A Comparison of Univariate and Multivariate Statistical and Data Mining Approaches to the Behavioural and Biochemical Effects of Vestibular Loss Related to the Hippocampus

A thesis submitted in partial fufilment of the requirements of the MApplStat in Applied Statistics, Massey University, Manawatu

Paul F. Smith

2013

To my wife Cynthia and my cats, Max, Poppy and Chloe, all of whom have had to endure my journey into statistics over the last 8 years.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Stephen Haslett, for agreeing to supervise me and for his support and guidance throughout my thesis. His incisive analytical mind has been an inspiration to me. I would also like to thank Dr. Siva Ganesh, who stimulated my interest in data mining with his 4th year course on the subject.

This thesis would not have been possible without the data sets, and I thank Dr. Yiwen Zheng, Lucy Stiles, Irene Cheung and Georgina Wilson for meticulously collecting the data used in this thesis, and also to Irene for the use of her maze photos. I particularly want to thank Yiwen for her endless encouragement and for taking care of so many things while I was doing this.

I have to thank my wife, Cynthia, for carefully reading and commenting on the entire thesis draft, for asking questions to help clarify my writing and for her unwavering support in my pursuit of a statistical education.

Lastly, I want to thank the Head of our Dept. of Pharmacology and Toxicology at the University of Otago School of Medical Sciences, Associate Professor Rhonda Rosengren, for her continuous support and encouragement to pursue the MApplStat.

Table of Contents

Published papers associated with this thesis:

Zheng, Y., Cheung, I., Smith, P.F. (2012) Performance in anxiety and spatial memory tests following bilateral vestibular loss in the rat and effects of anxiolytic and anxiogenic drugs. *Behavioural Brain Research*. 235; 21-29.

Smith, P.F. (2012) Statistical analysis in pharmacology is not always BO. *Trends in Pharmacological Sciences*. 33; 565-566.

Smith, P.F. (2012) A note on the advantages of using linear mixed model analysis with maximal likelihood estimation over repeated measures ANOVAs in psychopharmacology: Comment on Clark et al. (2012). *Journal of Psychopharmacology*. 26; 1605-1607.

Zheng, Y., Wilson, G., Stiles, L., Smith, P.F. (2013) Glutamate receptor subunit and calmodulin kinase II expression in the rat hippocampus, with and without T maze experience, following bilateral vestibular deafferentation. *PLoS One.* 8(2); e54527. doi:10.1371/journal.pone.0054527, pp. 1-10.

Smith, P.F., Haslett, S.J., Zheng, Y. (2013) A multivariate statistical and data mining analysis of spatial memory-related behavior following bilateral vestibular deafferentation in the rat. *Behavioural Brain Research.* 246; 15-23.

List of Figures

List of Tables

List of Abbreviations

A3:	3rd avoidance latency in the ETM
AIC:	Akaike's Information Criterion
AMPA:	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate
ANCOVA:	analysis of covariance
ANOVA:	analysis of variance
AR(1):	autoregressive order 1
ARMA:	autoregressive moving average
BIC:	Bayesian Information Criterion
Bus:	buspirone
BVD:	bilateral vestibular deafferentation
CA:	cluster analysis
CaMKIIα:	calmodulin kinase IIα
CE:	cerebellum
CI:	confidence interval
DG:	dentate gyrus
Dist:	distance travelled in the OFM
E3:	3rd escape latency in the ETM
EPM:	elevated plus maze
Epmdur:	duration of open arm entries in the EPM
Epmdist:	distance travelled in the EPM
Epmfreq:	frequency of open arm entries in the EPM
ETM:	elevated T maze
FG-7142:	N-methyl-β-carboline-3-carboxamide
Ln IO:	ln of ratio of time spent in the inner to the outer zone of the OFM
Ln percent:	ln of percent correct in the STM
i.p:	intraperitoneal
LDA:	linear discriminant analysis
LDF:	linear discriminant function
LMM:	linear mixed model
LOO:	leave one out
MANOVA:	multivariate analysis of variance
MCAR:	missing completely at random
MAR:	missing at random
MLE:	maximum likelihood estimation

