pair of ordered  $\mu_{(i)}$  values,

$$\delta_1 = \mu_{(2)} - \mu_{(1)}$$

$$\delta_2 = \mu_{(3)} - \mu_{(2)}$$

$$\cdot \quad \quad \cdot$$

$$\cdot \quad \quad \cdot$$

$$\cdot \quad \quad \cdot$$

$$\delta_{k-1} = \mu_{(k)} - \mu_{(k-1)}$$

The problem, as defined by Bechhofer (1954), is then to construct a decision rule such that the probability of a complete ordering of the vector  $\underline{\mu}$  is at least some probability  $P^*$  whenever each of the  $\delta_i$  is greater than a specified real constant  $\delta_i^*$ , that is, whenever

$$\delta_1 \geq \delta_1^*, \delta_2 \geq \delta_2^*, \dots, \delta_{k-1} \geq \delta_{k-1}^*.$$

To simplify matters, we will only consider the special case where  $\delta_1^* = \delta_2^* = \dots = \delta_{k-1}^* = \delta^*$ . That is, we want the probability of a correct ordering to be at least  $\delta^*$  whenever

$$\delta_1 \geq \delta^*, \delta_2 \geq \delta^*, \dots, \delta_{k-1} \geq \delta^*. \quad 2.1$$

The region of the parameter space satisfying (2.1) is referred to as the preference zone, and all points not in this parameter space are said to lie in the indifference zone. In other words, for all points lying inside the indifference zone, we are indifferent as to which decision is made. Furthermore, a configuration in the preference zone for which the probability of a correct ordering is at an infimum, is called the least favourable condition.

The best procedure (Eaton, 1966; Savage, 1957), based on a random sample of  $n$  observations from the population, is to rank the components of the multivariate mean vector and to assert that the population with the  $i$ 'th largest sample mean is the one with the  $i$ 'th largest  $\mu$  value. The probability of a correct ranking (PCR) can then be defined as

$$\Pr\{\bar{X}_{(1)} \leq \bar{X}_{(2)} \leq \dots \leq \bar{X}_{(k-1)} \leq \bar{X}_{(k)}\} = \Pr\{\bar{X}_{(2)} - \bar{X}_{(1)} \geq 0, \bar{X}_{(3)} - \bar{X}_{(2)} \geq 0, \dots, \bar{X}_{(k)} - \bar{X}_{(k-1)} \geq 0\}$$

where  $\bar{X}_{(i)}$  is the sample mean associated with the  $i$ 'th largest mean (i.e.  $\mu_{(i)}$ ).

Before considering the solution to this problem we first consider some preliminary results which will be used throughout this chapter.

LEMMA 2.1

$$\text{Let } Y_i^j = \frac{\bar{X}_j - \bar{X}_i - (\mu_j - \mu_i)}{\sigma(2/n)^{1/2} (1-\rho_{ij})^{1/2}} \quad (i \neq j)$$

Then, for  $i' = i+1$ , the  $\{Y_i^{i'}, i = 1, \dots, k-1\}$  have a multivariate normal distribution with

$$\text{corr } \{Y_i^{i'}, Y_t^{i'}\} = \gamma_{it} = \frac{\rho_{it} + \rho_{i't'} - \rho_{i't} - \rho_{it'}}{2(1-\rho_{ii'})^{1/2} (1-\rho_{tt'})^{1/2}} \quad 2.1.1$$

Further, if  $i' = t$  then 2.1.1 reduces to

$$\text{corr } \{Y_i^{i'}, Y_{i'}^{(i')'}\} = \gamma_{ii'} = \frac{\rho_{ii'} + \rho_{i'(i')'} - \rho_{i(i)'} - 1}{2(1-\rho_{ii'})^{1/2} (1-\rho_{i'(i')'})^{1/2}}$$

PROOF Using the above definition, the result follows since

$$\begin{aligned} \text{corr } (Y_i^{i'}, Y_t^{i'}) &= \frac{\text{cov } (\bar{X}_{i'} - \bar{X}_i, \bar{X}_{t'} - \bar{X}_t)}{\sigma^2(2/n) (1-\rho_{ii'})^{1/2} (1-\rho_{tt'})^{1/2}} \\ &= \frac{\rho_{it} + \rho_{i't'} - \rho_{i't} - \rho_{it'}}{2 (1-\rho_{ii'})^{1/2} (1-\rho_{tt'})^{1/2}} \end{aligned}$$

For simplicity let  $Y_i = Y_i^{i'}$  and  $\text{corr } \{Y_i^{i'}, Y_{i'}^{(i')'}\} = \gamma_{ii'}$ .

LEMMA 2.2

Let the  $\{Y_i, i = 1, \dots, k-1\}$  be as in Lemma 2.1. Then

$$\text{PCR} \geq \Pr (Y_i > -a(n)(1-\rho_{ii'})^{-1/2}, i=1, \dots, k-1)$$

where  $a(n) = (\delta^*/\sigma) (n/2)^{1/2}$

PROOF Using Lemma 2.1 we notice that

$$\begin{aligned}
 \text{PCR} &= \Pr(\bar{X}_{(i+1)} - \bar{X}_{(i)} \geq 0, i=1, \dots, k-1) \\
 &= \Pr(Y_i > \frac{-(\mu_{(i+1)} - \mu_{(i)})(n/2)^{1/2}}{\sigma(1-\rho_{ii'})^{1/2}}, i=1, \dots, k-1) \\
 &\geq \Pr(Y_i > \frac{-\delta^*(n/2)^{1/2}}{\sigma(1-\rho_{ii'})^{1/2}}, i=1, \dots, k-1). \quad 2.1.2
 \end{aligned}$$

Our task is to determine the least favourable and most favourable configuration for the right hand side of (2.1.2).

### 2.1.2 CASE OF EQUAL CORRELATIONS

When the off diagonal elements of the correlation matrix,  $R$ , are known to be equal to a common unknown  $\rho$  ( $-1/(k-1) \leq \rho \leq 1$ ),

$$\gamma_{it} = \begin{cases} -1/2 & \text{for } t = i, \\ 0 & \text{otherwise} \end{cases} \quad 1 \leq i < t \leq k-1$$

and the minimisation and maximisation of (2.1.2) occurs when  $\rho = -1/(k-1)$  and  $1$  respectively. Notice that in both these cases the multivariate distribution of  $\underline{X}$  is degenerate. Explicitly we may write

$$\inf \text{PCR} = \Pr(Y_i > -a(n)((k-1)/k)^{1/2}, i=1, \dots, k-1)$$

$$\text{and } \sup \text{PCR} = \Pr(Y_i > -\infty, i=1, \dots, k-1) = 1$$

### 2.1.3 CASE k=2

Although a particular case of the preceding section, since  $\underline{X}$  has a bivariate normal distribution, 2.1.2 reduces to a univariate normal integral, the minimum and maximum of which clearly occurs when  $\rho = -1$  and  $1$  respectively. Therefore

$$\inf \text{PCR} = \Pr(Y_1 > -a(n) 2^{-1/2}) \text{ and} \quad 2.1.3$$

$$\sup \text{PCR} = \Pr(Y_1 > -\infty) = 1.$$

where  $Y_1$  is a standard univariate normal random variable.

### 2.1.4 CASE k=3

In this particular instance, for all permissible values of  $\rho_{12}$ ,  $\rho_{13}$ , and  $\rho_{23}$ , we are interested in identifying the forms of the covariance matrix for which the probability of a correct ranking will be maximised and minimised. Thus, for  $k=3$ , we would like to identify the covariance matrix associated with the maximisation and minimisation of

$$\text{PCR} = \Pr(Y_1 > -a(n)(1-\rho_{12})^{-1/2}, Y_2 > -a(n)(1-\rho_{23})^{-1/2}) \quad 2.1.4$$

where  $\{Y_1, Y_2\}$  have a standard bivariate normal distribution with

$$\text{corr}\{Y_1, Y_2\} = \gamma = \frac{\rho_{12} + \rho_{23} - \rho_{13} - 1}{2(1 - \rho_{12})^{1/2}(1 - \rho_{23})^{1/2}} \quad 2.1.5$$

Since  $R$  is positive semi-definite, the region in 3 dimensional Euclidean space where  $R$  is defined is given by the ellipsoid

$\det R \geq 0$ . That is,  $\rho_{12}$ ,  $\rho_{13}$ ,  $\rho_{23}$  are constrained by the equation

$$1 + 2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 \geq 0 \quad 2.1.6$$

subject to  $\rho_{ij}^2 \leq 1$  ( $i \neq j$ ).

LEMMA 2.3

$$\frac{\partial \text{PCR}}{\partial \rho_{13}} < 0 \text{ for } \rho_{12} \neq 1 \text{ and } \rho_{23} \neq 1.$$

PROOF Let  $f_1(y_1, y_2)$  be the probability density function of  $\{Y_1, Y_2\}$ , i.e.

$$f_1(y_1, y_2) = (2\pi)^{-1} (1-\gamma^2)^{-1/2} \exp \{-1/2(1-\gamma^2)^{-1} (y_1^2 + y_2^2 - 2\gamma y_1 y_2)\}.$$

Then using the result given by Plackett (1954) that

$$\frac{\partial}{\partial \gamma} f_1(y_1, y_2) = \frac{\partial^2}{\partial y_1 \partial y_2} f_1(y_1, y_2), \text{ we have}$$

$$\begin{aligned} \frac{\partial \text{PCR}}{\partial \rho_{13}} &= \int_{-a_1}^{\infty} \int_{-a_2}^{\infty} \frac{\partial^2}{\partial y_1 \partial y_2} f_1(y_1, y_2) dy_1 dy_2 \frac{\partial \gamma}{\partial \rho_{13}} \\ &= f_1(-a_1, -a_2) (-1/2) (1-\rho_{12})^{-1/2} (1-\rho_{23})^{-1/2} < 0 \end{aligned}$$

$$\text{where } a_1 = a(n) (1 - \rho_{12})^{-1/2} \text{ and } a_2 = a(n) (1 - \rho_{23})^{-1/2}$$

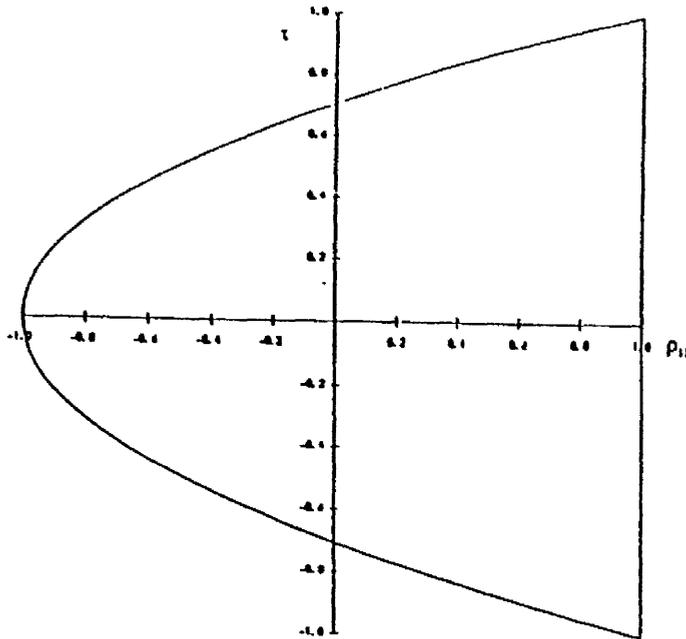
QED

For the case of equal correlation it was noted in section 2.1.2 that PCR is maximised when  $\rho_{12} = \rho_{23} = \rho_{13} = 1$ . To solve 2.1.3 for the more general case, it will be assumed, in the remaining part of this section, that neither  $\rho_{12}$  nor  $\rho_{23}$  equals unity.

Now suppose  $\rho_{12}$  and  $\rho_{23}$  are fixed, then by the previous Lemma, PCR is at a maximum when  $\rho_{13}$  takes its minimal possible value and at a minimum when  $\rho_{13}$  takes its maximal possible value, these being the smaller and larger roots respectively of the quadratic equation  $\det R = 0$ .

As an example, consider the special case where  $\rho_{12} = \rho_{23} = \tau$ . The region in which  $\det R \geq 0$  is illustrated in Figure 2.1.

FIGURE 2.1



Notice that as  $\tau \rightarrow 1$  the smaller root of the quadratic equation,  $\det R=0$ , tends to 1. In such regions of the parameter space, the configuration of the covariance matrix will be very close to the most favourable condition.

In the case for given  $\rho_{12}$  and  $\rho_{23}$  not equal to 1

$$\inf \text{PCR occurs when } \rho_{13} = \rho_{12} \rho_{23} + (1-\rho_{12}^2)^{1/2} (1-\rho_{23}^2)^{1/2} \text{ and}$$

$$\sup \text{PCR occurs when } \rho_{13} = \rho_{12} \rho_{23} - (1-\rho_{12}^2)^{1/2} (1-\rho_{23}^2)^{1/2}$$

To proceed with the minimisation and maximisation of PCR we define the Lagrangean function

$$F = \text{PCR} + \lambda \det R.$$

Now the correlation matrix  $R$ , subject to the constraint  $\det R = 0$ , must satisfy the following four equations:

$$\frac{\partial F}{\partial Q_{13}} = f_{\gamma}(-a_1, -a_2) (-1/2) (1 - Q_{12})^{-1/2} (1 - Q_{23})^{-1/2} + 2 \lambda (Q_{12} Q_{23} - Q_{13}) = 0 \quad 2.1.7$$

$$\begin{aligned} \frac{\partial F}{\partial Q_{12}} = f_{\gamma}(-a_1, -a_2) \frac{(1 + Q_{23} - Q_{12} - Q_{13})}{4(1 - Q_{12})^{3/2} (1 - Q_{23})^{1/2}} + \frac{a(n)}{2(1 - Q_{12})^{3/2}} \int_{-a_2}^{\infty} f_{\gamma}(-a_1, y_2) dy_2 \\ + 2\lambda(Q_{12} Q_{23} - Q_{13}) = 0 \end{aligned} \quad 2.1.8$$

$$\begin{aligned} \frac{\partial F}{\partial Q_{23}} = f_{\gamma}(-a_1, -a_2) \frac{(1 + Q_{12} - Q_{23} - Q_{13})}{4(1 - Q_{12})^{1/2} (1 - Q_{23})^{3/2}} + \frac{a(n)}{2(1 - Q_{23})^{3/2}} \int_{-a_1}^{\infty} f_{\gamma}(y_1, -a_2) dy_1 \\ + 2 \lambda (Q_{12} Q_{23} - Q_{13}) = 0 \end{aligned} \quad 2.1.9$$

where  $a_1 = a(n) (1 - Q_{12})^{-1/2}$  and  $a_2 = a(n) (1 - Q_{23})^{-1/2}$

$$\frac{\partial F}{\partial \lambda} = \det R = 0 \quad 2.1.10$$

By the symmetry of equations (2.1.8) and (2.1.9) with respect to  $Q_{12}$  and  $Q_{23}$ , one identifies a solution of the form

$$Q_{12} = Q_{23} = \tau$$

Using (2.1.10) in conjunction with Lemma 2.3, implies that for such solutions and fixed  $\tau$

- (a) PCR is minimised when  $Q_{13} = 1$  and
- (b) PCR is maximised when  $Q_{13} = 2\tau^2 - 1$

Furthermore, by substituting these values into (2.1.5) we can obtain the corresponding value of  $\gamma$  for the cases above

- (a)  $\gamma = -1$  and
- (b)  $\gamma = \tau$

Consider case (a) first. In this instance PCR is given by

$$\text{PCR} = \int_{-a}^{\infty} \int_{-a}^{\infty} f_{-1}(y_1, y_2) dy_1 dy_2 ,$$

where  $a = a(n) (1 - \tau)^{-1/2}$ , will be minimised when  $\tau = -1$ . Thus the least favourable configuration of R occurs when

$$\rho_{12} = \rho_{23} = -1 \quad \text{and} \quad \rho_{13} = 1$$

Now consider case (b). Since  $\rho_{12} = \rho_{23}$  equations (2.1.8) and (2.1.9) become identical and therefore to identify the correlation structure which will maximise the probability of a correct ranking we must eliminate the Lagrange multiplier between equation (2.1.7) and (2.1.8).

After a little algebra we arrive at the equation

$$\int_{-a}^{\infty} f_r(-a, y) dy = \frac{(1-\tau)^{3/2}}{a(n)} f_r(-a, -a) \quad (2.1.12)$$

where  $a = a(n) (1 - \tau)^{-1/2}$ .

To proceed to the solution, we use the factorisation

$$f(x, y) = f(y | x) f(x)$$

for the density inside the integral. The LHS of equation can then be rewritten as

$$\begin{aligned} & f(-a) \int_{-a}^{\infty} f(y | -a) dy \\ &= (2\pi)^{-1/2} \exp \{-1/2 a^2\} \times \int_{-b}^{\infty} (2\pi)^{-1/2} \exp \{-1/2 z^2\} dz \end{aligned}$$

where  $z = \frac{y + \tau a}{(1-\tau)^{1/2} (1+\tau)^{1/2}}$ ,  $a = a(n) (1-\tau)^{-1/2}$  and  $b = a(n) (1+\tau)^{-1/2}$ .

After further simplification (2.1.12) reduces to

$$\int_{-b}^{\infty} (2\pi)^{-1/2} \exp\{-z^2/2\} dz = \frac{(2\pi)^{-1/2}}{b} \frac{(1 - \tau)}{(1 + \tau)} \exp\{-b^2/2\} \quad 2.1.13$$

where  $b = a(n) (1 + \tau)^{-1/2}$

Therefore, for a given  $a(n)$  we can identify from (2.1.13) the correlation structure which will attain the maximum probability of a correct ranking.

Numerical evaluations of  $\tau$  and PCR are illustrated in Table 2.1, for given values of  $a(n)$ . Notice that the evaluation of  $\tau$  from (2.1.13) is greatly simplified if we actually fix values of  $b$  rather than  $a(n)$ . The value of  $a(n)$  can then be obtained after solving (2.1.13) for  $\tau$ .

As expected, as  $a(n)$  increases, the probability of a correct ranking increases:  $a(n)$  will increase if the sample size increases or the variance of the components decreases. Furthermore, if we allow  $a(n) \rightarrow \infty$  then equations (2.1.7), (2.1.8), (2.1.9) and (2.1.10) become

$$\begin{aligned} 2 \lambda ( \varrho_{12} \varrho_{23} - \varrho_{13} ) &= 0 , \\ 2 \lambda ( \varrho_{13} \varrho_{23} - \varrho_{12} ) &= 0 , \\ 2 \lambda ( \varrho_{12} \varrho_{13} - \varrho_{23} ) &= 0 , \text{ and} \\ \det R &= 0 \end{aligned}$$

which has solutions

$$\begin{aligned} \text{(a)} \quad \varrho_{12} &= \varrho_{23} = \varrho_{13} = 1 \\ \text{(b)} \quad \varrho_{12} &= \varrho_{23} = -1, \varrho_{13} = 1 \end{aligned} ,$$

which correspond to the most and least favourable configurations respectively.

Also notice that if we were simply to randomly rank the three components the associated probability of attaining a correct ranking would be  $1/6$ . However, as  $a(n)$  tends to zero, PCR\* can become less than  $1/6$  and so in such instances we would be better off just selecting any ranking.

**TABLE 2.1** Values of the probability of a correct ranking for given  $a(n)$  when  $k = 3$ .

$a(n)$	$\tau$	$Q_{13}$	PCR
6.00	0.997	0.998	$\cong 1.00$
2.03	0.83	0.38	0.81
1.24	0.55	-0.40	0.67
0.77	0.33	-0.89	0.55
0.51	0.05	-0.997	0.42
0.28	-0.29	-0.92	0.29
0.14	-0.54	-0.71	0.20
0.05	-0.76	-0.42	0.13
0.02	-0.88	0.55	0.01

### 2.1.5 CASE $k > 3$

In the previous section, we were able to derive results through numerical computations about the configurations of the correlation matrix, maximising and minimising the PCR. While tables of general multivariate normal integrals of dimension greater than 2 are not readily available, a lower bound on the PCR may be obtained using Bonferroni's inequality. Moreover, Bonferroni's inequality can be used to derive a lower bound for the PCR in the less restrictive case where not all the components of the multivariate mean vector have the same variance.

For  $p$  events,  $E_1, \dots, E_p$ , Bonferroni's inequality may be written explicitly as

$$\Pr \left( \bigcap_{i=1}^p E_i \right) \geq \sum_{i=1}^p \Pr (E_i) - (p-1)$$

Let  $i' = i + 1$ , as before, and defining

$$Y_i = \frac{\bar{X}_{i'} - \bar{X}_i - (\mu_{i'} - \mu_i)}{[\text{Var}(\bar{X}_{i'} - \bar{X}_i)]^{1/2}} = \frac{\bar{X}_{i'} - \bar{X}_i - (\mu_{i'} - \mu_i)}{[(\sigma_{ii}^2 + \sigma_{i'i}^2 - \sigma_{ii'})/n]^{1/2}}$$

we may use the Bonferroni inequality to derive a conservative estimate of the

$$\text{PCR} = \Pr \{ (Y_i > -a_i), i = 1, \dots, k-1 \}$$

$$\geq \sum_{i=1}^{k-1} \Pr (Y_i > -a_i) - (k-2) \quad 2.1.14$$

where  $a_i = \delta_i^* / [(\sigma_{ii}^2 + \sigma_{i'i}^2 - 2\sigma_{ii'})/n]^{1/2}$

The probability statement (2.1.14) can be evaluated for known  $\Sigma$  and  $\hat{\Sigma}$ , a sample estimate of the true covariance structure. If we assume that  $\Sigma = \sigma^2 \Gamma$ , where  $\Gamma$  denotes the correlation of  $\underline{X}$ , then

$$\text{PCR} = \Pr (Y_i > -a_i, i = 1, \dots, k-1)$$

$$\geq \sum_{i=1}^{k-1} \Pr (Y_i > -a_i) - (k-2)$$

$$\geq \sum_{i=1}^{k-1} \Pr (Y_i > -a(n)2^{-1/n}) - (k-2) \quad \text{by 2.1.3}$$

and of course, the maximum will occur when all correlations between

the components are equal to 1. Thus for any general multivariate normal vector we may obtain a lower bound for the probability of a correct ranking. In autoradiographic studies, where the dimension of the observation vector will often exceed 60, the lower bound of PCR, based on the Bonferroni inequality, is likely to be extremely small. Indeed for  $k = 60$ ,  $\Pr (Y_i > - a(n)2^{-1/2}, i=1, \dots, 59)$  would have to be at least 58/59 (or equivalently for  $a(n)$  to be greater than 3) for the lower bound to be positive. If on the other hand, a constant correlation model is a good working approximation then

$$\begin{aligned} \text{PCR} &= \Pr (Y_i > - a_i, i = 1, \dots, k-1) \\ &\geq \Pr (Y_i > - a(n) \left( \frac{k-1}{k} \right)^{1/2}; i = 1, \dots, k-1) \\ &\geq \sum_{i=1}^{k-1} \Pr (Y_i > - a(n) \left( \frac{k-1}{k} \right)^{1/2}) - (k-2) \end{aligned}$$

for which  $a(n)$  would have to be greater than 2.14 for the lower bound to be positive, assuming  $k=60$ , or  $a(n)$  to be greater than 2.22 for a guaranteed probability of at least 1/6.

If we confine ourselves to studies within particular neurological circuits or investigations, where the dimensionality of the problem is greatly reduced, and we are willing to assume the correlations between regions are positive, then there may be some possibility of obtaining a ranking with a certain degree of reliability. However, for general investigations a perfect ranking would appear to be an unrealistic aim. One possibility would be to focus on identifying extreme regions corresponding to drug side effects.

In the following section we address and study factors which affect the reliability of alternative approaches for quantifying the hierarchy of drug response.

## 2.2 INVESTIGATING THE PATTERN OF RESPONSE

In this section we consider the problem of investigating the hierarchy of regional responses, of the variable of interest, to treatment with a drug and possibly in comparing these hierarchies between drugs, using data from autoradiographic studies. For illustration, in the remainder of this chapter, the variable of interest will be assumed to be local cerebral glucose utilisation, measured in units of  $\mu\text{mol}/100\text{g}/\text{min}$ . As mentioned previously, a variety of methods, some of which may be potentially misleading, have evolved in the literature to quantify the regional responses. In this section we have a slight change in notation whereby we let  $Y_{i,j,k}$  denote the measurement of interest in the  $j^{\text{th}}$  brain region ( $j=1, \dots, J$ ) on the  $i^{\text{th}}$  animal ( $i=1, \dots, n_k$ ) in the  $k^{\text{th}}$  treatment group ( $k=1, \dots, k_1$ ;  $k=1$  representing the control group). Since, in autoradiographic studies, the variance of  $Y_{i,j,k}$  will tend to increase as its expected value increases, we will work with logged data to stabilise the variance, that is

$$X_{i,j,k} = \log_e (Y_{i,j,k}).$$

Let  $E(X_{i,j,k}) = \mu_{j,k}$ , then for each of the  $j$  regions we investigate two measures of response, namely

$$f_{1j} = \sum_{k=1}^K (\mu_{j,k} - \mu_{j,1})^2 \text{ and}$$

$$f_{2j} = \sum_{k=1}^K |\mu_{j,k} - \mu_{j,1}|$$

which we estimate by

$$\hat{f}_{1j} = \sum_{k=1}^K (\bar{X}_{.jk} - \bar{X}_{.j1})^2 \text{ and}$$

$$\hat{f}_{2j} = \sum_{k=1}^K |\bar{X}_{.jk} - \bar{X}_{.j1}| \text{ respectively}$$

Notice that both these approaches measure the response relative to the control group and allow for bidirectional changes in metabolic activity, hence providing a more sensible ordering of the regions. The divisor which is present in the F-statistic is also omitted since it may introduce spurious variability to the ranking problem.

One drawback with this type of approach is that there would not appear to be any exact procedure to assess the ordering of the  $f_{1j}$ 's or  $f_{2j}$ 's. Note that in the case of the  $f_{1j}$ 's with  $k=2$ , the problem at hand is concerned with ranking the elements of a multivariate noncentral chi-squared vector. However, although there are no exact procedures available, we may assess the reliability of the derived rankings using simulation techniques.

### 2.2.1 EXAMPLE DATA SETS

To investigate the applicability of  $f_{1j}$  and  $f_{2j}$  as measures of response in general experimental situations, we consider two data sets.

#### ELECTRICAL STIMULATION OF THE RAPHE NUCLEUS

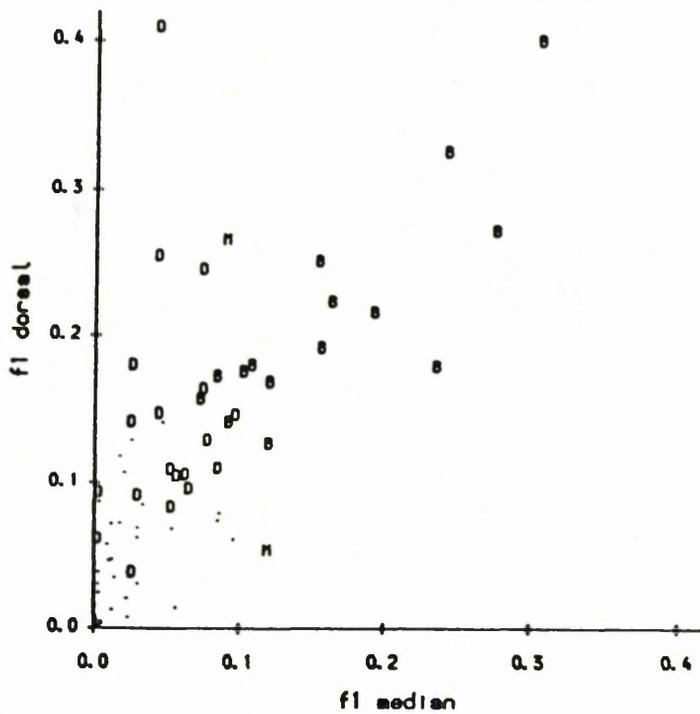
In an experiment to investigate the repercussions of activating the ascending serotonergic pathways, local cerebral glucose utilisation was measured in conscious rats after electrical stimulation of the

rostral (dorsal or median) raphe nuclei. Glucose use in 91 discrete brain regions was studied in three groups of rats, dorsal raphe stimulated (n=8); median raphe stimulated (n=5); and a control group of animals in which the electrode was placed in either the dorsal or median raphe nucleus but in which no current was applied (n=7).

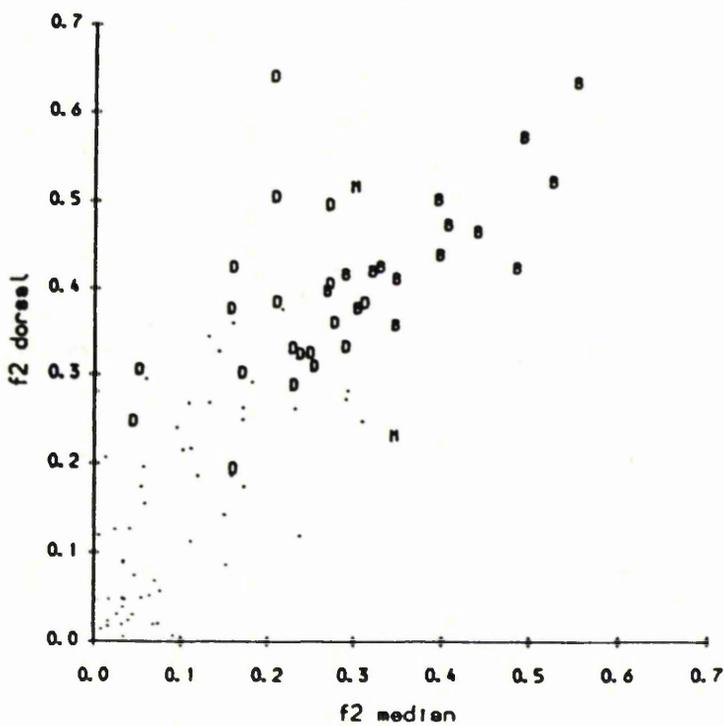
Furthermore, to investigate the specificity of the raphe-induced changes in integrated functional activity to the activation of serotonergic neurones, a group of rats that had underwent pretreatment with 5,7-dihydroxytryptamine (5,7 DHT), a serotonergic neurotoxin, was later subjected to stimulation of the dorsal raphe (n=6) and compared to (1) animals that received 5,7 DHT alone (n=5) - for illustration purposes label this comparison 5,7 DHT(1) - and (2) the control group of animals in which no 5,7 DHT was administered (labelled 5,7 DHT(2)). In Figures 2.2.1(i) - (viii) we plot the estimated  $f_{1,}$ 's and  $f_{2,}$ 's for each of the 91 brain regions studied, for the pairs of treatment (dorsal raphe stimulation, median raphe stimulation), (dorsal raphe stimulation, 5,7 DHT(1)), (median raphe stimulation, 5,7 DHT(1)) and (5,7 DHT(1), 5,7 DHT(2)). Brain regions in which a two sample t-test identified significant differences from the control group, using a significance level of 0.01, are coded D, M,  $\square$  and  $\Delta$  for the dorsal raphe, median raphe, 5,7 DHT(1) and 5,7 DHT(2) comparisons respectively. Regions in which both comparisons were significant are coded B in the figures.

Figures 2.2.1(i)-(viii) indicate electrical stimulation of the raphe nucleus increased local cerebral glucose utilisation in a number of brain regions, with dorsal raphe stimulation being systematically

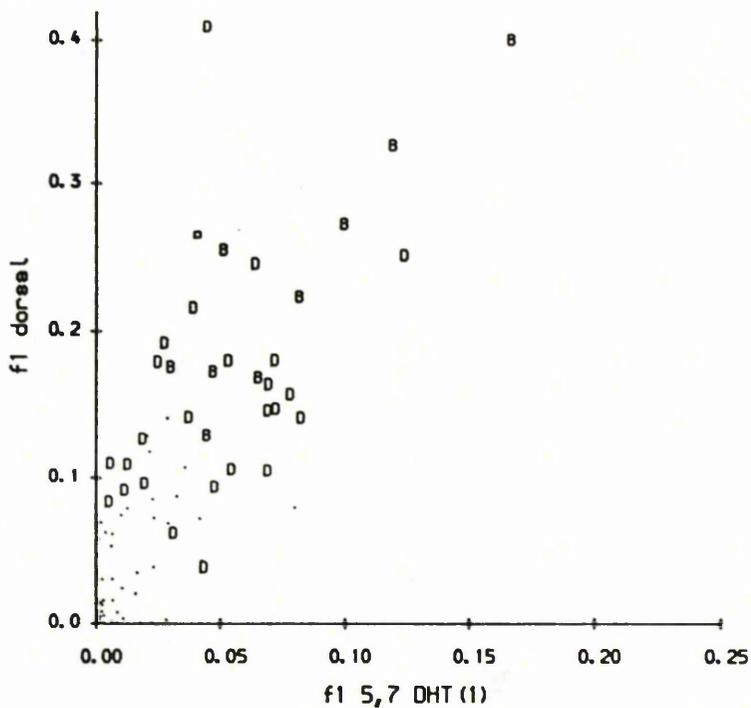
Figure 2.2.1 Relationship between the responsiveness of regional glucose use after stimulation of the dorsal median raphe and dorsal raphe after treatment with 5,7 - DHT.



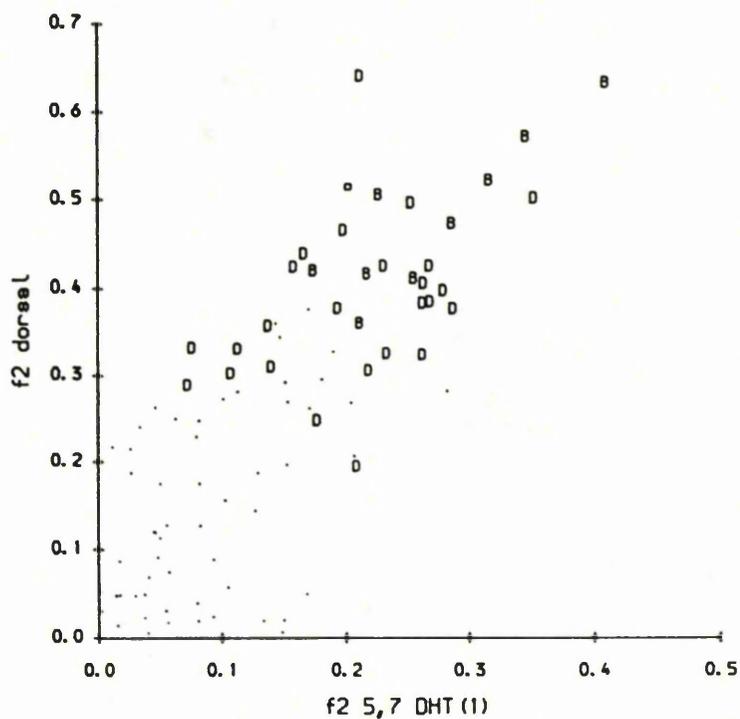
(i) f1 dorsal raphe and median raphe



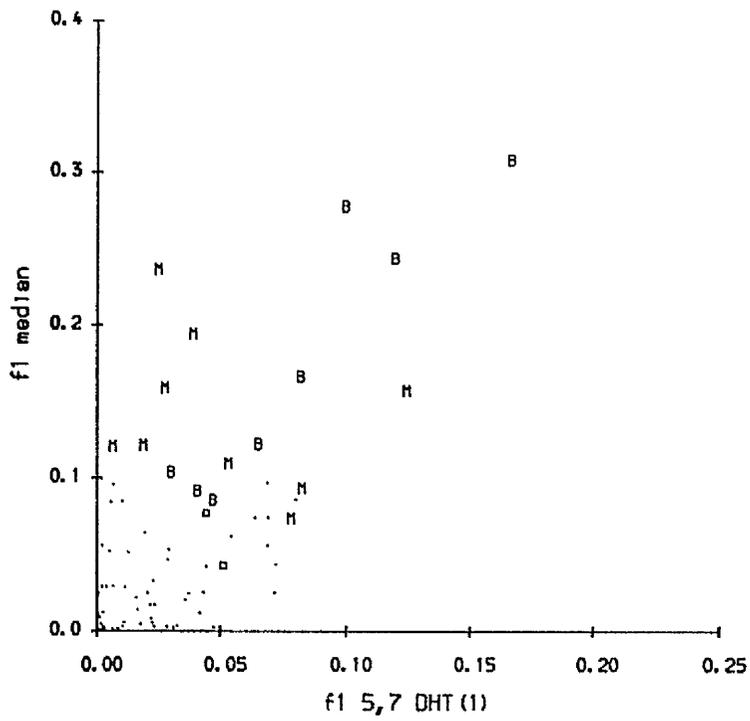
(ii) f2 dorsal and median raphe



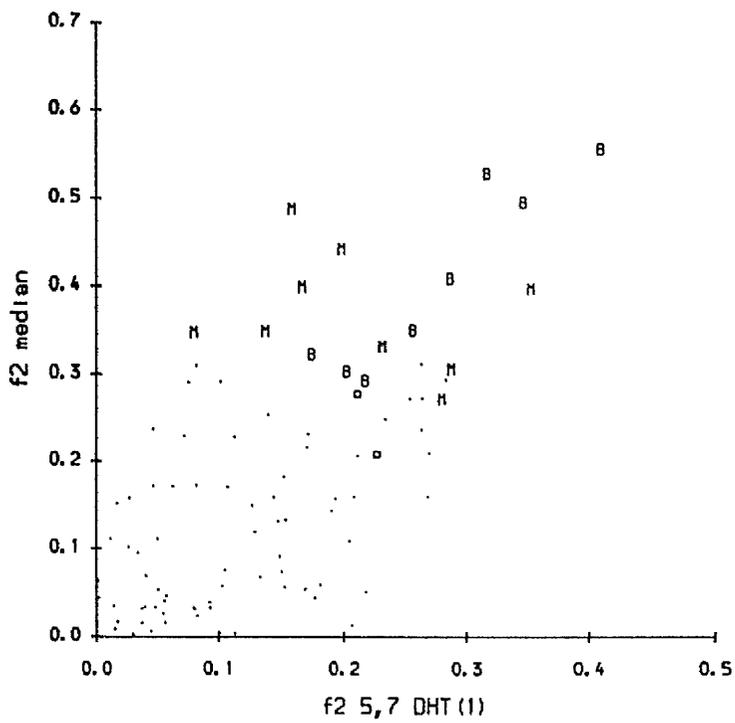
(iii) f1 dorsal raphe and 5,7 - DHT(1)



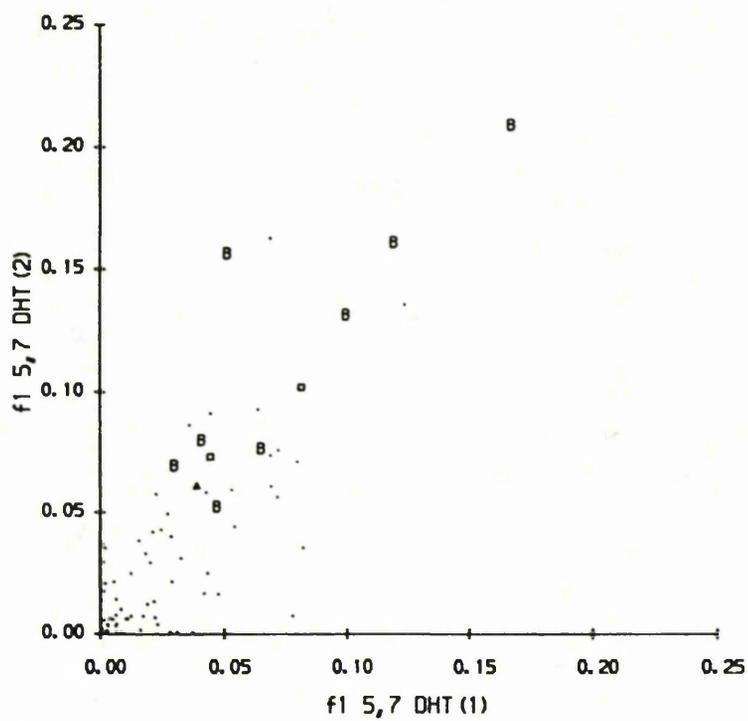
(iv) f2 dorsal raphe and 5,7 - DHT(1)



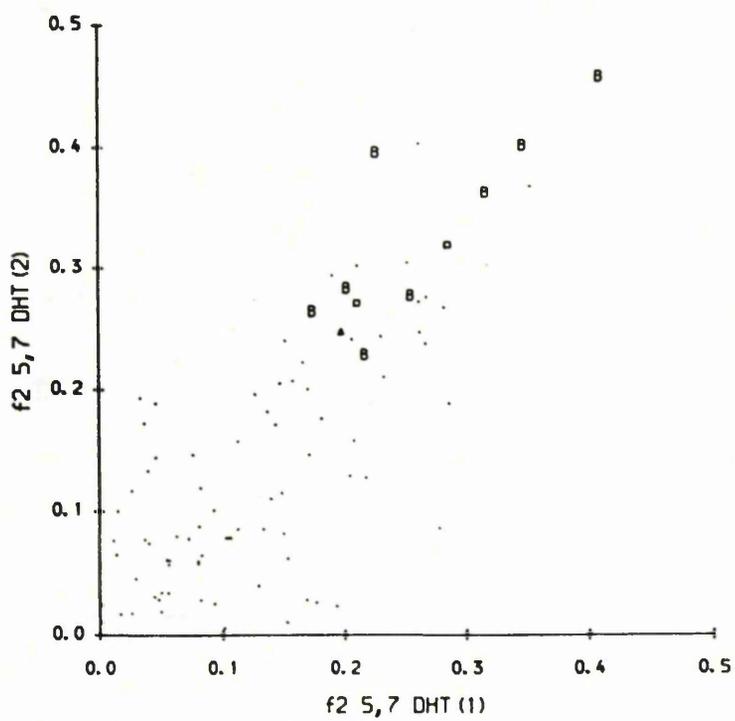
(v) f1 median raphe and 5,7 - DHT(1)



(vi) f2 median raphe and 5,7 - DHT(1)



(vii)  $f1\ 5,7 - DHT(2)$  and  $5,7 - DHT(1)$



(viii)  $f2\ 5,7 - DHT(2)$  and  $5,7 - DHT(1)$

more effective than median raphe stimulation. No brain regions showed a significant decrease in glucose use and with only a few exceptions, increases in local cerebral glucose utilisation were not observed by administering 5,7 DHT prior to stimulation of the dorsal raphe.

Although, there were fewer significant differences observed with certain comparisons, generally speaking the hierarchy or ordering of the response for all four comparisons were similar for both the  $f_{1j}$  and  $f_{2j}$  procedures (Pearson's product moment correlation coefficient, calculated on the  $\hat{f}_{1j}$ 's and  $\hat{f}_{2j}$ 's are provided in Table 2.2.1 as a useful distance measure between the rankings, for each pair of treatments). However, brain regions indicating an altered position in the hierarchy are easily identifiable. For instance, dorsal raphe stimulation increased local cerebral glucose use in both the dorsal raphe itself and in the median raphe, whilst in contrast median raphe stimulation or dorsal raphe stimulation after administration of 5,7 DHT failed to elicit significant changes in glucose use in the dorsal raphe.

Table 2.2.1 Pearson product moment correlation coefficients in raphe stimulation study.

	$f_1$	$f_2$
$\hat{p}$ (dorsal raphe, median raphe)	0.76	0.81
$\hat{p}$ (dorsal raphe, 5,7 DHT(1))	0.78	0.78
$\hat{p}$ (median raphe, 5,7 DHT(1))	0.70	0.67
$\hat{p}$ (5,7 DHT(1), 5,7 DHT(2))	0.83	0.78

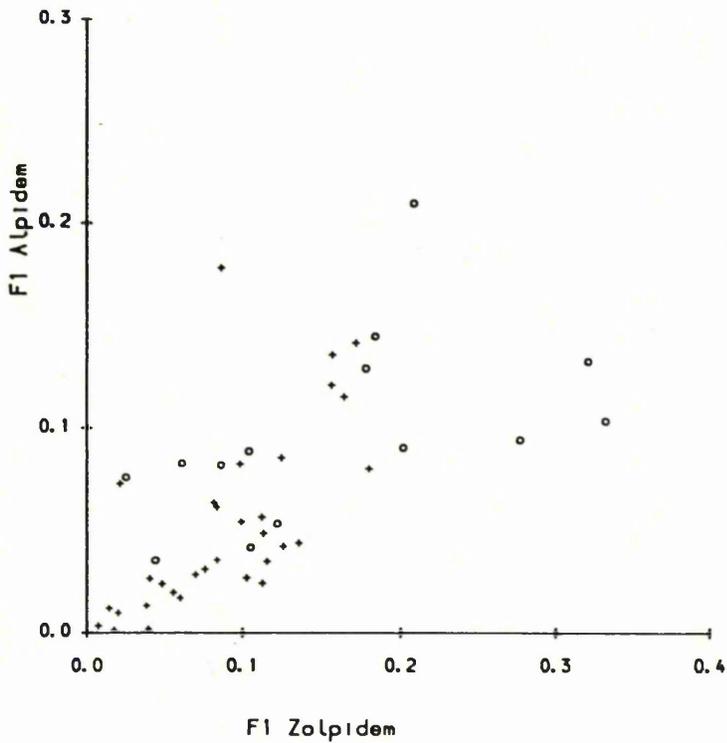
ALPIDEM/ZOLPIDEM/DIAZEPAM STUDY

In this situation we consider the investigation of 3 agents, namely Alpidem, Zolpidem and Diazepam, acting on benzodiazepam receptors. The effects of all three agents upon local cerebral glucose utilisation were analysed in 46 discrete brain regions at each of three doses (10-100 mg/kg i.v.), (0.1-1.0 mg/kg i.v) and (0.1-1.0 mg/kg i.v) respectively. Sample sizes in the (control, dose 1, dose 2 and dose 3) groups for each drug were respectively (2,5,5,5), (10,6,7,6) and (9,5,5,5,).