MLR:	multiple linear regression
MSE:	mean square error
MVA:	missing values analysis
NMDA:	N-methyl-D-aspartate
OFM:	open field maze
OOB:	out of bag
PCA:	principal component analysis
pCaMKIIα:	phosphorylated calmodulin kinase IIα
QDA:	quadratic discriminant analysis
REML:	restricted maximal likelihood estimation
RF:	random forest
RFR:	random forest regression
ROC:	receptor operating characteristic
RSE:	residual mean square error
s.c:	subcutaneous
Sdur:	duration of supported rearing
Sfreq:	frequency of supported rearing
SN:	spontaneous nystagmus
SSE:	sum of squares for the error
SST:	sum of squares for the treatments
STM:	spatial T maze
SVM:	support vector machine
Udur:	duration of unsupported rearing
Ufreq:	frequency of unsupported rearing
UVD:	unilateral vestibular deafferentation
VIF:	variance inflation factor
VNC:	vestibular nucleus complex
VOR:	vestibulo-ocular reflex
VSR:	vestibulo-spinal reflex

Abstract

Vestibular dysfunction is associated with a complex syndrome of cognitive and anxiety disorders. However, most studies have used simple univariate analyses of the effects of vestibular loss on behaviour and brain function. In this thesis, univariate statistical, and multivariate statistical and data mining approaches, to the behavioural and neurochemical effects of bilateral vestibular deafferentation (BVD), were compared. Using linear mixed model analyses, including repeated measures analyses of variance and analyses with the covariance structure of the repeated measures specified, rats with BVD were found to exhibit increased locomotor activity, reduced rearing and reduced thigmotaxis. By contrast, there were no significant differences between BVD and sham control animals in the elevated plus maze and the BVD animals exhibited a longer escape latency in the elevated T maze, with no change in avoidance latency. In the spatial T maze, the BVD animals demonstrated a significant decrease in accuracy compared to the sham control animals. Using linear discriminant analysis, cluster analysis, random forest classification and support vector machines, BVD animals could be distinguished from sham controls by their behavioural syndrome. Using multiple linear regression and random forest regression, the best predictors of performance in the spatial T maze were whether the animals had received a BVD or sham lesion, and the duration of rearing. In the neurochemical data set, the expression of 5-7 glutamate receptor subunits was measured in 3 different subregions of the rat hippocampus, at various times following BVD, using western blotting. In the 6 month group, half of the animals underwent training in a T-maze. Using multivariate analyses of variance, there was no significant effect of surgery for any hippocampal subregion. Linear discriminant analysis could not determine a linear discriminant function that could separate BVD from sham control animals. A random forest classification analysis was also unsuccessful in this respect. However, for the 6 month data set, T maze training had a significant effect independently of surgery. The results of these experiments suggest that BVD results in profound spatial memory deficits that are not associated with large changes in the expression of glutamate receptors in the hippocampus. The results of the multivariate statistical and data mining analyses, applied to both the

behavioural and neurochemical data sets, suggested that research in this field of neuroscience would benefit from analysing multiple variables in relation to one another, rather than simply conducting univariate analyses. Since the different behavioural and neurochemical variables do interact with one another, it is important to determine the nature of these interactions in the analyses conducted. However, this will require researchers to design experiments in which multiple variables can be measured under the one set of conditions.

1. Introduction

1.1 The vestibular system

The balance organs in the inner ear (the 'vestibular system'; Figure 1) encode head movement (strictly speaking, head acceleration) and generate rapid eye movement ('vestibulo-ocular') and postural ('vestibulo-spinal') reflexes that are important for maintaining clear vision and balance (Cullen, 2012). During unintentional head movement, which can include movement as small as that produced by the pulse beat, the image of the visual world on the retina shifts sufficiently, that without the compensatory eye movements generated by the vestibulo-ocular reflexes (VORs), the visual world would appear to smear, a condition known as 'oscillopsia'. Similarly, the vestibulo-spinal reflexes (VSRs) generate compensatory postural adjustments during head movement in order to maintain balance (Figure 2). Humans with deficient vestibular function experience difficulty maintaining balance and consequently cannot walk properly ('ataxia') (Cullen, 2012).

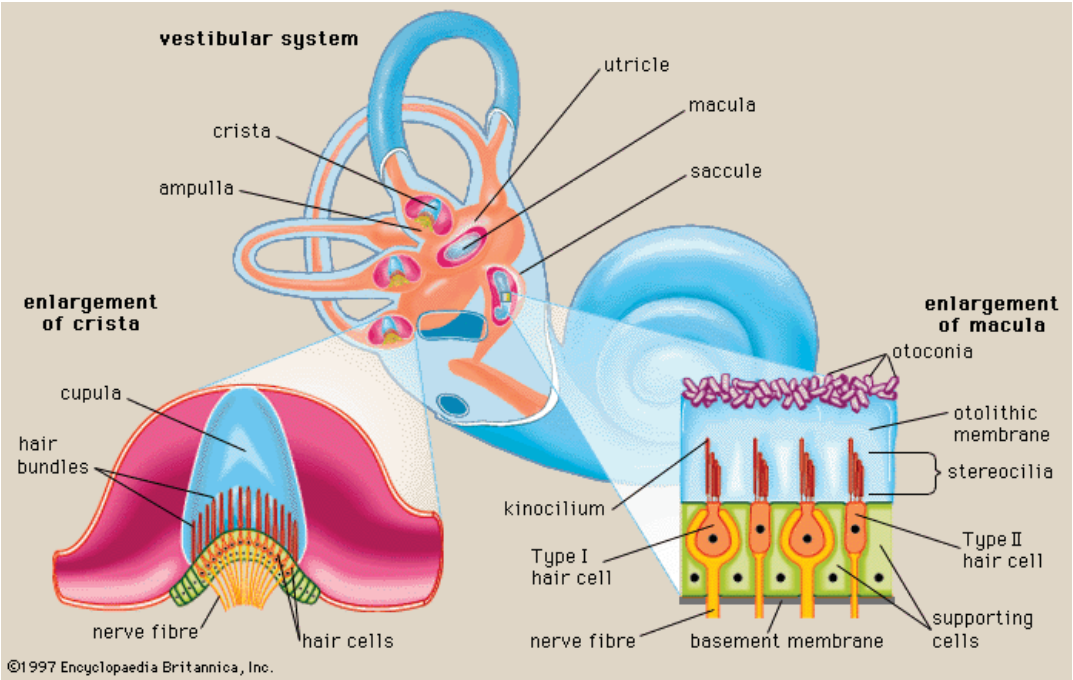


Figure 1: The peripheral vestibular system in the inner ear. The inset shows the hair cells that transduce movement of the head. From the *Encyclopaedia Britannica*, 1997.

Vestibular disorders are much more common than most people realise. Dizziness is one of the most common reasons for patients visiting either their general practitioner or a hospital emergency department, and the two most common causes are cardiovascular or vestibular dysfunction. Recent epidemiological studies suggest that vestibular disorders occur in more than 35% of adults aged 40 or older (Agrawal et al., 2009). Between the ages of 60 and 69, the prevalence increases to almost 50% and between 70 and 79, it is 69%. The odds of vestibular dysfunction are also 70% higher among people with diabetes mellitis (Agrawal et al., 2009).

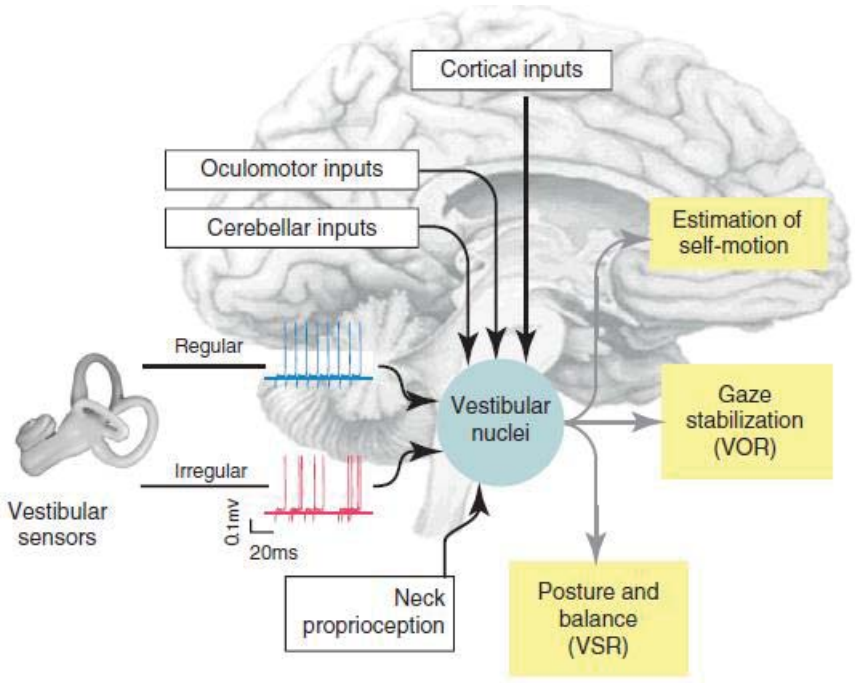


Figure 2: Schematic diagram illustrating the connections between vestibular sensation in the inner ear, the generation of eye movement (gaze stabilization, VOR), postural reflexes (posture and balance, VSR) and the estimation of self-motion. From Cullen, K.E. The vestibular system: Multimodal integration and encoding of self-motion for motor control. *Trends in Neuroscience*. 35 (2012) 185-196.