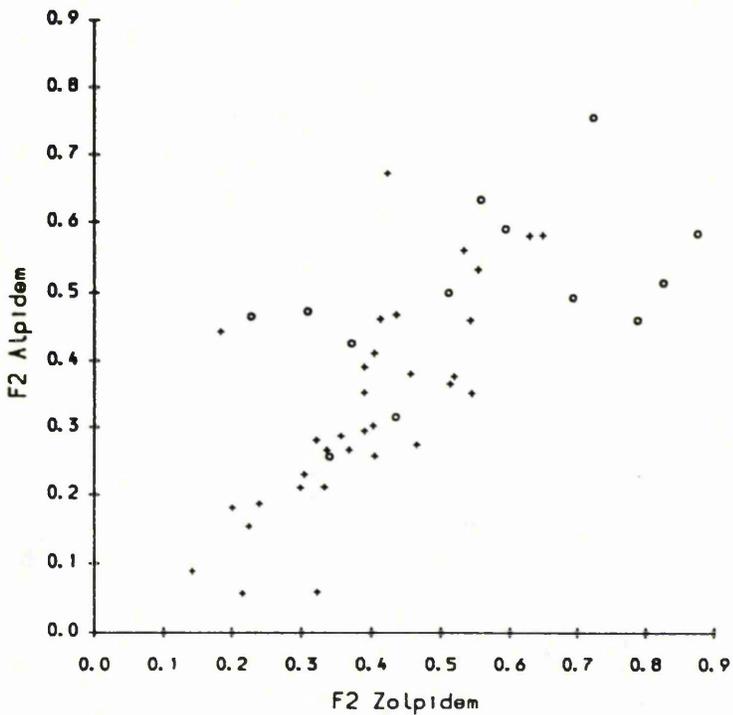
Experiments involving the agents Alpidem and Zolpidem were conducted at Lers Synthelabo, Bagneux with autoradiograms analysed on a Quantimet 720 image analyser. The corresponding investigations of Diazepam were conducted at the Wellcome Surgical Institute, Glasgow, with autoradiograms analysed on a Quantimet 970 image analyser. In both centres, levels of local cerebral glucose utilisation were calculated using identical software, utilising the operational equation and kinetic constants derived by Sokoloff et al (1977).

The estimated  $f_{1,}$ 's and  $f_{2,}$ 's, for each of the 46 regions studied, are displayed in Figures 2.2.2(i)-(vi), for the pairs of drugs (Alpidem, Zolpidem), (Alpidem, Diazepam) and (Zolpidem, Diazepam). Regions involved in the "Sleep" and "Anxiolytic/Angiogenic" circuits, where there were strong prior expectations about the relationships which should be observed are coded 0 in the figures. Pearson's product moment correlation coefficients are given in Table 2.2.2.

Figure 2.2.2 Relationship between the responsiveness of regional cerebral glucose use to Alpidem, Zolpidem and Diazepam.

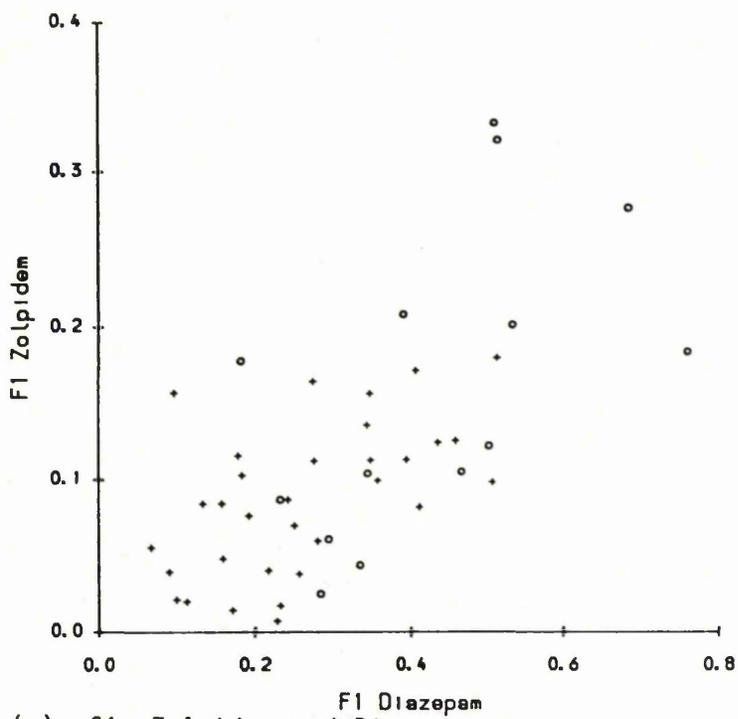


(i) f1 Alpidem and Zolpidem

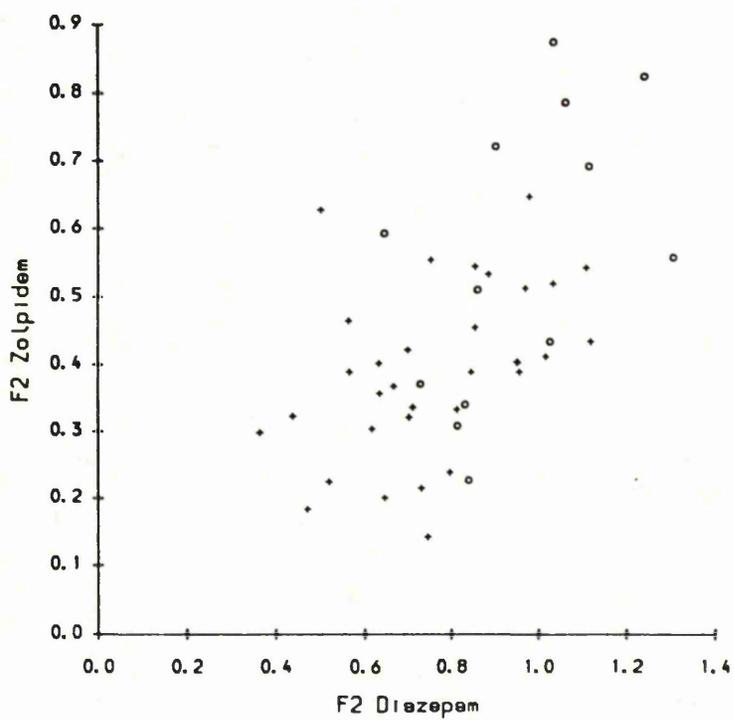


(ii) f2 Alpidem and Zolpidem





(v) f1 Zolpidem and Diazepam



(vi) f2 Zolpidem and Diazepam

Figures 2.2.2(i)-(vi) indicate that the pattern or hierarchy of response for the three drugs are similar, although the association between the  $f_{1j}$ 's and  $f_{2j}$ 's is not especially strong, particularly for Alpidem and Zolpidem (Figures 2.2.2(i) and (ii)). The regions involved in the "Sleep" and "Anxiolytic/Anxiogenic" circuits, denoted by 0, exhibit a tendency to possess higher  $f_{1j}$ 's and  $f_{2j}$ 's for all three drugs, although functional activity in these circuits would appear to have been affected less after the administration of Alpidem. Regions affected by Alpidem but largely unaffected by the other drugs could also be detected.

Table 2.2.2 Pearson's product moment correlation coefficients in Alpidem/Zolpidem/Diazepam Study.

	$f_1$	$f_2$
$\hat{p}$ (Alpidem, Zolpidem)	0.65	0.70
$\hat{p}$ (Alpidem, Diazepam)	0.40	0.44
$\hat{p}$ (Zolpidem, Diazepam)	0.65	0.56

### 2.2.2 RELIABILITY OF THE DERIVED RANKINGS

A simulation study was carried out to investigate the factors influencing the reliability of the estimated rankings of the  $f_{1j}$ 's and  $f_{2j}$ 's within a treatment, and the estimated correlation between rankings for different treatments or drugs. Our preliminary results are based on a constant correlation model which we believe will provide a useful approximation in the analysis of data of this type.

Explicitly, our model is

$$X_{ijk} = \mu_{jk} + a_{ik} + \epsilon_{ijk} \quad (2.2.2.1)$$

where  $X_{ijk}$  denotes the logged measure of response on individual  $i$ ,  $a_{ik}$  denotes a random animal effect assumed constant over regions in the same animal and  $\epsilon_{ijk}$  denotes independent random effects in each animal and region. In our simulations we will further assume that  $a_{ik} \sim N(0, \sigma^2)$  and  $\epsilon_{ijk} \sim N(0, \tau^2)$ . Notice that within this structure the variance covariance matrix for the observations on the vector of regional observations is

$$\Sigma = \sigma^2 \underline{1} \underline{1}^T + \Omega$$

where  $\underline{1}$  is a vector of ones, and  $\Omega$  represents the within animal covariance matrix.

Unbiased estimates of the variance components were obtained for each treatment in the rostral raphe stimulation study and each agent in the Alpidem/Zolpidem/Diazepam study using the observed data, and are given in Tables 2.2.3(i) and 2.2.3(ii) respectively. Tables 2.2.3 (i) and (ii) show that in many cases the estimates of  $\sigma$  and  $\tau$  are roughly of the same magnitude and are also similar for different drugs. Noticeable exceptions are the estimates of  $\sigma$  for the 5,7 DHT controls and alpidem, and the estimate of  $\tau$  from animals receiving both 5,7 DHT and dorsal raphe stimulation.

Table 2.2.3(i) Unbiased estimates of the variance components in the raphe stimulation study.

	$\hat{\sigma}$	$\hat{\tau}$
dorsal and median raphe controls	0.103	0.078
dorsal raphe	0.104	0.133
median raphe	0.102	0.103
5,7 DHT controls	0.042	0.110
5,7 DHT and dorsal raphe stimulation	0.072	0.186

Table 2.2.3(ii) Unbiased estimates of the variance components in the Alpidem/Zolpidem/Diazepam study

	$\hat{\sigma}$	$\hat{\tau}$
Alpidem	0.113	0.072
Zolpidem	0.067	0.071
Diazepam	0.077	0.089

### 2.2.3 SIMULATION ALGORITHM

To investigate the reliability of the estimated rankings of the  $f_{1j}$ 's and  $f_{2j}$ 's, based on their appropriate estimates, a simulation study, as detailed below, was carried out for each experimental paradigm.

- (i) The  $f_{1j}$ 's and  $f_{2j}$ 's are estimated by substituting  $\bar{X}_{.jk}$  for  $\mu_{jk}$  in the formulas for  $f_{1j}$  and  $f_{2j}$ , respectively, and are used to derive the observed ranking in the data set. Correlations between the sets of  $f_{1j}$ 's and  $f_{2j}$ 's are calculated for each relevant pair of drugs or physiological treatment.

- (ii) a data set is simulated based on the model (2.2.2.1) with parameters estimated from the observed data.
- (iii) a simulated set of estimated  $f_{1j}$ 's and  $f_{2j}$ 's are calculated. Denote these estimates by  $\hat{F}_{1j}$  and  $\hat{F}_{2j}$  respectively.
- (iv) Pearson's product moment correlation coefficient is calculated between the  $\hat{f}_{1j}$ 's and  $\hat{F}_{1j}$ 's and between the  $\hat{f}_{2j}$ 's and  $\hat{F}_{2j}$ 's, as a measure of the distance between the original ranking and the simulated ranking.

For each pair of drugs or treatment

- (v) the correlation between the two sets of  $\hat{F}_{1j}$ 's and  $\hat{F}_{2j}$ 's is calculated as a measure of the association between the two sets of rankings.
- (vi) steps (ii), (iii), (iv) are repeated 2000 times and the correlations stored.

To investigate the effect of increasing the magnitude of the variance components on the reliability of the ranking of the regions, a further simulation study was conducted similar to the one above with the exception that the general structure of the variance covariance matrix for the vector of regional observations was given by

$$\Sigma = a \sigma^2 \mathbf{1} \mathbf{1}^T + b \Omega$$

where  $\Omega$  is defined as above and  $a$  and  $b$  are scalars varying from study to study. As before, the  $\bar{X}_{.jk}$ 's have been taken as estimates of the  $\mu_{jk}$ 's.

In Figures 2.2.3(i)-(iv) the median correlation from the 2000 simulations is plotted against the combination of values used for  $a$  and  $b$  for the investigation of the within and between treatment

Figure 2.2.3(i) Median correlations for the within treatment rankings, using f1 (rostral raphe study).

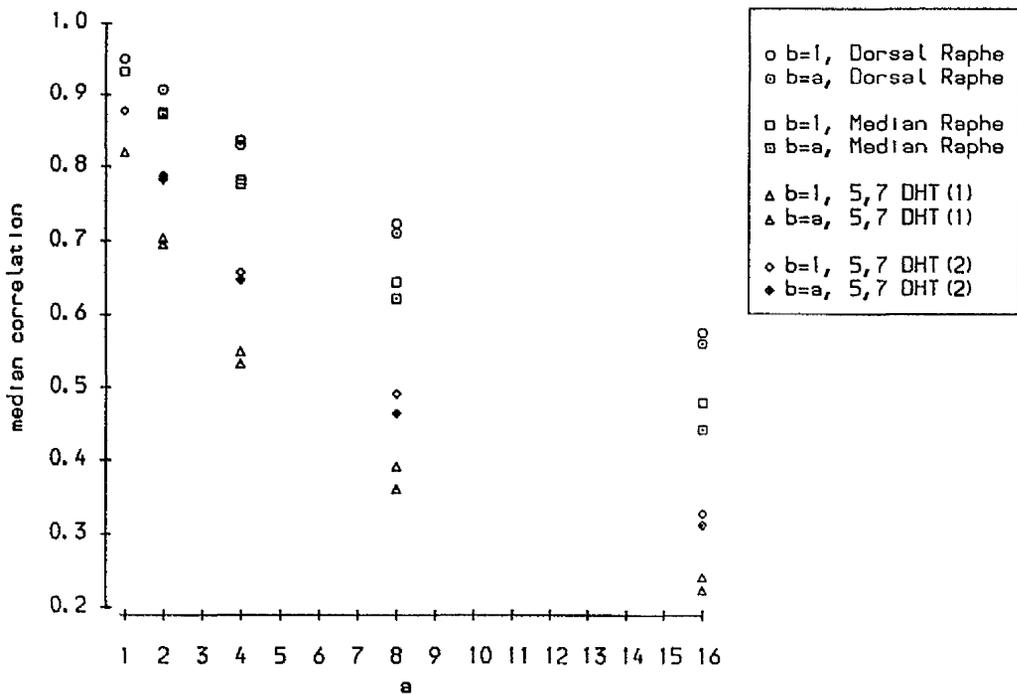


Figure 2.2.3(ii) Median correlations for the within treatment rankings, using f2 (rostral raphe study).

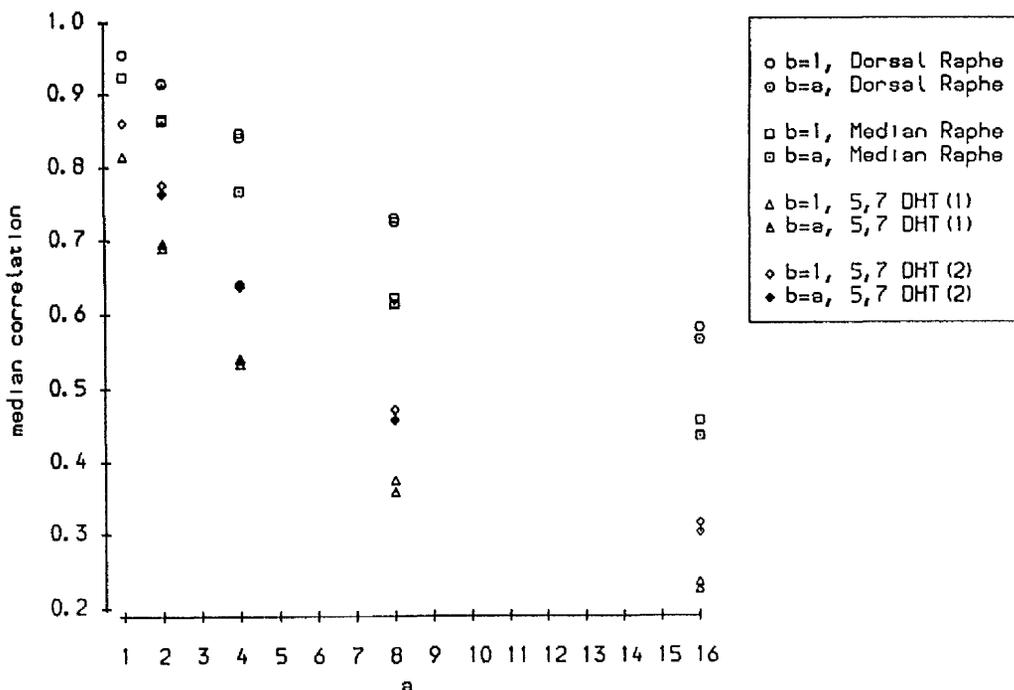


Figure 2.2.3 (iii) Median correlations for the between treatment rankings, using f1 (rostral raphe study).

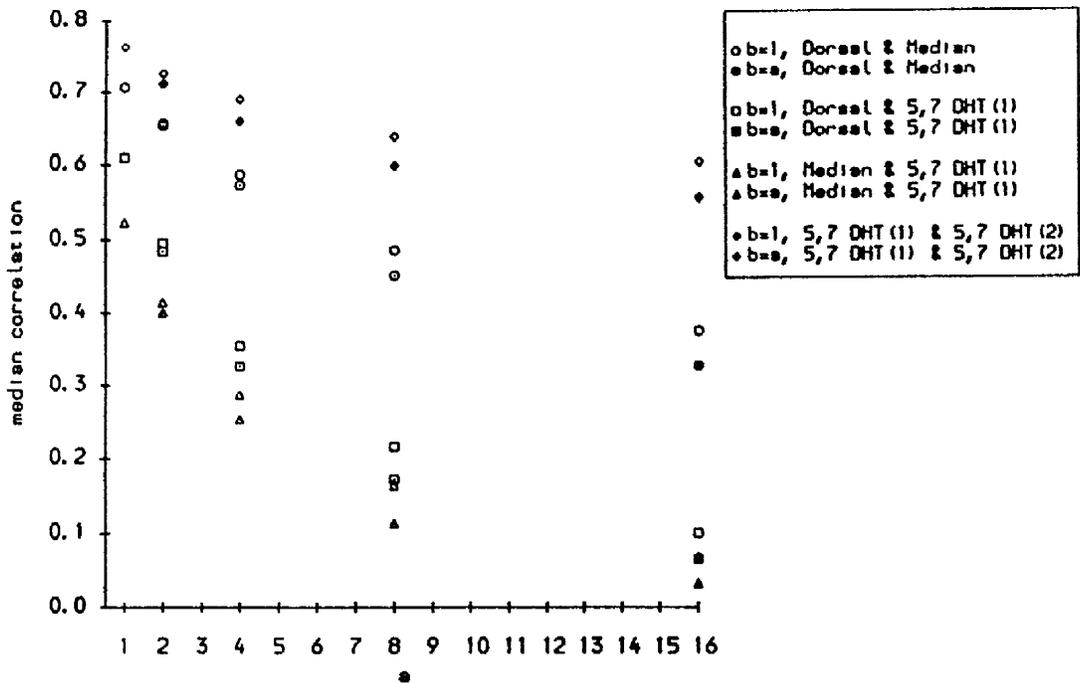


Figure 2.2.3(iv) Median correlations for the between treatment rankings, using f2 (rostral raphe study).

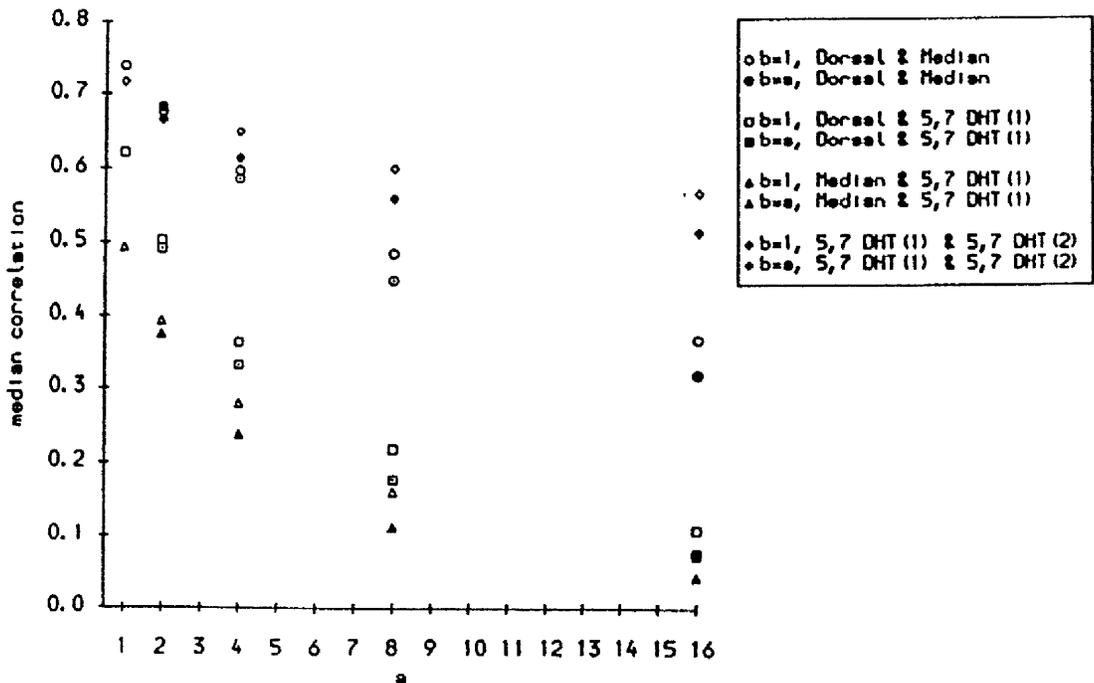


Figure 2.2.3(v) Median correlations for the within treatment rankings, using f1 (Alpidem/Zolpidem/Diazepam study).

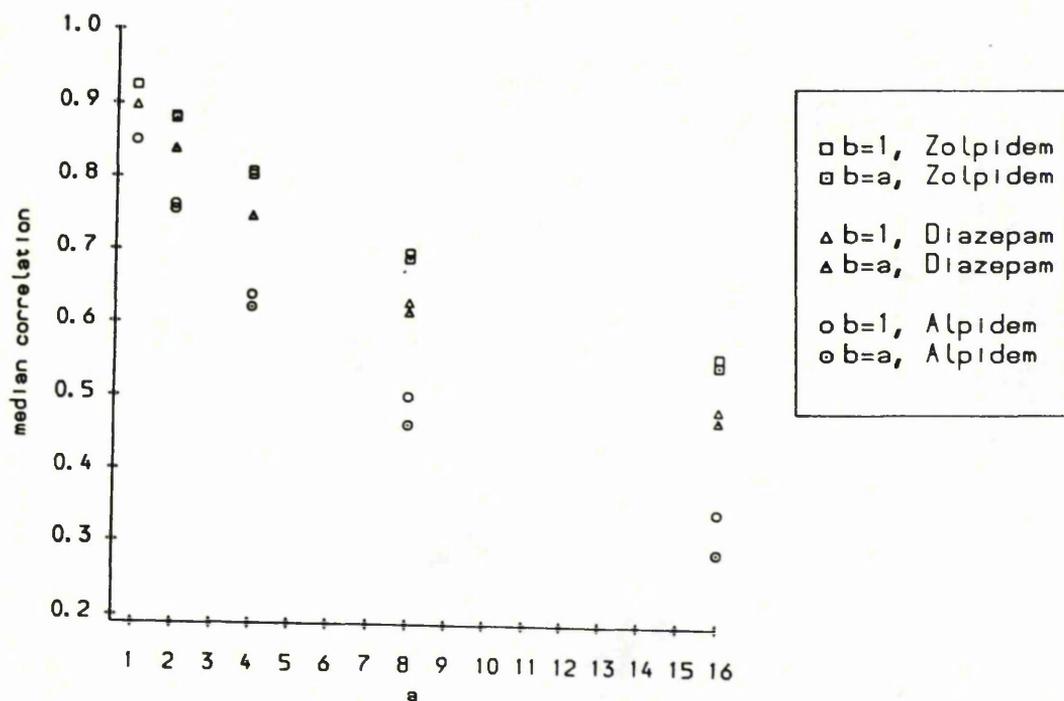


Figure 2.2.3(vi) Median correlations for the within treatment rankings, using f2 (Alpidem/Zolpidem/Diazepam study).

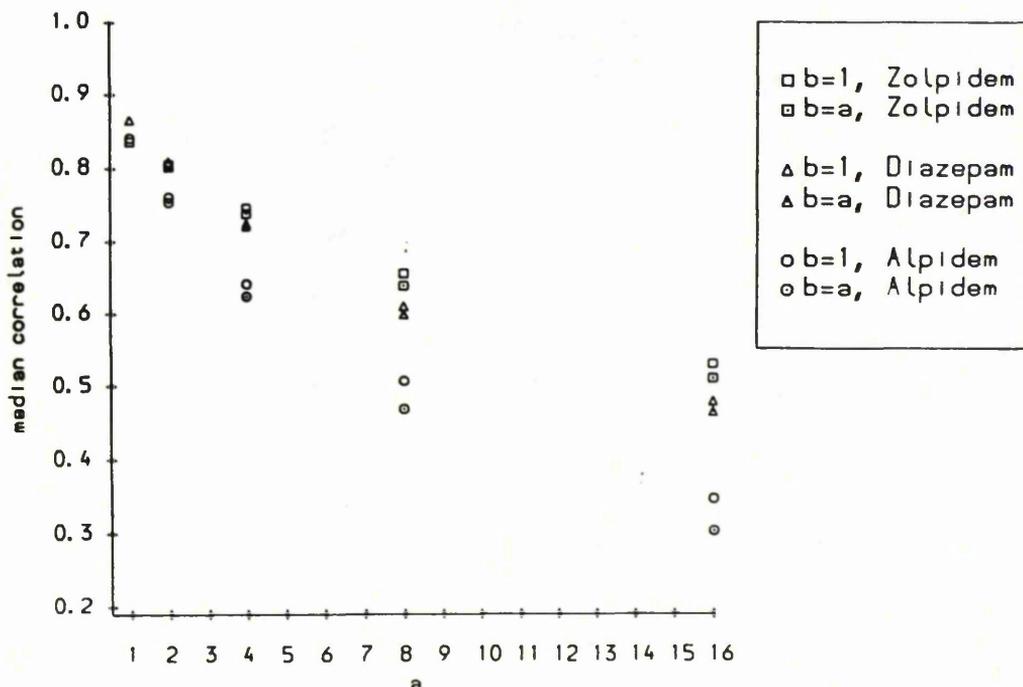


Figure 2.2.3(vii) Median correlations for the between treatment rankings, using f1 (Alpidem/Zolpidem/Diazepam study).

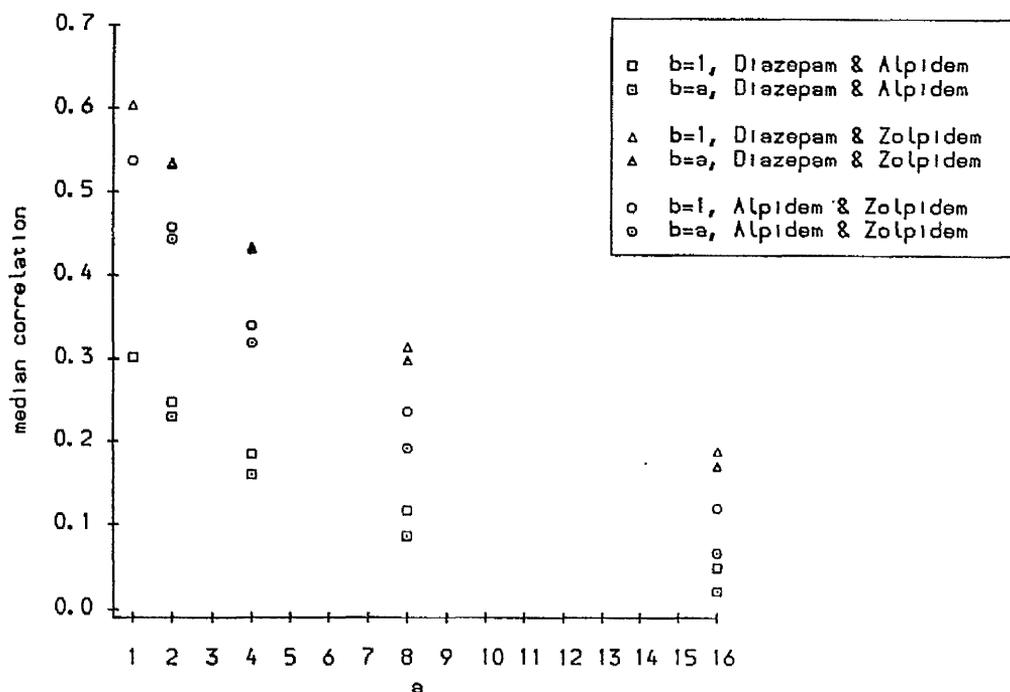
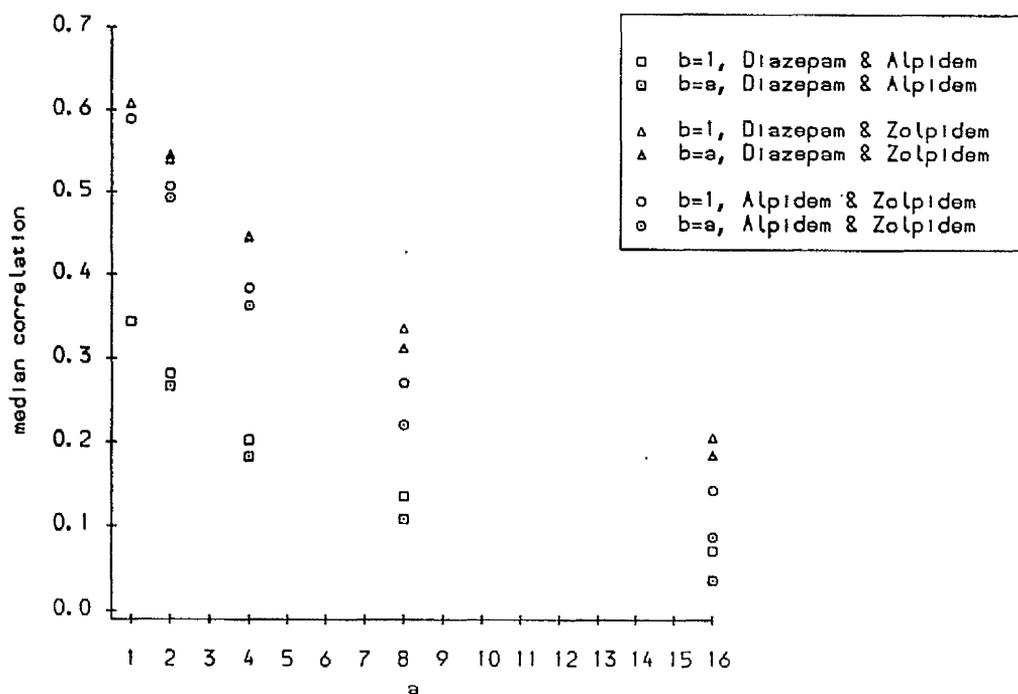


Figure 2.2.3(viii) Median correlations for the between treatment rankings, using f2 (Alpidem/Zolpidem/Diazepam study).



rankings for the rostral raphe stimulation study. The corresponding plots for the Alpidem/Zolpidem/Diazepam study are given in Figures 2.2.3(v)-(viii). The median and lower (Q1) and upper (Q3) quartiles of the estimated correlations for the simulation study for each approach are tabulated in Appendix A.1.

#### 2.2.4 RESULTS

Figures 2.2.3(i)-(viii) suggest that the interpretation of the point estimates given in Section 2.2.3 is fairly reliable. In general, it would seem that the correlations derived using the  $f_{1j}$  approach are, if anything, slightly more reliable when employing the estimated components of variability from the observed data. As expected, Figures 2.2.3(i)-(viii) show that increasing the magnitude of the variance components will result in a deterioration of the observed rankings. However, in particular cases (dorsal and median raphe stimulation, Alpidem and Zolpidem) the results suggest that doubling  $\sigma^2$  will have only a small effect on the reliability of the derived ranking of both the  $f_{1j}$ 's and  $f_{2j}$ 's. By increasing the variability to  $4\sigma^2$  the deterioration in the reliability becomes more evident in Figures 2.2.3(i)-(viii), and beyond that, a marked decrease in reliability is apparent. In all situations, the results would also suggest that there is little additional effect caused by increasing  $\tau^2$ .

An interesting feature from the between drug or treatment pairings is that the point estimate of the correlation coefficient derived from the observed data, in most cases exceeds the upper quartile of the simulated correlations. This conforms with our intuition that high correlations are likely to be underestimated, and hence, if

anything our point estimates will themselves be biased by underestimating the true association between the relevant pair of treatments or drugs.

### 2.2.5 THE RELIABILITY OF THE MODEL

In section 2.2.2 it was suggested that a constant correlation model might provide a useful approximation in the analysis of data of this type. In a final simulation study we investigate the effect of departures from this model by introducing variance-covariance structures corresponding to situations where the brain is considered to be made up of clusters of regions where the response is independent between clusters and correlated within clusters (Ford, 1986). Horwitz et al (1984, 1987) have shown that such patterns of the variance-covariance structure may be reasonable working approximations to the true functional interrelationships in the brain.

Consider the general covariance structure  $\Sigma = \sigma^2 \mathbf{1}\mathbf{1}^T + \Omega$ , where  $\Omega = \{\tau_{ij}\}$  represents the within animal covariance matrix. To illustrate the effect of regions clustering together it is useful to investigate different forms of the matrix  $\Omega$  corresponding to different models of the functional interrelationships in the brain.

For the purpose of this investigation we have selected a subset of 40 regions of interest (ROI's) from the dorsal and median raphe stimulated group of animals and 40 ROI's from the Zolpidem treated animals, together with their associated controls. In the simulation studies we have considered regions clustering into  $c$  clusters each of size  $r$  ( $r = 1, 2, 5, 10, 20, 40$ ) with functional

independence between clusters. We further assume that the within animal correlation  $\rho_{ij}$ , ( $i \neq j$ ), where

$$\rho_{ij} = \frac{\tau_{ij}}{\sqrt{\tau_{i1} \tau_{j1}}},$$

is equal (to  $\rho$  say) for all  $i$  and  $j$  within the same cluster.

The median correlations from the simulation studies for both the  $f_{1j}$  and  $f_{2j}$  approaches are plotted in Figures 2.2.4(i)-(vi) for the rostral raphe study and in Figures 2.2.4(vii), (viii) for the Zolpidem study. (The upper and lower quartiles, together with the median correlation from the simulation studies are tabulated in Appendix A.2). Pearson's product moment correlation coefficient between local cerebral glucose utilisation values, after dorsal raphe and median raphe stimulation, was estimated to be 0.88 from the observed data on the 40 ROI's selected. In all studies  $\sigma^2$  and  $\tau_{ii}$  were kept constant, both taking the value of 0.01 to reflect the typical experimental variance experienced in our autoradiographic example. Notice that when  $r=1$  all regions are functionally independent and hence  $\rho=0$  in this instance.

Figures 2.2.4 (i)-(viii) indicate that, at least within the limitations of this study, the structure of the correlation matrix, on the whole, does not greatly affect the variability of the rankings of the ROI's and that a fairly reliable measure of association between drug actions or treatments can be achieved. As would perhaps be expected, Figures 2.2.4(i)-(viii) show that there is some reduction in the reliability of the rankings as the number of clusters increases and the dependence between brain

Figure 2.2.4(i) Median correlations for the dorsal raphe rankings using f1 (40 ROIs)

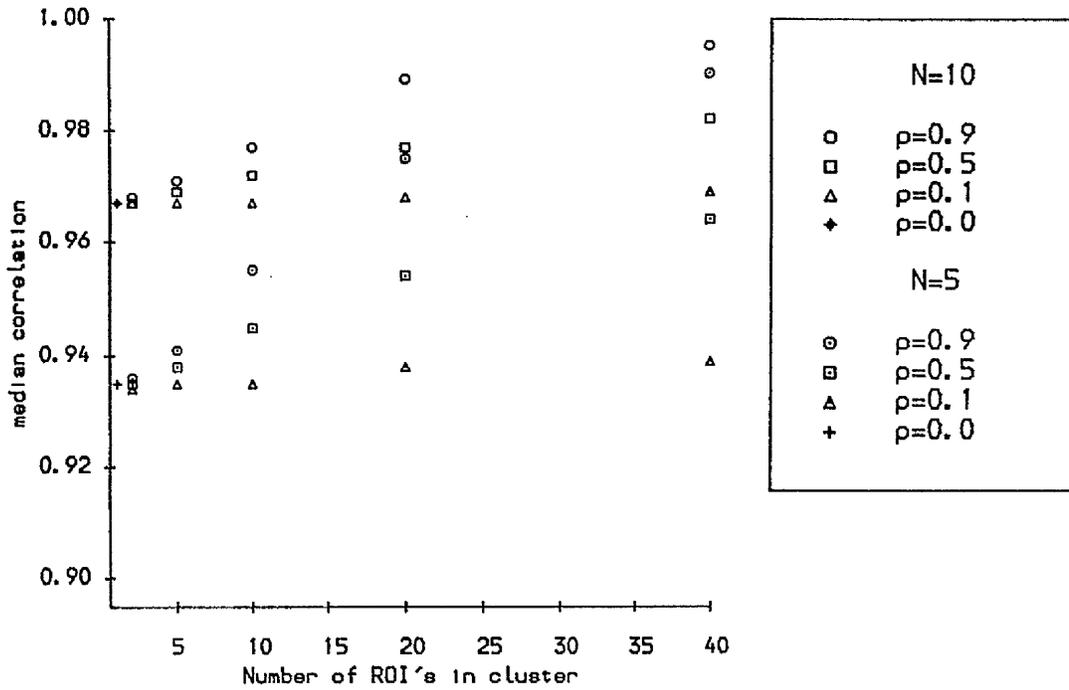


Figure 2.2.4(ii) median correlations for the dorsal raphe rankings using f2 (40 ROIs).

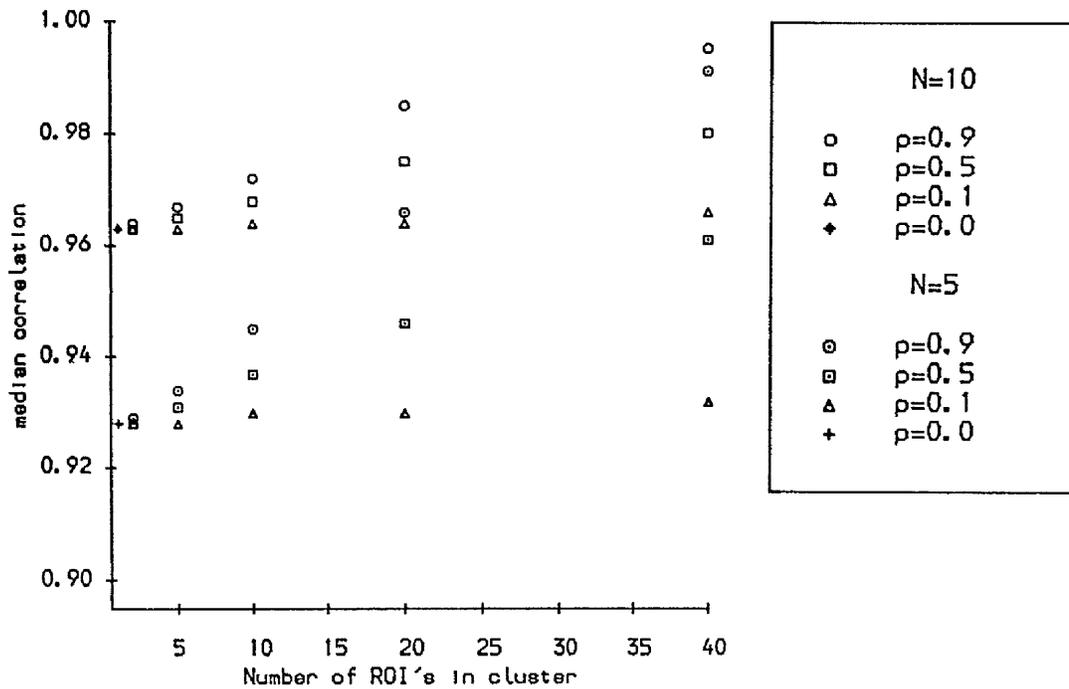


Figure 2.2.4(iii) Median correlations for the median raphe rankings using f1 (40 ROIs).

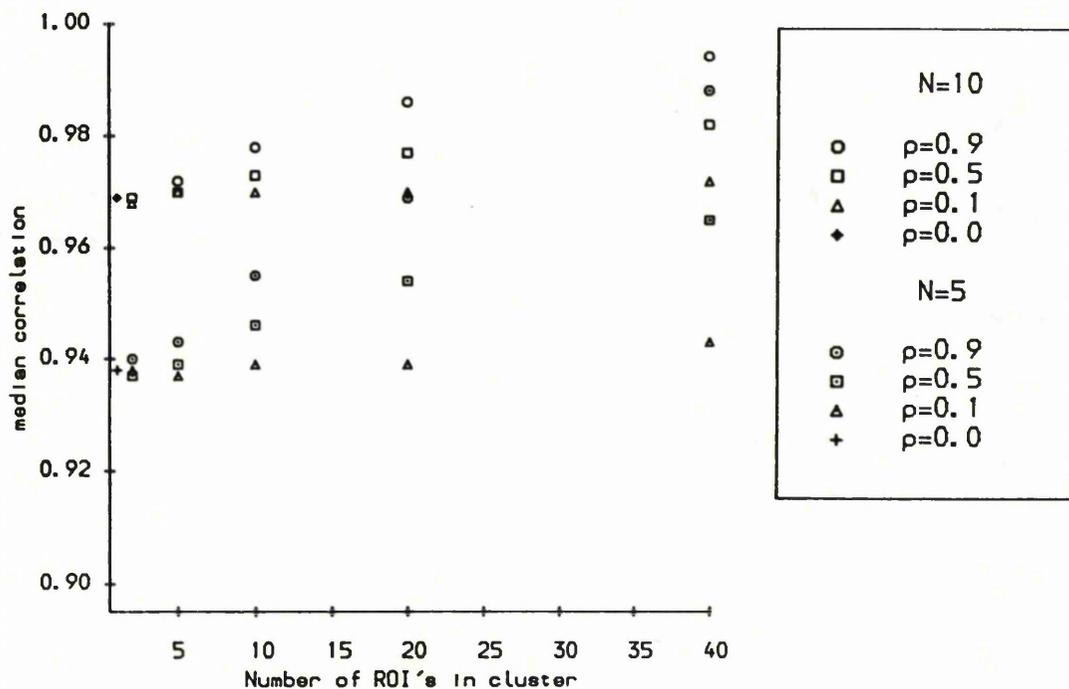


Figure 2.2.4(iv) Median correlations for the median raphe rankings using f2 (40 ROIs).