1.2 Vestibular contributions to cognition and emotion

Although the immediate and most obvious effects of poor vestibular function are oscillopsia and ataxia, vestibular dysfunction results in a complex

neurological syndrome characterized not only by reflex deficits, but also by spatial memory deficits, autonomic and anxiety disorders (Brandt et al., 2005; see Smith et al., 2010 for a recent review).

Figure 3: Balance control. Copyright NeuroCom, a division of Natus 2012.

It is clear that the loss of vestibular information results in the abnormal function of many brain regions, including the hippocampus, which is critical for remembering places in the environment ('spatial memory'). The hippocampus has been demonstrated to atrophy under conditions of complete vestibular loss (Brandt et al., 2005). However, exactly why this happens is unclear. The vestibular system encodes angular and linear acceleration of the head in 3 dimensions and, in addition to generating the vestibulo-ocular and vestibulo-spinal reflexes, provides the brain with information about self-motion that can be used to navigate through the environment and form memories for places in it (Smith et al., 2010; Cullen, 2012; Figures 3 and 4). However, the severity of the

symptoms of vestibular dysfunction suggests that there may be a critical dependence of the brain and body on vestibular input. The most primitive part of the vestibular system – the otoliths that transduce linear acceleration, including linear acceleration by gravity – is estimated to be more than 500 million years old and exists in primitive species such as sea squirts (Smith, 2010). These sensory organs evolved to provide information about gravitational vertical, before any other sense had evolved, and during development in mammals, the vestibular system is fully functional before the visual or auditory systems. Therefore, it is highly likely that human physiology has developed a special dependence upon the otolithic part of the vestibular system (the utricle and saccule in mammals) (Smith et al., 2010).

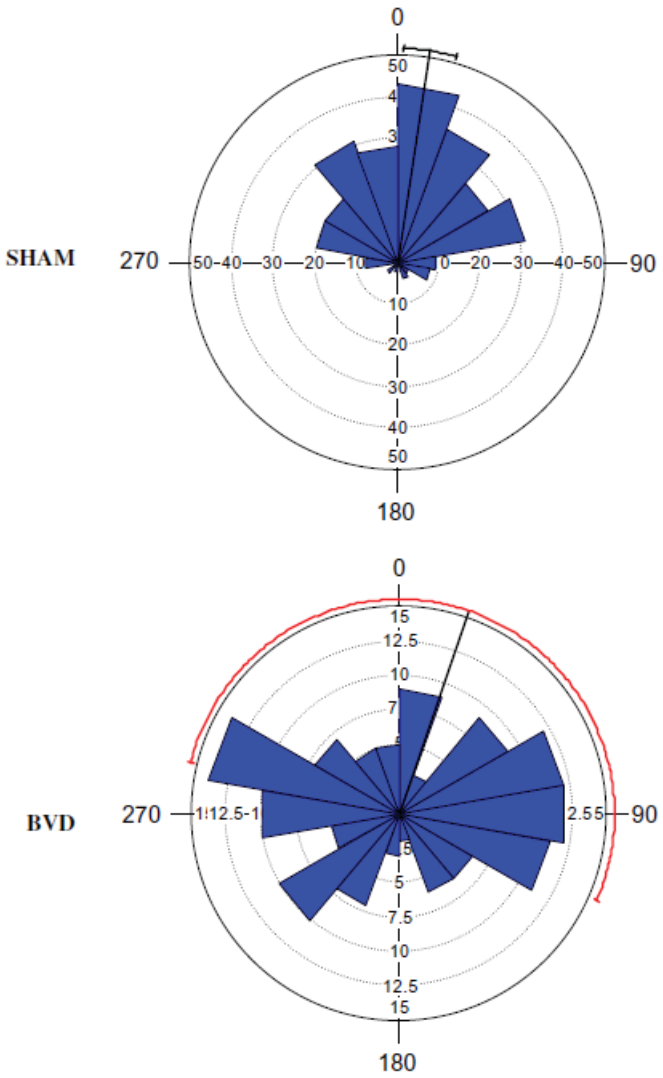


Figure 4: Rose diagram showing the initial heading angles of sham and bilateral vestibular deafferentation (BVD) animals in a foraging task in