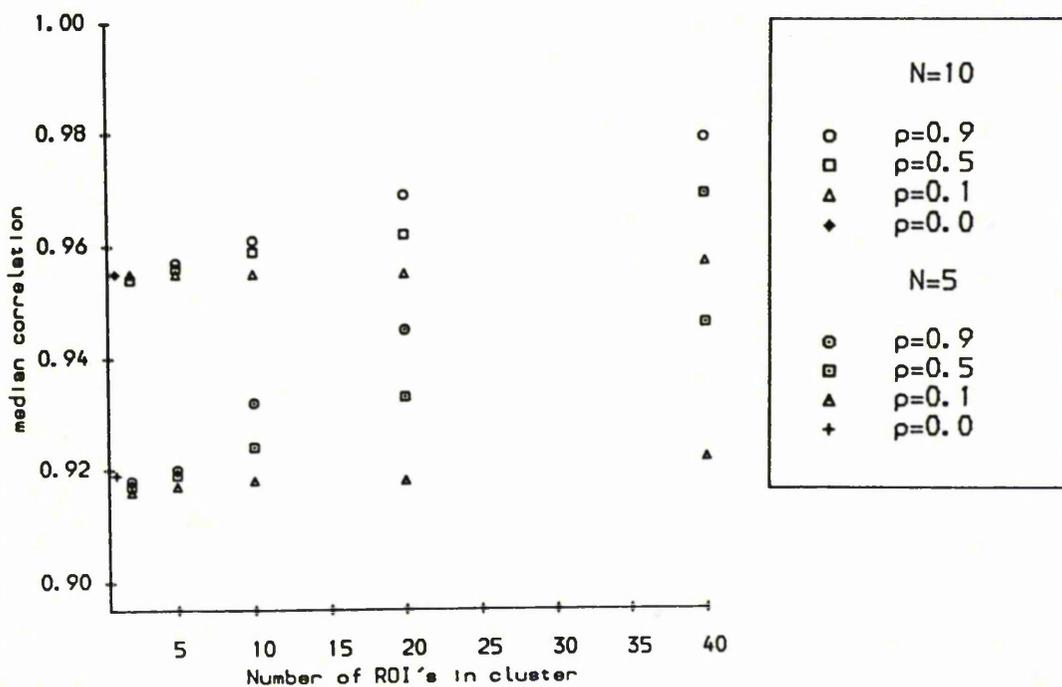


Figure 2.2.4(v) Median correlations for the association between the dorsal and median raphe stimulation rankings, using f1 (40 ROIs)

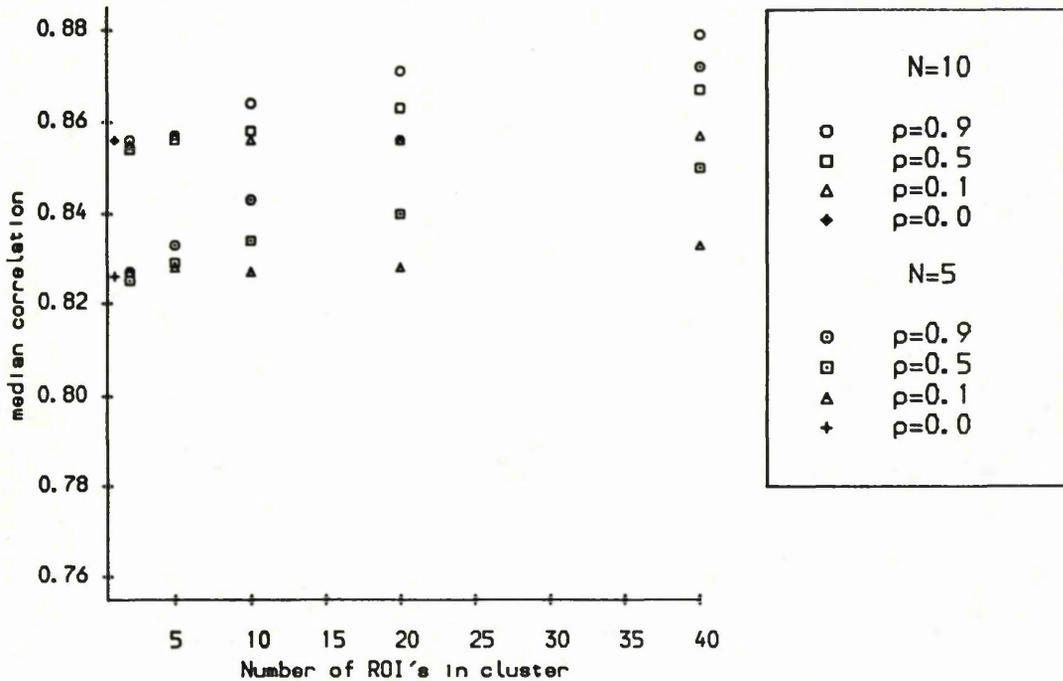


Figure 2.2.4(vi) median correlations for the association between the dorsal and raphe stimulation rankings, using f2 (40 ROIs)

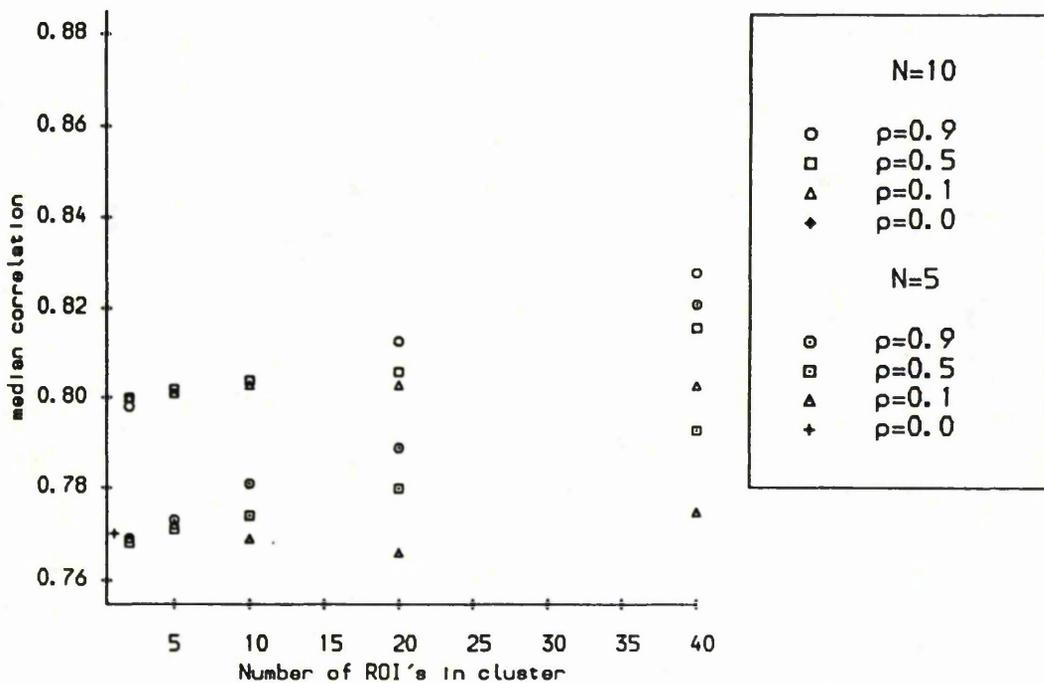


Figure 2.2.4(vii) Median correlations for the within treatment ranking of Zolpidem, using f1 (40 ROIs)

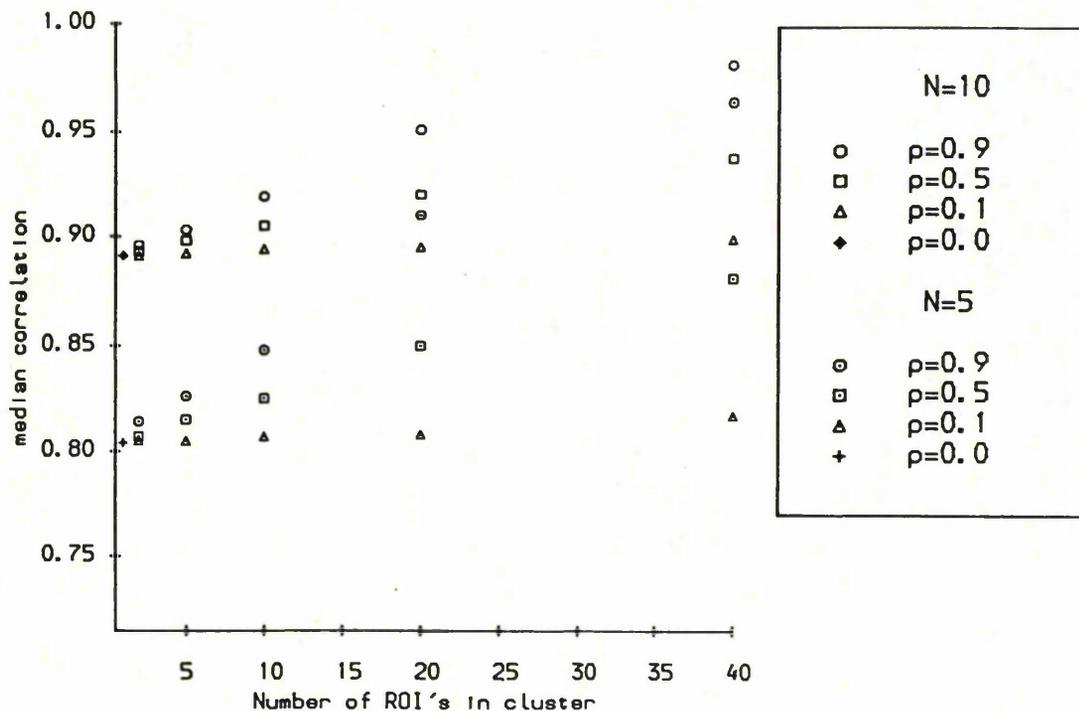
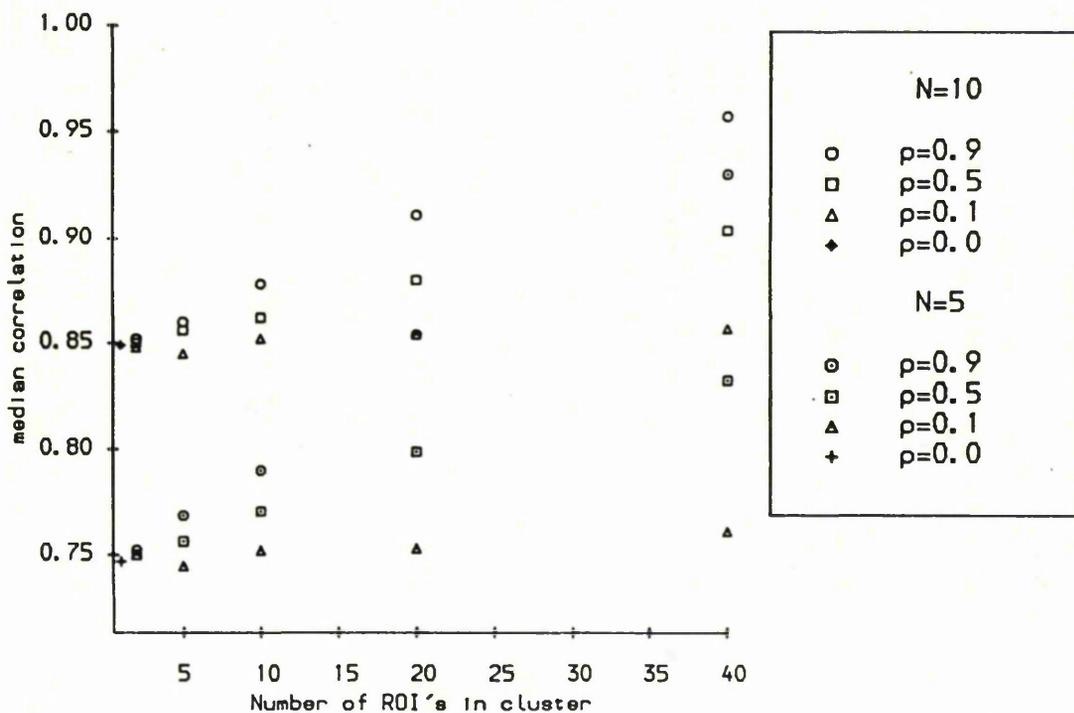


Figure 2.2.4(viii) Median correlations for the within treatment ranking of Zolpidem, using f2 (40 ROIs)



regions in the same cluster decreases. The performance of both approaches is also reduced by decreasing the sample size, although the effect of small sample size, on the reliability, is less a cause for concern in the extreme case when all regions have a strong functional dependence on all other regions ( $r=40$ ). In contrast, the strength of the dependence between regions in the same cluster becomes less of an issue if the brain is considered to be made up of a large number of regions, each containing a small number of functionally related clusters. Furthermore, the effect of clustering is negligible if only weak interrelationships between ROI's exist.

Figures 2.2.4 (v), (vi) show that the reliability of the correlation coefficient, as a measure of the association between drugs, has similar trends to those concerning the individual rankings. However, the effect, on the median correlation from the 2000 simulations, of strong regional interrelationships between all ROI's is not as great. In all instances, the point estimate derived from the observed data is greater than or equal to the upper quartile from the simulation studies, with equality being reached in the case where  $r=40$  and  $\rho=0.9$ .

### **2.3 SUMMARY OF RESULTS**

In this chapter we have set ourselves the rather difficult task of ranking a large number of brain regions with respect to their response to a drug or treatment. We remarked, in the introduction, that our ability to deal with such complex problems will depend on the strength of response being measured and the level of variability in the data being observed. In section 2.1 we addressed some of the

problems from a mathematical viewpoint and identified the form of the covariance structure for maximising and minimising the probability of correctly ranking a 3-dimensional observation vector, assuming multivariate normality. A lower bound for the probability of correctly ranking observation vectors with higher dimensions was investigated using the Bonferroni inequality. However, for realistic situations, these mathematical approaches will have severe limitations, due to the complex nature of the vector of responses.

In section 2.2, using a simple measure of response, we have seen that we can observe associations between brain regions which would be expected biologically, for two experimental data sets from quantitative autoradiography. Using simulation techniques we have established that, at least within the correlation matrices studied, that fairly reliable rankings of the regions can be obtained and fairly reliable measures of association between drugs can be computed. Increasing  $\sigma^2$  by a factor of two, in these studies, appeared to have no effect on the reliability of the derived rankings. However, a marked decrease was observed by increasing the variability to  $8\sigma^2$ . The results also suggested that the rankings will be robust with respect to the within animal variability and the structure of the correlation matrix, if we are willing to assume that all the correlations are greater than or equal to zero.

### CHAPTER 3 THE BLOOD FLOW-METABOLISM COUPLING PROBLEM

#### 3.1 INTRODUCTION

In cerebrovascular research it is generally accepted that, in a large number of experimental conditions, the level of blood flow is adjusted to meet the energetic demands of tissue. With the advent of tomographic and autoradiographic techniques for measuring local cerebral blood flow (Sakurada et al, 1978) and the rate of local cerebral glucose utilisation (Sokoloff et al, 1977; Phelps et al, 1979) neuroscientists have gained insight into the control of the coupling of local cerebral blood flow (LCBF) and local cerebral glucose utilisation (LCGU) and its breakdown in disease situations or after treatment with drugs (McCulloch et al, 1982; Kuchinsky et al, 1985). As the energetic demands of cerebral tissue are derived almost exclusively from the oxidation of glucose, alterations in the energetic demands of tissue are reflected by changes in LCGU.

Due to the basic cost of experimentation in both these areas, a common feature of the data from associated experiments, particularly in the higher resolution autoradiographic studies, is the high dimensionality of the observation vector corresponding to the set of measures of the response variable of interest in the brain regions studied. Consequently, if complex conclusions are to be drawn, the experimenter has the difficult task of analysing high dimensional data with a, relatively speaking, small number of experimental units. For example, McCulloch et al (1982), in an autoradiographic study,

analysed data on 36 regions of interest (ROIs), with a maximum of 6 animals in each group.

To answer questions of practical biological interest in ROI pattern analysis univariate repeated measures analysis of variance techniques are often suitable (McCulloch et al, 1982; Haxby et al, 1985). Experimental designs in which a random effect factor (e.g. subject) and a fixed effect factor (e.g. brain region or time) are crossed, with a single observed value for each cell, have become popular in medical and psychological research. However one major problem with this type of analysis, with ROI data, is that the correlation structure will display a certain degree of clustering (Horwitz et al, 1984, 1987) which inevitably results in a violation of the sphericity assumption of the repeated measures F-test causing inaccuracies in the calculated significance level (SL). Over the years, in the biostatistical and psychometrical literature, attention has focused on the proper method of analysis of such designs when this assumption is not met. Box (1954 a,b), Greenhouse and Geisser (1959) and Huynh and Feldt (1976) suggest referral of the test statistic to an F distribution with reduced degrees of freedom when the variance-covariance condition is violated. Greenhouse and Geisser (1959) suggest incorrectly, that these methods cannot be applied when the number of ROI's exceeds the number of experimental units. Maxwell and Avery (1982), in a limited study, investigated the applicability of these methods in such situations.

In the present study, the effect on the performance of these tests, induced by increasing the number of ROIs, has been

investigated in the case of a correlation structure with independence between ROI's (equivalent to no violation of the variance-covariance assumption). Departures from the constant covariance assumption, corresponding to situations where the brain is considered to be made up of clusters of regions where the response is independent between clusters and correlated within clusters (Ford, 1986), have also been considered.

### 3.2 NOTATION AND THE GENERAL STRUCTURE OF THE PROBLEM

Let  $Y_{ijk}$  denote the measurement of interest made in the  $j^{\text{th}}$  brain region ( $j=1, \dots, J$ ) on the  $i^{\text{th}}$  animal ( $i=1, \dots, n_k$ ) in the  $k^{\text{th}}$  treatment group ( $k=1, \dots, K$ ). Since all of the measurements made on each animal will be of the same basic quantity (for example LCGU or LCBF), we have, for each animal, a repeated measures vector of observations. As mentioned previously, the number of brain regions being investigated,  $J$ , will be large compared with the number of experimental units studied. For instance,  $J=60$  and  $\sum_{k=1}^K n_k=20$  would not be unusual in blood flow-metabolism coupling investigations. The success of any analysis of data of this type must depend critically on a combination of the strength of response which is present in the data and the structure and magnitude of the underlying variability in the data.

For investigating the relationship between LCBF and LCGU, the biological theory suggests that LCBF should be proportional to LCGU. That is, ignoring sources of variation

$$\text{LCBF} = \beta \times \text{LCGU}$$

where  $\beta$  is the constant of proportionality. In terms of the notation above, let  $Y_{i,j1}$  and  $Y_{i,j2}$  denote the measurements of LCBF and LCGU respectively. Until recently the standard approach to the problem was to regress the  $\bar{Y}_{.,j1}$ s on the  $\bar{Y}_{.,j2}$ s and then to compare regression coefficients and correlation coefficients between experimental subgroups (Des Rosiers et al, 1974; Sokoloff, 1977; Kuchinsky et al, 1981). These approaches are open to criticism for several reasons. Firstly, since the measurement of both LCBF and LCGU are subject to error, the problem is really of the linear functional relationship type and secondly, observations on different ROI's of the brain will not be independent. Furthermore, averaging over animals to obtain a mean response for each brain region conceals the between animal variability in the analysis. For these reasons  $\beta$  will be estimated with bias and with an estimated precision which grossly overstates the true precision. McCulloch, Kelly and Ford (1982) proposed an alternative approach to this problem. The proposal is that the underlying structure is written in the form

$$\log (\text{LCBF}) = \gamma + \log (\text{LCGU}) \quad , \quad \gamma = \log (\beta)$$

and that the log-transformed data  $X_{i,jk} = \log (Y_{i,jk})$  are analysed. This transformation is not only convenient mathematically but it also acts as a variance stabiliser since the variance of the  $Y_{i,jk}$ 's tends to increase with their expectation.

Considered in this form, the data may be analysed by repeated measures analysis of variance techniques with 'type of variable'

regarded as a grouping factor and 'brain region' regarded as a repeated measures trial factor. Although, in these particular types of experiments, there will not be sufficient data to conduct a full multivariate Profile Analysis (Seber, 1984), a univariate approach is possible. McCulloch, Kelly and Ford (1982) suggest that it is plausible that a large proportion of the correlation structure can be explained by an additive animal effect implying a constant correlation assumption may be a reasonable working approximation. Horwitz et al (1984, 1987) have indicated that there exists a certain degree of clustering in ROI correlation matrices.

Thus, to reduce potential inaccuracies in the assumed significance levels for tests involving within animal factors, it may be reasonable to proceed with an analysis based on a modified approach which can take into account estimated departures from the constant correlation structure.

### **3.3. REPEATED MEASURES ANALYSIS OF VARIANCE**

#### **3.3.1 THE MIXED TWO-WAY MODEL**

Repeated measures designs are essentially characterised by the presence of both fixed-effect factors (e.g. brain regions) and random-effect factors (e.g. subjects), with the fixed-effect factors usually of main interest. Consider the simplest case of a mixed two-way layout applied to ROI data, where subjects and brain regions are crossed, with a single replication in each cell, and we are interested in comparing the mean response of the measurement of interest between ROIs. In this instance the mixed two-way model reduces to that of a one-way repeated

measures analysis of variance. In terms of the earlier notation let  $X_{ij}$  denote the measurement of interest on the  $i^{\text{th}}$  patient in region  $j$ ,  $i=1, \dots, n$ ;  $j=1, \dots, J$ . Ford (1986) considered analysing data of this type using a linear model of the form

$$X_{ij} = \mu_j + p_i + \varepsilon_{ij}$$

where  $\mu_j$  is the average activity in region  $j$ ,  $p_i$  is the additive patient effect for patient  $i$ , and  $\varepsilon_{ij}$  is the random component representing the remaining variability within a patient.

Within this structure the variance-covariance matrix,  $\Sigma = \{\sigma_{ij}\}$ , for the observational vector  $\underline{X}_i = (X_{i1}, \dots, X_{iJ})^T$  can be written in the form  $\Sigma = \sigma^2 \underline{11}^T + T$  where  $\sigma^2$  is the between patient variance and  $T = \{T_{ij}\}$  represents the inpatient variance-covariance matrix.

Within this design, the hypothesis of homogeneity among the  $\mu_j$ 's is usually tested using the F-ratio,  $F$ , where

$$F = \frac{SS_J / (J-1)}{SS_{IJ} / (J-1)(n-1)} = \frac{(n-1)SS_J}{SS_{IJ}}$$

and  $SS_J$  and  $SS_{IJ}$  are the sums of squares associated with the brain regions (fixed effect) and interaction effect between individuals and brain regions respectively, and are given explicitly by the formulae

$$SS_J = \sum_{j=1}^J (\bar{X}_{.j} - \bar{X}_{..})^2 \quad , \quad \text{and}$$

$$SS_{IJ} = \sum_{j=1}^J \sum_{i=1}^n (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2$$

where  $\bar{X}_{.i} = \frac{1}{J} \sum_{j=1}^J X_{ij}$ ,  $\bar{X}_{.j} = \frac{1}{n} \sum_{i=1}^n X_{ij}$  and  $\bar{X}_{..} = \frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n X_{ij}$ .

The F-ratio, under the null hypothesis, has an exact F distribution with  $(J-1)$  and  $(J-1)(n-1)$  degrees of freedom, only for restricted forms of  $\Sigma$ . Huynh and Feldt (1970) and Rouanet and Lépine (1970) established the least restrictive form of  $\Sigma$  required for valid F-tests, for the one-way repeated measures design. This model assumes that  $\Sigma = \{\sigma_{ij}\}$  is given by

$$\sigma_{ij} = \begin{cases} 2\tau_j + \lambda & \text{if } i=j \\ \tau_i + \tau_j & \text{if } i \neq j \end{cases}$$

for some  $\tau_j$  ( $j=1, \dots, J$ ) and where all of the eigenvalues of  $\Sigma$  are equal to some positive  $\lambda$ . This pattern, referred to as sphericity, is equivalent to the condition that all pairwise differences of the mean regional responses,  $\bar{X}_{.j}$ , have the same variance, given by  $\text{Var} (\bar{X}_{.i} - \bar{X}_{.j}) = \frac{2\lambda}{n}$  ( $1 \leq i, j \leq J$ ).

A special case of sphericity is obtained when there is equality of the population variances of all brain regions and equality of the population covariances between all pairs of brain regions. Such covariance structures are often referred to as displaying 'compound symmetry' or satisfying the 'symmetry assumption'. Although it is easier to imagine contexts in which the compound symmetry condition would be satisfied, rather than the more general sphericity conditions, realistically a constant correlation structure will not be exactly representative of ROI data. However, it might be hoped that in some circumstances such a structure will be a good working approximation.

### 3.3.2 HIGHER ORDER MIXED MODELS

Now, consider the slightly more general mixed model corresponding to the the blood flow-metabolism coupling data structure described in Section 3.2 where patients are subdivided into K groups (e.g. 2 groups representing LCGU and LCBF).

Assume a model of the form

$$X_{ij,k} = \mu_{j,k} + p_{i(k)} + \epsilon_{ij,k}$$

where  $\mu_{j,k}$  is the average activity in region j of group k,  $p_{i(k)}$  is the additive patient effect of patient i in group k, and  $\epsilon_{ij,k}$  represents the remaining variability.

Within this framework, the mean square ratios in the analysis of variance table normally used to test the hypotheses of equality of response among ROI's,  $F_1$ , and no brain region x group interaction effects,  $F_2$ , are given by

$$F_1 = (N-J)SS_J/SS_E \quad \text{and} \quad F_2 = (N-J)SS_{IJ}/(J-1)S_E$$

where  $N = \sum_{k=1}^K n_k$  and  $SS_J$ ,  $SS_{IJ}$  and  $SS_E$  are the sums of squares

associated with the brain regions, brain region x group interaction and the pooled "split-plot" error and are defined as

$$SS_J = \sum_{j=1}^J N (\bar{X}_{.j.} - \bar{X}_{...})^2 \quad ,$$

$$SS_{IJ} = \sum_{j=1}^J \sum_{k=1}^K n_k (\bar{X}_{.jk} - \bar{X}_{.j.} - \bar{X}_{..k} + \bar{X}_{...})^2$$

$$\text{and} \quad SS_E = \sum_{k=1}^K \left[ \sum_{j=1}^J \sum_{i=1}^{n_k} (X_{ijk} - \bar{X}_{i.k} - \bar{X}_{.jk} + \bar{X}_{...k})^2 \right] \quad .$$

Huynh and Feldt (1970) extended their work to show that  $F_1$  and  $F_2$  have exact F distributions with degrees of freedom  $\{(J-1), (J-1)(N-K)\}$  and  $\{(J-1)(K-1), (J-1)(N-K)\}$  respectively if and only if

a) the variance-covariance matrices  $\Sigma_1, \Sigma_2, \dots, \Sigma_K$  associated with each level of the grouping factor satisfy the condition

$$C\Sigma_1C^T = C\Sigma_2C^T = \dots = C\Sigma_KC^T = \Lambda$$

where C is any matrix defining (J-1) orthonormal contrasts among the J ROI's and

b) the common matrix  $\Lambda$  is of the form  $\lambda I_{J-1}$ , where  $I_{J-1}$  is the identity matrix of rank J-1 and the eigenvalues of  $\Sigma$  are equal to some positive value  $\lambda$ .

An extension of this model, in a similar fashion, to incorporate a second grouping factor, results in the model proposed by McCulloch, Kelly and Ford (1982) to investigate differences between the blood flow-glucose use ratio between treatment groups. Mendoza et al (1976) describe a factorial design with two repeated factors and the subsequent necessary and sufficient conditions for F-ratios to have exact F distributions.

### 3.3.3. TESTING VALIDITY CONDITIONS

For repeated measures designs involving one or more grouping factors, Huynh and Feldt (1970) and Huynh and Mandeville (1979) suggest testing the validity conditions for traditional F-tests in two stages. Firstly, the authors suggest testing equality of covariance matrices, for the associated set of orthogonal contrasts across all levels of the groups, using a modified form

of Box's (1949) multivariate generalisation of Bartlett's (1937) homogeneity of variance tests. If Box's (1949) test does not reject equality of the covariance matrices over all levels of the grouping factor, Mauchly's (1940)  $W$  criterion can then be used to test sphericity in the pooled variance covariance matrix. If the test is nonsignificant one should proceed as if the validity conditions are satisfied.

Alternatively, if the equality conditions of the covariance matrices or sphericity are rejected, implying that the mean square ratios,  $F_1$  and  $F_2$ , do not have exact  $F$  distributions, the traditional  $F$  tests will have inflated Type I error rates (Box, 1954). That is, the actual probability of a Type I error will be more than the nominal level.

Several authors have investigated the appropriateness of either or both of these tests using simulation techniques. Davidson (1972) suggests that the modified Box's  $M$  test will not detect moderate departures from the null hypothesis when the sample size only just exceeds the number of repeated levels. Hopkins & Clay (1963), Korin (1972) and Olsen (1974) have shown that the test is not robust to departures from multivariate normality. Huynh and Mandeville (1979), on the other hand, examined the behaviour of Mauchly's  $W$  criterion under conditions of nonnormality. Empirical rates of Type I error from their simulation studies indicated that the test was conservative for light-tailed distribution and liberal for heavy-tailed distributions.

In instances where the number of repeated levels exceeds the sample size, both tests are inappropriate since calculations will

be based on the determinant of a singular sample covariance matrix.

Due to the sensitivity of both the Modified Box's M test and Mauchly's W criterion to all but very minor departures from their respective hypotheses, Rogan et al (1979) and Kesselman et al (1980) suggest that it is pointless to test the validity conditions of the traditional F-tests in repeated measures design using the aforementioned tests and recommend a modified univariate approach to repeated measures F-tests.

#### 3.3.4 MODIFIED UNIVARIATE APPROACHES

When the conditions on the variance-covariance matrix are violated, Box (1954b) showed that, under the null hypothesis, the true distribution of the traditional F-statistic, for the mixed two-way design with a single replication per cell, can be approximated by an F-distribution with degrees of freedom  $v_1 = (J-1)\epsilon$  and  $v_2 = (J-1)(n-1)\epsilon$ , where J and n are the number of levels of the fixed and random factor respectively and  $\epsilon$  denotes a correction factor reflecting the degree of departure of the population variance-covariance matrix from the sphericity assumption.

For variance-covariance matrix  $\Sigma = \{\sigma_{ij}\}$ ,  $\epsilon$  is defined explicitly by the formula

$$\varepsilon = \frac{J^2(\bar{\sigma}_{jj} - \bar{\sigma}_{..})^2}{(J-1) \left( \sum_{j=1}^J \sum_{i=1}^J \sigma_{ij}^2 - 2J \sum_{j=1}^J \bar{\sigma}_{j.}^2 + J^2 \bar{\sigma}_{..}^2 \right)}$$

where  $\bar{\sigma}_{jj}$  is the mean of the elements on the main diagonal of  $\Sigma$ ,

$\bar{\sigma}_{..}$  is the mean of all the elements of  $\Sigma$ ,

$\bar{\sigma}_{j.}$  is the mean of all the elements in row  $j$  of  $\Sigma$ ,

and  $\sigma_{ij}$  is the element in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $\Sigma$ .

If  $C$  is any matrix defining  $(J-1)$  orthonormal contrasts then Box's correction factor can be expressed alternatively as

$$\varepsilon = \frac{\text{trace}(C\Sigma C^T) \text{trace}(C\Sigma C^T)}{(J-1) \{ \text{trace}(C\Sigma C^T)^2 \}} = \frac{(\Sigma \lambda_j)^2}{(J-1) \Sigma \lambda_j^2} \quad 3.1$$

where  $\lambda_j$  ( $j=1, \dots, J-1$ ) are the non-negative eigenvalues of the matrix  $C\Sigma C^T$ .

Now consider the case where  $\Sigma$  displays sphericity. In such instances  $C\Sigma C^T = \lambda I_{J-1}$  and subsequently all  $J-1$  eigenvalues will be equal. Moreover  $\varepsilon$  will equal its upper limit of 1 and as a result Box's approximation will yield exact results.

Geisser and Greenhouse (1958) extended the work of Box (1954b) to the two-way mixed model with a grouping factor. If the standard assumptions do not hold, the authors showed that all within subject mean square ratios have approximate  $F$  distributions, with degrees of freedom reduced from the

traditional ones by the multiplicative factor  $\epsilon$ . Geisser and Greenhouse (1958) also demonstrated that for a general variance-covariance matrix the lower bound for  $\epsilon$  was equal to  $(J-1)^{-1}$  since

$$\sum_{j=1}^{J-1} \lambda_j^2 \leq \frac{(\sum_{j=1}^{J-1} \lambda_j)^2}{J-1}$$

with equality holding when  $C\Sigma C^T$  has only one positive eigenvalue, all others being zero. Wallenstein and Fleiss (1979) derived the lower bound for  $\epsilon$  when the variance-covariance matrix has a serial correlation pattern.

Since the true population covariance matrix is seldom known, Greenhouse and Geisser (1959) proposed using the sample estimator,  $\hat{\epsilon}$ , to adjust the degrees of freedom of the F-test, if the variance-covariance matrix is estimated with a large number of degrees of freedom. Otherwise the use of the most conservative test, with  $\epsilon$  taking its lower bound, was suggested. In general, the authors proposed the following three step approach for conducting traditional F-tests:

- (i) Compare the computed F statistic with the critical point from the F distribution with unadjusted degrees of freedom. Stop if the result is not significant since reducing the degrees of freedom will not alter the outcome.
- (ii) Compare the computed F statistic with the critical point from the F distribution with degrees of freedom reduced by the most conservative estimate of  $\epsilon$ . Stop if the result is significant.

(iii) Conduct the approximate F-test with  $\varepsilon$  estimated from the sample variance-covariance structure.

Collier et al (1967), Stoloff (1970), Huynh and Feldt (1976) and Huynh (1978) have investigated, using Monte Carlo studies, the use of  $\hat{\varepsilon}$  to adjust the degrees of freedom. Empirical results suggest that  $\hat{\varepsilon}$  underestimates the true value of  $\varepsilon$  producing a more conservative estimate of the significance level than the nominal level being used, particularly when there is minimal departure from sphericity or the sample size is small. Indeed when the number of repeated levels,  $J$ , exceeds the sample size,  $n$ , it can be shown that the upper bound for  $\hat{\varepsilon}$  is  $(n-1)/(J-1)$  since the sample variance-covariance matrix will be singular.

To prove that  $\hat{\varepsilon}$  has an upper bound of  $(n-1)/(J-1)$  in such instances, first consider any positive semi-definite  $(J-1) \times (J-1)$  matrix  $\Omega$  of rank  $n-1$  with  $n-1 < J-1$ . Without loss of generality, assume the eigenvalues  $\lambda_j, (j=1, \dots, J-1)$ , of  $\Omega$  sum to 1. Since  $\Omega$  has rank  $n-1$ ,  $\Omega$  will have  $n-1$  positive eigenvalues ( $\lambda_1, \dots, \lambda_{n-1}$ , say) and  $J-n$  eigenvalues ( $\lambda_n, \dots, \lambda_{J-1}$ ), with the value zero. Then

$$\sum_{j=1}^{J-1} \lambda_j^2 = \sum_{j=1}^{n-1} \lambda_j^2$$

is minimised when all the  $\lambda_j$  are equal to the value  $(n-1)^{-1}$ . The result then follows since

$$\hat{\varepsilon} = \frac{\sum_{j=1}^{J-1} (\lambda_j)^2}{(J-1) \sum_{j=1}^{J-1} \lambda_j^2}$$

where the  $\lambda_j$  ( $j=1, \dots, J-1$ ) are the eigenvalues of the matrix  $CSC^T$ ,  $S$  is a  $J \times J$  singular matrix of rank  $n-1$  and  $C$  is a  $(J-1) \times J$  matrix of orthonormal contrasts of rank  $J-1$ .

To compensate for the conservativeness of tests using the sample estimate  $\hat{\varepsilon}$ , Huynh and Feldt (1976) proposed adjusting the degrees of freedom of the traditional F-tests using an alternative estimator

$\tilde{\varepsilon}$ , where

$$\tilde{\varepsilon} = \frac{n(J-1)\hat{\varepsilon} - 2}{(J-1)(n-1-(J-1)\hat{\varepsilon})} \quad \text{for a one factor design and}$$

$$\tilde{\varepsilon} = \frac{n(J-1)\hat{\varepsilon} - 2}{(J-1)(n-K-(J-1)\hat{\varepsilon})} \quad \text{for a two way mixed model with}$$

individuals split into  $K$  groups. According to the authors' Monte Carlo simulations,  $\tilde{\varepsilon}$  is less biased and less dependent on large sample sizes when the variance-covariance matrix deviates only moderately from uniformity ( $\varepsilon > 0.75$ ). Furthermore, it can be verified that  $\tilde{\varepsilon} \geq \hat{\varepsilon}$  for any values of  $n$  and  $J$ , with equality holding when  $\tilde{\varepsilon} = \hat{\varepsilon} = (J-1)^{-1}$ . However, it is feasible that for certain sample correlation matrices  $\tilde{\varepsilon}$  will exceed unity. In such instances the authors suggest equating  $\tilde{\varepsilon}$  with 1, since it would not be reasonable to actually increase the degrees of freedom for the F distribution.

Huynh (1978), based on a theorem due to Box (1954b), proposed two additional test procedures (the General Approximate (GA) and the Improved General Approximate (IGA) for assessing the significance of within-subject effects when the variance-covariance matrices between groups are heterogeneous. Although computationally more complex, Huynh (1978) demonstrated in a limited simulation investigation that the IGA procedure showed good control over Type I error, but did not necessarily outperform the simpler  $\tilde{\epsilon}$  approximation.

### 3.4 AN EXAMPLE

At this stage it is useful to introduce a data set. Consider the simple case of an investigation of the relationship between local cerebral blood flow (LCBF) and local cerebral glucose utilisation (LCGU) in control conscious rats, LCBF and LCGU being measured in separate groups of five animals. The relationship between the mean logged LCBF and the mean logged LCGU is plotted in Figure 3.1 for 20 brain regions of interest.

Generally speaking the levels of LCBF and LCGU display the same regional hierarchies, with the level of LCBF (measured in units of  $\mu\text{mol}/100\text{g}$ ) approximately 1.5 times the level of LCGU (measured in units of  $\mu\text{mol}/100\text{g}/\text{min}$ ). Although in most regions the ratio of LCBF to LCGU is close to 1.5 the repeated measures ANOVA (Table 3.1), based on the equal correlation model (unadjusted p-value) identifies the presence of a significant region x variable interaction, signifying regional inconsistencies in the LCBF to LCGU ratio. Correcting the degrees of freedom, to take account

Table 3.1 Repeated Measures Anova Table for Flow-Metabolism  
Coupling data.

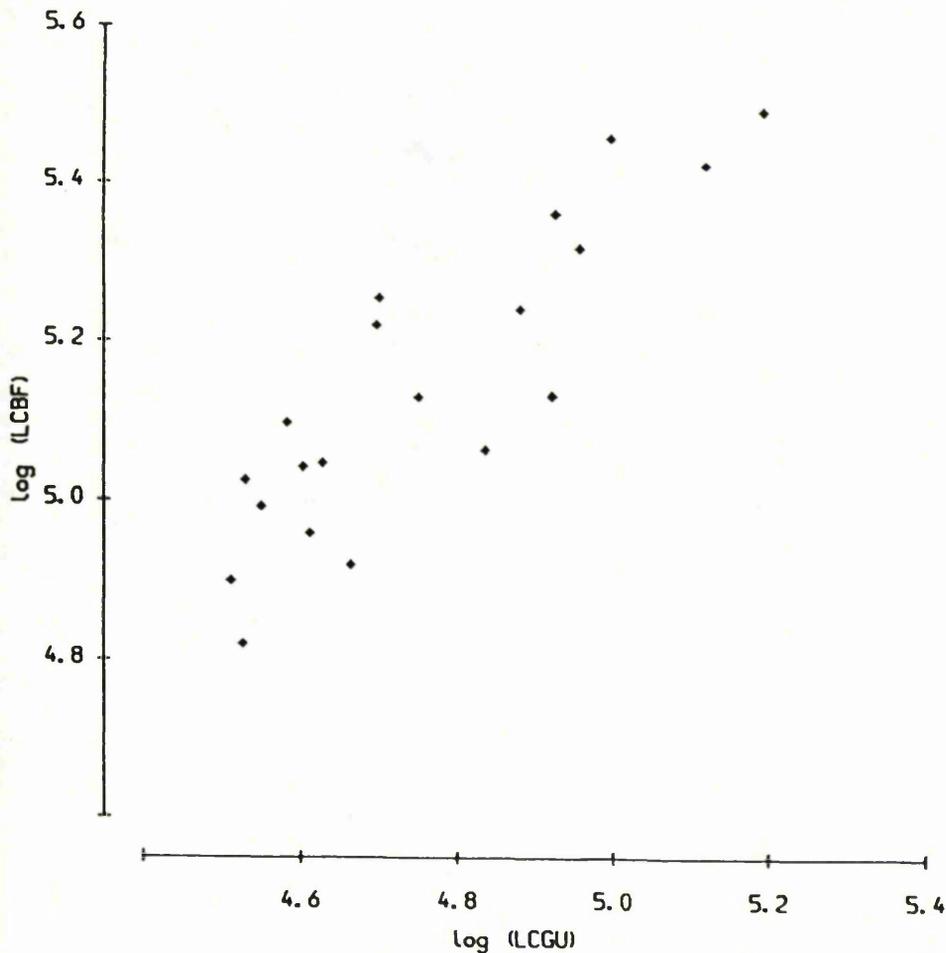
Source	F	p value		
		unadjusted	GG	HF
Variable	35.38	0.0003		
Region	24.03	0.0000	0.0000	0.0000
Region x Variable	2.33	0.0025	0.0720	0.0146

most conservative test  $F(1,19;0.9)=2.99$

GG = Greenhouse-Geisser adjusted test

HF = Huynh-Feldt adjusted test

Figure 3.1 Comparison of the relationship between LCGU and LCBF  
in control animals.



of the departure from the constant correlation model, using the Huynh and Feldt ( $\hat{\epsilon}$ -adjustment) correction factor also rejects the hypothesis of equality among LCBF to LCGU ratios. However, using the more conservative approach of Greenhouse and Geisser ( $\hat{\epsilon}$ -adjustment), the test fails to reject the hypothesis at the 5% significance level. Moreover, the most conservative lower bound approach fails to reject the hypothesis at the 10% level. This example displays the not too uncommon feature of analyses of this type, whereby the three point approach suggested by Greenhouse and Geisser coupled with the inability to identify which correction factor is most appropriate, results in regional alterations in the LCBF/LCGU ratios not being established conclusively.

However, by conducting suitable follow-up tests employing the Bonferroni correction, a number of significant regional differences in the LCBF to LCGU ratio could be identified. The appropriateness of conducting pairwise comparisons in such circumstances will be dealt with in a later section.

### 3.5 SIMULATION STUDY

In order to assess how the Greenhouse-Geisser ( $\hat{\epsilon}$ ), the Huynh-Feldt ( $\hat{\epsilon}$ ) and the Greenhouse-Geisser lower bound correction factors effect the nominal significance level of the F-test when the number of repeated levels exceed the number of experimental animals, a simulation study was conducted varying sample sizes, numbers of ROI's and correlation matrices. Since the mathematical properties of all within animal tests are similar

(Collier et al, 1967), we have concentrated on the simpler problem of testing for equality of mean response between regions in a study involving only a single group of animals, on whom a single variable (say LCBF) is measured. The mixed two-way model is specified in section 3.3.1.

The true significance level of a test of equality of mean response among ROI's was estimated as detailed below:

- (i) data are generated for specific sample sizes ( $n=5,10$ ) and number of ROI's ( $J=10,20$ ), using pseudo random number generators, for multivariate normal distributions, with zero mean vectors and specified covariance matrices.
- (ii) the F statistic for testing equality of means across regions is calculated from the data.
- (iii) significance of the test of equality of mean regional response (with nominal significance levels of 0.05, 0.025, and 0.01) is determined for the unadjusted approach together with the  $\tilde{\epsilon}$ -adjusted,  $\hat{\epsilon}$ -adjusted and lower bound procedures.
- (iv) steps (i), (ii) and (iii) are repeated 10,000 times and the relative frequency of significance is taken as an estimate of the true significance level of the test.

The effect on the performance of the test of the number of ROI's was investigated in the case of a correlation structure with independence between ROI's (equivalent to no violation of the constant variance-covariance assumption) for sample sizes of 5 and 10. Departure from the constant correlation assumption was investigated by varying the correlation in hypothetical

situations where the brain is considered to be composed of clusters of regions. Details of these correlation patterns are summarised in Table 3.2.

### 3.6 RESULTS

Empirical results of the estimated significance levels for the unadjusted,  $\tilde{\epsilon}$ -adjusted,  $\hat{\epsilon}$ -adjusted and the most conservative lower bound approach are presented in Figures 3.2-3.7 for a nominal significance level of 0.05. Results with nominal significance levels of 0.025 and 0.01, together with those for 0.05, are tabulated in Appendices B.1 to B.7. The standard errors associated with the estimated significance levels are all less than 0.005 since each estimate is based on 10000 simulations. Of course, since most of the estimated significance levels are less than 0.1, the standard error of these estimates will undoubtedly be much smaller. It is also worth noting that the results from this study will not be truly independent since within each simulation the test statistic for all four procedures is the same. Thus if the most conservative approach is significant, the remaining three procedures must also be significant.

#### 3.6.1 INCREASING THE NUMBER OF ROI'S

Figure 3.2 contains the results of increasing the number of repeated levels within the independence context. The unadjusted approach serves as a check on the pseudo random number generator used in the study. As would be expected, the results display a close agreement between the estimated significance level and the nominated significance level, with any discrepancies being

Table 3.2 Summary of the correlation structures in the simulation study

No. clusters	No. ROI's	Description	
a)	2	20	Equal correlation within clusters; independence between regions in different clusters.
b)	4	20	Equal correlation within clusters; independence between regions in different clusters.
c)	10	20	Equal correlation within clusters; independence between regions in different clusters.
d)	2	10	Equal correlation within clusters; independence between regions in different clusters.
e)	2	20	Correlation in the first cluster equal to $2/3$ ; equal correlation within second cluster; independence between regions in different clusters.
f)	2	20	Correlation in both clusters equal to $2/3$ ; equal correlation between regions in different clusters.

**Figure 3.2** Empirical results of the effect of increasing the number of ROI's within an independence covariance structure.

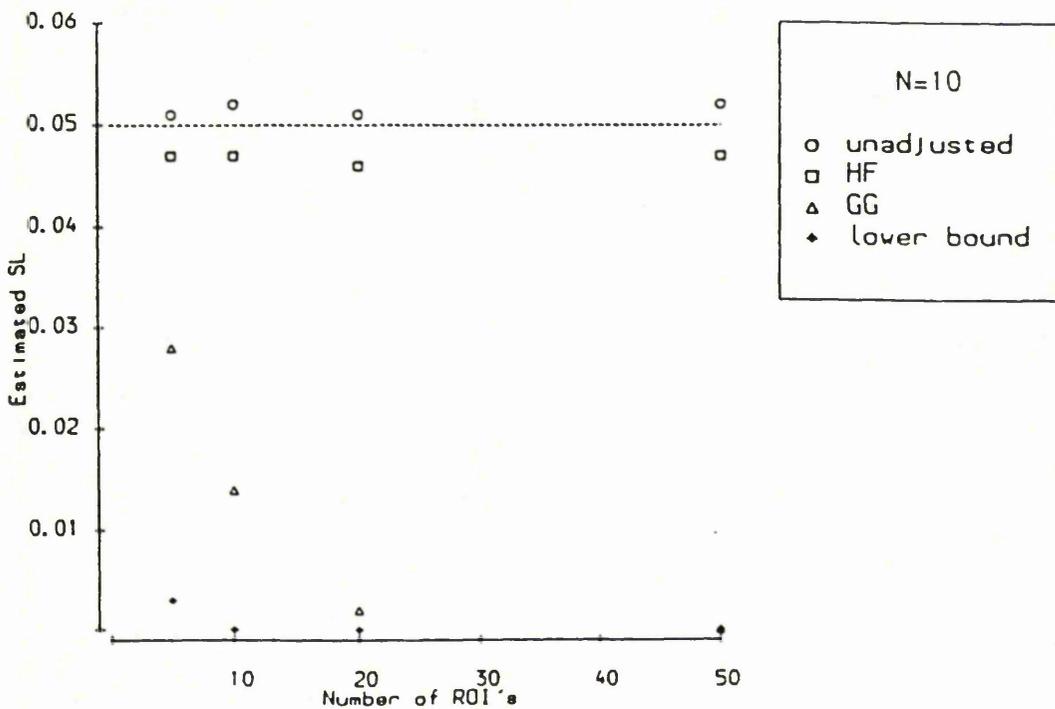
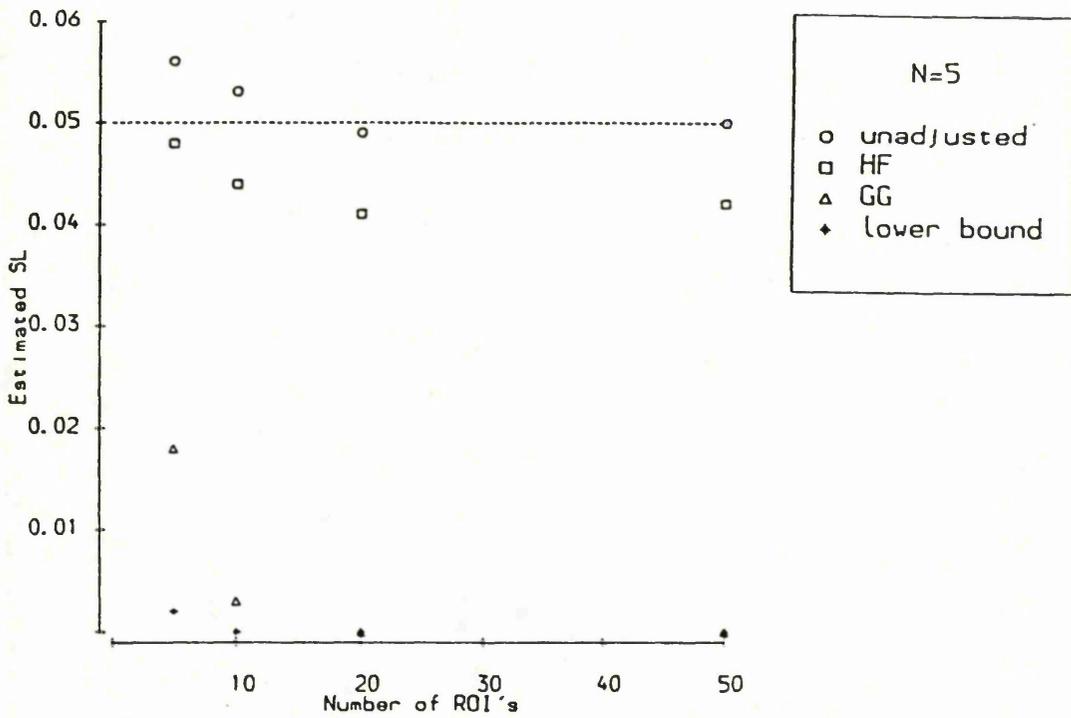
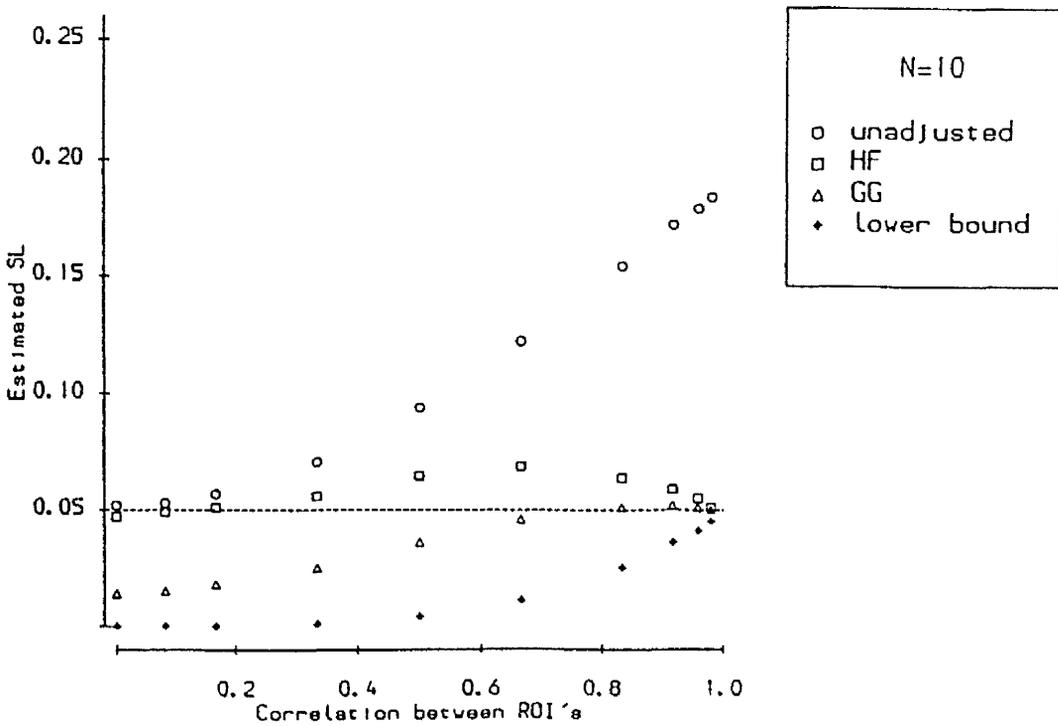
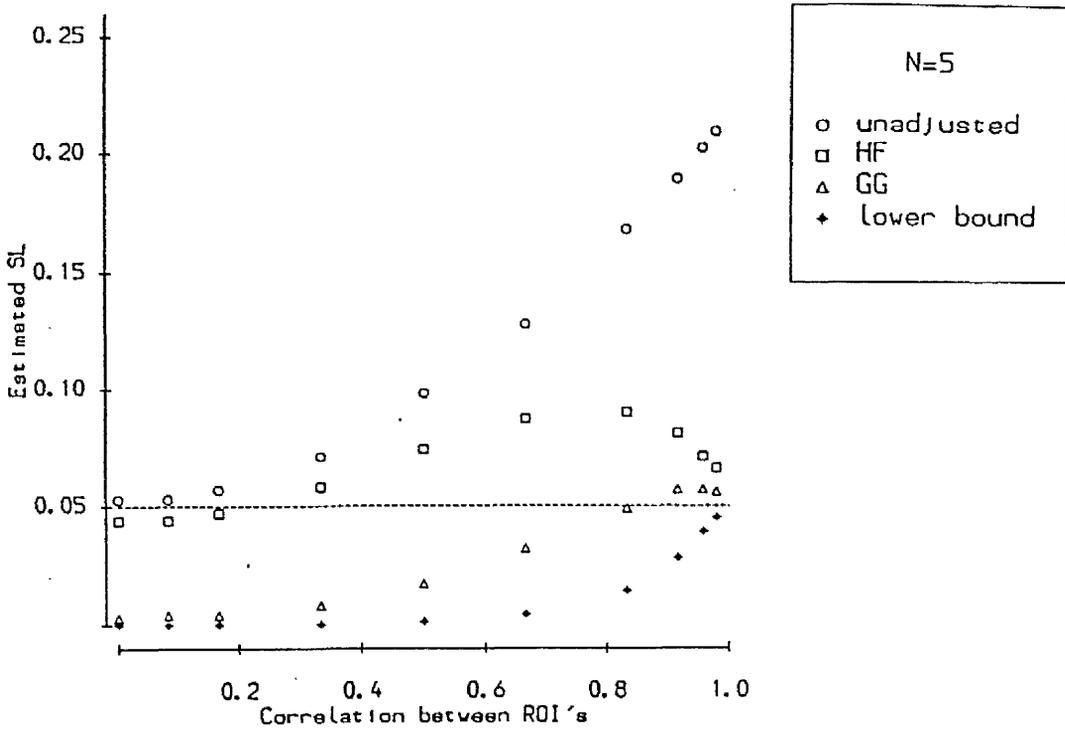


Figure 3.3 Empirical results of the estimated significance level corresponding to situation (a) in Table 3.2.



**Figure 3.4** Empirical results of the estimated significance level corresponding to situation (b) in Table 3.2.

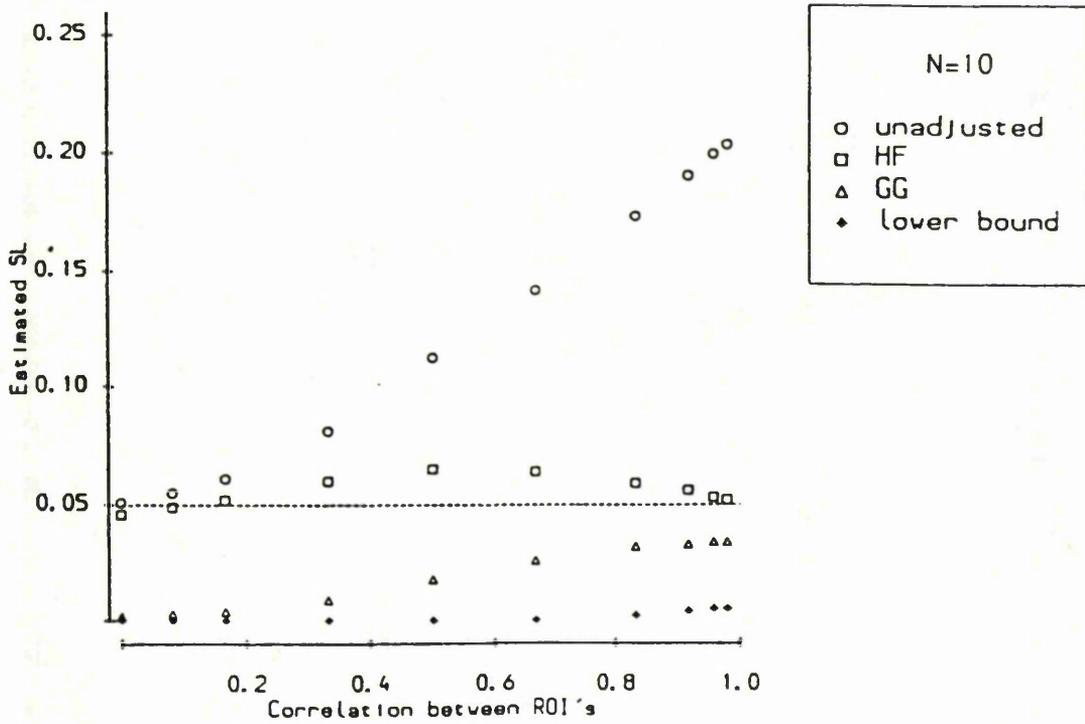
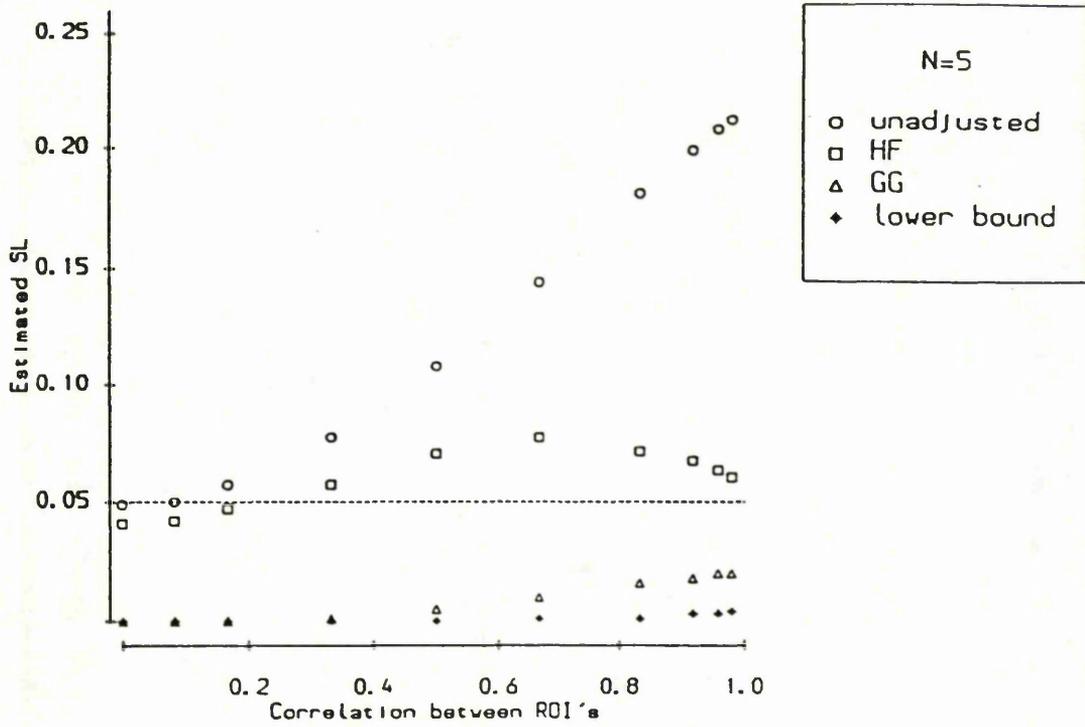
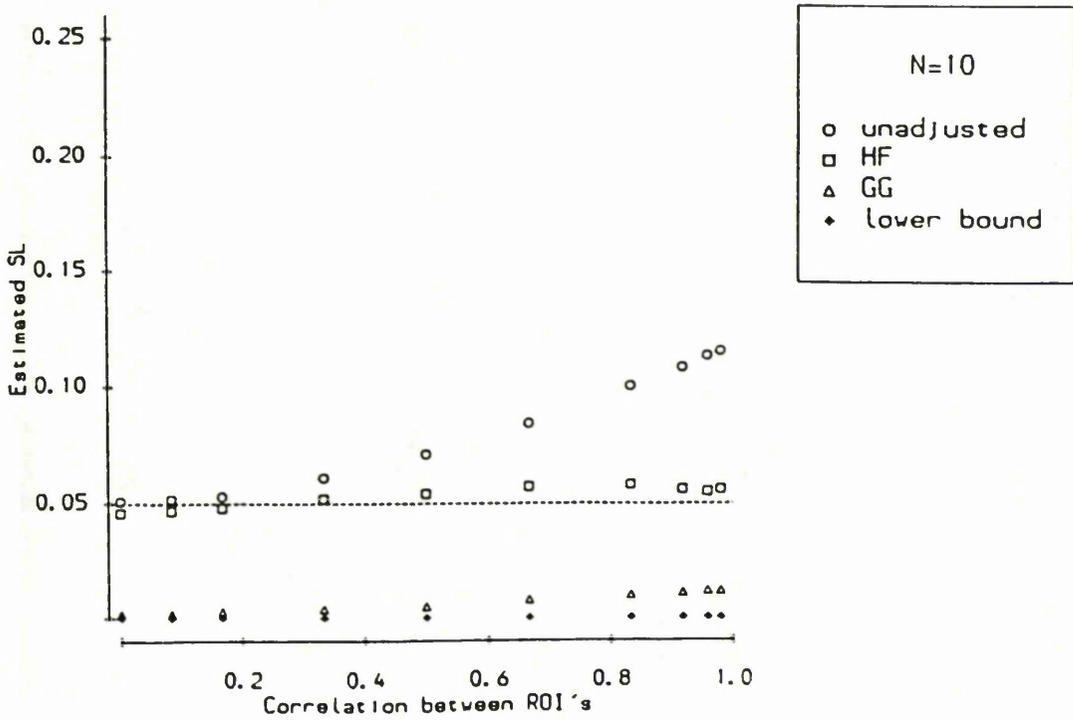
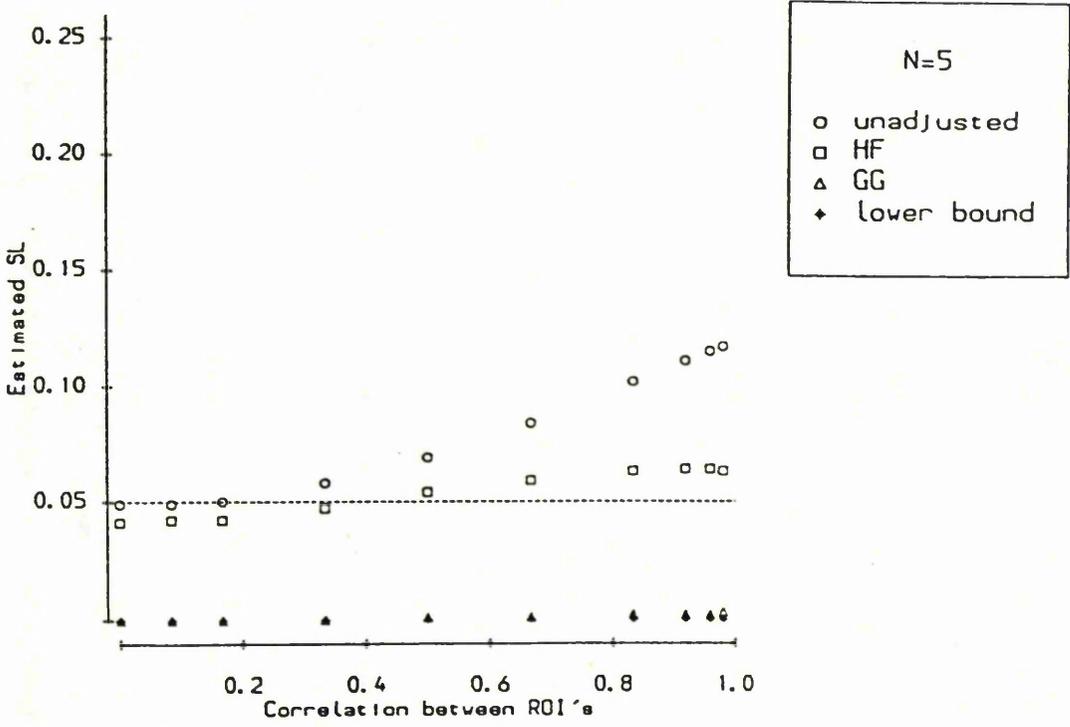
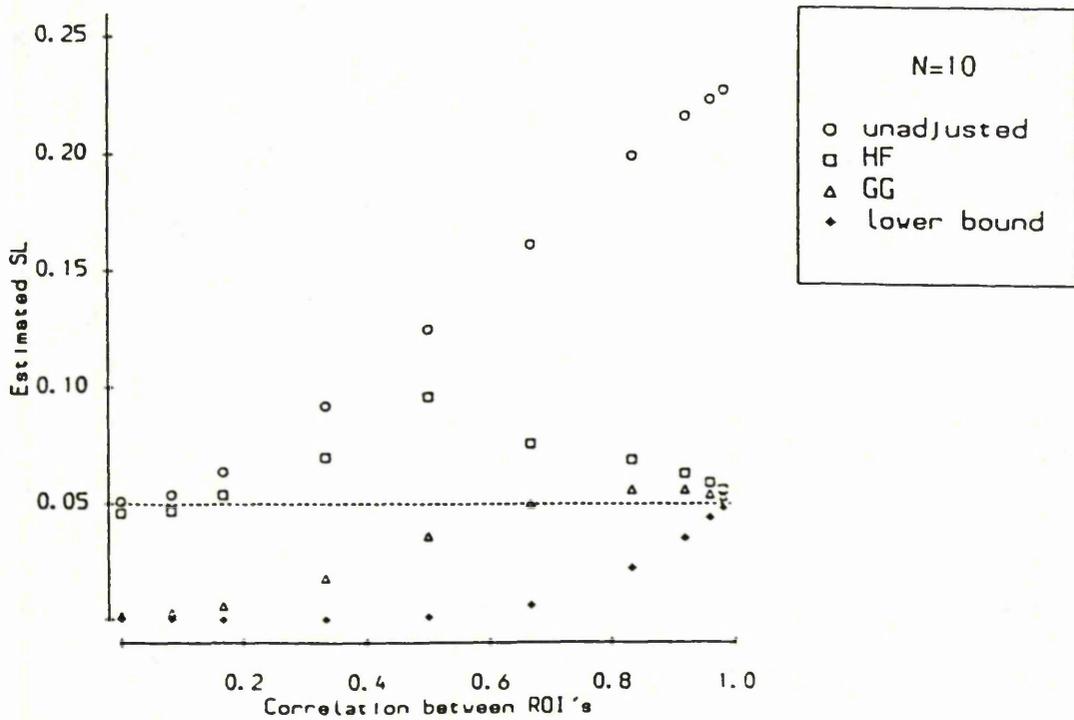
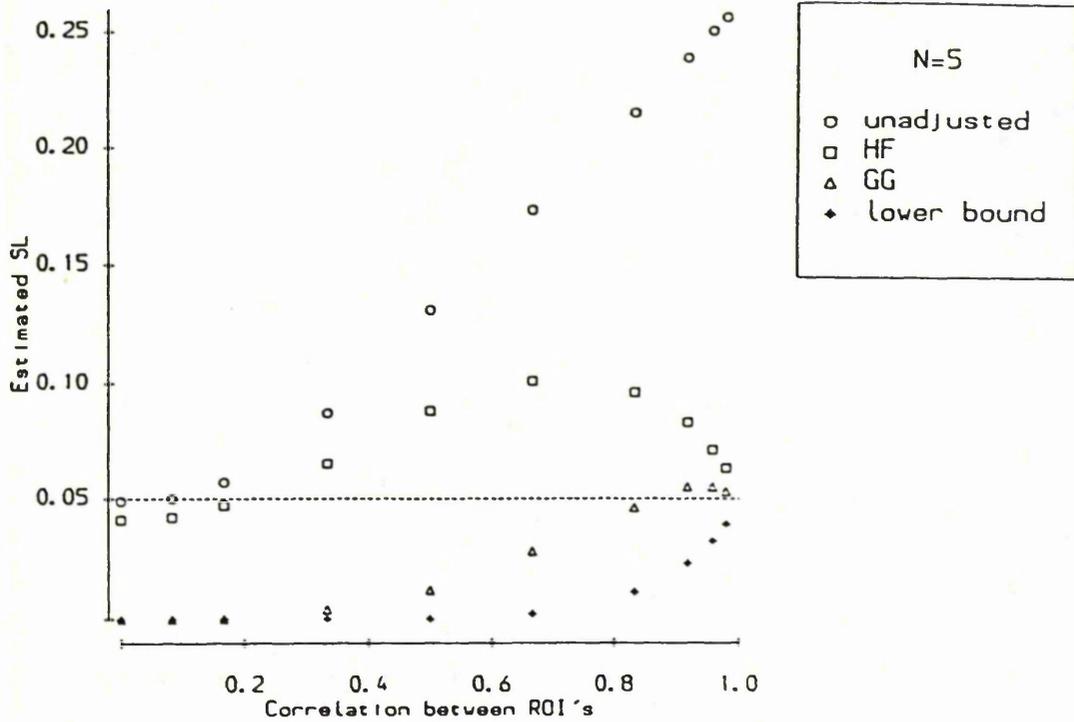


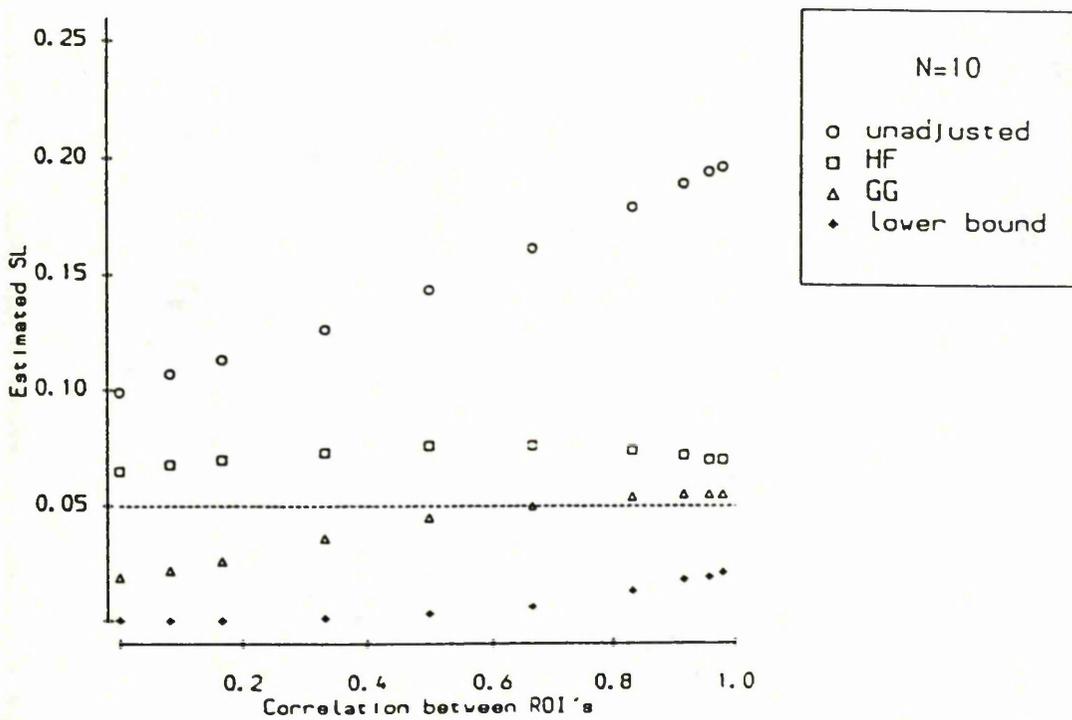
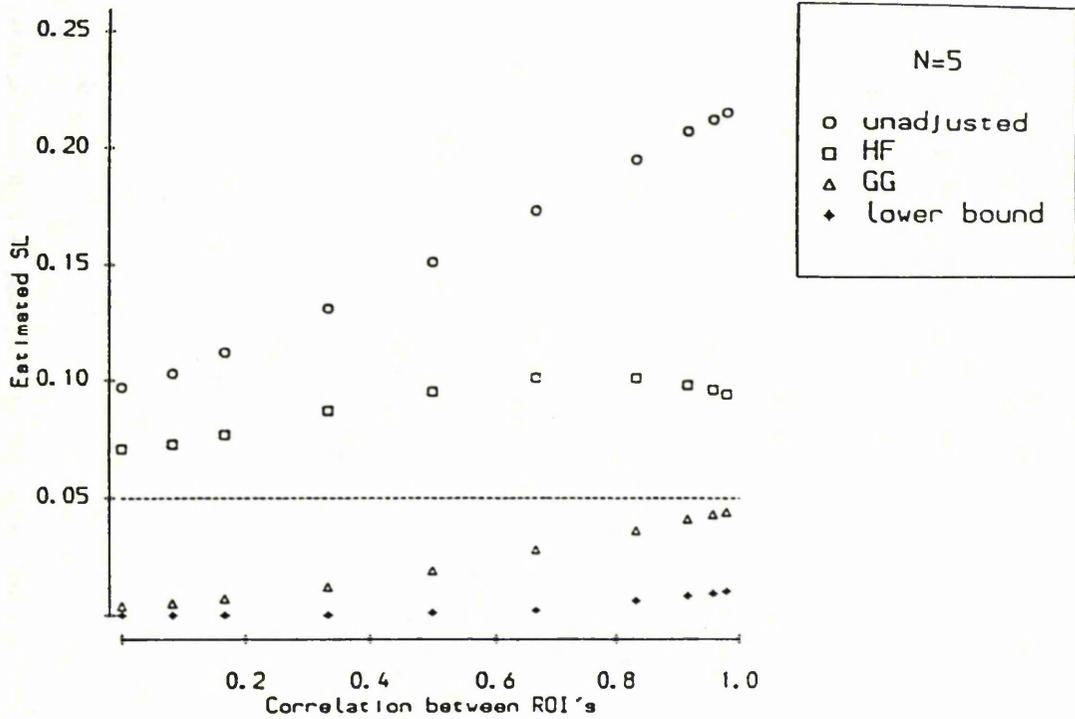
Figure 3.5 Empirical results of the estimated significance level corresponding to situation (c) in Table 3.2.



**Figure 3.6** Empirical results of the estimated significance level corresponding to situation (d) in Table 3.2.



**Figure 3.7** Empirical results of the estimated significance level corresponding to situation (e) in Table 3.2.



accounted for by sampling variation. Moreover, if a correction were to be unnecessarily applied, then the  $\tilde{\epsilon}$ -adjusted test would reliably reflect the nominal SL regardless of the number of ROI's being studied, particularly in the case of the larger sample size. On the other hand the  $\hat{\epsilon}$ -adjusted test deteriorates rapidly as the number of ROI's increases. Again the departure from the nominal SL is greater for the smaller sample sizes. Finally, the most conservative test performs poorly for any number of ROI's studied at both chosen sample sizes, rejecting the null hypothesis on almost no occasions.

### 3.6.2 VARYING THE CORRELATION STRUCTURE

Figures 3.3-3.5 display the results for the four different approaches using two, four and ten cluster correlation structures with 20 ROI's. Figure 3.6 contains estimates of the SL of the test using a two cluster correlation structure and 10 ROI's, whilst Figure 3.7 contains the estimated SL for the two cluster correlation matrix with 20 ROI's, all correlations in the first cluster equal to  $2/3$ , varying correlation in the second cluster and independence between clusters. Figures 3.3-3.7 indicate that, as expected, the estimated SL from the unadjusted test was equal to, or larger than, the nominal  $\alpha$ -level for both sample sizes, ( $n=5$  and  $n=10$ ) with the difference from the nominal  $\alpha$ -level increasing as the value of  $\epsilon$  decreases, or alternatively, as the correlation within clusters increases. This contrasts with the estimate of the SL obtained by the most conservative approach which in all hypothetical situations investigated, was less than the nominal  $\alpha$ -level for both sample sizes with

performance increasing as the true value of  $\epsilon$  decreased (or correlation within cluster increased). In general, the results for the  $\hat{\epsilon}$ -adjustment procedure suggest that the true SL will be considerably lower than the nominal level for low within cluster correlation, increasing to the nominal level with increasing correlation. The results of the  $\check{\epsilon}$  adjusted test, with regard to increasing the within cluster correlation, are more complex. Agreement is good, with the nominal SL, when the within cluster correlation is low (high values of  $\epsilon$ ). As the correlation increases, the estimated SL increases initially and then decreases towards the nominal 5% level.

In all cases, the nominal 5% significance level lay between the estimated significance levels of the Greenhouse-Geisser and Huynh-Feldt tests: the test based on the Huynh-Feldt estimator giving an inflated significance level whilst the Greenhouse-Geisser test was unduly conservative.

Figures 3.3-3.5 also indicate that as the number of clusters in the variance-covariance matrix increases the estimated SL for all four procedures decreases. Thus, the unadjusted and Huynh-Feldt tests will become less liberal, whilst the Greenhouse-Geisser and the most conservative test will become more conservative. Furthermore, Figure 3.3 and 3.6 show that the estimated SL for the unadjusted and Huynh-Feldt approaches will be further reduced, albeit only slightly for the Huynh-Feldt procedure, if the number of ROI's decreases. The Greenhouse-Geisser and the most conservative approach, on the other hand, show a slight increase in the estimated SL when the number of ROI's is reduced.

Thus, if the number of ROIs in the study is reduced, the performance of all four approaches will increase.

Examination of Figures 3.3-3.7 reveals that as the sample size decreases the performance of all four procedures deteriorates: the Greenhouse-Geisser and the most conservative approach become more conservative; the unadjusted and Huynh-Feldt procedures, in general, becoming more liberal.

The results for the two cluster correlation structure with 20 ROIs, all within cluster correlations equal to  $2/3$  and increasing correlation between ROIs not in the same cluster are given in Table B.7 . As expected, Table B.7 indicates that as the correlation between ROIs not in the same cluster increases up to a  $1/2$  the variance-covariance matrix departs less from sphericity and the corresponding estimated SL for all four approaches is reduced.

Generally speaking, Tables B.1 to B.7 reveal that the pattern of the results for nominal levels of 0.025 and 0.01 are very similar.

Finally, although the test based on the Huynh-Feldt correction factor comes closer to the nominal  $\alpha$  level over a wide range of  $\epsilon$  values, the performance of such a test can never be guaranteed to be closer to the nominal value than the GG approach for all of the variance-covariance matrices considered in this study. For example, while Figures 3.4 and 3.5 show that the Huynh-Feldt approach has a SL closer to the nominal value for all values of

the correlation between ROIs in the same cluster, Figures 3.3, 3.6 and 3.7 (those based on the two cluster correlation pattern) demonstrate that the Greenhouse-Geisser approach has a SL closer to the nominal value for high values of the within cluster correlation. Moreover, if a conservative test is required the researcher has no choice but to select the test based on the Greenhouse-Geisser correction factor.

Relating the results of the simulation study to the data set described in Section 3.4 does not establish conclusively the existence of a difference between the ratios of LCBF to LCGU. The more complex nature of the sample correlation structure, coupled with the small sample sizes, does not overwhelmingly suggest using the Huynh-Feldt estimator. However, considering the highly significant nature of the Huynh-Feldt test for this particular data set together with the Greenhouse-Geisser adjusted test almost reaching significance (Table 3.1), it may be reasonable to suggest that there is evidence of uncoupling.

### 3.6.3 THE DISTRIBUTION OF $\hat{\epsilon}$ AND $\tilde{\epsilon}$

Table 3.3.1 displays the mean and variances of  $\hat{\epsilon}$  and  $\tilde{\epsilon}$  for the independence covariance structure with 5, 10, 20 and 50 ROI's. As would be expected from Figure 3.2,  $\hat{\epsilon}$  grossly underestimates the true value of  $\epsilon$ , particularly for a large number of ROIs and small sample sizes. These results can be explained, at least in part, by the dependence of the maximum possible value of  $\hat{\epsilon}$  on the number of repeated levels and sample size.

The mean value of  $\tilde{\epsilon}$  also appears to be affected by sample size but is much closer to the true value of  $\epsilon$  than the corresponding Greenhouse-Geisser estimate. Increasing the number of ROIs appears to have little to no effect on the mean of  $\tilde{\epsilon}$ .

Tables 3.3.2-3.3.4 contain the mean and variances of  $\tilde{\epsilon}$  and  $\hat{\epsilon}$  for the two, four and ten cluster correlation pattern, with independence between clusters. Briefly, as expected from Figures 3.3-3.5, an examination of the results in Tables 3.3.2-3.3.4 show that  $\hat{\epsilon}$  tends to underestimate the true value of  $\epsilon$ , particularly when the sample size is small. Bias is reduced when the number of clusters in the population covariance matrix is small and the correlation between ROI's in the same cluster is high. In contrast, Tables 3.3.2-3.3.4 indicate that  $\tilde{\epsilon}$  tends to overestimate  $\epsilon$  when the variance-covariance structure departs from sphericity with the bias being reduced if the number of clusters present in the variance-covariance matrix is increased and the correlation between ROI's in the same cluster decreased.

In general, when the number of ROI's equals 20, the true value of  $\epsilon$  lies between the mean values of  $\hat{\epsilon}$  and  $\tilde{\epsilon}$ , with the mean value of  $\tilde{\epsilon}$  being closer to the population value of  $\epsilon$  whenever the population variance-covariance matrix exhibits a minor to moderate departure from sphericity. However, the reduction in bias by using the  $\tilde{\epsilon}$  estimator is offset by a larger variance.

Table 3.3.1 Means and variances ( $\hat{\epsilon}$  and  $\tilde{\epsilon}$ ) using an independence covariance matrix ( $\epsilon = 1.0$  in all situations)

J	n = 5		n = 10	
	$\hat{\epsilon}$	$\tilde{\epsilon}$	$\hat{\epsilon}$	$\tilde{\epsilon}$
5	0.520 (0.0086)	0.894 (0.0263)	0.691 (0.0079)	0.937 (0.0010)
10	0.307 (0.0016)	0.888 (0.0274)	0.490 (0.0026)	0.934 (0.0099)
20	0.172 (0.0002)	0.889 (0.0270)	0.316 (0.0005)	0.934 (0.0097)
50	0.075 ( <del>0.0001</del> )	0.888 (0.0276)	0.153 ( <del>0.0001</del> )	0.935 (0.0092)

Table 3.3.2 Means and variances ( ) of  $\hat{\varepsilon}$  and  $\tilde{\varepsilon}$  using the two cluster correlation matrix (20 ROI's, equal correlation within cluster, independence between clusters)

$\rho$	$\varepsilon$	$n = 5$		$n = 10$	
		$\hat{\varepsilon}$	$\tilde{\varepsilon}$	$\hat{\varepsilon}$	$\tilde{\varepsilon}$
0	1.000	0.172 (0.0002)	0.889 (0.0270)	0.316 (0.0005)	0.934 (0.0097)
1/12	0.964	0.171 (0.0002)	0.878 (0.0295)	0.312 (0.0005)	0.919 (0.0121)
1/6	0.860	0.169 (0.0002)	0.846 (0.0378)	0.302 (0.0007)	0.867 (0.0209)
1/3	0.561	0.159 (0.0004)	0.722 (0.0646)	0.265 (0.0017)	0.669 (0.0440)
1/2	0.318	0.143 (0.0007)	0.549 (0.0803)	0.212 (0.0025)	0.441 (0.0406)
2/3	0.174	0.121 (0.0009)	0.364 (0.0655)	0.153 (0.0020)	0.253 (0.0186)
5/6	0.096	0.092 (0.0007)	0.189 (0.0274)	0.097 (0.0006)	0.125 (0.0032)
11/12	0.071	0.074 (0.0004)	0.113 (0.0091)	0.073 (0.0002)	0.083 (0.0005)
23/24	0.061	0.064 (0.0001)	0.080 (0.0025)	0.062 (0.0001)	0.066 (0.0001)
49/50	0.057	0.058 (0.0001)	0.065 (0.0005)	0.057 (0.0001)	0.059 (0.0001)

Table 3.3.3 Means and variances of  $\hat{\varepsilon}$  and  $\tilde{\varepsilon}$  using the four cluster correlation structure (20 ROI's, equal correlation within cluster, independence between clusters)

$\rho$	$\varepsilon$	n = 5		n = 10	
		$\hat{\varepsilon}$	$\tilde{\varepsilon}$	$\hat{\varepsilon}$	$\tilde{\varepsilon}$
0	1.000	0.172 (0.0002)	0.889 (0.0270)	0.316 (0.0005)	0.934(0.0097)
1/12	0.977	0.172 (0.0002)	0.882 (0.0286)	0.313 (0.0005)	0.925 (0.0111)
1/6	0.910	0.170 (0.0002)	0.860 (0.0336)	0.307 (0.0006)	0.893 (0.0156)
1/3	0.701	0.163 (0.0003)	0.770 (0.0502)	0.281 (0.0008)	0.749 (0.0270)
1/2	0.491	0.151 (0.0004)	0.624 (0.0596)	0.243 (0.0010)	0.550 (0.0230)
2/3	0.333	0.135 (0.0004)	0.451 (0.0453)	0.200 (0.0008)	0.374 (0.0098)
5/6	0.228	0.117 (0.0004)	0.295 (0.0192)	0.157 (0.0005)	0.248 (0.0029)
11/12	0.189	0.106 (0.0004)	0.233 (0.0104)	0.138 (0.0003)	0.201 (0.0015)
23/24	0.173	0.101 (0.0003)	0.206 (0.0077)	0.128 (0.0003)	0.181 (0.0012)
49/50	0.165	0.098 (0.0003)	0.193 (0.0068)	0.124 (0.0003)	0.172 (0.0010)

Table 3.3.4 Means and variances of  $\hat{\varepsilon}$  and  $\tilde{\varepsilon}$  using the ten cluster correlation structure (20 ROI's, equal correlation within cluster, independence between clusters)

$\rho$	$\varepsilon$	$n = 5$		$n = 10$	
		$\hat{\varepsilon}$	$\tilde{\varepsilon}$	$\hat{\varepsilon}$	$\tilde{\varepsilon}$
0	1.000	0.172 (0.0002)	0.889 (0.0270)	0.316 (0.0005)	0.934 (0.0097)
1/12	0.993	0.172 (0.0002)	0.886 (0.0275)	0.314 (0.0005)	0.932 (0.0100)
1/6	0.973	0.172 (0.0002)	0.880 (0.0289)	0.313 (0.0005)	0.923 (0.0111)
1/3	0.897	0.169 (0.0002)	0.854 (0.0345)	0.305 (0.0005)	0.885 (0.0156)
1/2	0.792	0.165 (0.0002)	0.807 (0.0425)	0.292 (0.0006)	0.814 (0.0204)
2/3	0.677	0.160 (0.0003)	0.739 (0.0503)	0.275 (0.0006)	0.711 (0.0207)
5/6	0.569	0.153 (0.0003)	0.651 (0.0535)	0.254 (0.0006)	0.599 (0.0157)
11/12	0.519	0.150 (0.0003)	0.603 (0.0526)	0.244 (0.0006)	0.545 (0.0128)
23/24	0.496	0.148 (0.0003)	0.578 (0.0518)	0.238 (0.0006)	0.519 (0.0115)
49/50	0.484	0.146 (0.0003)	0.565 (0.0513)	0.235 (0.0006)	0.506 (0.0110)

### 3.7 THE RELATIVE POWER OF THE PROCEDURES

Rogan et al (1979) have investigated the relative power of the four univariate test procedures and the Hotellings  $T^2$  test using a range of covariance structures reflecting differing degrees of departure from sphericity. In the simulation studies, the unadjusted univariate test was more powerful than the multivariate test on all occasions, whilst the Huynh-Feldt and Greenhouse-Geisser adjusted tests were more powerful than the multivariate test when the covariance structure exhibited a minor violation of sphericity ( $\epsilon > 0.75$ ). In all simulated conditions the most conservative univariate test was less powerful than the multivariate test.

Davidson (1972) has also compared the power of the multivariate test with the unadjusted and the Greenhouse Geisser univariate tests in a simulation study. For an example variance-covariance matrix with  $\epsilon = 0.52$ , Davidson concluded that the multivariate test will be more powerful than the Greenhouse-Geisser test if the sample size exceeds the number of repeated levels. However, the power of the multivariate test will drop rapidly as the sample size approaches the number of repeated levels.

In a limited simulation study, we investigate the power of the four univariate tests when the number of ROIs exceeds the sample size. In such circumstances the multivariate test cannot be computed.

Data were generated for specific sample sizes ( $n=5,10$ ) and 20

ROIs using pseudo random number generators for multivariate normal distributions, with a two cluster correlation structure and specified mean vector. Two values of the within cluster correlation,  $\rho$ , were selected ( $\rho=1/3$  and  $\rho=2/3$ ). Mean vectors were chosen such that 5 of the 10 ROI's in each cluster took the value A, the remaining 5 taking the value of 0. For each value of A, either 0.5, 1.0, 1.5 or 2.0, the estimated power was derived in a similar fashion to the simulation study in section 3.5, with the relative frequency of significance taken as an estimate of the true power. The results are given in Table 3.4 for nominal significant levels of 0.05, 0.025 and 0.01.

For the test of equality of mean ROI values, the four univariate tests all have the same test statistic and only differ in that the degrees of freedom for the critical values are different. Hence, under all circumstances, if the test statistic is compared to the same quantile of the respective F distributions, the unadjusted test would be deemed the most powerful approach and the most conservative approach the least powerful. However, since the Type I error rate of the unadjusted test is greatly inflated when the sphericity assumption is violated, the power of this test will also be inflated. Thus, although it would appear that some tests have much greater power than others, we are not really comparing like with like and therefore any increase in power must be contrasted with the instability of the significance levels.

As expected, the data from this investigation indicate that as A and n increase there will be a substantial increase in power. For example, when A=2.0 and n=10 the power of all four tests,

Table 3.4 Power for the test of equality of mean response

n	$\rho$	$\varepsilon$	A	Unadjusted			$\hat{\varepsilon}$ -adjusted			$\hat{\varepsilon}$ -adjusted			Lower bound			
				0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	
5	1/3	0.561	0.5	0.21	0.16	0.12	0.17	0.13	0.09	0.02	0.01	0.00	0.00	0.00	0.00	
				1.0	0.53	0.47	0.39	0.46	0.38	0.31	0.13	0.06	0.02	0.00	0.00	0.00
				1.5	0.85	0.81	0.74	0.78	0.72	0.62	0.42	0.25	0.11	0.03	0.00	0.00
				2.0	0.98	0.96	0.95	0.95	0.92	0.87	0.75	0.57	0.36	0.16	0.01	0.00
5	2/3	0.174	0.5	0.27	0.23	0.20	0.17	0.13	0.10	0.07	0.04	0.01	0.01	0.00	0.00	
				1.0	0.52	0.48	0.43	0.36	0.29	0.24	0.19	0.12	0.06	0.04	0.01	0.00
				1.5	0.78	0.75	0.70	0.59	0.51	0.43	0.41	0.29	0.17	0.14	0.03	0.00
				2.0	0.93	0.92	0.89	0.79	0.72	0.63	0.65	0.51	0.36	0.34	0.12	0.01
10	1/3	0.561	0.5	0.35	0.29	0.23	0.28	0.22	0.16	0.14	0.09	0.05	0.01	0.00	0.00	
				1.0	0.83	0.79	0.73	0.77	0.70	0.63	0.61	0.49	0.36	0.11	0.02	0.00
				1.5	0.99	0.99	0.98	0.98	0.97	0.95	0.95	0.91	0.82	0.56	0.24	0.04
				2.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.74	0.33
10	2/3	0.174	0.5	0.37	0.33	0.28	0.21	0.16	0.12	0.16	0.11	0.07	0.04	0.01	0.00	
				1.0	0.76	0.72	0.68	0.40	0.48	0.57	0.50	0.39	0.28	0.24	0.11	0.03
				1.5	0.96	0.95	0.94	0.87	0.81	0.73	0.83	0.75	0.64	0.63	0.42	0.18
				2.0	1.00	1.00	1.00	0.98	0.97	0.93	0.98	0.95	0.89	0.91	0.78	0.53

using a significance level of 0.05, exceeds 0.90, whilst in the corresponding situation with  $n$  reduced to 5 only the power of the unadjusted test remains above 0.90.

In general, as the correlation within clusters increases the unadjusted and Huynh-Feldt tests become less powerful. In contrast, the most conservative test becomes more powerful as the correlation within clusters increases.

The results for the Greenhouse-Geisser test are more complex. Table 3.4 indicates that the test becomes less powerful as the within cluster correlation increases, especially for large values of  $A$  with  $n=10$ . On the other hand, for  $n=5$  an increase in power is observed for the 0.025 and 0.01 significance levels at values of  $A$  not exceeding 1.5. Employing a significance level of 0.05 an increase in power is only observed for values of  $A$  equal to 0.5 and 1.0.

### **3.8 MULTIPLE COMPARISON PROCEDURES IN REPEATED MEASURES DESIGNS**

To distinguish whether there has been any significant uncoupling between LCBF and LCGU across regions, an alternative approach to the repeated measures F-test could be built up by considering all possible pairwise comparisons between the ROI's. Maxwell (1980) investigated the performance of Type I and Type II error rates using five different multiple comparisons procedures, for repeated measures designs with 3, 4 and 5 repeated levels and covariance structures exhibiting various degrees of departure from sphericity. Mitzel and Games (1981), in a more limited study, investigated the effect of employing a separate error

term for each comparison, as opposed to a constant estimate of the variance for all contrasts based on the interaction mean square from the analysis of variance table. For  $J$  ROIs, the latter approach is equivalent to assuming that the  $J(J-1)/2$  variances of the estimated differences of the pairwise ROI means are equal. This is precisely the sphericity assumption.