darkness in which animals have to learn to remember where their home base is and navigate back to it. The mean vector is indicated by the black line and the 95% confidence interval (C.I.) for the mean is indicated by the line extending either side. The 95% C.I. values for the BVD animals were unreliable due to the low concentration of vectors, hence the red line. The inner circles (dotted line) indicate the number of observations for the given vectors (blue triangles). From Baek, J.H., Zheng, Y., Darlington, C.L., Smith, P.F. Evidence that spatial memory deficits following bilateral vestibular deafferentation in rats are probably permanent. *Neurobiology of Learning and Memory*. 94 (2010) 402-413.

Vestibular dysfunction in humans is associated with a variety of cognitive deficits, including spatial memory and attention deficits (Black et al., 2004; Redfern et al., 2004; Brandt et al., 2005; Jauregui-Renaud et al., 2008a; Jauregui-Renaud et al., 2008b; see Smith et al., 2005 for a review). In addition, anxiety disorders, including panic attacks and phobias, are common (Balaban and Thayer, 2001; Furman and Jacob, 2001; Balaban, 2002; Staab, 2006). While it is possible that anxiety is a direct consequence of vestibular dysfunction, it has also been reported that anxiety disorders can cause dizziness (Bolmont et al., 2002; Best et al., 2006; Furman et al., 2006) and antidepressants such as selective serotonin reuptake inhibitors (SSRIs) have been reported to relieve dizziness associated with psychiatric symptoms (Staab et al., 2002; Simon et al., 2005). Therefore, it is possible that emotional disorders arise indirectly from cognitive impairment. However, Halberstadt and Balaban (2006) have reported that the same neurons in the dorsal raphe nucleus, that release the neurotransmitter, serotonin, send projections into the amygdala, an area of the brain concerned with fear and panic, as well as the brainstem vestibular nucleus. This finding suggests that changes in emotional tone may directly influence the vestibular system.

The relatively few studies of anxiety-related behaviour in animals with vestibular damage have yielded contradictory results. Paradoxically, Lindemann et al. (2007) reported that *ci2* mutant rats, which have cochlear and vestibular deficits, exhibited *less* anxiety-related behaviour compared to controls, manifested as reduced exploration of the outer areas of an open field maze (reduced 'thigmotaxis'), and an increased amount of time spent in the open arms of the

elevated plus maze and in the light compartment in the light-dark exploration test. Zheng et al. observed similar results in rats with complete bilateral peripheral vestibular lesions (surgical bilateral vestibular deafferentation, BVD) (Zheng et al., 2008; Goddard et al., 2008; Neo et al., 2012). By contrast, other animal studies have suggested that vestibular deficiency is associated with *increased* anxiety. Avni et al. (2009) reported that mutant headbanger (*Hdb*) mice, which also exhibit vestibular deficits, demonstrated increased signs of anxiety, manifested as an increased number of stretch-attends (i.e., orientation with the top of the back lower than the ears, suggesting 'risk assessment'). Shefer et al. (2010) also reported that *Hdb* mice exhibited increased signs of anxiety in the open field and elevated plus maze tests. Machado et al. (2012) have recently reported evidence for increased anxiety in the black and white box test in rats that received bilateral chemical vestibular lesions. Regarding the possibility that anxiety can lead to vestibular dysfunction, Lepicard et al. (2000) reported that the highly anxious mouse strain, *BALB/cByJ*, performed poorly in the rotating beam balance test compared to the non-anxious strain, *C57BL/6J*, an effect that could be reduced with the administration of the anxiety-reducing ('anxiolytic') drug, diazepam. Using the same anxious strain, Venault et al. (2001) reported that balance was improved by the administration of the selective serotonin reuptake inhibitors (SSRIs), fluoxetine and paroxetine.