In both these studies, the authors advocate that Bonferroni pairwise multiple comparisons are preferable to maintain the Type I error rate. However, the performance of multiple comparison procedures, when the number of repeated levels is large and exceeds the sample size is unknown. To assess the appropriateness of pairwise comparisons in such circumstances, we investigate, empirically, the properties of several post hoc pairwise comparison techniques in a number of hypothetical situations similar to those described in section 3.5, where the brain is considered to be made up of clusters of regions. But first, we consider a number of approaches for conducting simultaneous pairwise comparisons, four of which require the sphericity assumption (Methods 1-4), two based on the interaction mean square but which attempt to adjust for departures from sphericity (Methods 5 and 6) and four based on separate estimates of the variance for each contrast (Methods 7-10).

### **3.8.1 MULTIPLE COMPARISON PROCEDURES**

We will consider only the mixed-two way model, with a single replication in each cell, where we are interested in comparing the mean response of the measurement of interest between ROIs.

Let  $X_{ij}$  denote the measurement of interest on the  $i^{\text{th}}$  patient in region  $j$  ( $i=1, \dots, n; j=1, \dots, J$ ) as before. We further assume that the vector of responses,  $\underline{X}_i = (X_{i1}, X_{i2}, \dots, X_{iJ})^T$ , has a  $J$ -dimensional normal distribution with mean vector  $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_J)^T$  and covariance matrix  $\Sigma = \{\sigma_{rs}\}$ . Now, let  $\psi_{rs} = \mu_r - \mu_s$  ( $1 \leq r < s \leq J, r \neq s$ ) denote the contrast between the  $r^{\text{th}}$  and  $s^{\text{th}}$  elements of the mean vector  $\underline{\mu}$ . An unbiased estimate of  $\psi_{rs}$  is given by  $\hat{\psi}_{rs} = \bar{X}_{.r} - \bar{X}_{.s}$ . Consequently, if  $\hat{V}(\hat{\psi}_{rs})$  is an unbiased estimate of the variance of  $\hat{\psi}_{rs}$  then we can construct confidence intervals for all  $J(J-1)/2$  pairwise contrasts of the form

$$\hat{\psi}_{rs} \pm CV \times [\hat{V}(\hat{\psi}_{rs})]^{1/2} \quad (1 \leq r < s \leq J, r \neq s) \quad (3.7.1)$$

where CV denotes the critical value matching the type of multiple comparison required.

#### METHOD 1 : TUKEY T-PROCEDURE (TUKEY, 1953)

If  $\Sigma$  displays sphericity then a Tukey-type multiple comparison procedure exists which gives exact  $(1-\alpha)$  level joint confidence intervals for the set of differences of the form  $\mu_r - \mu_s$ . These intervals are given explicitly by

$$\mu_r - \mu_s \in [\bar{X}_{.r} - \bar{X}_{.s} \pm Q_{J, (J-1)(n-1)}^\alpha (MS/n)^{1/2}] ; 1 \leq r < s \leq J$$

and satisfy the probability statement:

$$P\{\mu_r - \mu_s \in [\bar{X}_{.r} - \bar{X}_{.s} \pm Q_{J, (J-1)(n-1)}^\alpha (MS/n)^{1/2}] ; 1 \leq r < s \leq J, r \neq s\} = 1 - \alpha \quad (3.7.2)$$

where

$$MS = \frac{\sum_{i=1}^n \sum_{j=1}^J (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2}{(J-1)(n-1)}$$

denotes the mean square for the interaction between individuals and brain regions and  $Q_{J, (J-1)(n-1)}^{(\alpha)}$  is the upper  $\alpha$ -point of the Studentised range distribution with  $J$  and  $(J-1)(n-1)$  degrees of freedom. In terms of the notation in (3.7.1)  $\hat{V}(\hat{\psi}_{rs}) = (2MS/n)^{1/2}$  and  $CV = Q_{J, (J-1)(n-1)}^{(\alpha)} / \sqrt{2}$ . Thus if  $\Sigma$  exhibits sphericity, this procedure, often referred to in the literature as the honestly significant difference or the wholly significant difference test, must provide the shortest possible intervals which guarantee a joint confidence level of at least  $1-\alpha$ . However, if  $\Sigma$  departs from sphericity the probability statement (3.7.2) will not necessarily be preserved.

#### METHOD 2 : SCHEFFÉ S-METHOD (SCHEFFÉ)

As an alternative approach, the Scheffé (1953) S-procedure will, under the sphericity assumption, provide exact  $(1-\alpha)$ -level joint confidence intervals for all possible contrasts. Thus if sphericity holds, the joint confidence level for all pairwise contrasts of the form

$$\mu_r - \mu_s \in [\bar{X}_{.r} - \bar{X}_{.s} \pm \{(J-1)F_{(J-1), (J-1)(n-1)}^{(\alpha)}\}^{1/2} (2MS/n)^{1/2}]; \quad 1 \leq r < s \leq J$$

will be conservative, i.e.

$$P\{\mu_r - \mu_s \in [\bar{X}_{.r} - \bar{X}_{.s} \pm \{(J-1)F_{(J-1), (J-1)(n-1)}^{(\alpha)}\}^{1/2} (2MS/n)^{1/2}\} ;$$

$$1 \leq r < s \leq J\} \geq 1-\alpha$$

where  $MS$  is defined above and  $F_{(J-1), (J-1)(n-1)}^{(\alpha)}$  is the upper  $\alpha$  point from the  $F$  distribution with  $J-1$  and  $(J-1)(n-1)$  degrees of freedom.

METHOD 3 : BONFERRONI PROCEDURE (BONF(1))

When the number of ROI's is large, a Bonferroni type procedure may well provide shorter intervals than the Scheffe S-procedure, as well as preserving a joint confidence level of at least  $(1-\alpha)$ .

Let  $T_{rs} = (\bar{X}_{.r} - \bar{X}_{.s}) (2MS/n)^{-1/2}$  then the Bonferroni procedure is based on the inequality:

$$\Pr \left\{ \max_{1 \leq r < s \leq J} |T_{rs}| \leq \xi \right\} \geq 1 - \sum_{1 \leq r < s \leq J} \Pr \left\{ |T_{rs}| \leq \xi \right\} \quad (2.7.3.)$$

Conservative Bonferroni  $(1-\alpha)$  level joint confidence intervals for all pairwise comparisons can then be given by

$$\mu_r - \mu_s \in [\bar{X}_{.r} - \bar{X}_{.s} \pm T_{(J-1)(n-1)}^{(\alpha/2J^*)} (2MS/n)^{1/2}] ; \quad 1 \leq r < s \leq J$$

where  $T_{(J-1)(n-1)}^{(\alpha/2J^*)}$  denotes the upper  $(\alpha/2J^*)$  percentage point from the Studentised  $t$  distribution with  $(J-1)(n-1)$  degrees of freedom, and  $J^* = J(J-1)/2$ . Note that these intervals have been computed with the assumption that all pairwise differences in the means have a common variance.

METHOD 4: DUNN SIDAK PROCEDURE (SIDAK (1))

Using the sharper inequality (Sidak, 1967):

$$\Pr \left\{ \max_{1 \leq r < s \leq J} |T_{rs}| \leq \xi \right\} \geq \prod_{1 \leq r < s \leq J} \Pr \left\{ |T_{rs}| \leq \xi \right\} \quad (2.7.4),$$

a less conservative upper bound on the upper  $\alpha$  point of the distribution of  $\max_{1 \leq r < s \leq J} |T_{rs}|$  can be obtained, namely

$$T_{(J-1)(n-1)}^{(\alpha'/2)} \text{ where } \alpha' = 1 - (1 - \alpha)^{1/J^*} \text{ and } J^* = J(J-1)/2. \quad \text{Thus,}$$

shorter intervals of the form

$$\mu_r - \mu_s \in [\bar{X}_{.r} - \bar{X}_{.s} \pm T_{(J-1)(n-1)}^{(\alpha/2)} (2MS/n)^{1/2}]; 1 \leq r < s \leq J$$

can be computed which maintain a nominal joint confidence level of at least  $1-\alpha$  under the sphericity assumption. When the number of comparisons is large  $T_{(J-1)(n-1)}^{(\alpha/2)}$  will be very close to  $T_{(J-1)(n-1)}^{(\alpha/2J^*)}$ . Therefore we would expect the Sidak procedure to yield intervals of roughly the same length as the Bonferroni procedure.

As mentioned earlier in this chapter, the sphericity assumption will almost certainly be violated in ROI studies, and hence the joint confidence level associated with methods 1-4 may not necessarily be greater than  $1-\alpha$ . In Section 3.3.4 we discussed the use of adjustment factors to correct the degrees of freedom of the F distribution when conducting F tests in repeated measures ANOVA. If the covariance matrix deviates from the sphericity assumption, we propose that Scheffe type pairwise intervals, based on the Greenhouse-Geisser (1959) and Huynh-Feldt (1976) correction factors may exert better control on the joint confidence level than the Scheffe S-procedure. However, alternative post hoc pairwise comparisons in repeated measures designs can be derived, based on the estimated variance of each contrast. These preserve the nominal coverage probability of the simultaneous confidence intervals.

#### METHOD 5 GREENHOUSE-GEISSER ADJUSTED INTERVALS (GGADJ)

Let  $\hat{\epsilon}$  denote the Greenhouse-Geisser adjustment factor, estimated from the sample covariance matrix. Then we propose pairwise intervals of the form

$$\mu_r - \mu_s \in [\bar{X}_{..r} - \bar{X}_{..s} \pm \{(J-1)F^{(\alpha)}_{(J-1)\hat{\epsilon}, (J-1)(n-1)\hat{\epsilon}}\}^{1/2} (2MS/n)^{1/2}] \quad (1 \leq r < s \leq J)$$

as an alternative to the Scheffe S-procedure.

#### METHOD 6 HUYNH-FELDT ADJUSTED INTERVALS (HFADJ)

Let  $\tilde{\epsilon}$  denote the estimated Huynf-Feldt adjustment factor from the sample covariance matrix. We have shown earlier that  $\tilde{\epsilon} \geq \hat{\epsilon}$  for any given variance-covariance matrix. Thus intervals of the form

$$\mu_r - \mu_s \in [\bar{X}_{..r} - \bar{X}_{..s} \pm \{(J-1)F^{(\alpha)}_{(J-1)\tilde{\epsilon}, (J-1)(n-1)\tilde{\epsilon}}\}^{1/2} (2MS/n)^{1/2}] \quad (1 \leq r < s \leq J)$$

will be shorter than those based on the Greenhouse-Geisser adjustment factor and hence the joint confidence level for this procedure will be less. However, there is no guarantee that the actual simultaneous confidence level for these approaches will be equal to or greater than the stated confidence level of  $1-\alpha$ .

#### METHOD 7 BONFERRONI PROCEDURE (BONF (2))

Let  $d_{rs} = (\sigma_{rr} + \sigma_{ss} - 2\sigma_{rs})/n$  ( $1 \leq r < s \leq J$ ) denote the variance of the contrast between the  $r^{\text{th}}$  and  $s^{\text{th}}$  ROIs. This can be estimated from the sample covariance matrix,  $S = \{s_{rs}\}$ , giving

$$\hat{d}_{rs} = (s_{rr} + s_{ss} - 2s_{rs})/n .$$

Then, using the inequality (2.7.3), conservative pairwise confidence intervals of the form

$$\mu_r - \mu_s \in [\bar{X}_{..r} - \bar{X}_{..s} \pm T^{(\alpha/2J)}_{(n-1)} (\hat{d}_{rs})^{1/2}]; \quad 1 \leq r < s \leq J$$

can be computed which satisfy the joint confidence level, i.e.

$$P\{\mu_r - \mu_s \in [\bar{X}_{..r} - \bar{X}_{..s} \pm T^{(\alpha/2J)}_{(n-1)} (\hat{d}_{rs})^{1/2}]; \quad 1 \leq r < s \leq J \} \geq 1-\alpha .$$

METHOD 8 DUNN SIDAK PROCEDURE (SIDAK (2))

Similarly, a less conservative approach can be obtained using the inequality (2.7.4.). The resulting pairwise intervals are of the form

$$\mu_r - \mu_s \in [\bar{X}_{\cdot r} - \bar{X}_{\cdot s} \pm T^{(\alpha/2)}_{(n-1)} (\hat{d}_{rs})^{1/2}] ; 1 \leq r < s \leq J$$

where  $\alpha$  is as before.

Notice that for methods 7 and 8 the degrees of freedom from the Studentised t distribution has been reduced to  $n-1$  since each element of the covariance structure has been estimated with  $n-1$  degrees of freedom.

METHOD 9 GREENHOUSE-GEISSER SCHEFFE TYPE INTERVALS (SCHGG)

Let  $\hat{\epsilon}$  denote the Greenhouse-Geisser adjustment factor, estimated from the sample covariance structure. Although we give no justification, we have investigated the control on the nominal coverage probability exerted by pairwise intervals of the form

$$\mu_r - \mu_s \in [\bar{X}_{\cdot r} - \bar{X}_{\cdot s} \pm \{(J-1)F^{(\alpha)}_{(J-1)\hat{\epsilon}, (J-1)(n-1)\hat{\epsilon}}\}^{1/2} (\hat{d}_{rs})^{1/2}] ; (1 \leq r < s \leq J) .$$

METHOD 10 HUYNH-FELDT SCHEFFE TYPE INTERVALS (SCHHF)

Similarly, we can construct intervals of the form

$$\mu_r - \mu_s \in [\bar{X}_{\cdot r} - \bar{X}_{\cdot s} \pm \{(J-1)F^{(\alpha)}_{(J-1)\hat{\epsilon}, (J-1)(n-1)\hat{\epsilon}}\}^{1/2} (\hat{d}_{rs})^{1/2}] ; (1 \leq r < s \leq J)$$

which will be shorter in length than the corresponding intervals based on the Greenhouse-Geisser adjustment factor. Again, we provide no justification for this approach.

### 3.8.2 SIMULATION STUDY

In the present study, we compare these various approaches to pairwise comparisons, with respect to the estimated joint confidence level, in the mixed model with a single replication per cell. For the purpose of this investigation, we have considered correlation patterns where the brain is considered to be composed of either two or four clusters of regions with a total of 20 ROI's, and the two cluster correlation pattern with 10 ROI's. Six levels were selected for the within cluster correlation ( $\rho=0, 1/6, 1/3, 1/2, 2/3, 5/6$ ) with independence between regions in different clusters. Notice that when  $\rho=0$  we have a correlation pattern with independence between all ROI's, corresponding to the situation where there is no violation of the sphericity assumption.

Estimation of the true joint confidence is detailed below:

- (i) data are generated with zero mean vector and specified covariance structure for specific sample sizes ( $n=10,30$ ) and number of ROI's using a pseudo random number generator for multivariate normal distributions.
- (ii) the sample covariance matrix, the average sample variance ( $\bar{S}^2$ ) and the average sample covariance ( $\bar{C}$ ) are calculated from the data.
- (iii)  $MS = \bar{S}^2 - \bar{C}$  (Winer, 1971) is then calculated.
- (iv) For each of the  $J(J-1)/2$  comparisons the following statistics are calculated:

$$t_1 = (\bar{X}_{.r} - \bar{X}_{.s}) / (2MS/n)^{1/2} \quad \text{and}$$

$$t_2 = (\bar{X}_{.r} - \bar{X}_{.s}) / (\hat{d}_{rs})^{1/2}$$

- (v) The maximum  $t_1$  and  $t_2$  value,  $t_1^*$  and  $t_2^*$  say, from the  $J(J-1)/2$  statistics is determined and stored in an array.
- (vi)  $t_1^*$  is compared to each of the critical values from methods 1-6 and  $t_2^*$  to each of the critical values from methods 7-10 and significance determined ( $\alpha = 0.05$ )
- (vii) Steps (i)-(vi) are repeated 10000 times and the relative frequency of significance, for each approach, is taken as an estimate of the joint significance level for that particular method.

In all the studies, we also determine, from the stored values of  $t_1^*$  and  $t_2^*$ , the critical values for which a joint confidence level of 0.95 would be attained. Finally, for the two cluster correlation pattern, we also investigate the frequency of occurrence of a significant  $t_1^*$  or  $t_2^*$  statistic coming from a between cluster comparison.

### 3.8.3. RESULTS

Empirical joint confidence levels from the 10000 simulations are plotted in Figures 3.8(i)-(vi) for each combination of covariance structure, sample size and number of ROIs studied. Since the critical value of the Bonferroni procedure will be very similar to the corresponding Sidak procedure when the number of pairwise comparisons is large, only the estimated joint confidence level of Sidak procedure has been plotted. The estimated joint confidence level from both procedures were equal (to two decimal places) for all combinations of factors. For presentation purposes the GGADJ and HFADJ procedures have also been omitted from the figures. Tabulated results for all approaches are given

Figure 3.8(i) Estimated joint confidence levels for the four cluster correlation matrix with  $J=20$  and  $n=10$ .

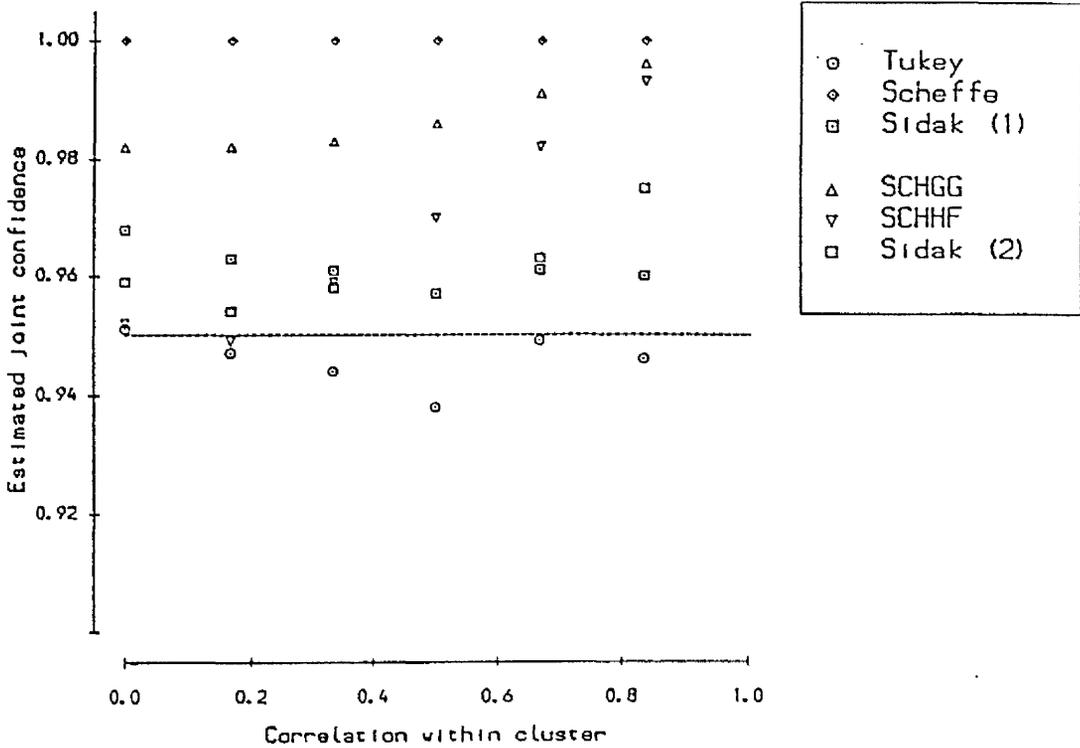


Figure 3.8(ii) Estimated joint confidence levels for the four cluster correlation matrix with  $J=20$  and  $n=30$ .

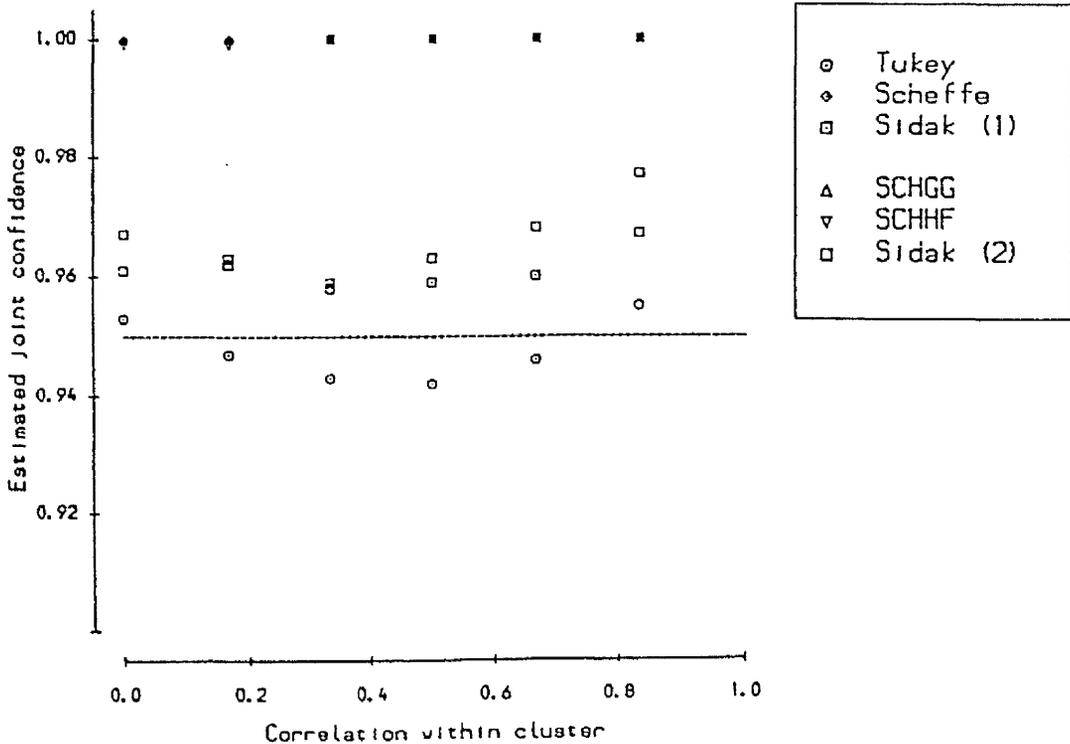


Figure 3.8(iii) Estimated joint confidence levels for the two cluster correlation matrix with  $J=20$  and  $n=10$ .

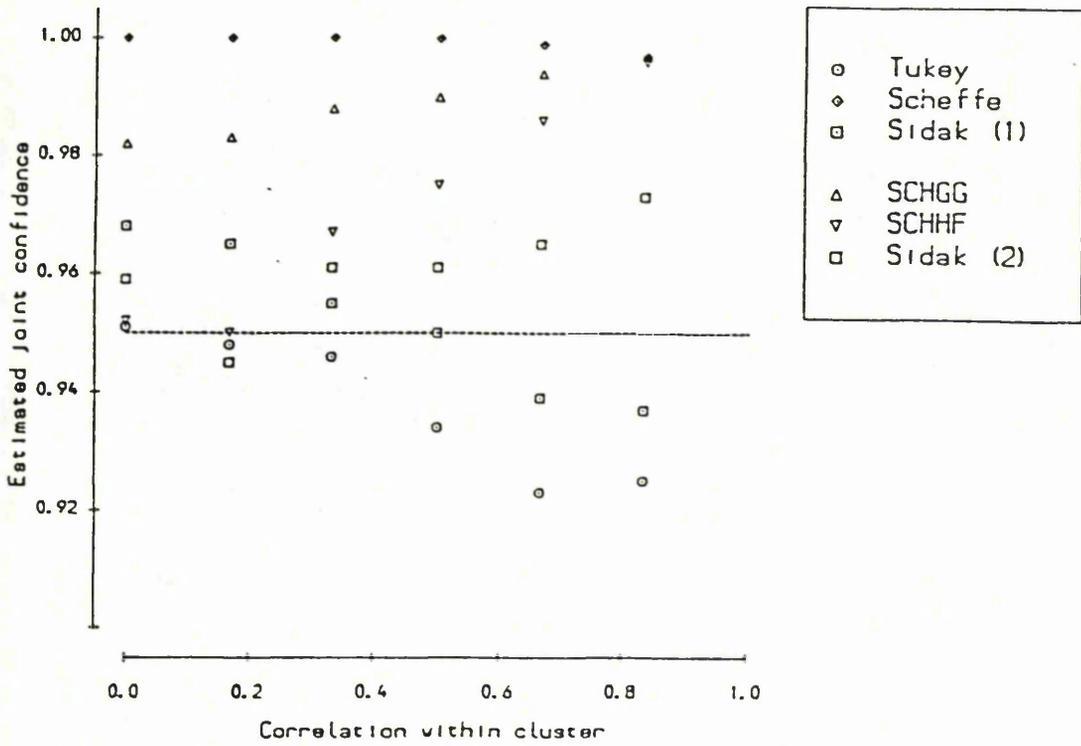


Figure 3.8(iv) Estimated joint confidence levels for the two cluster correlation matrix with  $J=20$  and  $n=30$ .

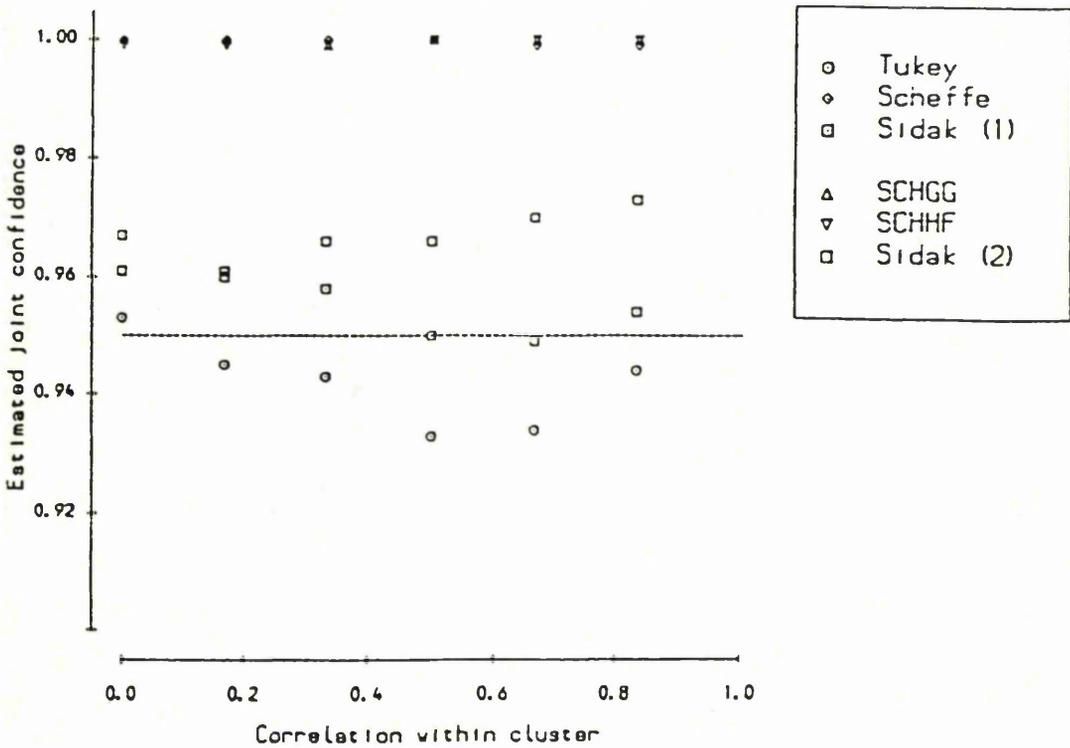


Figure 3.8(v) Estimated joint confidence levels for the two cluster correlation matrix with  $J=10$  and  $n=10$ .

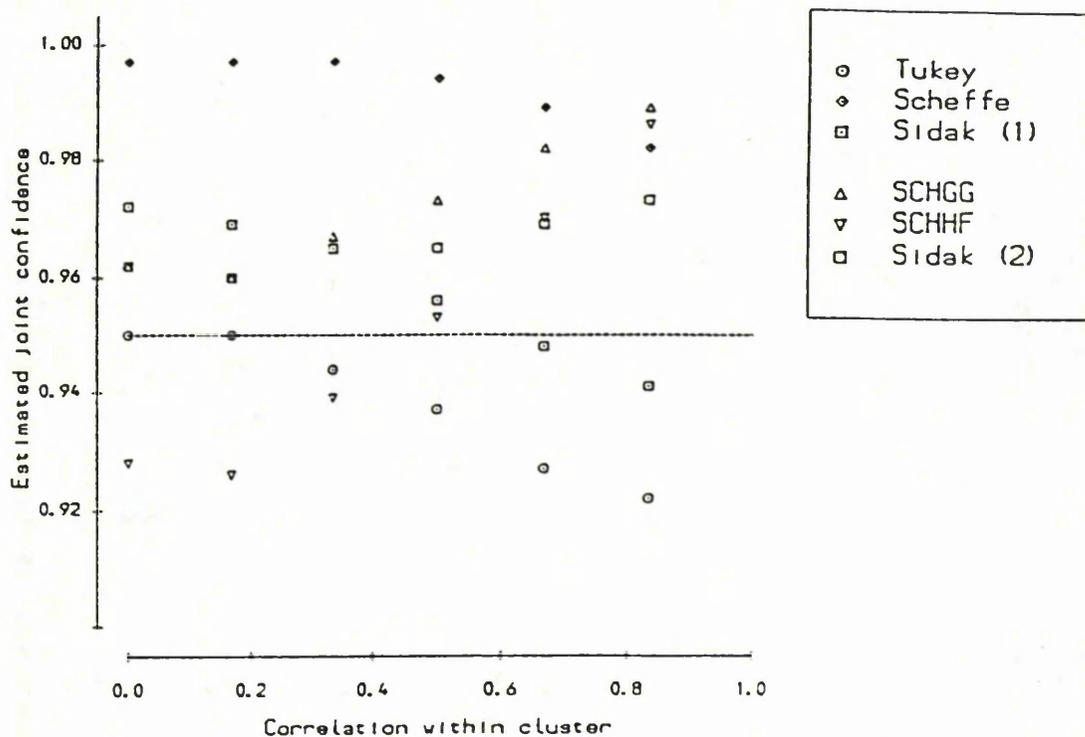
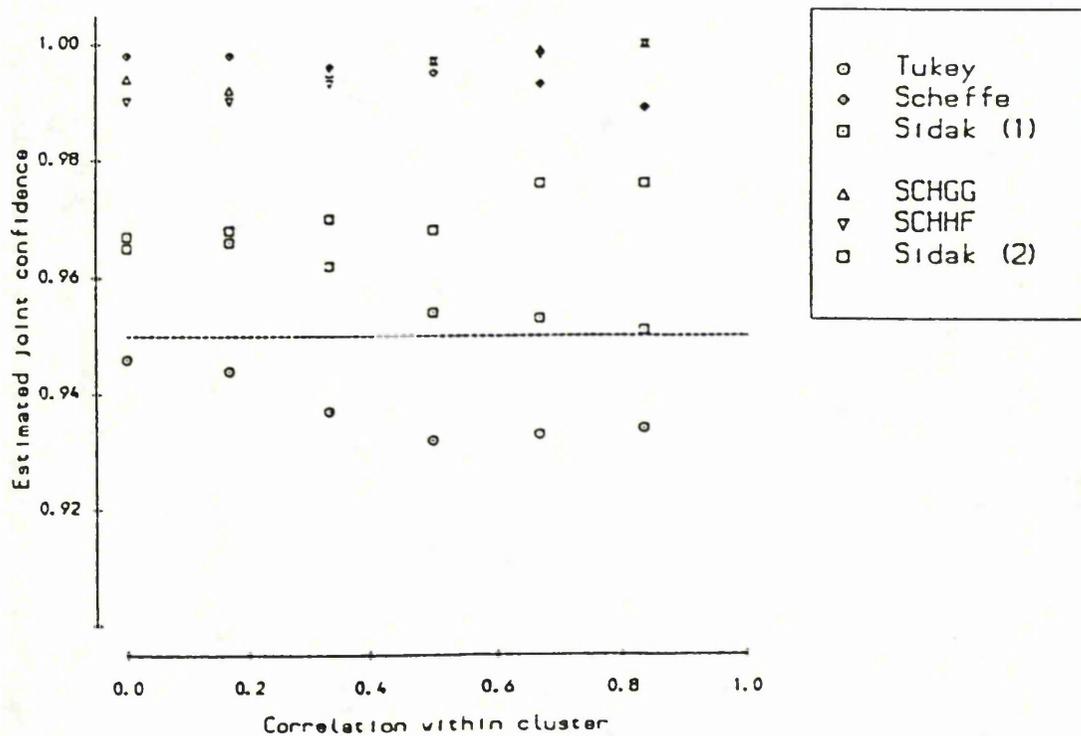


Figure 3.8(vi) Estimated joint confidence levels for the two cluster correlation matrix with  $J=10$  and  $n=30$ .



in Appendix B.8.

Several conclusions can be drawn from these results. Firstly, the SCHEFFE, HFADJ and GGADJ procedures are always extremely conservative, if only pairwise comparisons are of interest and the number of ROIs is large (Figures 3.8(i)-(iv)). A slight decrease in the estimated joint confidence level of the SCHEFFE procedure was observed with 10 ROIs and a high correlation between regions in the same cluster (Figures 3.8(v) and 3.8(vi)). Secondly, the use of Tukey's procedure will not guarantee an overall joint confidence level of at least 0.95, when the covariance structure departs from the sphericity condition. The Bonferroni and Sidak procedures, assuming a constant correlation model, are at their most conservative when  $\rho=0$  (when the sphericity assumption is valid) and become less conservative as  $\rho$  increases. However, Figures 3.8(iii) and 3.8(v) indicate that for the two cluster correlation structure with a sample size of ten (10 and 20 ROI's), the estimated simultaneous confidence level will dip below the stated confidence level of 0.95.

By comparison, as must be the case, the Bonferroni and Sidak procedures, using a separate term for each comparison, remain conservative throughout the range of values for the correlation between ROIs, in all paradigms studied. In contrast to the associated procedures assuming a constant correlation structure, the BONF (2) and SIDAK (2) procedures are at their most conservative when there is a strong interrelation between regions in the same cluster, although in general, the estimated simultaneous confidence level remains more stable across the

range of conditions investigated.

Of the two remaining procedures only the SCHGG procedure maintained an actual simultaneous confidence level greater than or equal to the stated confidence level. Although this procedure is comparable with the BONF (2) and SIDAK (2) procedures when  $n=10$ ,  $J=10$  and for low interregional correlation, the conservativeness becomes greater as  $n$  increases,  $J$  increases and the number of clusters decreases.

The SCHHF procedure, although less conservative than the SCHGG procedure fails to maintain the stated confidence level throughout all combinations of the factors studied and therefore may be unacceptable for use in conducting post hoc pairwise comparisons.

The critical values for which a joint confidence level of 0.95 would be attained, estimated from the set of 10000 simulated values of  $t_1^*$  and  $t_2^*$ , are given in Tables 3.5(i) and (ii) respectively.

Table 3.5(i) indicates that the critical values, obtained from the set of maximum  $t$  statistics utilising the sphericity assumption, increases as the sample size decreases and the number of pairwise comparisons increases. In most instances, the critical values also increase as the within cluster correlation increases. However, for  $n=30$  and  $J=20$ , Table 3.5(i) shows that the critical value initially increases as the covariance structure departs more from sphericity and then decreases in the more extreme cases where there is a high within cluster

Table 3.5(i) Critical values for which a joint confidence level of 0.95 would be obtained using  $t_1^*$ .

n	J	c	within cluster correlation					
			0	1/6	1/3	1/2	2/3	5/6
10	20	2	3.58	3.61	3.68	3.73	3.83	3.87
30	20	2	3.53	3.58	3.60	3.67	3.68	3.61
10	10	2	3.25	3.26	3.31	3.38	3.47	3.56
30	10	2	3.20	3.19	3.25	3.31	3.32	3.33
10	20	4	3.58	3.61	3.63	3.66	3.72	3.62
30	20	4	3.53	3.57	3.59	3.60	3.59	3.52

Table 3.5(ii) Critical values for which a joint confidence level of 0.95 would be obtained using  $t_2^*$ .

n	J	c	within cluster correlation					
			0	1/6	1/3	1/2	2/3	5/6
10	20	2	5.61	5.69	5.58	5.54	5.50	5.30
30	20	2	4.03	4.06	4.00	3.99	3.94	3.88
10	10	2	4.59	4.64	4.53	4.55	4.44	4.37
30	10	2	3.52	3.51	3.49	3.51	3.41	3.40
10	20	4	5.61	5.69	5.61	5.62	5.52	5.19
30	20	4	4.03	4.04	4.07	4.00	3.93	3.82

correlation. Indeed, within the four cluster variance-covariance structure, the critical value when  $\rho=5/6$  is 3.52 compared with the critical value of 3.53 for the independence covariance structure. Note that in the independence case, the Tukey procedure with a corresponding value of 3.55 will give an exact coverage probability of 0.95. Furthermore, for high values of within cluster correlation, the number of clusters present in the variance-covariance matrix would also appear to influence the critical value.

In contrast, given the dimensionality of the measurement vector, the critical values of the  $t_2^*$  statistics in Table 3.5(ii) would not appear, on the basis of this limited study, to be as dependent upon the number of clusters present in the variance-covariance structure. Furthermore, Table 3.5(ii) indicates that the critical value required to achieve a joint confidence level of 0.95 decreases as the within cluster correlation increases. Once again, as expected, the critical value decreases as  $n$  increases and the number of pairwise comparisons decreases. It is also worth noting that the estimated critical values, from the approaches which assume sphericity, are smaller in magnitude than the estimated critical values based on separate error terms. This is completely in accordance with our intuition since the variance in the former method has been estimated with greater degrees of freedom.

The average width of the pairwise intervals for all ten approaches are given in Tables 3.6(i), (ii), for the two cluster correlation structure based on 20 and 10 ROI's, with sample sizes of 10 and 30. For methods based on a separate error term for

Table 3.6(i) Average width of pairwise intervals for the two cluster correlation matrix with 20 ROI's.

n		Method									
		TUKEY	SCHEFFE	BONF1	SIDAK1	GGADJ	HFADJ	BONF2	SIDAK2	SCHGG	SCHHF
10	0	3.93	6.12	4.08	4.07	7.19	-6.17	6.17	6.15	7.00	6.01
								6.17	6.14	7.00	6.01
	1/6	3.77	5.88	3.92	3.90	6.96	5.99	5.63	5.61	6.45	5.55
								6.17	6.15	7.06	6.08
	1/3	3.60	5.61	3.74	3.73	6.83	5.97	5.03	5.02	5.91	5.16
								6.17	6.15	7.27	6.35
	1/2	3.42	5.34	3.56	3.55	6.82	6.12	4.36	4.35	5.37	4.80
							6.18	6.16	7.65	6.87	
	2/3	3.23	5.04	3.36	3.35	6.93	6.40	3.56	3.55	4.70	4.33
								6.18	6.15	8.23	7.61
	5/6	3.01	4.69	3.13	3.12	7.15	6.83	2.52	2.51	3.67	3.50
								6.15	6.13	9.09	8.68
30	0	2.24	3.49	2.32	2.32	3.70	3.50	2.60	2.60	3.67	3.47
								2.60	2.60	3.67	3.47
	1/6	2.15	3.35	2.23	3.23	3.60	3.40	2.38	2.37	3.39	3.21
								2.60	2.60	3.72	3.52
	1/3	2.06	3.20	2.13	2.13	3.57	3.41	2.12	2.12	3.15	3.01
								2.60	2.60	3.86	3.69
	1/2	1.96	3.05	2.03	2.03	3.62	3.50	1.82	1.82	2.90	2.80
							2.60	2.60	4.12	3.98	
	2/3	1.85	2.88	1.92	1.92	3.72	3.64	1.50	1.50	2.58	2.52
								2.60	2.60	4.48	4.38
	5/6	1.74	2.71	1.80	1.80	3.88	3.84	1.06	1.06	2.02	2.00
								2.60	2.60	4.97	4.91

Note: the upper and lower entries in the table for the BONF2, SIDAK2, SCHGG and SCHHF methods correspond to average within and between cluster width respectively.

Table 3.6(ii) Average width of pairwise intervals for the two cluster correlation matrix with 10 ROI's.

n	q	Method										
		TUKEY	SCHEFFE	BONF1	SIDAK1	GGADJ	HFADJ	BONF2	SIDAK2	SCHGG	SCHHF	
10	0	3.56	4.63	3.75	3.75	5.22	3.75	5.16	5.14	5.09	4.56	
								5.16	5.14	5.09	4.56	
	1/6	3.43	4.46	3.62	3.62	5.06	4.54	4.72	4.70	4.68	4.20	
								5.17	5.15	5.14	4.60	
	1/3	3.28	4.27	3.46	3.46	4.96	4.47	4.22	3.64	4.27	3.85	
								5.17	5.15	5.25	4.74	
	1/2	3.13	4.07	3.31	3.31	4.93	4.50	3.65	3.64	3.84	3.50	
								5.17	5.15	5.48	5.01	
	2/3	2.97	3.86	3.13	3.13	4.97	4.63	2.98	2.97	3.33	3.10	
								5.17	5.15	5.82	5.44	
	5/6	2.79	3.62	2.94	2.94	5.11	4.90	2.12	2.10	2.57	2.46	
								5.16	5.14	6.37	6.11	
	30	0	2.00	2.62	2.12	2.12	2.73	2.63	2.32	2.31	2.71	2.61
									2.32	2.31	2.71	2.61
1/6		1.92	2.52	2.04	2.04	2.65	2.55	2.12	2.11	2.49	2.40	
								2.32	2.32	2.73	2.63	
1/3		1.84	2.42	1.96	1.95	2.61	2.52	1.89	1.89	2.29	2.21	
								2.32	2.31	2.80	2.71	
1/2		1.76	2.31	1.87	1.87	2.62	2.55	1.64	1.64	2.08	2.03	
								2.32	2.32	2.96	2.88	
2/3		1.67	2.19	1.77	1.77	2.67	2.62	1.34	1.34	1.82	1.79	
								2.31	2.31	3.16	3.10	
5/6		1.58	2.19	1.68	1.67	2.76	2.73	0.95	0.94	1.41	1.40	
								2.32	2.32	3.47	3.43	

Note: the upper and lower entries in the table for the BONF2, SIDAK2, SCHGG and SCHHF methods correspond to average within and between cluster widths respectively.

each contrast, we have tabulated the average width of both the within and between clusters comparisons. For example, with 20 ROIs, in each of the 10000 simulations, the average width of a between cluster interval is estimated utilising all 100 between cluster comparisons. In the case of the independence covariance structure, the average of all 190 comparisons are tabulated. For methods 1-6, in each simulation the average width of all 190 comparisons is calculated for all covariance structures.

As would be expected, Tables 3.6(i),(ii), indicate that for methods 1-4 the average width of the pairwise comparisons decreases as the correlation within each cluster increases and the sample size increases. Note, however, that although the average width decreases as the correlation within each cluster increases, this is no guarantee that the critical value required to obtain the stated confidence level will necessarily increase - we need only reconsider the case  $n=30$  and  $J=20$  discussed earlier in this section. For the GGADJ and HFADJ methods, as the within cluster correlation increases from 0 to  $5/6$  the average width of the pairwise comparisons first decrease then increases. Note that as  $\rho$  increases the variance of the within cluster contrasts decreases, while the critical value increases. The trade off between these two factors may explain, at least in part, the results of these procedures. The results, for the BONF (2) and SIDAK (2) methods, in Tables 3.6(i),(ii), indicate that the average width of within cluster comparisons decreases as the within cluster correlation decreases, while the average width of the between cluster comparisons remains constant. Both widths decrease as  $n$  increases and the number of ROIs decrease.

Although the average width of the within cluster comparisons decreases for the SCHGG and SCHHF procedures, the average width of the between cluster intervals increases as the correlation within clusters increases. The later result is expected since the critical values of the SCHGG and SCHHF methods both take account of the departure from sphericity. Thus as  $\epsilon$  decreases, the critical value will increase and hence we observe an increase in the width of between cluster intervals. However, although the critical values of these methods increase as  $\rho$  increases, the variance of the within cluster contrasts will decrease. The results in Tables 3.6(i) and 3.6(ii) suggest that the decrease in variance is having a larger effect than the increase in the critical value.

Finally, in this section we have investigated the frequency of occurrence of a significant  $t_1^*$  and  $t_2^*$  from a comparison based on 2 ROI's not within the same cluster. Again, we have only considered the two cluster correlation structure. Table 3.7(i),(ii) give the percentage from the 10000 simulations for 20 and 10 ROI's and sample sizes of 10 and 30. Note that if each of the  $J(J-1)/2$  pairwise comparisons has an equal chance of attaining the maximum  $t$  statistic we would expect approximately 53% of the significant  $t_1^*$  and  $t_2^*$  values coming from between cluster comparisons when  $J=20$  and approximately 56% when  $J=10$ .

For methods 1-6, Tables 3.7(i),(ii) indicate that as the correlation between ROI's in the same cluster increases the proportion of significant between clusters comparisons increases.