BVD has been reported to cause a variety of functional changes in the hippocampus, an area of the brain concerned with emotion as well as spatial memory (Figures 5 and 6). Brain cells (neurons) that normally increase their electrical activity in response to particular places in the environment where food has been obtained (known as 'place cells'), exhibit disordered responses following peripheral vestibular lesions (Stackman et al., 2002; Russell et al., 2003a; Figure 6). Furthermore, a brain electroencephalographic (EEG) rhythm, known as 'theta rhythm', which is thought to be important for coordinating the responses of place cells in the hippocampus, becomes severely disordered following BVD (Russell et al., 2006; Neo et al., 2012; Tai et al., 2012).

While it is clear that functional changes occur in the hippocampus that might explain spatial memory and emotional impairment following bilateral vestibular loss, the neurochemical bases of these changes remain unknown. Relatively few data are available on the neurochemical changes that occur in the

hippocampus following BVD, in particular those relating to synaptic transmission involving the neurotransmitter, glutamate, that might be important for spatial memory. Previous studies involving unilateral vestibular deafferentation (UVD) in rats, which elicits a severe imbalance in vestibulo-ocular and vestibulo-spinal reflexes that gradually abates over time, showed that the expression of the NR1 and NR2A subunits of the N-methyl-D-aspartate (NMDA) subtype of glutamate receptor, decreased in the ipsilateral CA2/3 region of the hippocampus at 2 weeks following UVD, while the NR2A subunit was also reduced in the contralateral CA2/3 region at the same time point (Liu et al., 2003). On the other hand, the expression of the NR2A subunit was increased in the CA1 region at 10 hs following UVD (Liu et al., 2003). This study did not investigate the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) glutamate receptor subunits, GluR1-GluR4, and the longest post-operative time point was 2 weeks. The only study to date to investigate glutamate receptors in the hippocampus following BVD, measured NMDA receptor density and affinity using receptor autoradiography. In this study, Besnard et al. (2012) used a sequential UVD procedure, involving intratympanic sodium arsanilate injections (i.e., one ear, followed several weeks later by the other ear), and observed a significant increase in the number of NMDA receptors (increased B_{max}) and a decrease in affinity (increased K_d) in the hippocampus. This sequential UVD procedure has the advantage of relevance to paroxysmal vestibular disorders in humans in which the right vestibular labyrinth malfunctions, and then the left, or vice versa, e.g. some types of Meniere's disease (Besnard et al., 2012). Nonetheless, because sequential UVDs which ultimately result in BVD, produce vestibular reflex imbalances, first ipsilateral to the initial UVD, and then ipsilateral to the second one, the behavioural syndrome is different from that produced by a simultaneous BVD procedure in which both peripheral vestibular systems are lesioned in rapid succession under anaesthesia.

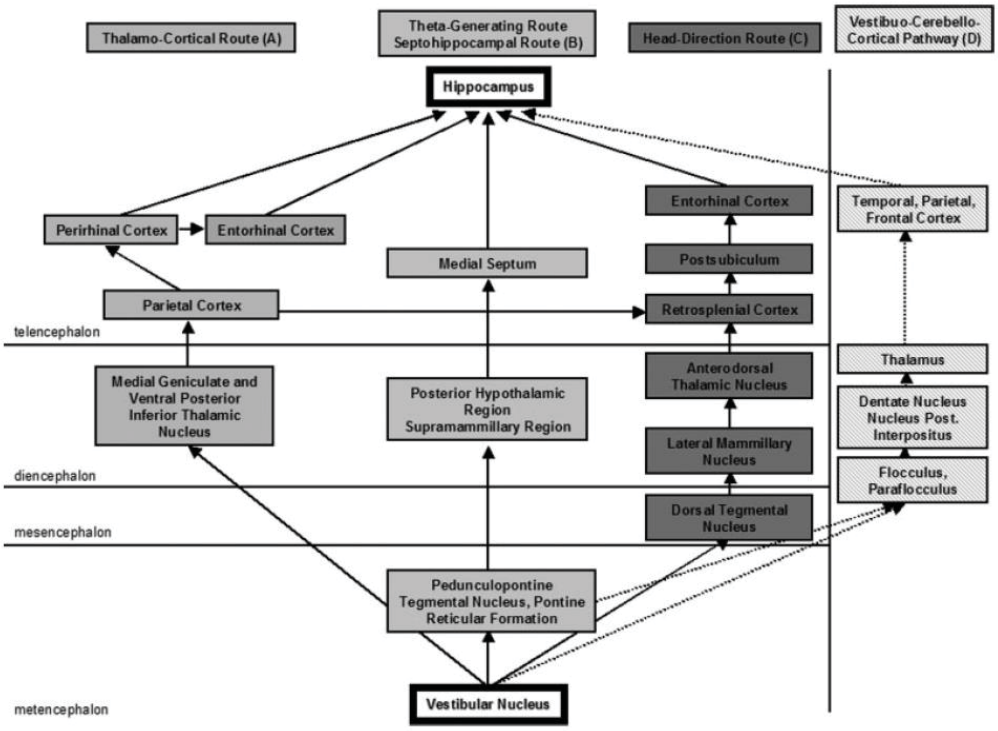


Figure 5: Putative pathways from the brainstem vestibular nucleus to the hippocampus. From Hüfner. K., Hamilton, D.A., Kalla, R., Stephan, T., Glasauer, S., Ma, J., Brüning, R., Markowitsch, H.J., Labudda, K., Schichor, C., Strupp, M., Brandt, T. Spatial memory and hippocampal volume in humans with unilateral vestibular deafferentation. *Hippocampus.* 17 (2007) 471-485.

1.3. Data sets used in this thesis

The results of the behavioural studies present a very complex picture of the cognitive and emotional state of animals with vestibular dysfunction (see Kalueff et al., 2008 for a review), which is not easy to relate to the clinical evidence from humans. Because of the possibility that the performance of rats with vestibular deficits in cognitive tests is affected by their emotional state (Kalueff et al., 2007; Kalueff et al., 2008), it was necessary to re-investigate the effects of BVD on the performance of rats in both cognitive and anxiety tests under the same conditions. Furthermore, the non-sedating, anxiolytic drug, buspirone, and the anxiety-inducing ('anxiogenic') drug, FG-7142, were used in an attempt to alter anxiety levels in these rats. The first data set in this thesis (the behavioural study) relates to this experiment.

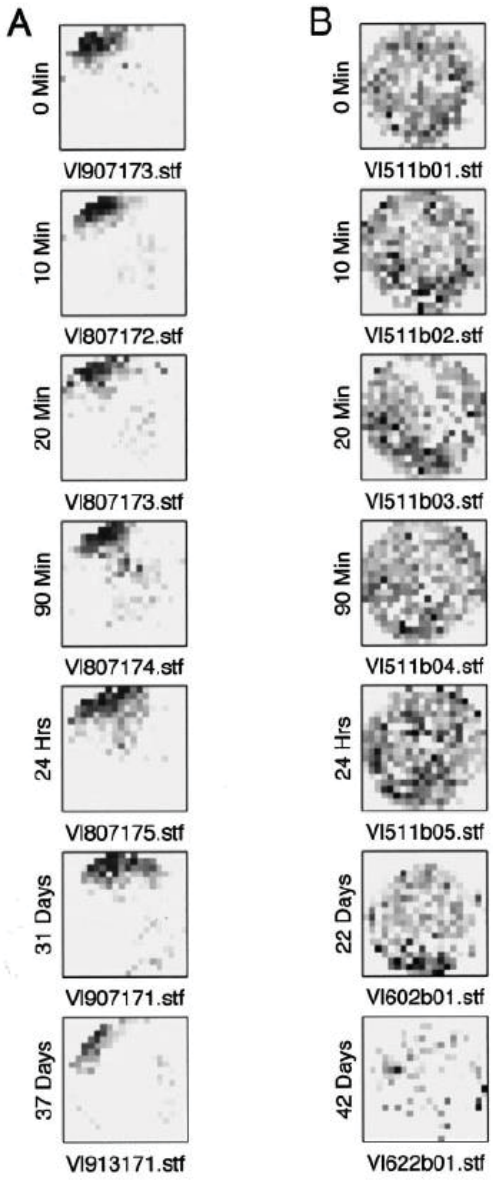


Figure 6: Example of place cell dysfunction in the rat hippocampus following bilateral vestibular lesions. Each plot shows a firing rate map for a single neuron, in which the number of pixels is proportional to the cell's firing rate, as the animal moves through the spatial environment. The neuron in A is from a sham-lesioned animal; B is from an animal with bilateral vestibular lesions. Recordings were made in alert animals up to 37-42 days post-op. Note that in B, the neuron does not have a clear spatial preference for firing. From Russell N.A., Horii A, Smith P.F., Darlington C.L., Bilkey, D.K. Long-

term effects of permanent vestibular lesions on hippocampal spatial firing. *Journal of Neuroscience* 23 (2003) 6490-6498.