Table 3.7(i) Proportion of significant  $t_1^*$  and  $t_2^*$  statistics from between cluster comparisons (20 ROI's).

n	q	Method									
		TUKEY	SCHEFFE	BONF1	SIDAK1	GGADJ	HFADJ	BONF2	SIDAK2	SCHGG	SCHHF
10	0	49.7	*	48.2	48.1	*	*	52.9	52.5	54.5	51.9
	1/6	80.7	*	78.9	78.7	*	*	49.2	49.2	50.3	51.6
	1/3	91.4	100.0	94.5	94.4	*	100.0	50.5	50.5	63.0	55.0
	1/2	97.4	100.0	97.8	97.8	*	100.0	42.8	43.3	54.7	54.1
	2/3	99.9	100.0	100.0	100.0	*	100.0	37.2	37.1	59.4	59.9
	5/6	100.0	100.0	100.0	100.0	*	100.0	26.3	25.6	53.8	52.3
30	0	50.2	*	47.7	47.7	*	*	51.7	51.4	66.7	55.6
	1/6	80.7	*	80.4	80.4	*	*	51.8	51.5	100.0	66.7
	1/3	93.7	100.0	94.6	94.6	*	100.0	48.0	48.0	80.0	75.0
	1/2	98.2	*	98.8	98.8	*	*	39.4	39.6	0.0	0.0
	2/3	100.0	100.0	100.0	100.0	*	*	28.8	29.0	*	*
	5/6	100.0	100.0	100.0	100.0	*	*	17.2	17.8	*	*

Table 3.7(ii) Proportion of significant  $t_1^*$  and  $t_2^*$  statistics from between cluster comparisons (10 ROI's).

n	q	Method									
		TUKEY	SCHEFFE	BONF1	SIDAK1	GGADJ	HFADJ	BONF2	SIDAK2	SCHGG	SCHHF
10	0	54.9	51.9	61.9	61.9	75.0	40.0	54.2	54.1	54.5	55.2
	1/6	75.0	84.4	73.5	73.5	71.4	84.6	54.5	54.4	54.6	52.7
	1/3	90.9	100.0	92.2	92.2	100.0	100.0	48.8	48.4	52.0	53.6
	1/2	97.3	100.0	97.5	97.5	100.0	100.0	51.0	51.1	56.8	56.1
	2/3	99.5	100.0	99.6	99.6	100.0	100.0	39.4	40.1	49.7	52.8
	5/6	99.9	100.0	99.8	99.8	100.0	100.0	32.6	33.2	50.0	50.7
30	0	58.6	52.6	58.5	58.3	54.5	50.0	58.0	58.0	55.6	57.4
	1/6	75.5	87.5	76.4	76.0	83.3	85.0	55.8	56.1	61.2	59.0
	1/3	90.0	97.5	93.2	93.3	100.0	96.6	55.9	55.2	55.3	55.6
	1/2	98.1	100.0	98.0	98.0	100.0	100.0	43.8	44.2	42.9	48.6
	2/3	99.6	100.0	99.6	99.6	100.0	100.0	34.7	34.8	36.4	63.2
	5/6	100.0	100.0	100.0	100.0	100.0	100.0	27.0	27.1	100.0	100.0

As expected, as the correlation between ROI's in the same cluster increases the proportion of significant between cluster comparisons using the BONF(2) and SIDAK(2) methods decreases. However the proportion of significant between cluster comparisons using the SCHGG and SCHHF methods, which are also based on a separate estimate of the variance for each comparison, remains fairly constant over the range of within cluster correlations investigated. As mentioned previously, the SCHEFFE, HFADJ and GGADJ procedures are very conservative. In instances where there were no intervals identified which contained zero, a \* has been placed in the table. It is also worth noting that for methods which are particularly conservative, the proportions in the tables will not be accurately determined.

### 3.9 CONCLUSIONS FROM THE SIMULATION STUDIES

Empirical results of the performance of four test procedures - the traditional F-test, the Huynh-Feldt and Greenhouse-Geisser adjusted tests and a conservative test based on the lower bound of Box's correction factor,  $\epsilon$ , to the degrees of freedom - were obtained in a repeated measures framework to investigate the effect of the departure from sphericity on the Type I error rate, in hypothetical situations representing the blood flow metabolism coupling problem. In general, the results of these investigations suggest that the use of the Huynh-Feldt adjusted test is more robust than the Greenhouse-Geisser adjusted test when the sample size is small and the number of ROIs is large. The Huynh-Feldt correction factor also reflected the nominal alpha level with less bias in all conditions for the four and ten cluster correlation pattern or in situations where the within

cluster correlation was low. With the exception of the independence variance-covariance structure, the Huynh-Feldt correction factor overestimated the true value of  $\epsilon$ , giving somewhat liberal results. In contrast, the Greenhouse-Geisser adjustment factor was unduly conservative in most of the hypothetical situations covered in this investigation. Moreover, the Huynh-Feldt test often has a substantial power advantage over the corresponding Greenhouse-Geisser test.

Although the results from these investigations do not suggest a clear cut rule for choosing between the Huynh-Feldt and Greenhouse-Geisser modified approaches over the whole range of simulated conditions, the Greenhouse-Geisser approach is preferable if a conservative approach is necessary.

In terms of post hoc multiple pairwise comparisons, the BONF(2) and SIDAK(2) approaches perform consistently well, giving joint confidence levels close to the nominal 95% level. In all but a few cases the BONF(1) and SIDAK(1) methods have an actual simultaneous confidence level greater than or equal to the nominal confidence level. Of the other procedures, the SCHEFFE, HFADJ, GGADJ, SCHHF and SCHGG tend to be too conservative, particularly for large sample sizes, whilst the Tukey procedure generally produces intervals with an actual joint confidence level less than 95%.

Often, in practice, multiple comparisons between the ROI's would only be pursued when the F-test for the equality in mean response is significant. However, if the inferences of interest are pursued using the BONF(2) or SIDAK(2) approaches, regardless of

the outcome of the preliminary F-test, the familywise error rate (the probability of making any error in the given family of inferences) will be less than the nominal  $\alpha$  level.

Moreover, in such cases, not only will the test be conservative, but compared with the Huynh-Feldt and Greenhouse-Geisser adjusted tests, the nominal  $\alpha$  level will be achieved more reliably in all but a few extreme cases.

## CHAPTER 4 STATISTICAL ANALYSIS OF IMAGE DIFFERENCES

### 4.0 INTRODUCTION

Recently there has been considerable interest (Fox et al, 1986; Fox et al, 1987; Fox et al, 1988) in the analysis and interpretation of subtracted positron emission tomography (PET) and single photon emission computerised tomography (SPECT) images, particularly in activation studies where the tomographic image of the patient at rest (control state) is subtracted from the image obtained when the patient is performing a specific task. Often these types of analyses will be concerned with detecting subtle differences in metabolic activity or blood flow on an individual subject basis or in a group of subjects. However, when images are subtracted on a pixel by pixel basis there are statistical difficulties in the assessment of any apparent effects observed in the difference image due to the large number of comparisons and the spatial correlation of the pixels.

One proposal, to determine whether any pixels within the difference image are significant distributional outliers, has been based on the kurtosis of the grey values of the image (Fox et al, 1988). Although this procedure assesses whether significant changes in grey level value from control to activation exist, it does not however identify which or how many are significant. Moreover, this approach also assumes that the pixels in the difference image are independent which is probably

invalid. An alternative model of change distribution analysis to identify where the significant changes have occurred would be to assume that the distribution of the grey values or change parameter in the difference image is normal. Areas of 'high change' could then be identified by highlighting pixels in the difference image which were greater than 1.96 standard deviations from the mean difference in the whole image. This would constitute approximately 5% of the total pixels if the data were normal.

Although these proposals have much to recommend them as an exploratory step in image analysis, both assume that image noise is spatially random and free of any artefacts from the measurement process. As a baseline for the assessment of apparent differences seen in activation type studies, we examine, in this pilot study, the spatial patterns which are observed in 'control-minus-control' subtractions in the case of a NOVO SPECT tomograph situated at the Southern General Hospital, Glasgow.

#### 4.1 DATA AND METHODS

To assess whether the noise in a subtracted image is spatially random, we have confined our investigation to a data set consisting of a series of six 128 x 128 pixel reconstructions of an artificial image or phantom. The phantom was designed specifically as a simple imitation of a human brain using perspex sheets. The images are of the same standard slice and were obtained sequentially with no movement between scans. An appealing property of the phantom is that each area in the model

will have constant intensity with fairly large differences in intensities between regions. One of the six reconstructed images of the phantom is shown in Figure 4.1 using 256 grey levels to reflect the emission intensity. Lighter pixels represent areas of high emission intensities while darker areas represent areas of low intensity.

The use of the phantom represents a 'best case' scenario. For assessing noise in a subtracted image, it has several advantages over patient data. Firstly, and perhaps most importantly, for image subtraction, by eliminating the possibility of head movement, pairs of images can be placed in exact spatial register with little difficulty. Even slight head movements of a few millimetres can cause movement artefacts in the difference image. These are seen as rings of apparent functional change at brain and regional boundaries. Secondly, differences in the level of anxiety or stress during the period of scanning could hinder the investigation by producing patterns of functional change in brain regions associated with the respective neuroanatomical pathways.

These and other factors, such as head repositioning and time of day, which affect the reproductability of functional activity in resting human patients have been investigated by Bartlett et al (1988).

To reduce the variability in global metabolic activity (if any) in the six images due to any slight decay of the radioactive substance in the phantom over the scanning period or any reconstruction artefact, all images were normalised by dividing the grey level in each pixel by the average grey level in that

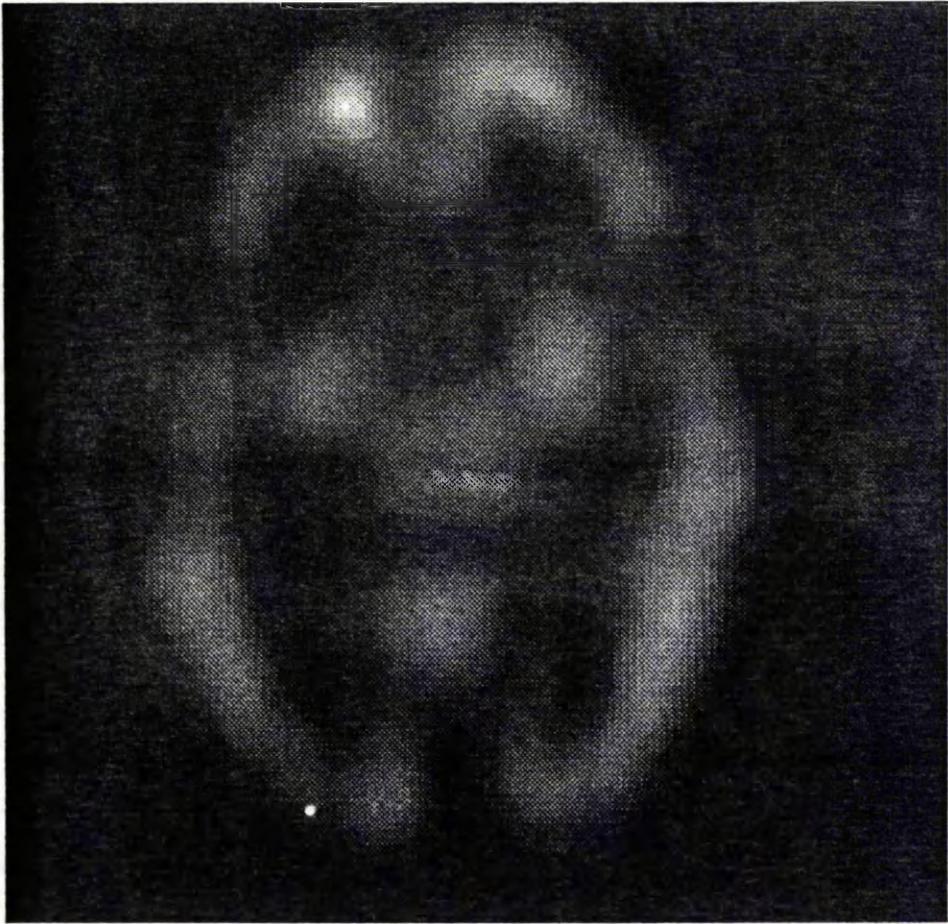


image. Figures 4.2a and 4.2b show the average normalised image and the variance in grey levels across the series of normalised scans respectively. Figure 4.2b illustrates that a number of clusters of pixels vary over the series of scans to a greater extent than other pixels.

To determine whether the noise, as measured by the subtraction of one image from another, is spatially random we have limited our investigation to the central 64 x 64 pixel submatrix in order to avoid problems with the shape of the phantom. The corresponding 64 x 64 submatrices of Figures 4.2a and 4.2b are given in Figures 4.3a and 4.3b respectively. These Figures show that the variance across the series of images is not related to the mean activity (correlation = -0.033).

In this pilot study, we have also investigated the noise after reducing the resolution of the image by one half. The reduction in resolution to a 64 x 64 pixel image was achieved by averaging the grey levels in four pixels of the 128 x 128 pixel image. The average normalised 64 x 64 image and variance of the 64 x 64 pixel images are given in Figures 4.4a and 4.4b. The corresponding central 32 x 32 submatrices for these figures are given in Figures 4.5a and 4.5b. For the 32 x 32 pixel images no association between the variance and mean value of the normalised images was detected (correlation = -0.036).

In order to investigate whether the difference images using the phantom data are totally random with no spatial structure, three images were obtained by subtracting the second image of the

Figure 4.2a Average normalised image



Figure 4.2b Variance in normalised images

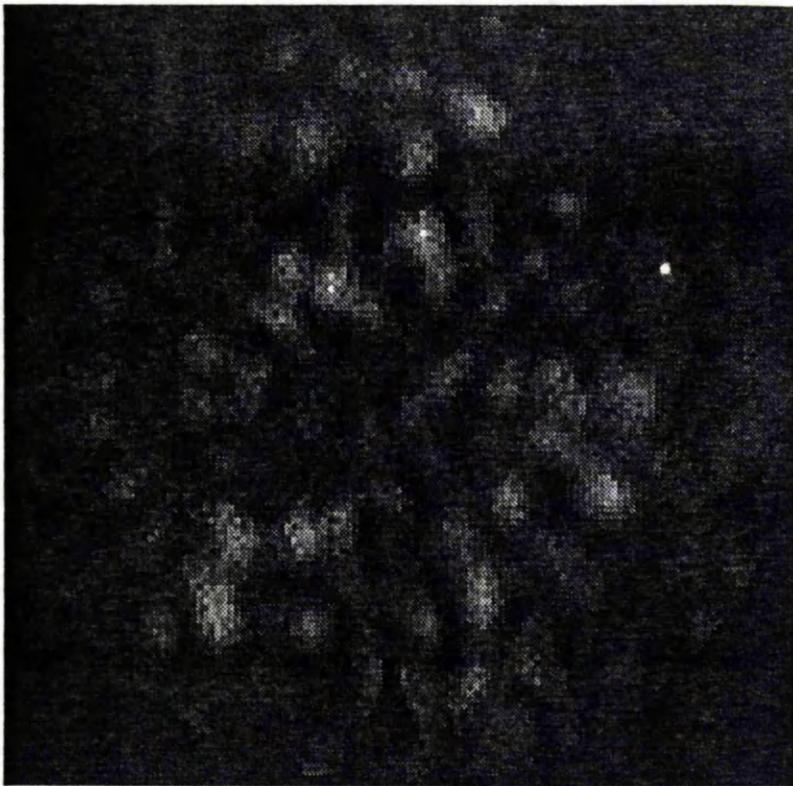


Figure 4.3a 64 x 64 central submatrix of figure 4.2a

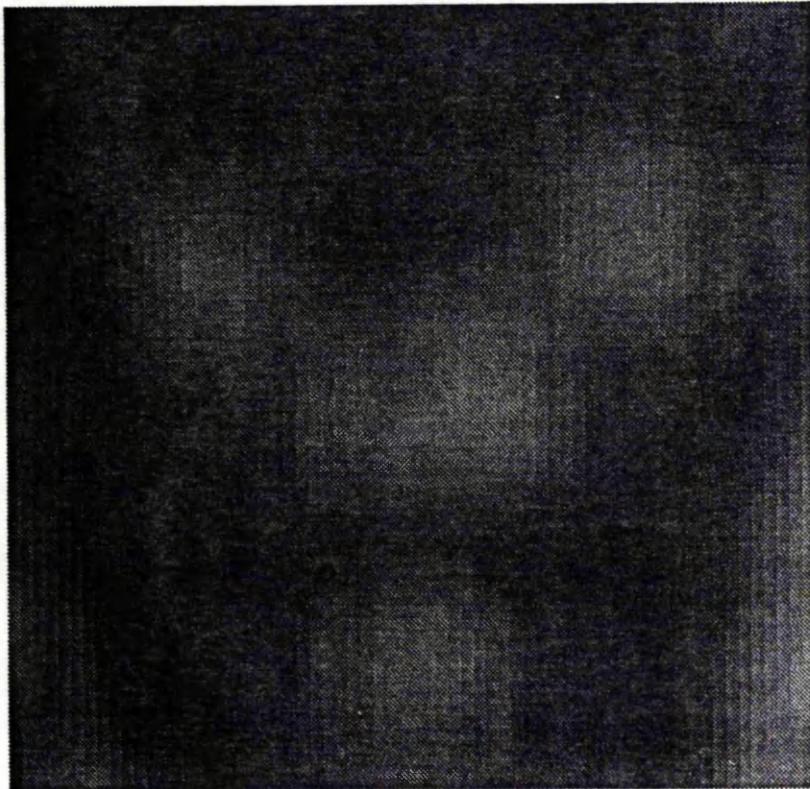


Figure 4.3b 64 x 64 central submatrix of figure 4.2b

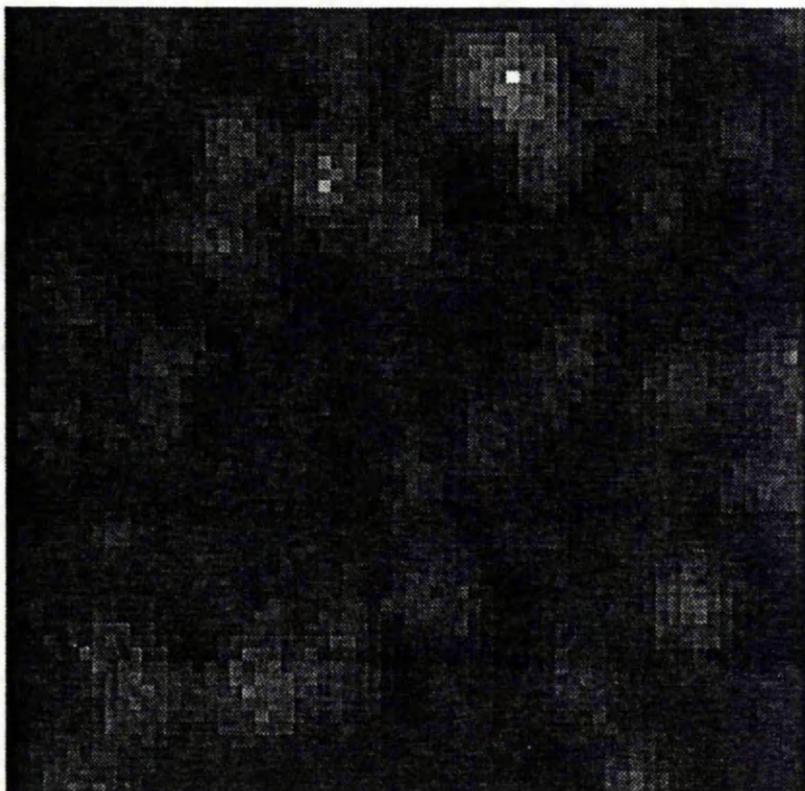


Figure 4.4a Average normalised image after a reduction in resolution

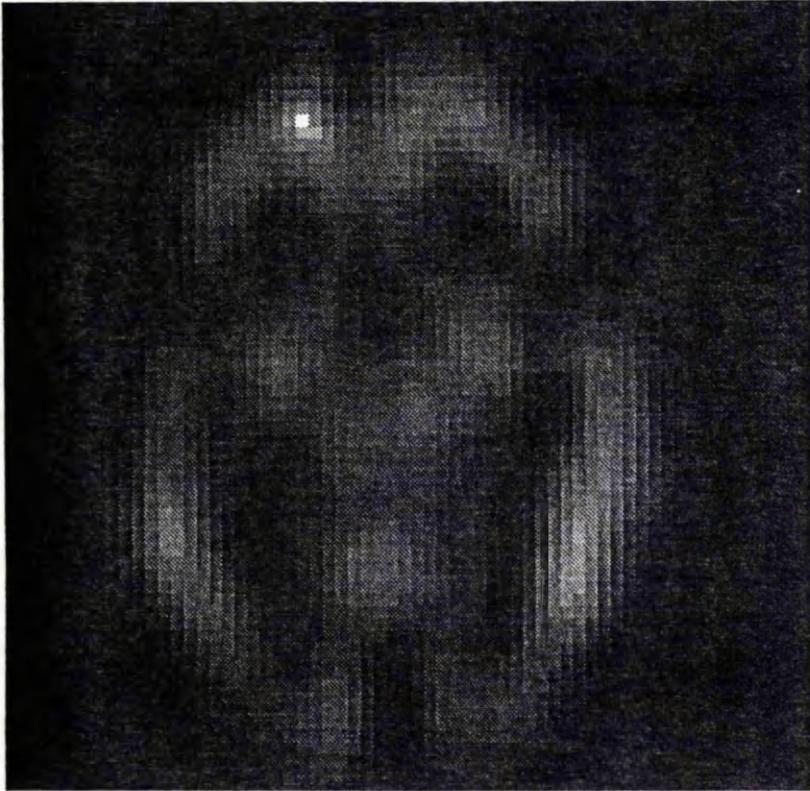


Figure 4.4b Variance in normalised images after a reduction in resolution

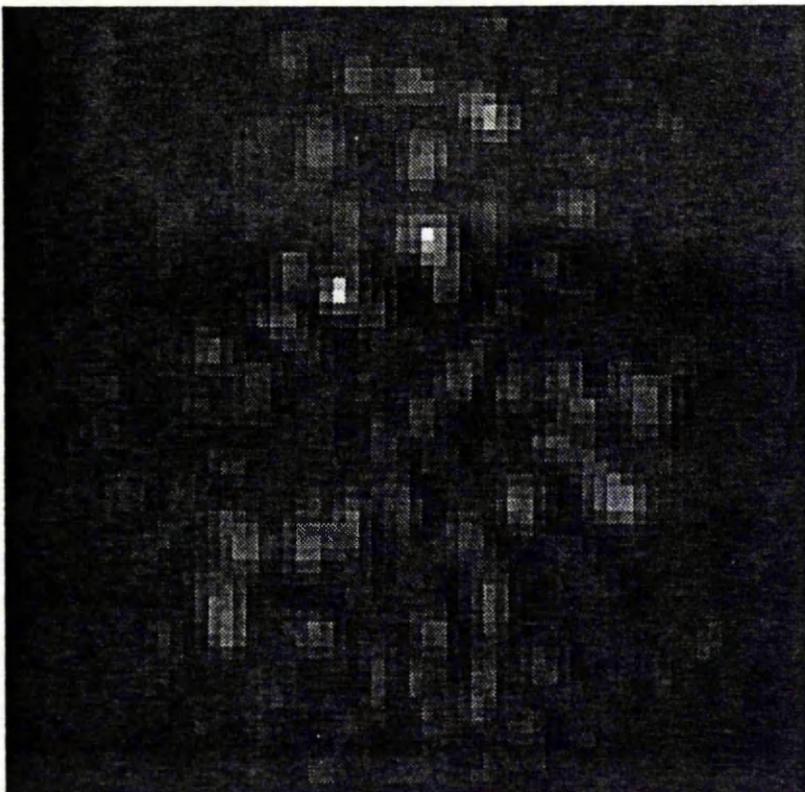


Figure 4.5a 32 x 32 central submatrix of Figure 4.4a

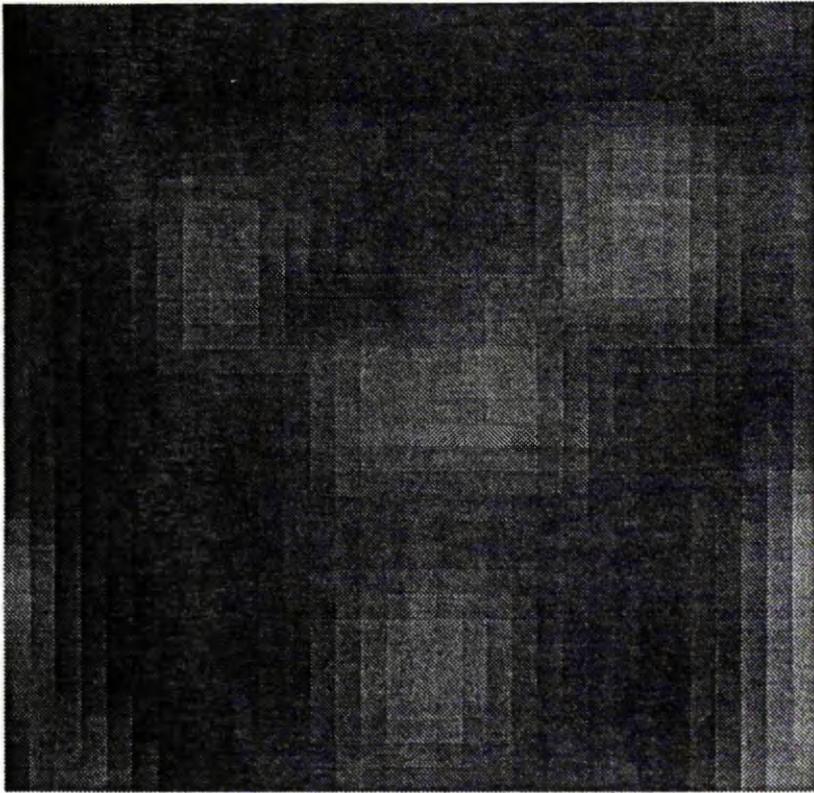
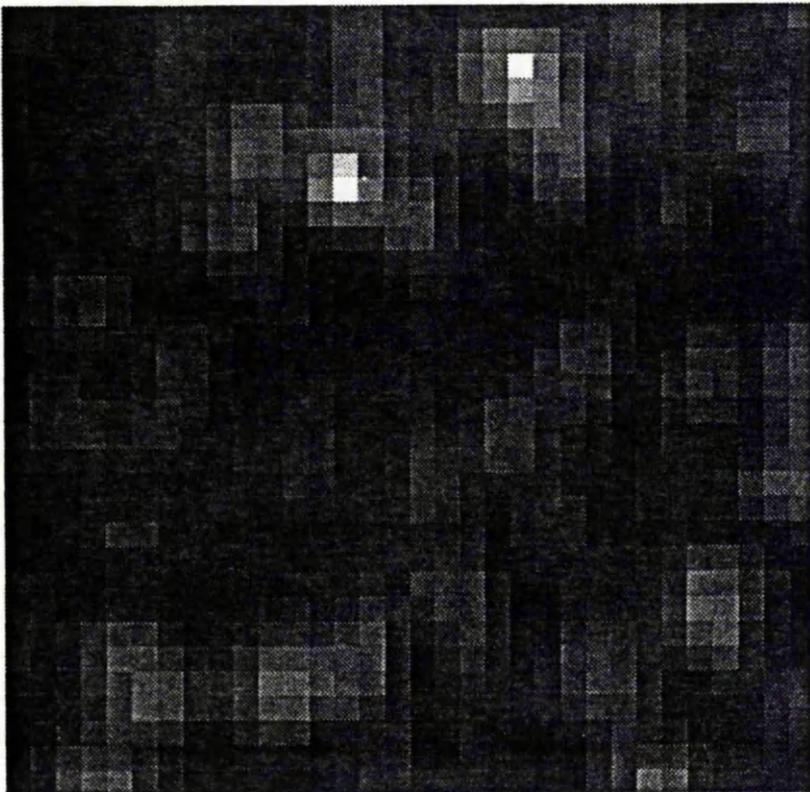


Figure 4.5b 32 x 32 central submatrix of Figure 4.4b



series of six from the first image, the fourth from the third and the sixth from the fifth. For the purpose of this investigation, areas in the 64 x 64 pixel and 32 x 32 pixel submatrices corresponding to high change in metabolic activity were identified by highlighting all pixels with an absolute change in grey value of greater than 1.96 standard deviations from the mean.

## 4.2 RESULTS

The difference image obtained by subtracting the second 128 x 128 pixel image in the series of six from the first is given in Figure 4.6a. The central 64 x 64 pixel submatrix corresponding to this image is given in Figure 4.6b. Lighter pixels in these images represent areas where the first image had higher metabolic activity whilst darker areas represent areas where the second image had higher metabolic activity. Mid-grey levels similar to those in the four corners of Figure 4.6a correspond to pixels in which no change has occurred.

It is apparent from Figures 4.6a and 4.6b that particular areas of the difference image indicate higher change than others. For the purpose of identifying pixels reflecting large differences in metabolic activity, irrespective of sign, it seems more appropriate to consider only absolute change. The corresponding figures to 4.6a and 4.6b showing absolute change are given in Figures 4.7a and 4.7b respectively. The advantage of contrasting no change with high change is immediately obvious: areas of no change (represented by dark grey pixels) in Figures

Figure 4.6a Difference between Image 1 and Image 2

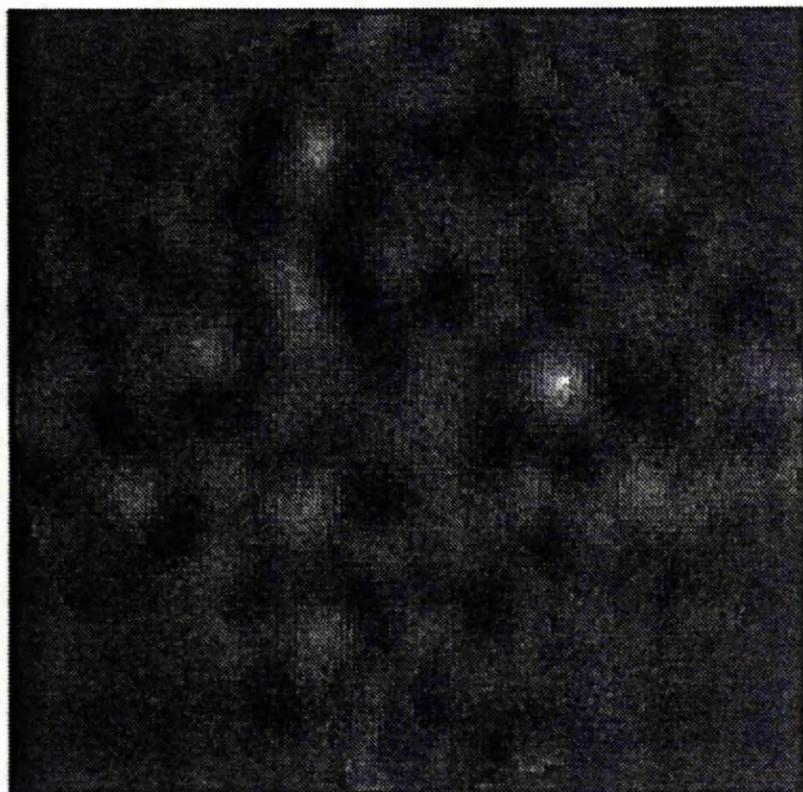


Figure 4.6b 64 x 64 central submatrix of Figure 4.6a

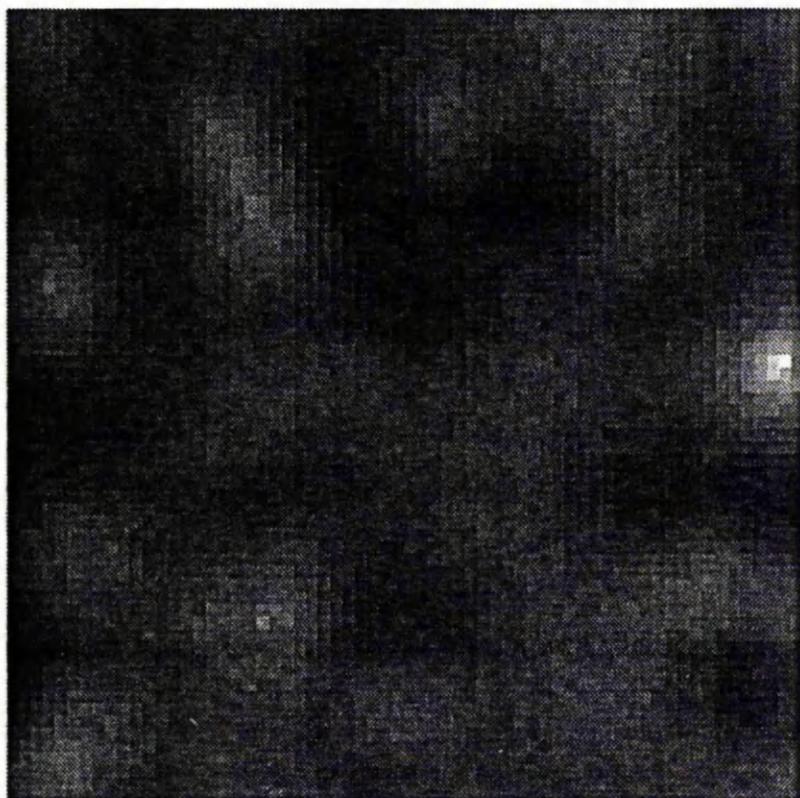


Figure 4.7a Absolute difference between Image 1 and Image 2

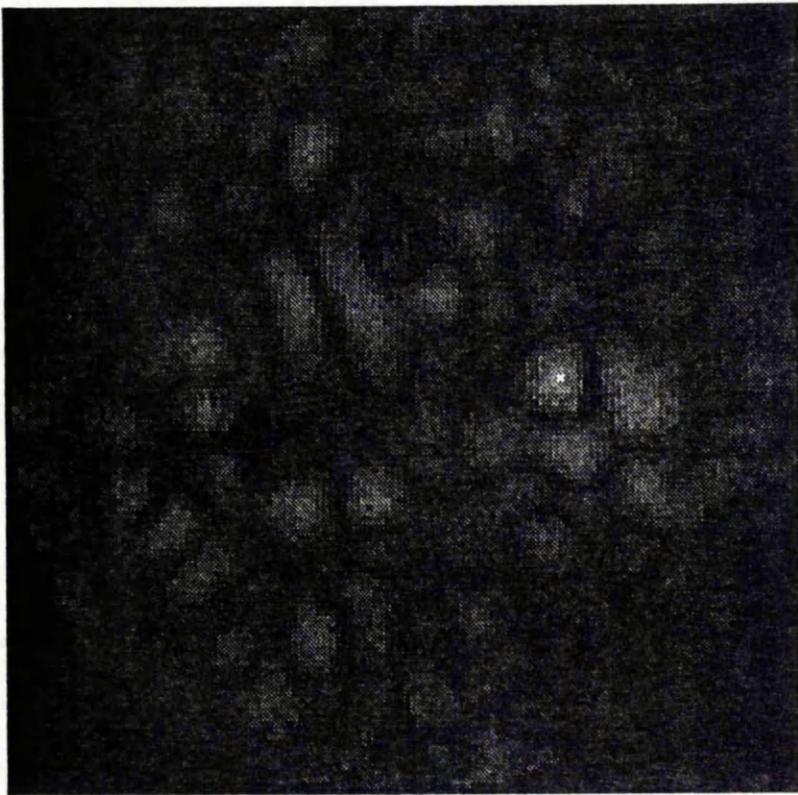
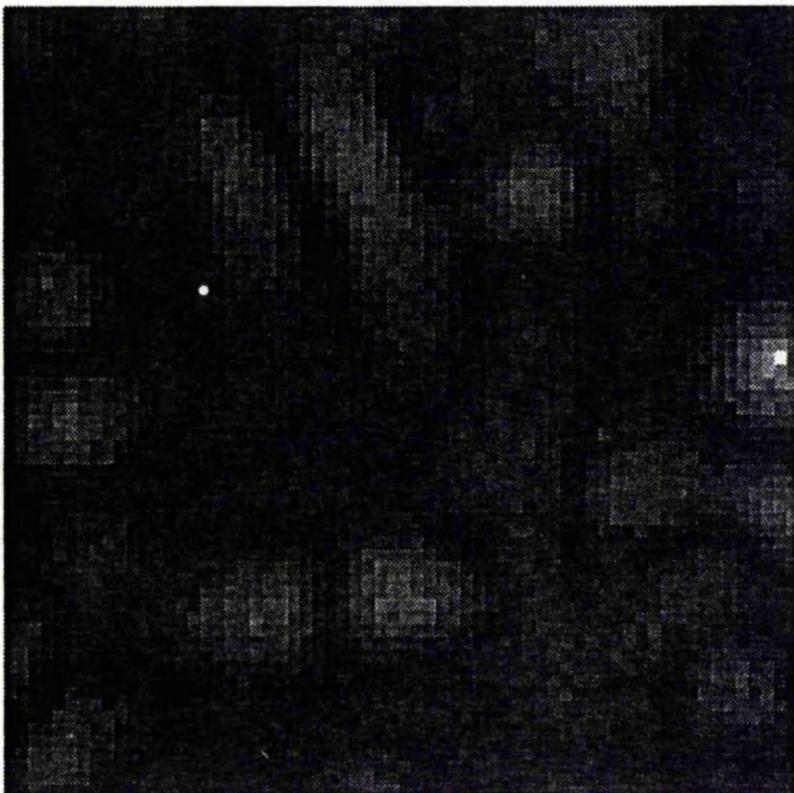


Figure 4.7b 64 x 64 central submatrix of Figure 4.7a



4.7a and 4.7b are more easily identifiable.

Figure 4.8a gives the image corresponding to the absolute difference between the third and fourth reconstructed images of the phantom. The central  $64 \times 64$  submatrix is given in Figure 4.8b. Corresponding figures for the absolute difference between the fifth and sixth images are given in Figures 4.9a and 4.9b respectively.

The structure of the difference images in Figures 4.6a to 4.9b reveal several striking patterns. Firstly, all images are highly fragmented: discrete areas, often roughly similar in shape, are separated by pixels registering no change in metabolic activity. Secondly, pixels corresponding to a high change in metabolic activity group into a number of clusters which are predominantly circular in shape.

Figures 4.10a and 4.10b give the  $64 \times 64$  pixel image and the  $32 \times 32$  pixel central submatrix of the absolute difference in emission intensities of the first and second images, after reducing the resolution of the images. Absolute difference images for the subtraction of the fourth image from the third image (Figure 4.11a) in the series of reconstructions of the phantom and the sixth image from the fifth image (Figure 4.12a) are given with their central  $32 \times 32$  pixel submatrices (Figures 4.11b and 4.12b respectively).

As would be expected, the  $64 \times 64$  pixel difference images are more 'hazy' than their  $128 \times 128$  pixel counterparts. Moreover,

Figure 4.8a Absolute difference between Image 3 and Image 4

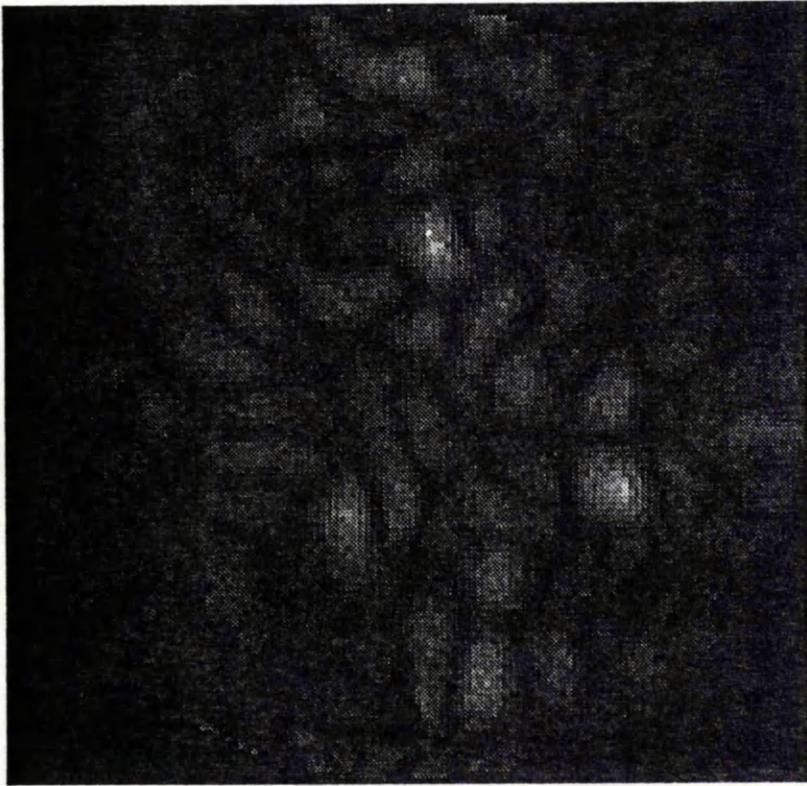


Figure 4.8b 64 x 64 central submatrix of Figure 4.8a

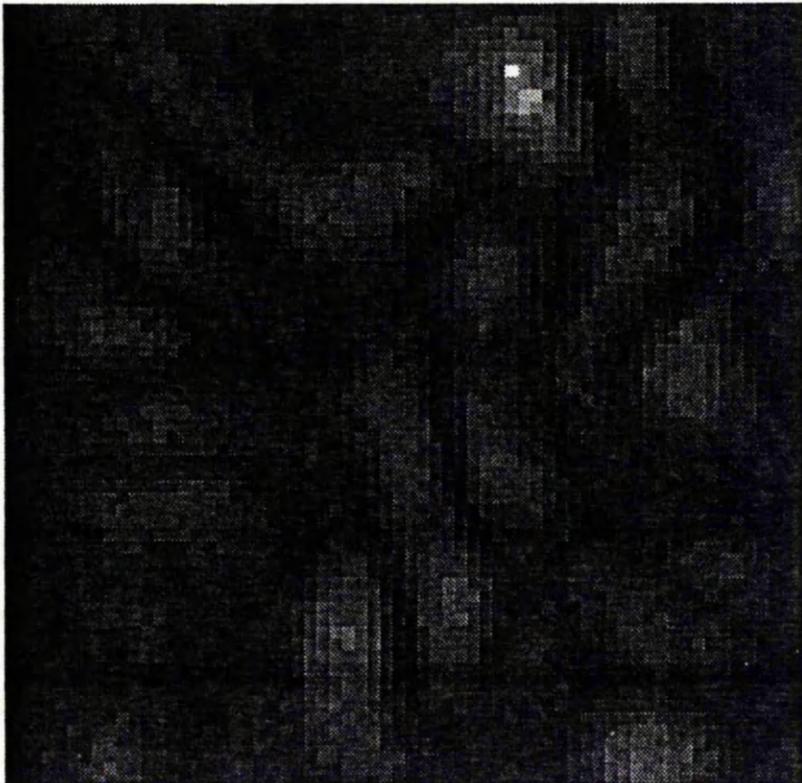


Figure 4.9a Absolute difference between Image 5 and Image 6

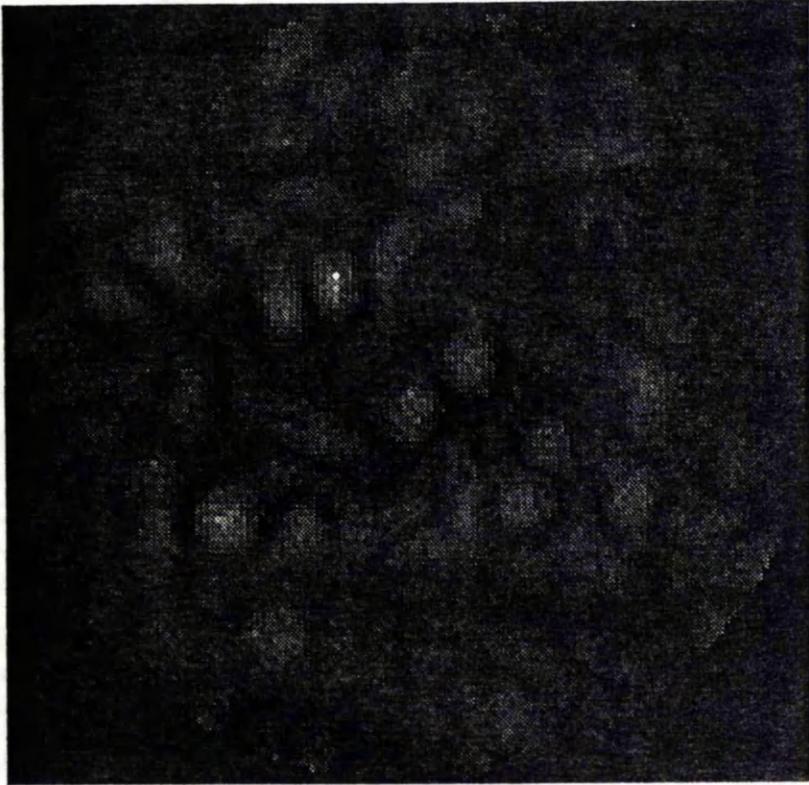


Figure 4.9b 64 x 64 central submatrix of Figure 4.9a

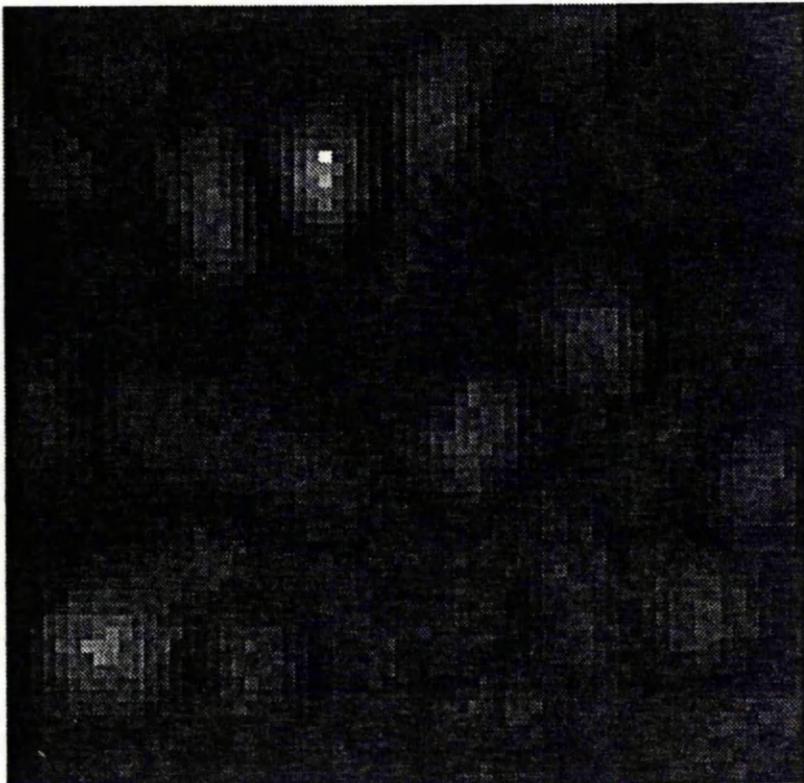


Figure 4.10a Absolute difference between Image 1 and Image 2 after a reduction in resolution

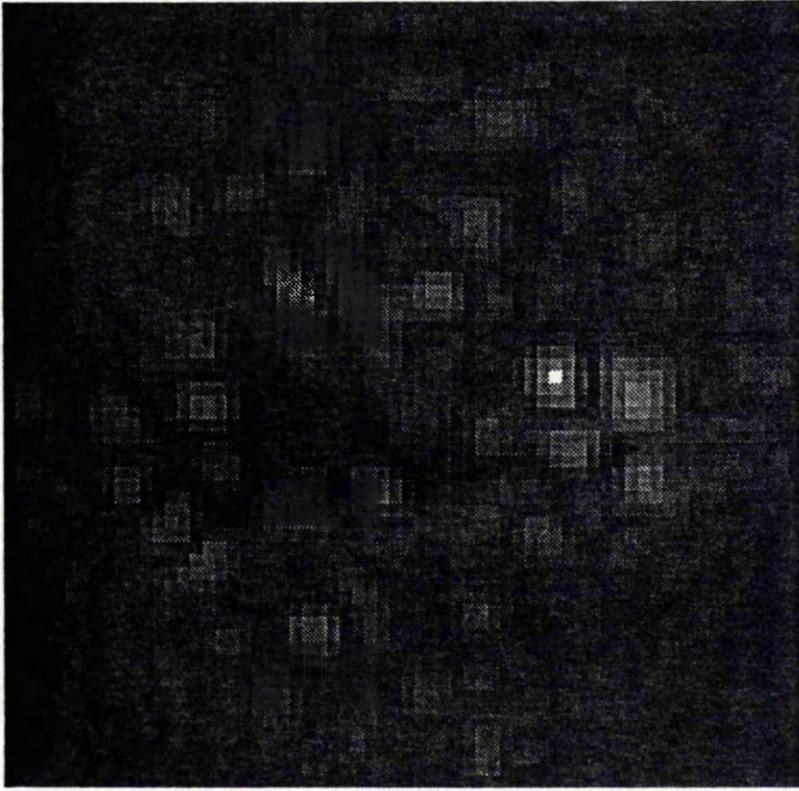


Figure 4.10b 32 x 32 central submatrix of Figure 4.10a

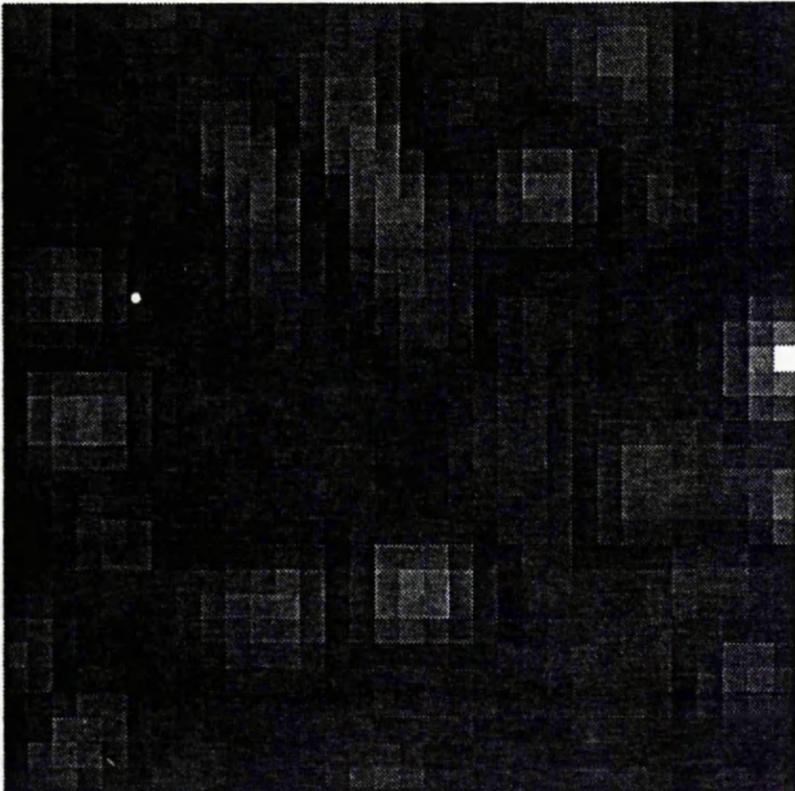


Figure 4.11a Absolute difference between Image 3 and Image 4 after a reduction in resolution

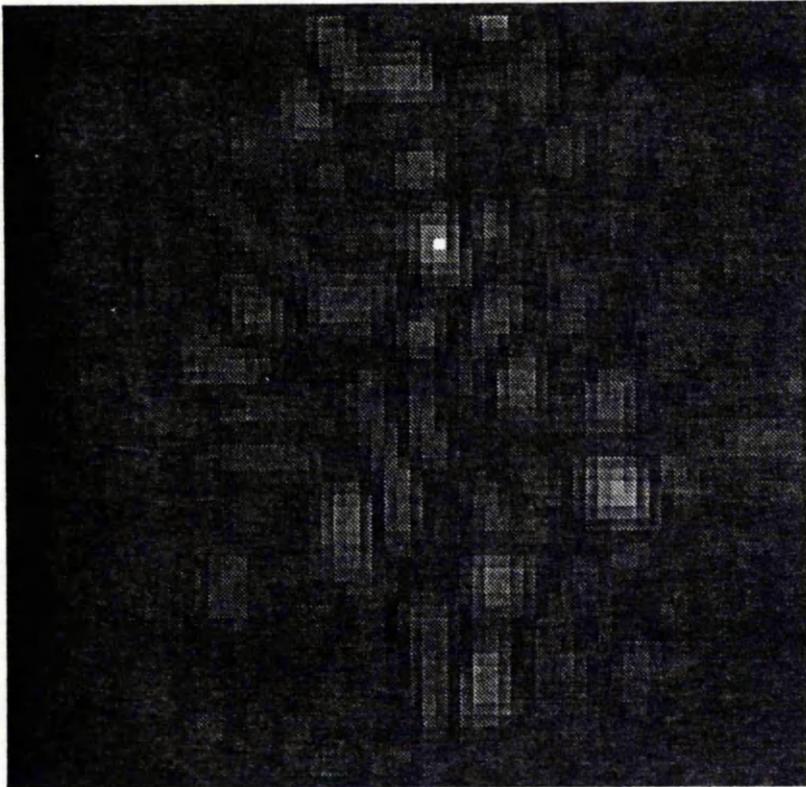


Figure 4.11b 32 x 32 central submatrix of Figure 4.11a

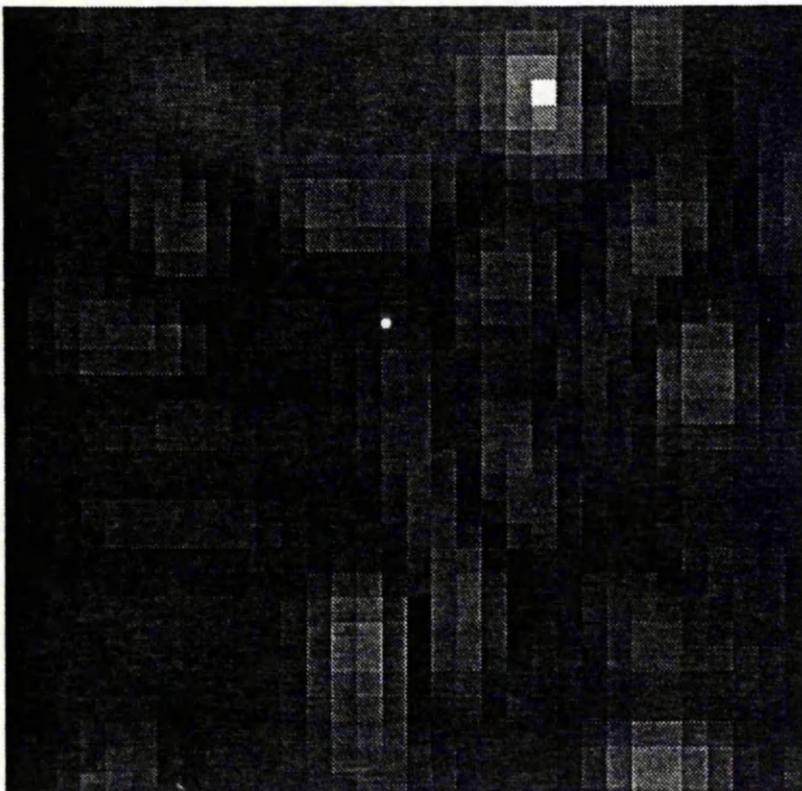


Figure 4.12a Absolute difference between Image 5 and Image 6 after a reduction in resolution

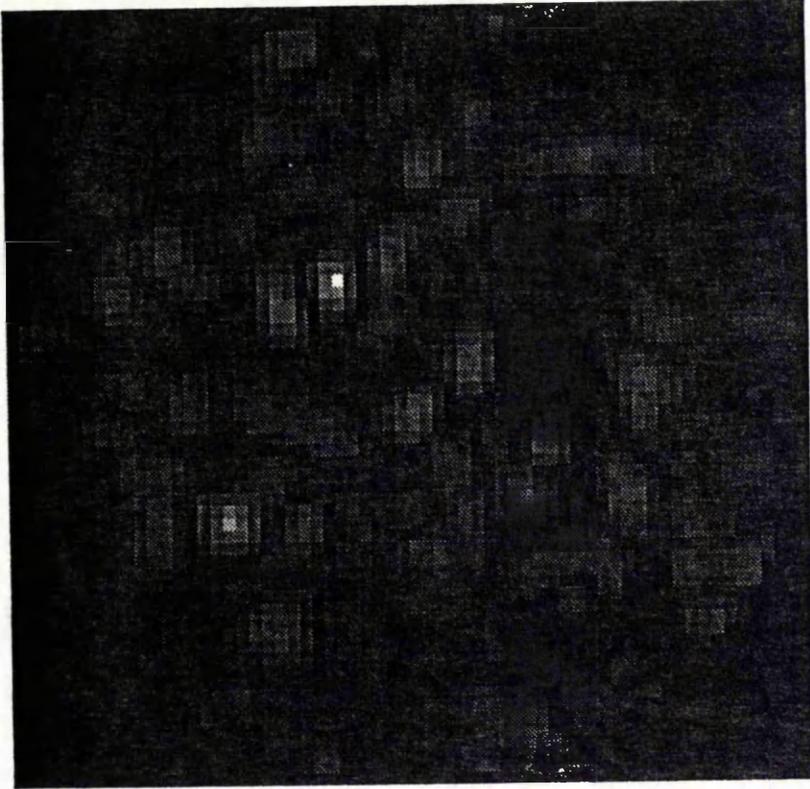
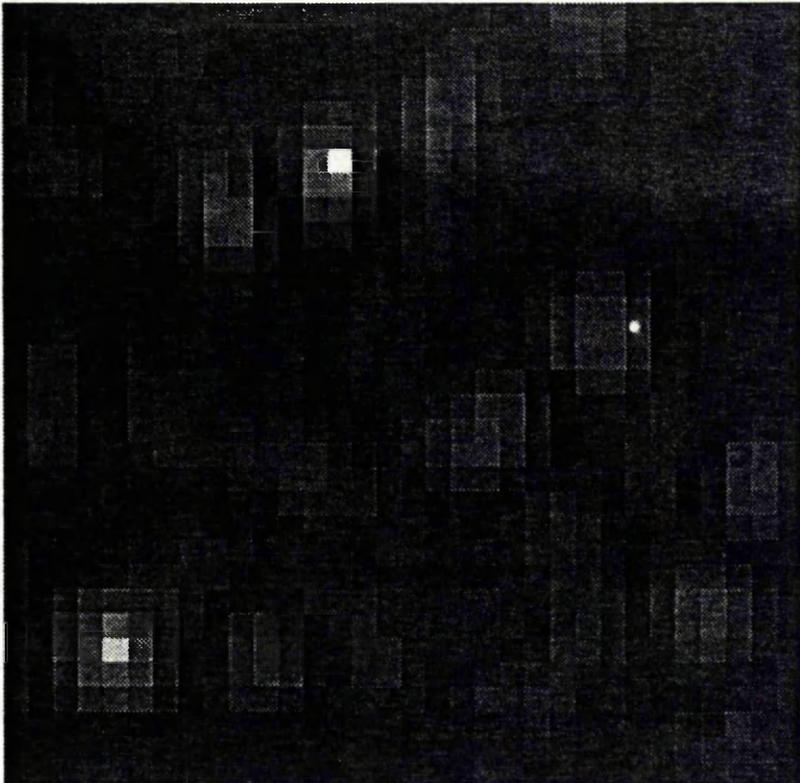


Figure 4.12b 32 x 32 central submatrix of Figure 4.12a



Figures 4.10a-4.12b appear to have less fragmentation. This is particularly noticeable for the central  $32 \times 32$  pixel submatrices. However, areas of high change remain grouped in several clusters.

Using a threshold of 1.96 standard deviations from the mean grey value in each central submatrix, areas of high change have been highlighted. Figure 4.13a displays the pixels representing 'significant' change for the central  $64 \times 64$  submatrix subtraction of the second image from the first. Figure 4.13b gives the corresponding central  $32 \times 32$  submatrix subtraction under poorer resolution. Similar difference images are given in Figures 4.14a,b for the subtraction of the fourth from the third image and Figures 4.15a,b for the subtraction of the sixth from the fifth image.

It can be seen from Figures 4.13a to 4.15b that the pixels highlighted in the image do not occur totally at random with no spatial structure but in general appear in clusters. In the  $64 \times 64$  pixel images, clusters vary in size from 7 to roughly 70 pixels and from 7 to 16 pixels in the  $32 \times 32$  images. However, with the exception of a few extreme cases, most clusters are roughly similar in size. Moreover, the resolution of the image appears to have very little influence on the location of areas of predominantly high change. Although decreasing the resolution tends to average out isolated pixels of high change in the  $128 \times 128$  pixel, as would be expected, clusters of pixels indicating high change remain. Furthermore the number of clusters in all the  $64 \times 64$  pixel and  $32 \times 32$  pixel images

Figure 4.13a 'Significant' change between Image 1 and Image 2

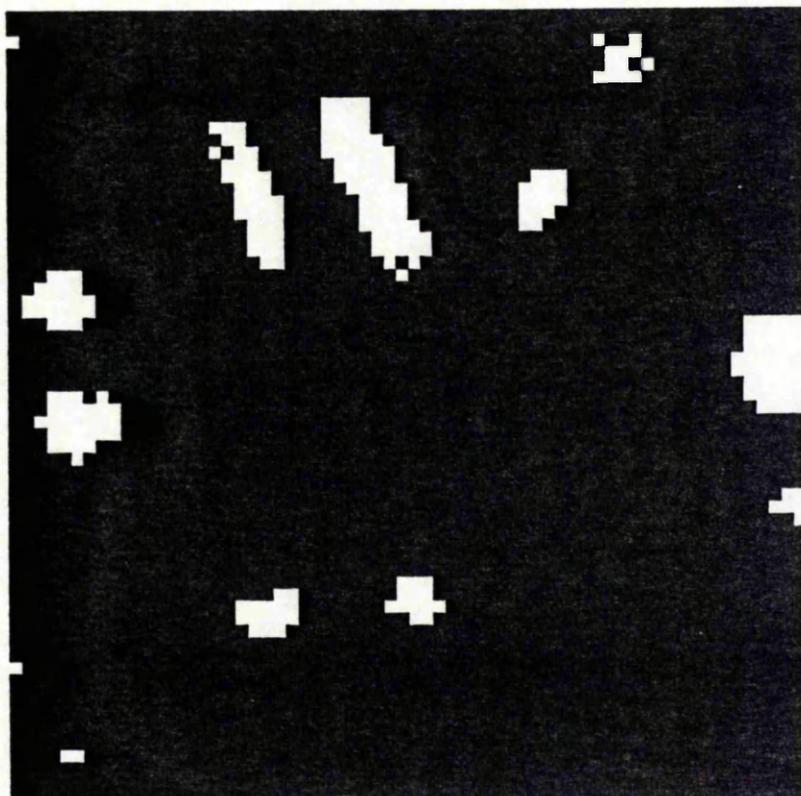


Figure 4.13b 'Significant' change between Image 1 and Image 2 after a reduction in resolution

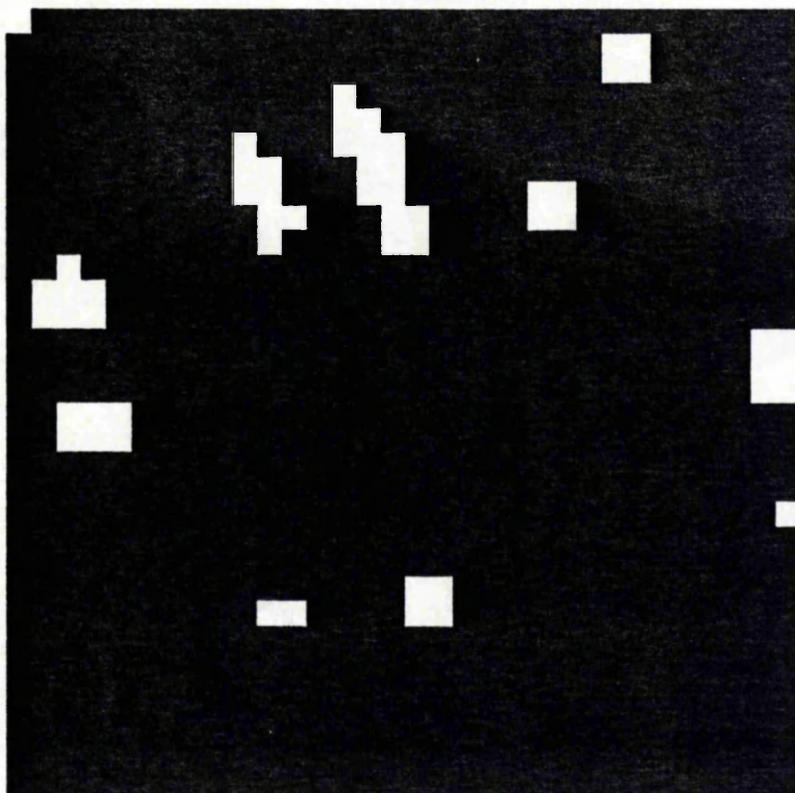


Figure 4.14a 'Significant' change between Image 3 and Image 4

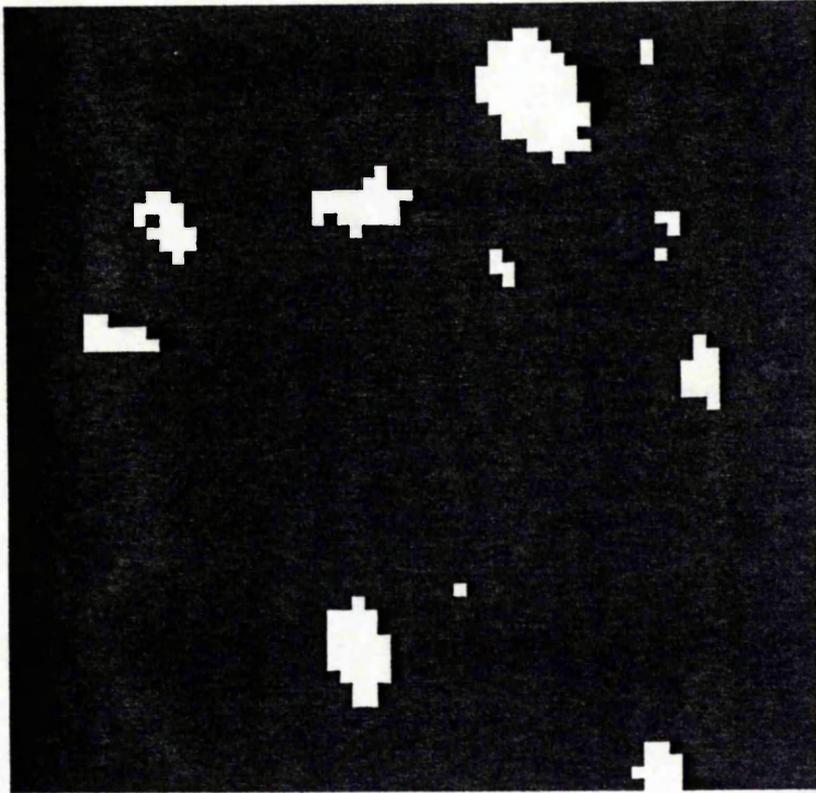


Figure 4.14b 'Significant' change between Image 3 and Image 4  
after a reduction in resolution

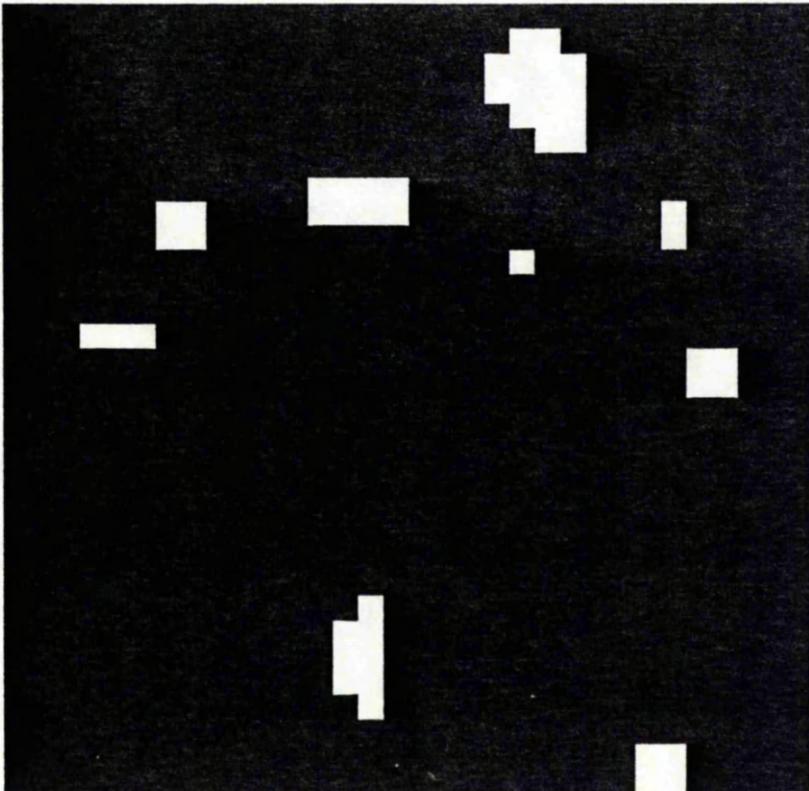


Figure 4.15a 'Significant' change between Image 5 and Image 6

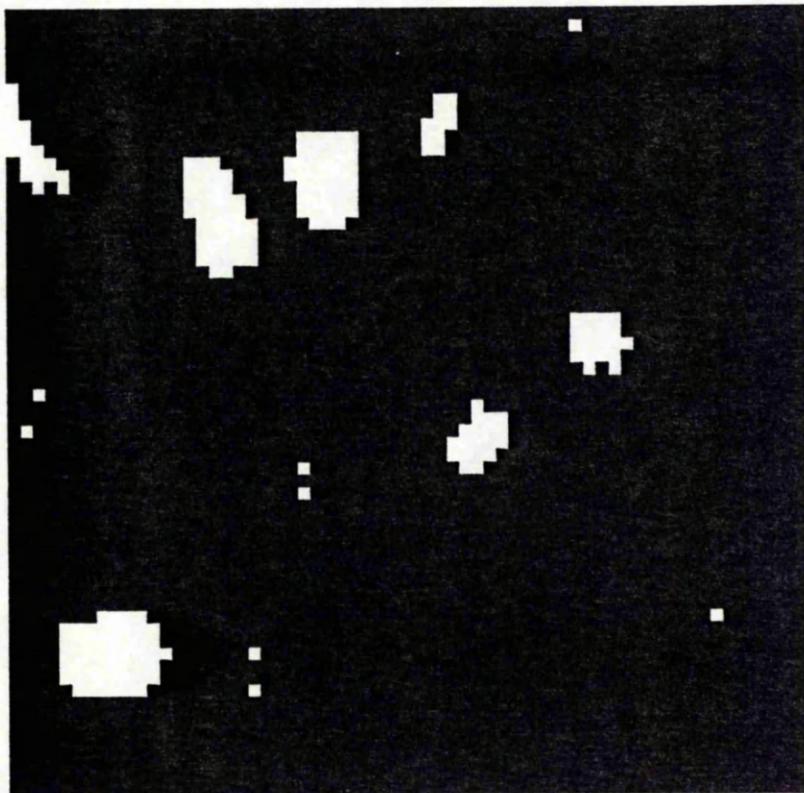
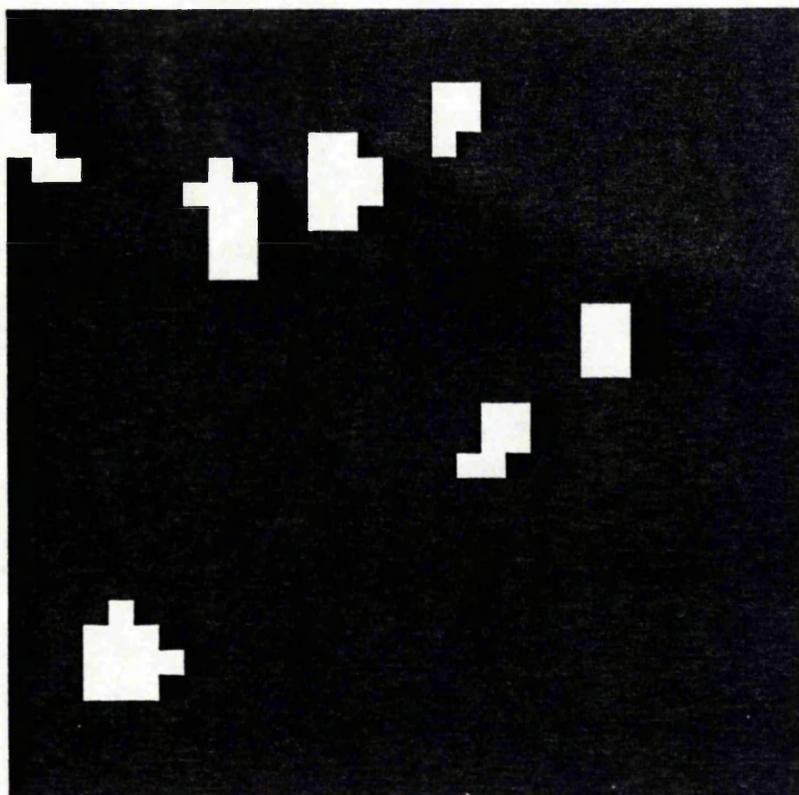


Figure 4.15b 'Significant' change between Image 5 and Image 6  
after a reduction in resolution



appears to be roughly similar, with 7 to 10 clusters being present.

### 4.3 DISCUSSION

The findings from this pilot study, using SPECT reconstructed phantom images, indicated that noise in differenced images obtained by subtraction on a pixel by pixel basis was not completely spatially random. Regardless of the spatial resolution of the reconstructed images, clusters of pixels reflecting high noise were clearly visible and often appeared to be circular in shape. Factors which may contribute to the presence of clusters of 'high noise' include the filtering and smoothing process used in the image reconstruction, the reconstruction algorithm and perhaps the initial position of the detectors of the imaging equipment. These features are likely to be machine and software dependent. Hence the same machine with different software or other machines may not display these features.

The consequences of image noise having a spatial structure are far reaching, particularly in activation studies if the response is within the range of image noise. Fox et al (1988) suggest that averaging difference images over a group of patients will enhance the detection of activation dependent changes in brain activity, since focal changes will sum whilst measurement noise will cancel. However, such an approach to functional brain mapping introduces a further technical problem. Highly accurate anatomical standardisation is required which is capable of

correcting for individual differences in slice orientation, brain size and shape. Moreover, it remains to be seen what ratio of signal to noise is required to have at least some assurance that the differences observed reflect areas of high activation.

Although the relatively simplistic approach employed in this analysis revealed that it may be erroneous to make biological interpretations of reconstructed SPECT images for the NOVO system and current software, the procedure used for detecting change in activation is by no means perfect itself. One disadvantage is that inferences about specific changes in activation can only be made with respect to the whole brain. For example, rather than saying that functional activity has changed in the hippocampus, we can only say that functional activity has changed in the hippocampus relative to the rest of the brain. Furthermore, in circumstances where metabolic activity has increased uniformly over the whole brain, highlighting approximately five per cent of the pixels would be very misleading.

A further approach for imaging change would be to construct a probability map using t-statistics. For example in activation type studies, a t-statistic could be calculated by comparing the grey level information contained in the same pixel in anatomically standardised task minus control images of different subjects. However, the use of significant probability maps in this way introduces the problem of multiple comparisons. If each of the t-tests are applied separately with a significance level of 0.05 then the probability of concluding that at least one significant difference exists, when in fact there is no

difference, will be very close to 1 for a sufficiently large number of comparisons. Thus, to control for the multiplicity effect and remove the possibility of erroneous conclusions the significance level of each comparison must be carefully chosen to preserve the familywise error rate.

**CHAPTER 5    CONCLUSIONS AND FURTHER WORK**

The principal aim of this thesis was to develop and evaluate statistical methods which would provide greater and more reliable insight into the biological processes which are illustrated in functional maps of the brain. These maps are generated by positron emission tomography (PET), single photon emission computerised tomography (SPECT) and quantitative autoradiography.

In chapter one we derived an exploratory mathematical model to describe the use of MK-801 as an *in vivo* ligand for studying glutamenergic mechanisms and to gain insight into the anti-ischaemic effect of this compound. A nonlinear regression analysis of data from normal and ischaemic rat brain tissue revealed that a two compartment was adequate to describe the kinetic properties of MK-801 in normal tissue while a three compartment model could be fitted to four of the nine regions in the ischaemic brain, where there is low cerebral blood flow and enhanced glutamate release.

A severe limitation of any data that arises from compartmental modelling of receptor binding, or any other biochemical process in the brain, is that a point estimate of the parameter of interest is derived, using a nonlinear function of a number of underlying constants. These constants in turn are usually estimated from experimental data using nonlinear regression, and are used as if they are measured without error. No work has been done to determine how the error in the estimation of these constants propagates through the nonlinear function to effect the parameter of interest. Ideally, for each of the models used,

interval estimates of the parameters of interest should be derived and the reliability of the quoted values assessed, using simulation.

In chapter two we introduced the problem of ranking brain regions with respect to the strength of response, in a set of brain regions selected for study. In section 2.1 we identified the forms of the covariance structure for maximising and minimising the probability of correctly ranking a 3-dimensional observation vector, assuming multivariate normality. The Bonferroni inequality was used to derive a lower bound for the probability of ranking the components of vectors with higher dimensions. We have seen that these mathematical approaches to the problem will have severe limitations in more realistic situations, due to the complex nature of the observational vector. One difficulty is that the variances of the regional measurements may not be necessarily equal. To solve problems of practical interest, in section 2.2 we have considered an alternative approach for characterising response to a drug which might have some potential for identifying regions having extreme or atypical responses and for comparing the pattern of response with respect to different drugs. Using a simple measure of response, we investigated, for specific data sets, rankings of the regions and measures of association between drugs. These reflect patterns which would be expected biologically. Moreover, these rankings should be fairly robust for moderate sample sizes, if we are willing to assume that all the correlations between regions are greater than or equal to zero. This should give us some confidence in interpreting these measures when studying new drugs of unknown action.

In chapter three we compared, in a simulation study, four univariate repeated measures analysis of variance techniques. Particular attention was paid to the performance of these techniques in hypothetical situations representing the blood flow-metabolism coupling problem, where the number of brain regions was greater than the number of experimental animals. Under these conditions we have seen that the F-test or any of the adjusted tests will not be robust, with respect to the Type I error rate, when the covariance matrix departs from sphericity. However, the Greenhouse-Geisser test is recommended if a conservative approach is necessary. In section 3.8, the performance of pairwise comparisons, based on the Bonferroni or Sidak inequality and using a separate estimate of the variance for each contrast, was very encouraging. If inferences of interest are pursued using either of these approaches, regardless of the outcome of the F-test, then the probability of making any error in the given family of inferences will be less than the nominal value and will be achieved more reliably than the Huynh-Feldt and Greenhouse-Geisser tests in all but a few extreme cases.

In chapter four we examined the spatial patterns which were observed when image reconstructions of a phantom using SPECT, were subtracted on a pixel by pixel basis. The presence of spatially distributed noise will impose severe limitations on the assessment of any apparent effects observed in the difference image. There is much scope for further work here. For instance, it remains to be seen what ratio of signal to noise would be required to have at least some assurance, that any differences in the images of patients are not merely artefacts of the

measurement process. We should also investigate further whether or not the noise from other imaging modalities, or similar machines with different software display the same features.

Appendix A.1(i) Correlations between the estimated and simulated within treatment rankings for the rostral raphe study, using f1.

		DORSAL			MEDIAN			5,7-DHT(1)			5,7-DHT(2)		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.951	0.942	0.958	0.933	0.917	0.945	0.820	0.781	0.850	0.878	0.839	0.900
2	1	0.908	0.892	0.921	0.876	0.847	0.898	0.704	0.644	0.751	0.788	0.731	0.825
4	1	0.837	0.810	0.859	0.782	0.733	0.819	0.550	0.473	0.616	0.657	0.578	0.716
8	1	0.724	0.682	0.761	0.644	0.572	0.703	0.392	0.303	0.468	0.492	0.391	0.569
16	1	0.576	0.511	0.634	0.480	0.400	0.558	0.242	0.150	0.334	0.328	0.233	0.419
1	1	0.951	0.942	0.958	0.933	0.917	0.945	0.820	0.781	0.850	0.878	0.839	0.900
2	2	0.907	0.888	0.922	0.873	0.834	0.898	0.696	0.620	0.749	0.783	0.697	0.826
4	4	0.830	0.789	0.857	0.777	0.691	0.820	0.534	0.426	0.614	0.647	0.491	0.718
8	8	0.711	0.629	0.761	0.622	0.437	0.700	0.362	0.221	0.470	0.465	0.246	0.579
16	16	0.561	0.404	0.635	0.433	0.161	0.557	0.224	0.078	0.343	0.313	0.090	0.437

Appendix A.1(ii) Correlations between the estimated and simulated within treatment rankings for the rostral raphe study, using f2.

		DORSAL			MEDIAN			5,7-DHT(1)			5,7-DHT(2)		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.955	0.947	0.960	0.924	0.894	0.936	0.815	0.778	0.840	0.861	0.818	0.883
2	1	0.916	0.901	0.925	0.867	0.825	0.886	0.696	0.644	0.737	0.776	0.720	0.807
4	1	0.847	0.823	0.866	0.767	0.703	0.805	0.541	0.473	0.596	0.641	0.563	0.695
8	1	0.730	0.690	0.764	0.622	0.542	0.679	0.374	0.302	0.448	0.472	0.381	0.538
16	1	0.582	0.524	0.629	0.457	0.376	0.525	0.235	0.150	0.307	0.315	0.230	0.389
1	1	0.955	0.947	0.960	0.924	0.894	0.936	0.815	0.778	0.840	0.861	0.818	0.883
2	2	0.914	0.890	0.926	0.864	0.779	0.886	0.690	0.614	0.734	0.764	0.672	0.802
4	4	0.841	0.783	0.864	0.767	0.616	0.807	0.534	0.409	0.604	0.637	0.467	0.695
8	8	0.725	0.608	0.770	0.614	0.349	0.691	0.358	0.211	0.454	0.458	0.238	0.558
16	16	0.566	0.350	0.643	0.436	0.108	0.554	0.226	0.073	0.333	0.302	0.086	0.418

Appendix A.1(iii) Correlations between the simulated between treatment rankings for the rostral raphe study, using f1.

		DORS.,MED.			DORS.,5,7-DHT(1)			MED.,5,7-DHT(1)			5,7-DHT(1),(2)		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.707	0.664	0.738	0.610	0.545	0.661	0.524	0.463	0.579	0.763	0.703	0.804
2	1	0.657	0.608	0.702	0.496	0.421	0.562	0.415	0.342	0.488	0.726	0.663	0.775
4	1	0.587	0.518	0.644	0.355	0.275	0.435	0.287	0.205	0.375	0.690	0.617	0.749
8	1	0.486	0.402	0.560	0.216	0.128	0.300	0.164	0.072	0.255	0.639	0.560	0.703
16	1	0.376	0.285	0.466	0.101	0.017	0.190	0.069	-0.009	0.145	0.604	0.519	0.676
1	1	0.707	0.664	0.738	0.610	0.545	0.661	0.524	0.463	0.579	0.763	0.703	0.804
2	2	0.654	0.589	0.702	0.485	0.390	0.560	0.401	0.297	0.483	0.712	0.624	0.776
4	4	0.573	0.473	0.643	0.327	0.215	0.422	0.253	0.134	0.354	0.660	0.545	0.731
8	8	0.452	0.276	0.559	0.172	0.050	0.278	0.114	0.002	0.218	0.599	0.470	0.695
16	16	0.328	0.138	0.451	0.066	-0.039	0.173	0.033	-0.050	0.135	0.557	0.414	0.661

Appendix A.1(iv) Correlations between the simulated between treatment rankings for the rostral raphe study, using f2.

		DORS.,MED.			DORS., 5,7-DHT(1)			MED., 5,7-DHT(1)			5,7-DHT(1),(2)		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.739	0.682	0.772	0.621	0.536	0.676	0.494	0.423	0.547	0.717	0.652	0.765
2	1	0.683	0.621	0.725	0.503	0.426	0.571	0.394	0.324	0.459	0.683	0.618	0.733
4	1	0.597	0.517	0.652	0.364	0.286	0.434	0.279	0.198	0.350	0.650	0.584	0.701
8	1	0.485	0.403	0.552	0.218	0.142	0.300	0.160	0.084	0.237	0.600	0.540	0.658
16	1	0.367	0.288	0.442	0.107	0.028	0.182	0.073	-0.003	0.147	0.567	0.504	0.625
1	1	0.739	0.682	0.772	0.621	0.536	0.676	0.494	0.423	0.547	0.765	0.652	0.765
2	2	0.677	0.576	0.726	0.492	0.382	0.577	0.376	0.266	0.455	0.666	0.575	0.738
4	4	0.586	0.438	0.658	0.332	0.212	0.436	0.238	0.114	0.340	0.614	0.502	0.692
8	8	0.448	0.245	0.559	0.177	0.044	0.279	0.111	0.005	0.215	0.559	0.433	0.651
16	16	0.318	0.122	0.444	0.075	-0.030	0.176	0.043	-0.048	0.134	0.514	0.378	0.617

Appendix A.1(v) Correlations between the estimated and simulated within treatment rankings for the Alpidem/Zolpidem/Diazepam study, using f1.

		ALPIDEM			ZOLPIDEM			BENZODIAZEPAM		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.850	0.818	0.876	0.925	0.912	0.937	0.898	0.882	0.911
2	1	0.762	0.705	0.803	0.883	0.859	0.902	0.839	0.812	0.862
4	1	0.638	0.563	0.700	0.808	0.770	0.842	0.746	0.704	0.786
8	1	0.501	0.403	0.582	0.697	0.640	0.753	0.629	0.565	0.684
16	1	0.344	0.235	0.495	0.558	0.479	0.631	0.486	0.399	0.564
1	1	0.850	0.818	0.876	0.925	0.912	0.937	0.898	0.882	0.911
2	2	0.755	0.686	0.801	0.880	0.857	0.901	0.838	0.811	0.861
4	4	0.621	0.493	0.695	0.803	0.765	0.838	0.747	0.704	0.786
8	8	0.462	0.267	0.570	0.689	0.622	0.740	0.614	0.552	0.676
16	16	0.289	-0.003	0.432	0.546	0.442	0.626	0.471	0.379	0.555

Appendix A.1(vi) Correlations between the estimated and simulated within treatment rankings for the Alpidem/Zolpidem/Diazepam study, using f2.

		ALPIDEM			ZOLPIDEM			DIAZEPAM		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.841	0.806	0.867	0.836	0.821	0.852	0.866	0.849	0.883
2	1	0.761	0.706	0.799	0.805	0.779	0.826	0.809	0.782	0.834
4	1	0.641	0.575	0.699	0.746	0.709	0.777	0.721	0.678	0.757
8	1	0.507	0.413	0.581	0.655	0.600	0.699	0.610	0.547	0.663
16	1	0.347	0.244	0.436	0.530	0.457	0.591	0.479	0.401	0.544
1	1	0.841	0.806	0.867	0.836	0.821	0.852	0.866	0.849	0.883
2	2	0.754	0.677	0.795	0.802	0.779	0.824	0.808	0.782	0.834
4	4	0.624	0.482	0.692	0.738	0.702	0.773	0.725	0.679	0.759
8	8	0.469	0.253	0.576	0.638	0.576	0.689	0.592	0.540	0.656
16	16	0.303	-0.025	0.431	0.510	0.413	0.586	0.465	0.376	0.534

Appendix A.1(vii) Correlations between the simulated between treatment rankings for the Alpidem/Zolpidem/Diazepam study, using f1.

		ALPIDEM, ZOLPIDEM			ALPIDEM, DIAZEPAM			ZOLPIDEM, DIAZEPAM		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.537	0.467	0.595	0.302	0.232	0.368	0.603	0.558	0.646
2	1	0.457	0.376	0.531	0.248	0.162	0.332	0.533	0.469	0.591
4	1	0.340	0.249	0.430	0.184	0.083	0.275	0.434	0.363	0.510
8	1	0.236	0.135	0.337	0.116	0.019	0.226	0.314	0.227	0.402
16	1	0.119	0.008	0.233	0.050	-0.052	0.169	0.188	0.083	0.294
1	1	0.537	0.467	0.595	0.302	0.232	0.368	0.603	0.558	0.646
2	2	0.443	0.343	0.523	0.230	0.130	0.316	0.535	0.477	0.591
4	4	0.319	0.191	0.426	0.159	0.047	0.267	0.431	0.348	0.508
8	8	0.191	0.031	0.307	0.086	-0.046	0.204	0.298	0.201	0.399
16	16	0.067	-0.058	0.203	0.022	-0.089	0.137	0.170	0.062	0.280

Appendix A.1(viii) Correlations between the simulated between treatment rankings for the Alpidem/Zolpidem/Diazepam study, using f2.

		ALPIDEM, ZOLPIDEM			ALPIDEM, DIAZEPAM			ZOLPIDEM, DIAZEPAM		
a	b	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
1	1	0.590	0.521	0.643	0.344	0.275	0.403	0.609	0.571	0.644
2	1	0.507	0.428	0.572	0.282	0.203	0.354	0.540	0.491	0.592
4	1	0.384	0.303	0.464	0.203	0.118	0.291	0.446	0.381	0.513
8	1	0.271	0.170	0.362	0.135	0.039	0.228	0.336	0.256	0.410
16	1	0.143	0.038	0.247	0.072	-0.036	0.174	0.207	0.109	0.303
1	1	0.590	0.521	0.643	0.344	0.275	0.403	0.609	0.571	0.644
2	2	0.493	0.391	0.565	0.267	0.166	0.350	0.546	0.495	0.590
4	4	0.363	0.220	0.459	0.183	0.069	0.283	0.447	0.372	0.512
8	8	0.221	0.046	0.331	0.108	-0.026	0.218	0.313	0.225	0.400
16	16	0.088	-0.051	0.210	0.037	-0.078	0.144	0.186	0.081	0.283

Appendix A.2(i) Correlations for the dorsal raphe rankings using  
f1 (40 ROIs).

n	r	e								
		0.1			0.5			0.9		
		median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.939	0.921	0.953	0.964	0.953	0.972	0.990	0.985	0.993
	20	0.938	0.919	0.951	0.954	0.934	0.966	0.975	0.948	0.987
	10	0.935	0.915	0.950	0.945	0.923	0.959	0.955	0.925	0.974
	5	0.935	0.916	0.949	0.938	0.917	0.953	0.941	0.918	0.961
	2	0.934	0.917	0.954	0.935	0.916	0.950	0.936	0.916	0.953
	1	(independence)		median=0.935	Q1=0.917	Q3=0.949				
10	40	0.969	0.961	0.976	0.982	0.976	0.986	0.995	0.992	0.996
	20	0.968	0.960	0.975	0.977	0.969	0.983	0.989	0.977	0.994
	10	0.967	0.958	0.974	0.972	0.961	0.979	0.977	0.963	0.987
	5	0.967	0.958	0.974	0.969	0.958	0.977	0.971	0.958	0.980
	2	0.967	0.958	0.974	0.967	0.958	0.974	0.968	0.957	0.975
	1	(independence)		median=0.967	Q1=0.958	Q3=0.974				

Appendix A.2(ii) Correlations for the dorsal raphe rankings using  
f2 (40 ROIs).

n	r	e								
		0.1			0.5			0.9		
		median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.932	0.914	0.945	0.961	0.945	0.968	0.991	0.979	0.993
	20	0.930	0.911	0.943	0.946	0.918	0.961	0.966	0.923	0.985
	10	0.930	0.909	0.942	0.937	0.910	0.954	0.945	0.910	0.969
	5	0.928	0.911	0.941	0.931	0.910	0.946	0.934	0.904	0.954
	2	0.928	0.910	0.940	0.928	0.907	0.943	0.929	0.907	0.945
	1	(independence)		median=0.928	Q1=0.909	Q3=0.940				
10	40	0.966	0.959	0.972	0.980	0.974	0.984	0.995	0.989	0.996
	20	0.964	0.955	0.971	0.975	0.962	0.980	0.985	0.967	0.993
	10	0.964	0.956	0.970	0.968	0.957	0.976	0.972	0.954	0.984
	5	0.963	0.955	0.970	0.965	0.955	0.972	0.967	0.954	0.977
	2	0.963	0.954	0.970	0.955	0.955	0.970	0.964	0.953	0.971
	1	(independence)		median=0.963	Q1=0.955	Q3=0.969				

Appendix A.2(iii) Correlations for the median raphe rankings

using f1 (40 ROIs).

n	r	e								
		0.1			0.5			0.9		
		median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.943	0.921	0.958	0.965	0.949	0.974	0.988	0.978	0.993
	20	0.939	0.915	0.956	0.954	0.926	0.968	0.969	0.939	0.984
	10	0.939	0.914	0.955	0.946	0.919	0.962	0.955	0.922	0.975
	5	0.937	0.915	0.955	0.939	0.915	0.957	0.943	0.914	0.962
	2	0.938	0.913	0.954	0.937	0.913	0.954	0.940	0.914	0.957
	1	(independence)			median=0.938	Q1=0.914		Q3=0.954		
10	40	0.972	0.962	0.979	0.982	0.974	0.987	0.994	0.989	0.996
	20	0.970	0.960	0.978	0.977	0.967	0.984	0.986	0.971	0.993
	10	0.970	0.958	0.978	0.973	0.961	0.981	0.978	0.962	0.987
	5	0.970	0.958	0.977	0.970	0.958	0.978	0.972	0.958	0.981
	2	0.968	0.957	0.977	0.969	0.957	0.977	0.969	0.957	0.977
	1	(independence)			median=0.969	Q1=0.959		Q3=0.977		

Appendix A.2(iv) Correlations for the median raphe rankings using

f2 (40 ROIs).

n	r	e								
		0.1			0.5			0.9		
		median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.922	0.888	0.940	0.946	0.913	0.960	0.969	0.935	0.984
	20	0.918	0.879	0.936	0.933	0.885	0.952	0.945	0.885	0.969
	10	0.918	0.880	0.934	0.924	0.884	0.944	0.932	0.876	0.958
	5	0.917	0.881	0.935	0.919	0.882	0.939	0.920	0.878	0.944
	2	0.916	0.881	0.934	0.917	0.877	0.936	0.918	0.883	0.938
	1	(independence)			median=0.919	Q1=0.886		Q3=0.935		
10	40	0.957	0.942	0.966	0.969	0.953	0.978	0.979	0.965	0.991
	20	0.955	0.939	0.965	0.962	0.944	0.973	0.969	0.941	0.982
	10	0.955	0.939	0.965	0.959	0.938	0.969	0.961	0.936	0.975
	5	0.955	0.940	0.964	0.956	0.939	0.966	0.957	0.937	0.969
	2	0.955	0.939	0.964	0.954	0.937	0.965	0.954	0.936	0.971
	1	(independence)			median=0.955	Q1=0.940		Q3=0.963		

Appendix A.2(v) Correlations for the association between the dorsal and median raphe stimulation rankings, using f1 (40 ROIs).

n	r	e								
		0.1			0.5			0.9		
		median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.833	0.795	0.862	0.850	0.815	0.873	0.872	0.851	0.885
	20	0.828	0.785	0.862	0.840	0.798	0.870	0.856	0.814	0.882
	10	0.827	0.785	0.860	0.834	0.791	0.869	0.843	0.788	0.881
	5	0.828	0.784	0.861	0.829	0.787	0.864	0.833	0.791	0.866
	2	0.827	0.782	0.859	0.825	0.784	0.859	0.827	0.784	0.860
1	(independence)		median=0.826			Q1=0.786		Q3=0.858		
10	40	0.857	0.834	0.878	0.867	0.848	0.882	0.879	0.866	0.888
	20	0.856	0.833	0.877	0.863	0.838	0.883	0.871	0.846	0.888
	10	0.856	0.830	0.879	0.858	0.829	0.883	0.864	0.829	0.888
	5	0.857	0.831	0.877	0.856	0.829	0.879	0.857	0.830	0.881
	2	0.855	0.829	0.877	0.854	0.828	0.876	0.856	0.828	0.878
1	(independence)		median=0.856			Q1=0.832		Q3=0.878		

Appendix A.2(vi) Correlations for the association between the dorsal and median raphe stimulation rankings, using f2 (40 ROIs).

n	r	e								
		0.1			0.5			0.9		
		median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.775	0.717	0.812	0.793	0.729	0.830	0.821	0.740	0.850
	20	0.766	0.705	0.809	0.780	0.713	0.824	0.789	0.711	0.841
	10	0.769	0.710	0.809	0.774	0.712	0.818	0.781	0.704	0.832
	5	0.772	0.708	0.810	0.771	0.712	0.812	0.773	0.709	0.816
	2	0.769	0.712	0.810	0.768	0.705	0.808	0.769	0.712	0.811
1	(independence)		median=0.770			Q1=0.712		Q3=0.808		
10	40	0.803	0.763	0.831	0.816	0.770	0.840	0.828	0.780	0.852
	20	0.803	0.763	0.830	0.806	0.766	0.837	0.813	0.762	0.847
	10	0.803	0.764	0.832	0.804	0.758	0.834	0.804	0.757	0.839
	5	0.801	0.764	0.829	0.802	0.760	0.831	0.801	0.761	0.831
	2	0.800	0.765	0.829	0.800	0.760	0.829	0.798	0.758	0.827
1	(independence)		median=0.801			Q1=0.761		Q3=0.827		

Appendix A.2(vii) Correlations for the within treatment ranking  
of Zolpidem, using f1 (40 ROIs).

		e								
		0.1			0.5			0.9		
n	r	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.817	0.775	0.853	0.881	0.851	0.904	0.964	0.948	0.973
	20	0.808	0.765	0.848	0.850	0.803	0.886	0.910	0.837	0.951
	10	0.807	0.763	0.842	0.825	0.774	0.862	0.848	0.781	0.901
	5	0.805	0.759	0.843	0.815	0.764	0.850	0.826	0.767	0.873
	2	0.805	0.760	0.843	0.807	0.757	0.846	0.814	0.762	0.852
1		(independence)		median=0.804	Q1=0.760	Q3=0.838				
10	40	0.899	0.875	0.918	0.938	0.923	0.949	0.981	0.973	0.986
	20	0.895	0.871	0.914	0.920	0.894	0.939	0.905	0.951	0.974
	10	0.894	0.871	0.914	0.905	0.876	0.927	0.919	0.878	0.948
	5	0.892	0.869	0.912	0.898	0.871	0.921	0.903	0.866	0.931
	2	0.891	0.867	0.911	0.893	0.867	0.911	0.896	0.866	0.917
1		(independence)		median=0.891	Q1=0.868	Q3=0.911				

Appendix A.2(viii) Correlations for the within treatment ranking  
of Zolpidem, using f2 (40 ROIs).

		e								
		0.1			0.5			0.9		
n	r	median	Q1	Q3	median	Q1	Q3	median	Q1	Q3
5	40	0.761	0.706	0.806	0.833	0.784	0.865	0.930	0.885	0.952
	20	0.753	0.696	0.801	0.799	0.728	0.839	0.854	0.741	0.913
	10	0.752	0.692	0.794	0.770	0.704	0.819	0.790	0.699	0.858
	5	0.745	0.686	0.793	0.756	0.697	0.804	0.768	0.691	0.822
	2	0.750	0.694	0.796	0.750	0.690	0.796	0.752	0.688	0.801
1		(independence)		median=0.747	Q1=0.688	Q3=0.791				
10	40	0.857	0.828	0.882	0.904	0.877	0.922	0.957	0.926	0.973
	20	0.854	0.820	0.879	0.880	0.839	0.907	0.911	0.842	0.949
	10	0.852	0.818	0.877	0.862	0.822	0.894	0.878	0.824	0.917
	5	0.845	0.820	0.876	0.856	0.817	0.885	0.860	0.814	0.895
	2	0.848	0.814	0.876	0.850	0.814	0.876	0.853	0.816	0.880
1		(independence)		median=0.849	Q1=0.819	Q3=0.874				



Appendix B.2 Empirical results of the estimated significance level corresponding to situation (a) in Table 3.2.

n	$\rho$	$\epsilon$	Unadjusted			$\tilde{\epsilon}$ -adjusted			$\hat{\epsilon}$ -adjusted			Lower bound		
			0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01
5	0	1.000	0.049	0.024	0.010	0.041	0.020	0.007	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.964	0.050	0.026	0.011	0.042	0.021	0.008	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.860	0.057	0.032	0.015	0.047	0.026	0.012	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.561	0.087	0.058	0.034	0.065	0.042	0.024	0.004	0.001	0.0	0.0	0.0	0.0
	$\frac{1}{2}$	0.318	0.131	0.098	0.070	0.088	0.064	0.041	0.012	0.005	0.001	0.0	0.0	0.0
	$\frac{2}{3}$	0.174	0.173	0.142	0.112	0.101	0.075	0.053	0.028	0.012	0.006	0.002	0.0	0.0
	$\frac{5}{6}$	0.096	0.215	0.189	0.162	0.096	0.072	0.051	0.046	0.025	0.012	0.011	0.003	0.0
	$\frac{11}{12}$	0.071	0.239	0.213	0.187	0.083	0.059	0.041	0.054	0.030	0.015	0.023	0.009	0.002
	$\frac{23}{24}$	0.061	0.251	0.227	0.203	0.071	0.047	0.030	0.055	0.030	0.014	0.032	0.013	0.004
	$\frac{49}{50}$	0.057	0.257	0.235	0.210	0.063	0.038	0.022	0.053	0.028	0.013	0.039	0.017	0.006
10	0	1.000	0.051	0.025	0.008	0.046	0.020	0.007	0.002	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.964	0.054	0.027	0.010	0.047	0.023	0.008	0.003	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.860	0.064	0.033	0.016	0.054	0.028	0.013	0.006	0.001	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.561	0.092	0.065	0.041	0.070	0.044	0.025	0.018	0.007	0.002	0.0	0.0	0.0
	$\frac{1}{2}$	0.316	0.125	0.096	0.073	0.076	0.054	0.036	0.036	0.020	0.007	0.001	0.0	0.0
	$\frac{2}{3}$	0.174	0.161	0.133	0.109	0.076	0.054	0.037	0.050	0.030	0.015	0.006	0.0	0.0
	$\frac{5}{6}$	0.096	0.199	0.174	0.149	0.069	0.048	0.028	0.056	0.033	0.017	0.022	0.007	0.001
	$\frac{11}{12}$	0.071	0.216	0.194	0.171	0.063	0.039	0.020	0.056	0.031	0.015	0.035	0.014	0.004
	$\frac{23}{24}$	0.061	0.223	0.203	0.182	0.059	0.033	0.015	0.054	0.028	0.013	0.044	0.019	0.007
	$\frac{49}{50}$	0.057	0.227	0.208	0.188	0.056	0.028	0.013	0.053	0.027	0.012	0.048	0.022	0.009

Appendix B.3 Empirical results of the estimated significance level corresponding to situation (b) in Table 3.2.

n	ρ	ε	Unadjusted			ε-adjusted			A ε-adjusted			Lower bound		
			0.05	0.005	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01
5	0	1.000	0.049	0.024	0.010	0.041	0.020	0.007	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.977	0.050	0.024	0.010	0.042	0.020	0.007	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.910	0.057	0.029	0.011	0.047	0.022	0.009	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.701	0.077	0.047	0.025	0.057	0.035	0.017	0.001	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{2}$	0.491	0.108	0.074	0.043	0.070	0.044	0.026	0.005	0.001	0.0	0.0	0.0	0.0
	$\frac{2}{3}$	0.333	0.144	0.107	0.073	0.077	0.049	0.029	0.010	0.003	0.001	0.0	0.0	0.0
	$\frac{5}{6}$	0.228	0.181	0.143	0.107	0.071	0.046	0.027	0.016	0.006	0.002	0.001	0.0	0.0
	$\frac{11}{12}$	0.189	0.199	0.162	0.125	0.067	0.040	0.022	0.018	0.008	0.003	0.002	0.0	0.0
	$\frac{23}{24}$	0.173	0.208	0.171	0.134	0.063	0.037	0.020	0.020	0.008	0.003	0.003	0.001	0.0
	$\frac{49}{50}$	0.165	0.212	0.178	0.139	0.060	0.034	0.018	0.020	0.008	0.003	0.004	0.001	0.0
10	0	1.000	0.051	0.025	0.008	0.046	0.020	0.007	0.002	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.977	0.055	0.024	0.009	0.049	0.020	0.007	0.003	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.910	0.061	0.031	0.012	0.052	0.026	0.010	0.004	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.701	0.081	0.051	0.027	0.060	0.035	0.016	0.009	0.002	0.0	0.0	0.0	0.0
	$\frac{1}{2}$	0.491	0.112	0.077	0.048	0.065	0.039	0.021	0.018	0.007	0.002	0.0	0.0	0.0
	$\frac{2}{3}$	0.333	0.141	0.110	0.078	0.064	0.039	0.020	0.026	0.012	0.004	0.0	0.0	0.0
	$\frac{5}{6}$	0.228	0.173	0.141	0.109	0.059	0.034	0.017	0.032	0.014	0.005	0.002	0.0	0.0
	$\frac{11}{12}$	0.189	0.190	0.155	0.125	0.056	0.031	0.015	0.033	0.015	0.005	0.004	0.0	0.0
	$\frac{23}{24}$	0.173	0.199	0.163	0.131	0.053	0.029	0.014	0.034	0.015	0.005	0.005	0.001	0.0
	$\frac{49}{50}$	0.165	0.203	0.168	0.136	0.052	0.028	0.013	0.034	0.015	0.005	0.005	0.001	0.0

Appendix B.4 Empirical results of the estimated significance level corresponding to situation (c) in Table 3.2.

n	p	ε	Unadjusted			ε-adjusted			A ε-adjusted			Lower bound		
			0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01
5	0	1.000	0.049	0.024	0.010	0.041	0.020	0.007	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.993	0.049	0.024	0.009	0.042	0.020	0.007	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.973	0.050	0.025	0.010	0.042	0.020	0.008	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.897	0.058	0.030	0.013	0.047	0.023	0.009	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{2}$	0.792	0.069	0.038	0.018	0.054	0.029	0.013	0.001	0.0	0.0	0.0	0.0	0.0
	$\frac{2}{3}$	0.677	0.084	0.050	0.027	0.059	0.034	0.016	0.001	0.0	0.0	0.0	0.0	0.0
	$\frac{5}{6}$	0.569	0.102	0.065	0.036	0.063	0.037	0.018	0.002	0.0	0.0	0.0	0.0	0.0
	$\frac{11}{12}$	0.519	0.111	0.074	0.044	0.064	0.037	0.018	0.002	0.0	0.0	0.0	0.0	0.0
	$\frac{23}{24}$	0.496	0.115	0.078	0.047	0.064	0.036	0.017	0.002	0.001	0.0	0.0	0.0	0.0
$\frac{49}{50}$	0.484	0.117	0.080	0.048	0.063	0.036	0.017	0.003	0.001	0.0	0.0	0.0	0.0	
10	0	1.000	0.051	0.025	0.008	0.046	0.020	0.007	0.002	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.993	0.052	0.025	0.009	0.047	0.020	0.007	0.002	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.973	0.053	0.025	0.010	0.048	0.021	0.008	0.003	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.897	0.061	0.030	0.013	0.052	0.024	0.010	0.004	0.001	0.0	0.0	0.0	0.0
	$\frac{1}{2}$	0.792	0.071	0.030	0.019	0.054	0.028	0.012	0.005	0.001	0.0	0.0	0.0	0.0
	$\frac{2}{3}$	0.677	0.084	0.054	0.027	0.057	0.031	0.013	0.008	0.002	0.0	0.0	0.0	0.0
	$\frac{5}{6}$	0.569	0.100	0.066	0.037	0.058	0.029	0.013	0.010	0.003	0.001	0.0	0.0	0.0
	$\frac{11}{12}$	0.519	0.108	0.073	0.043	0.056	0.028	0.012	0.011	0.003	0.001	0.0	0.0	0.0
	$\frac{23}{24}$	0.496	0.113	0.077	0.047	0.055	0.027	0.012	0.012	0.003	0.001	0.0	0.0	0.0
$\frac{49}{50}$	0.484	0.115	0.079	0.048	0.056	0.026	0.012	0.012	0.003	0.001	0.0	0.0	0.0	

Appendix B.5 Empirical results of the estimated significance level corresponding to situation (d) in Table 3.2.

n	$\rho$	$\epsilon$	Unadjusted			$\tilde{\epsilon}$ -adjusted			$\hat{\epsilon}$ -adjusted			Lower bound			
			0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	
5	0	1.000	0.053	0.027	0.012	0.044	0.021	0.009	0.003	0.0	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.982	0.053	0.027	0.012	0.044	0.022	0.008	0.004	0.001	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.926	0.057	0.031	0.014	0.047	0.024	0.010	0.004	0.001	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.726	0.071	0.043	0.022	0.058	0.033	0.016	0.008	0.003	0.001	0.0	0.0	0.0	0.0
	$\frac{1}{2}$	0.495	0.098	0.064	0.040	0.074	0.046	0.027	0.017	0.006	0.002	0.001	0.0	0.0	0.0
	$\frac{2}{3}$	0.311	0.128	0.097	0.069	0.087	0.061	0.038	0.032	0.015	0.005	0.004	0.001	0.0	0.0
	$\frac{5}{6}$	0.188	0.168	0.136	0.108	0.090	0.065	0.045	0.049	0.026	0.012	0.014	0.004	0.004	0.001
	$\frac{11}{12}$	0.145	0.189	0.161	0.131	0.081	0.057	0.036	0.057	0.031	0.014	0.028	0.008	0.008	0.002
	$\frac{23}{24}$	0.127	0.202	0.176	0.143	0.071	0.047	0.029	0.057	0.031	0.015	0.039	0.013	0.013	0.005
	$\frac{49}{50}$	0.118	0.209	0.184	0.151	0.066	0.038	0.019	0.056	0.029	0.013	0.045	0.018	0.018	0.006
10	0	1.000	0.052	0.027	0.011	0.047	0.024	0.009	0.014	0.004	0.0	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.982	0.053	0.029	0.012	0.049	0.025	0.009	0.015	0.004	0.002	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.926	0.057	0.032	0.015	0.051	0.028	0.012	0.018	0.005	0.002	0.0	0.0	0.0	0.0
	$\frac{1}{3}$	0.726	0.071	0.045	0.025	0.056	0.033	0.018	0.025	0.012	0.005	0.001	0.0	0.0	0.0
	$\frac{1}{2}$	0.495	0.094	0.065	0.043	0.065	0.040	0.026	0.036	0.020	0.007	0.004	0.001	0.0	0.0
	$\frac{2}{3}$	0.311	0.122	0.094	0.068	0.069	0.045	0.028	0.046	0.028	0.013	0.011	0.003	0.0	0.0
	$\frac{5}{6}$	0.188	0.154	0.126	0.099	0.064	0.041	0.023	0.051	0.032	0.016	0.025	0.009	0.009	0.003
	$\frac{11}{12}$	0.145	0.172	0.144	0.117	0.059	0.035	0.019	0.052	0.030	0.015	0.036	0.015	0.015	0.005
	$\frac{23}{24}$	0.127	0.179	0.153	0.127	0.055	0.031	0.016	0.051	0.028	0.013	0.041	0.020	0.020	0.007
	$\frac{49}{50}$	0.118	0.184	0.159	0.133	0.051	0.028	0.013	0.050	0.026	0.012	0.045	0.022	0.022	0.009

Appendix B.6 Empirical results of the estimated significance level corresponding to situation (e) in Table 3.3.

n	ρ	ε	Unadjusted			ε-adjusted			A ε-adjusted			Lower bound		
			0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01
5	0	0.518	0.097	0.065	0.041	0.071	0.046	0.025	0.004	0.001	0.0	0.0	0.0	0.0
	$\frac{1}{12}$	0.464	0.103	0.072	0.046	0.073	0.050	0.029	0.005	0.002	0.0	0.0	0.0	0.0
	$\frac{1}{6}$	0.411	0.112	0.080	0.052	0.077	0.054	0.033	0.007	0.002	0.001	0.0	0.0	0.0
	$\frac{1}{3}$	0.314	0.131	0.098	0.069	0.087	0.063	0.042	0.012	0.004	0.001	0.0	0.0	0.0
	$\frac{1}{2}$	0.235	0.151	0.120	0.089	0.095	0.070	0.049	0.019	0.008	0.003	0.001	0.0	0.0
	$\frac{2}{3}$	0.174	0.173	0.142	0.112	0.101	0.075	0.053	0.028	0.012	0.006	0.002	0.0	0.0
	$\frac{5}{6}$	0.129	0.195	0.167	0.138	0.101	0.076	0.052	0.036	0.019	0.009	0.006	0.0	0.0
	$\frac{11}{12}$	0.111	0.207	0.179	0.150	0.098	0.073	0.050	0.041	0.021	0.010	0.008	0.002	0.0
	$\frac{23}{24}$	0.103	0.212	0.186	0.157	0.096	0.071	0.047	0.043	0.022	0.011	0.009	0.002	0.0
	$\frac{49}{50}$	0.099	0.215	0.190	0.161	0.094	0.070	0.046	0.044	0.022	0.011	0.010	0.002	0.0
10	0	0.518	0.099	0.070	0.045	0.065	0.042	0.024	0.019	0.008	0.002	0.0	0.0	0.0
	$\frac{1}{12}$	0.464	0.107	0.077	0.051	0.068	0.043	0.027	0.022	0.001	0.004	0.0	0.0	0.0
	$\frac{1}{6}$	0.411	0.113	0.083	0.060	0.070	0.048	0.029	0.026	0.013	0.005	0.0	0.0	0.0
	$\frac{1}{3}$	0.314	0.126	0.100	0.073	0.073	0.052	0.034	0.036	0.019	0.008	0.001	0.0	0.0
	$\frac{1}{2}$	0.235	0.143	0.116	0.091	0.076	0.054	0.037	0.045	0.025	0.012	0.003	0.0	0.0
	$\frac{2}{3}$	0.174	0.161	0.133	0.109	0.076	0.054	0.037	0.050	0.030	0.015	0.006	0.0	0.0
	$\frac{5}{6}$	0.129	0.179	0.153	0.126	0.074	0.051	0.033	0.054	0.031	0.016	0.013	0.003	0.0
	$\frac{11}{12}$	0.111	0.189	0.162	0.138	0.072	0.049	0.030	0.055	0.032	0.016	0.018	0.005	0.0
	$\frac{23}{24}$	0.103	0.194	0.168	0.142	0.070	0.047	0.028	0.055	0.032	0.015	0.019	0.006	0.0
	$\frac{49}{50}$	0.099	0.196	0.171	0.145	0.070	0.046	0.027	0.055	0.032	0.015	0.021	0.006	0.0

Appendix B.7 Empirical results of the estimated significance level corresponding to situation (f) in Table \*.2.

n	$\rho$	$\epsilon$	Unadjusted			B E-adjusted			A E-adjusted			Lower bound		
			0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01
5	0	0.174	0.173	0.142	0.112	0.101	0.075	0.053	0.028	0.012	0.006	0.002	0.0	0.0
	$\frac{1}{6}$	0.222	0.156	0.126	0.094	0.097	0.072	0.049	0.020	0.009	0.004	0.001	0.0	0.0
	$\frac{1}{3}$	0.318	0.131	0.098	0.070	0.088	0.064	0.041	0.012	0.005	0.001	0.0	0.0	0.0
	$\frac{1}{2}$	0.561	0.087	0.058	0.034	0.065	0.042	0.024	0.004	0.001	0.0	0.0	0.0	0.0
10	0	0.174	0.161	0.133	0.109	0.076	0.054	0.037	0.050	0.030	0.015	0.006	0.0	0.0
	$\frac{1}{6}$	0.222	0.146	0.018	0.094	0.077	0.055	0.037	0.046	0.025	0.013	0.003	0.0	0.0
	$\frac{1}{3}$	0.318	0.125	0.096	0.073	0.076	0.054	0.036	0.036	0.020	0.007	0.001	0.0	0.0
	$\frac{1}{2}$	0.561	0.092	0.065	0.041	0.070	0.044	0.025	0.018	0.007	0.002	0.0	0.0	0.0

B.8.(i) Estimated Joint Confidence Levels for the Four Cluster  
Correlation Matrix with  $J = 20$  and  $n = 10$

	$\alpha$					
	0	1/6	1/3	1/2	2/3	5/6
SCHEFFE	1.0	1.0	1.0	1.0	1.0	1.0
TUKEY	0.951	0.947	0.944	0.938	0.949	0.946
BONF(1)	0.970	0.964	0.962	0.958	0.962	0.961
SIDAK(1)	0.968	0.963	0.961	0.957	0.961	0.960
GGADJ	1.0	1.0	1.0	1.0	1.0	1.0
HFADJ	1.0	1.0	1.0	1.0	1.0	1.0
BONF(2)	0.960	0.955	0.959	0.958	0.964	0.975
SIDAK(2)	0.959	0.954	0.958	0.957	0.963	0.975
SCHGG	0.982	0.982	0.983	0.986	0.991	0.996
SCHHF	0.952	0.949	0.959	0.970	0.982	0.993

B.8.(ii) Estimated Joint Confidence Levels for the Four Cluster  
Correlation Matrix with  $J = 20$  and  $n = 30$

	$\alpha$					
	0	1/6	1/3	1/2	2/3	5/6
SCHEFFE	1.0	1.0	1.0	1.0	1.0	1.0
TUKEY	0.953	0.947	0.943	0.942	0.946	0.955
BONF(1)	0.967	0.963	0.959	0.959	0.960	0.967
SIDAK(1)	0.967	0.963	0.959	0.959	0.960	0.967
GGADJ	1.0	1.0	1.0	1.0	1.0	1.0
HFADJ	1.0	1.0	1.0	1.0	1.0	1.0
BONF(2)	0.962	0.963	0.959	0.964	0.969	0.978
SIDAK(2)	0.961	0.962	0.958	0.963	0.968	0.977
SCHGG	1.0	1.0	1.0	1.0	1.0	1.0
SCHHF	0.999	0.999	1.0	1.0	1.0	1.0

B.8.(iii) Estimated Joint Confidence Levels for the Two Cluster  
Correlation Matrix with  $J = 20$  and  $n = 10$

	$\rho$					
	0	1/6	1/3	1/2	2/3	5/6
SCHEFFE	1.0	1.0	1.0	1.0	0.999	0.997
TUKEY	0.951	0.948	0.946	0.934	0.923	0.925
BONF(1)	0.970	0.966	0.957	0.951	0.940	0.938
SIDAK(1)	0.968	0.965	0.955	0.950	0.939	0.937
GGADJ	1.0	1.0	1.0	1.0	1.0	1.0
HFADJ	1.0	1.0	1.0	1.0	1.0	1.0
BONF(2)	0.960	0.955	0.962	0.962	0.965	0.973
SIDAK(2)	0.959	0.945	0.961	0.961	0.965	0.973
SCHGG	0.982	0.983	0.988	0.990	0.994	0.997
SCHHF	0.952	0.950	0.967	0.975	0.986	0.996

B.8.(iv) Estimated Joint Confidence Levels for the Two Cluster  
Correlation Matrix with  $J = 20$  and  $n = 30$

	$\rho$					
	0	1/6	1/3	1/2	2/3	5/6
SCHEFFE	1.0	1.0	1.0	1.0	0.999	0.999
TUKEY	0.953	0.945	0.943	0.933	0.934	0.944
BONF(1)	0.967	0.961	0.958	0.950	0.949	0.954
SIDAK(1)	0.967	0.961	0.958	0.950	0.949	0.954
GGADJ	1.0	1.0	1.0	1.0	1.0	1.0
HFADJ	1.0	1.0	1.0	1.0	1.0	1.0
BONF(2)	0.962	0.961	0.967	0.967	0.971	0.974
SIDAK(2)	0.961	0.960	0.966	0.966	0.970	0.973
SCHGG	1.0	1.0	0.999	1.0	1.0	1.0
SCHHF	0.999	0.999	0.999	1.0	1.0	1.0

B.8.(v) Estimated Joint Confidence Levels for the Two Cluster  
Correlation Matrix with  $J = 10$  and  $n = 10$

	$\alpha$					
	0	1/6	1/3	1/2	2/3	5/6
SCHEFFE	0.997	0.997	0.997	0.994	0.989	0.982
TUKEY	0.950	0.950	0.944	0.937	0.927	0.922
BONF(1)	0.972	0.969	0.965	0.956	0.948	0.941
SIDAK(1)	0.972	0.969	0.965	0.956	0.948	0.941
GGADJ	1.0	0.999	1.0	0.999	0.998	0.997
HFADJ	0.998	0.997	0.998	0.996	0.995	0.995
BONF(2)	0.964	0.961	0.966	0.966	0.970	0.974
SIDAK(2)	0.962	0.960	0.965	0.965	0.969	0.973
SCHGG	0.962	0.960	0.967	0.973	0.982	0.989
SCHHF	0.928	0.926	0.939	0.953	0.970	0.986

B.8.(vi) Estimated Joint Confidence Levels for the Two Cluster  
Correlation Matrix with  $J = 10$  and  $n = 30$

	$\alpha$					
	0	1/6	1/3	1/2	2/3	5/6
SCHEFFE	0.998	0.998	0.996	0.995	0.993	0.989
TUKEY	0.946	0.944	0.937	0.932	0.933	0.934
BONF(1)	0.969	0.967	0.963	0.955	0.954	0.952
SIDAK(1)	0.967	0.966	0.962	0.954	0.953	0.951
GGADJ	0.999	0.999	0.998	0.999	0.999	0.999
HFADJ	0.998	0.998	0.997	0.998	0.998	0.999
BONF(2)	0.966	0.969	0.971	0.969	0.976	0.977
SIDAK(2)	0.965	0.968	0.970	0.968	0.976	0.976
SCHGG	0.994	0.992	0.995	0.997	0.999	1.0
SCHHF	0.990	0.990	0.993	0.997	0.998	1.0

REFERENCES

- ARNETT, C.D., SHIUE, C-Y., WOLF, A.P., FOWLER, J.S., LOGAN, J., and WATANABE, M. (1985) Comparison of three  $^{18}\text{F}$  - labelled butyrophenone neuroleptic drugs in the baboon using positron emission tomography. Journal of Neurochemistry, 44, 835-844.
- BAHN, M.M., HUANG, S-C., HAWKINS, R.A., SATYAMURTHY, N., HOFFMAN, J.M., BARRIO, J.R., MAZZIOTTA, J.C., and PHELPS, M.E. (1989). Models for In Vivo kinetic interactions of dopamine D2-neuroreceptors and 3-(1'-[ $^{18}\text{F}$ ] Fluoroethyl) spiperone examined with positron emission tomography. Journal of Cerebral Blood Flow and Metabolism, 9, 840-849.
- BARTLETT, E.J., BRODIE, J.D., WOLF, A.F., CHRISTMAN, D.R., LASKA E., and MEISSNER, M. (1988). Reproducibility of cerebral glucose metabolic measurements in resting human subjects. Journal of Cerebral Blood Flow and Metabolism, 8, 502-512.
- BARTLETT, M.S. (1937). Some examples of statistical methods of research in agriculture. Journal of the Royal Statistical Society Supplement, 4, 137-183.
- BECHHOFER, R.E. (1954). A single sample multiple decision procedure for ranking means of normal populations with known variances. Annals of Mathematical Statistics, 25, 16-39.
- BLASBERG, R.G., FENSTERMACHER, J.D. and PATLAK, C.S. (1983). Transport of  $\alpha$ -Aminoisobutyric acid across brain capillary and cellular membranes. J. Cereb. Blood Flow Metabol., 3, 8-32.
- BMDP (1987) Biomedical computer programs p-series. Berkeley, California, University of California Press.
- BOX, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36, 317-346.
- BOX, G.E.P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance in the one-way classification. The Annals of Mathematical Statistics, 25, 290-302.
- BOX, G.E.P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. effects of inequality of variance and of correlation between errors in the two-way classification. The Annals of Mathematical Statistics, 25, 484-498.

- COLLIER, R.O., Jr., BAKER, F.B., MANDEVILLE, G.K. and HAYES, T.F. (1967). Estimates of test size for several test procedures based on variance ratios in the repeated measures design. Psychometrika, 32, 339-353.
- DAVIDSON, M.L. (1972). Univariate versus multivariate tests in repeated-measures experiments. Psychological Bulletin, 77, 446-452.
- DES ROSIERS, M.H., KENNEDY, C., PATLAK, C.S., PETTIGREW, K.D., SOKOLOFF, L. and REIVICH, M. (1974). Relationship between local cerebral blood flow and glucose utilization in the rat. Neurology, 24, 389.
- DUNN, A.J., STEELMAN, S. and DELANOY, R. (1980). Intraventricular ACTH and vasopressin cause regionally specific changes in cerebral deoxyglucose uptake. Journal of Neuroscience Research, 5, 485-495.
- EATON, M.L. (1966). Some optimum properties of ranking procedures. Annals of Mathematical Statistics, 37, 124-137.
- ECKARDT, M.J., CAMPBELL, G.A., MARIETTA, C.A., MAJCHROWICZ, E., WIXON, H.N. and WEIGHT, F.F. (1986). Cerebral 2-Deoxyglucose uptake in rats during ethanol withdrawal and postwithdrawal. Brain Research, 366, 1-9.
- FARDE, L., HALL, H., EHRIN, E. and SEDVALL, G. (1986). Quantitative analysis of D2 dopamine receptor binding in the living human brain by PET. Science, 231, 258-261.
- FICK, A. (1870). Uber die Messung des Blutquantums in den Herzventrikeln. Verhanbl. d. phys.-med. Ges. zu Wurzburg, 2:xvi (quoted in its entirety by HOFF, H.E., SCOTT, H.J. (1948). New England Journal of Medicine, 239:122.
- FORD, I. (1986). Confounded correlations: Statistical limitations in the analysis of interregional relationships of cerebral metabolic activity. Journal of Cerebral Blood Flow and Metabolism, 6, 385-388.
- FOX, P.T., MINTUN, M.A., RAICHLE, M.E., MIEZIN, F.M., ALLMAN, J.M., and VAN ESSEN, D.C. (1986). Mapping human visual cortex with positron emission tomography. Nature, 323, 806-809.
- FOX, P.T., BURTON, H., and RAICHLE, M.E. (1987). mapping human visual cortex with positron emission tomography. Journal of Neurosurgery, 67, 34-43.

- FOX, P.T., MINTUN, M.A., RIEMAN, E.M., and RAICHLE, M.E. (1988). Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. Journal of Cerebral Blood Flow and Metabolism, 8, 642-653.
- FREY, K.A., HICHA, R.D., EHRENKAUFER, R.L.E. and AGRANOFF, B.W. (1985). Quantitative in vivo receptor binding. III: Tracer kinetic modelling of muscarinic cholinergic receptor binding. Proc. Natl. Acad. Sc., 82, 6711-6715.
- FROST, J.J. and WAGNER, H.N., Jr. (1984). Kinetics of binding to opiate receptors. In Vivo predicted from in Vitro parameters. Brain Research, 305, 1-11.
- FRISCHTAK, R.M. (1973). Statistical multiple-decision procedures for some multivariate selection problems. Technical Report No. 187, Department of Operations Research, Cornell University, Ithaca, New York.
- GALLANT, A.R. (1987). Nonlinear statistical models. Wiley, New York.
- GEISSER, S. and GREENHOUSE, S.W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. The Annals of Mathematical Statistics, 29, 885-891.
- GIBBONS, J.D., OLKIN, I. and SOBEL, M. (1977). Selecting and Ordering Populations. a New Statistical Methodology. John Wiley and Sons.
- GIBSON, R.E., WECKSTEIN, D.J., JAGODA, E.M., RZESZOTARSKI, W.J., REBA, R.C., and ECKELMAN, W.C. (1984). The characteristics of I-125 4-IQNB and H-3 QNB in Vivo and in Vitro. Journal of Nuclear Medicine, 25, 214-222.
- GJEDDE, A., WEINHARD, K., HEISS, W.D., KLOSTER, G., DIEMER, N.H., HERHOLZ, K. and PAWLIK, G. (1985). Comparative regional analysis of 2-fluorodeoxyglucose and methylglucose uptake in brain of four stroke patients. With special reference to the regional estimation of the lumped constant. Journal of Cerebral Blood Flow and Metabolism, 5, 163-178.
- GLASS, H.I. and DE GARRETA, A.C. (1967). Quantitative analysis of experimental curve fitting for biological applications. Phys. Med. Biol., 12, 379-388.

- GLASS, H.I. and DE GARRETA, A.C. (1971). The quantitative limitations of exponential curve fitting. Phys. Med. Biol., 16, 119-130.
- GODFREY, K. (1983). Compartmental Models and their application. New York, Academic Press.
- GREENHOUSE, S.W. and GEISSER, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, No. 2, June.
- GUPTA, S.S. (1956). On a decision rule for a problem in ranking means. Inst. Stat. Mimeo. Ser. No. 150, University of N.C., Chapel Hill, N.C.
- GUPTA, S.S. and PANCHAPAKESAN, S. (1979). Multiple decision procedures: theory and methodology of selecting and ranking populations. John Wiley and Sons.
- HAXBY, J.V. DUARA, R., GRADY, C.L., CUTLER, N.R. and RAPOPORT, S.I. (1985). Relations between neuropsychological and cerebral metabolic asymmetries in early alzheimer's disease. Journal of Cerebral Blood Flow and Metabolism, 5, 193-200.
- HOPKINS, J.W. and CLAY, P.P.F. (1963). Some empirical distributions of bivariate  $T^2$  and homoscedasticity criterion  $M$  under unequal variance and leptokurtosis. Journal of the American Statistical Association, 58, 1048-1053.
- HORWITZ, B., DUARA, R. and RAPOPORT, S.I. (1984). Intercorrelations of glucose metabolic rates between brain regions: application to healthy males in a state of reduced sensory input. Journal of Cerebral Blood Flow and Metabolism, 4, 484-499.
- HORWITZ, B., GRADY, C.L., SCHLAGETER, N.L., DUARA, R. and RAPOPORT, S.I. (1987). Intercorrelations of regional cerebral glucose metabolic rates in Alzheimer's disease. Brain Research, 407, 294-306.
- HUANG, S-C., PHELPS, M.E., HOFFMAN, E.J., SIDERIS, K., SELIN, C.J. and KUHL, D.E. (1980). Noninvasive determination of local cerebral metabolic rate of glucose in man. American Journal of Physiology, 238, E69-E82.
- HUANG, S-C., BARRIO, J.R. and PHELPS, M.E. (1986). Neuroreceptor assay with positron emission tomography: equilibrium versus dynamic approaches. Journal of Cerebral Blood Flow and Metabolism, 6, 515-521.

- HUCKER, H.B., HUTT, J.E., WHITE, S.D., ARISON, B.H. and ZACCHEI, A.G. (1983). Disposition and metabolism of (+)-5-methyl-10,11-dihydro-5H-dibenzo [a,d] cyclohepten-5,10-Inine in rats, dogs and monkeys. Drug Metabolism and Disposition, 11, No. 1.
- HUYNH, H. (1978). Some approximate tests for repeated measurement designs. Psychometrika, 43, 161-175.
- HUYNH, H. and FELDT, L.S. (1970). Conditions under which mean square ratios in repeated measurement designs have exact F-distributions. Journal of the American Statistical Association, 65, 1582-1589.
- HUYNH, H. and FELDT, L.S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomised block and split-plot designs. Journal of Education Statistics, 1, 69-82.
- HUYNH, H and MANDEVILLE, G.K. (1979). Validity conditions in repeated measures designs. Psychological Bulletin, 86, 964-973.
- JENNRICH, R.I., and SAMPSON, P.F. (1968). Application of stepwise regression to nonlinear least squares estimation. Technometrics, 10, 63-67.
- KESELMAN, H.J., ROGAN, J.C., MENDOZA, J.L. and BREEN, L.J. (1980). Testing the validity conditions of repeated measures F tests. Psychological Bulletin, 87, 479-481.
- KETY, S.S. (1951). The theory and applications of the exchange of inert gas at lungs and tissue. Pharmacol Rev, 3, 1-41.
- KOBATAKE, K., SAKO, K., IZAWA, M., YAMAMOTO, Y.L. and HAKIM, A.M. (1984). Autoradiographic determination of brain pH following middle cerebral artery occlusion in the rat. Stroke, 15, 540.
- KORIN, B.P. (1972). Some comments on the homoscedasticity criterion M and the multivariate analysis of variance tests  $T^2$ , W and R. Biometrika, 59, 215-216.
- KUSCHINSKY, W., SUDA, S. and SOKOLOFF, L. (1981). Local cerebral glucose utilization and blood flow during metabolic acidosis. American Journal of Physiology, 241, H772-H777.
- KUSCHINSKY, W., SUDA, S. and SOKOLOFF, L. (1985). Influence of  $\gamma$ -Hydroxybutyrate on the relationship between local cerebral glucose utilization and local cerebral blood flow in the rat brain. Journal of Cerebral Blood Flow and Metabolism, 5, 58-64.

- LOGAN, J., WOLF, A.P., SHINE, C.-Y. and FOWLER, J.S. (1987). Kinetic modelling of receptor-ligand binding applied to positron emission tomographic studies with neuroleptic tracers. Journal of Neurochemistry, 48, No. 1.
- McCULLOCH, J., KELLY, P.A.T. and FORD, I. (1982). Effect of apomorphine on the relationship between local cerebral glucose utilization and local cerebral blood flow (with an appendix on its statistical analysis). Journal of Cerebral Blood Flow and Metabolism, 2, 487-499.
- MARQUART, D.W. (1964). I.B.M. Share Programme, No. 3094.
- MAUCHLY, J.W. (1940). Significance test for sphericity of a normal n-variate distribution. The Annals of Mathematical Statistics, 29, 204-209.
- MAXWELL, S.E. (1980). Pairwise Multiple Comparisons in repeated measures designs. Journal of Educational Statistics, 5, 269-287.
- MAXWELL, S.E. and ARREY R.D. (1982). Small sample profile analysis with many variables. Psychological Bulletin, 29, 778-785.
- MENDOSA, J.L., TOOTHAKER, L.E. and CRAIN, B.R. (1976). Necessary and sufficient conditions for F ratios in the LxJxK factorial design with two repeated factors. Journal of the American Statistical Association, 71, 992-993.
- MITZEL, H.C. and GAMES, P.A. (1981). Circularity and multiple comparisons in repeated measures designs. The British Journal of Mathematical and Statistical Psychology, 34, 253-259.
- MINTUN, M.A., RAICHLE, M.E., KILBOURNE, M.R., WOOTEN, G.F. and WELCH, M.J. (1984). A quantitative model for the in vivo assessment of drug binding sites with positron emission tomography. Annals of Neurology, 15, 217-227.
- OLSEN, C.L. (1974). Comparative robustness of six tests in multivariate analysis of variance. Journal of the American Statistical Association, 69, 894-907.
- OSBORNE, M.R. (1975). Some special nonlinear least square problems. Journal of Numerical Analysis, 12, No. 4.
- PATLAK, C.S., BLASBERG, R.G. and FENSTERMACHER, J.D. (1983). Graphical evaluation of blood-to-brain transfer constants for multiple-time uptake data. Journal of Cerebral Blood Flow and Metabolism, 3, 1-7.

- PERLMUTTER, J.S., LARSON, K.B., RAICHLE, M.E., MARKHAM, J., MINTUN, M.A., KILBOURNE, M.R. and WELCH, M.J. (1986). Strategies for in vivo measurement of receptor binding using positron emission tomography. Journal of Cerebral Blood Flow and Metabolism, 6, 154-169.
- PHELPS, M.E., MAZZIOTTA, J.C. and SCHELBERT, H.R. (1979). Positron Emission Tomography: Principles and Applications for the Brain and Heart. Raven Press, New York.
- PLACKETT, R.L. (1954). A reduction formula for normal multivariate integrals. Biometrika, 41, 351-360.
- ROGAN, J.C., KESELMAN, H.J., MENDOZA, J.L. (1979). Analysis of repeated measurements. The British Journal of Mathematical and Statistical Psychology, 32, 269-286.
- ROUANET, H. and LEPINE, D. (1970). Comparison between treatments in a repeated-measurement design : ANOVA and multivariate methods. The British Journal of Mathematical and Statistical Psychology, 23, 147-163.
- SAKURADA, O., KENNEDY, C., JEHL, J., BROWN, J.D., CARBIN, G.L. and SOKOLOFF, L. (1978). Measurement of local cerebral blood flow with iodo [ $^{14}\text{C}$ ] antipyrine. American Journal of Physiology, 234, H59-H66.
- SAVAGE, I.R. (1957). Contributions to the theory of rank order statistics in the "trend" case. Annals of Mathematical Statistics, 28, 968-977.
- SCHEFFE, H. (1953). A method for judging all contrasts in the analysis of variance. Biometrika, 40, 87-104.
- SEBER, G.A.F. (1984). Multivariate Observations. John Wiley and Sons.
- SIDAK, Z. (1967). Rectangular confidence regions for the mean of multivariate normal distributions. Journal of the American Statistical Association, 62, 626-633.
- SMITH, C.B., CRANE, A.M., KADEKARO, M., AGRANOFF, B.W. and SOKOLOFF, L. (1984). Stimulation of protein synthesis and glucose utilization in the hypoglossal nucleus induced by axotomy. Journal of Neuroscience, 4, 2489-2496.
- SOKOLOFF, L., REIVICH, M., KENNEDY, C., DES ROSIERS, M.H., PATLAK, C.S., PETTIGREW, K.D., SAKURADA, O. and SHINOHARA, M. (1977). The [ $^{14}\text{C}$ ] Deoxyglucose method for the measurement of local cerebral glucose utilisation: Theory, procedure and normal values in the conscious and anesthetized albino rat. Journal of Neurochemistry, 28, 897-916.

- STOLOFF, P.H. (1970). Correcting for heterogeneity of covariance for repeated measures designs of the analysis of variance. Educ. and Psych. Measurements, 30, 909-924.
- TUKEY, J.W. (1953). The problem of multiple comparisons. Unpublished report, Princeton University, Princeton, New Jersey.
- WALLENSTEIN, S. and FLEISS, J.L. (1979). Repeated measurements analysis of variance when the correlations have a certain pattern. Psychometrika, 44, 229-233.
- WHITEHOUSE, P.J. (1985). Receptor autoradiography: applications in neuropathology. TINS, 8, 434.
- WINER, B.J. (1971). Statistical Principles in Experimental Design. (Second Edition) McGraw-Hill.
- WONG, D.F., GJEDDE, A. and WAGNER, H.N., Jr. (1986a). Quantification of neuroreceptors in the living human brain. I. Irreversible binding of ligands. Journal of Cerebral Blood Flow and Metabolism, 6, 137-146.
- WONG, D.F., GJEDDE, A., WAGNER, H.N. Jr., DANNALS, R.F., DOUGLAS, K.H., LINKS, J.M., and KUCHAR, M.J. (1986b). Quantification of neuroreceptors in the living human brain. II. Inhibition studies of receptor density and affinity. Journal of Cerebral Blood Flow and Metabolism, 6, 147-153.
- ZEEBERG, B.R. and WAGNER, H.N., Jr. (1987). Analysis of three - and four - compartment models for in vivo radioligand-neuroreceptor interaction. Bulletin of Mathematical Biology, 49, 469-489.