Because of the dearth of neurochemical data relating to synaptic transmission involving glutamate in the hippocampus following vestibular loss, the aim of the biochemical study, the second data set with which this thesis is concerned, was to investigate the expression of several glutamate receptor subunits and the protein kinase, calmodulin kinase IIα (CaMKIIα), in the CA1, CA2/3 and dentate gyrus subregions of the hippocampus, at various time points following BVD, using western blotting. For the NMDA receptor, the NR1 subunit was analysed because it is necessary for NMDA receptor function, binding the co-agonist, glycine, while the NR2 subunit binds glutamate (Pertralia et al., 1994). The NR2A and NR2B subunits were measured because they have an important impact on the receptor's channel function (i.e., channel conductance, ligand affinity and sensitivity to Mg^{2+}) (Kutsuwada et al., 1992; Ishii et al., 1993; Krupp et al., 1998; Williams et al., 1994). For the AMPA receptor, all four GluR subunits were measured, GluR1 and GluR2 being the most commonly expressed in the hippocampus, with lower levels of GluR3 and GluR4 (Craig et al., 1993; Geiger et al., 1995; Wentold et al., 1996). There is a close relationship between CaMKIIα and NMDA and AMPA receptor subunits. CaMKIIα binds to the NR1 and NR2B subunits, and phosphorylates AMPA receptors, thereby altering their channel conductances (Leonard et al., 1999; Barria et al., 1997). Furthermore, activation of NMDA receptors increases the activation of CaMKIIα, leading to autophosphorylation (Giese et al., 1998). Therefore, in the biochemical study, CaMKIIα and phosphorylated CaMKIIα (pCaMKIIα) expression were measured in the same hippocampal subregions.

1.4 Literature and rationale for statistical approach

1.4.1 Univariate statistical methods

The analysis of research data in this field of neuroscience has employed univariate statistical methods almost exclusively. Of the nine studies that are directly relevant to the data sets analysed in this thesis, listed in Table 1, none

employed any kind of multivariate statistical analysis, even though all of them measured multiple dependent variables.

Most analyses of behavioural studies, related to cognition or anxiety, or neurochemical studies following vestibular lesions, have employed a general linear model analysis of variance (ANOVA) (see Table 1), which, for independent measures with a single factor, has the general form:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \qquad (1)$$

Where Y_{ij} = the *i*th subject's score for the dependent variable in the *j*th experimental condition; μ = a constant representing the dependent variable independent of any experimental condition; α_j = the effect of the *j*th experimental condition; and ε_{ij} = the error term, describing variation from uncontrolled sources (Doncaster and Davey, 2007). In such cases the single factor is usually surgery, with 2 levels: vestibular lesions and sham, control lesions.

A two-factor factorial model is often used where the objective is to investigate the effects of two independent factors on the dependent variable, for example, surgery and drug (Small et al., 2011):

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \qquad (2)$$

Where Y_{ijk} = the *i*th subject's score for the dependent variable in the experimental condition which is the *j*th level of Factor A, where $j = 1,\ldots,p$, and the *k*th level of Factor B, where $k = 1,\ldots,q$; μ = a constant representing the dependent variable independent of any experimental condition; α_j = the effect of the *j*th level of Factor A; β_k = the effect of the *k*th level of Factor B; $(\alpha\beta)_{jk}$ = the effect of the interaction between the *j*th level of Factor A and the *k*th level of Factor B; and ε_{ijk} = the error term (Doncaster and Davey, 2007).

Sometimes, if the change in the dependent variable over a period of time is of interest, and repeated measurements can be made in the same animals (i.e. one observation per animal per time point), a repeated measures design with a single factor may be used:

$$Y_{ij} = \mu + \pi_i + \alpha_j + (\pi\alpha)_{ij} \qquad (3)$$

Where Y_{ij} = the *i*th subject's score for the dependent variable in the *j*th experimental condition; μ = a constant representing the dependent variable independent of any experimental condition; π_i = the random effect of the *i*th subject; α_j = the effect of the *j*th experimental condition; and $(\pi\alpha)_{ij}$ = the error term (Doncaster and Davey, 2007).

One of the most commonly used designs is a factorial design including one fixed, between group, measure, for example surgery, and one repeated measure, over time, a so-called 'mixed' or 'split-plot' design (Festing, 2003):

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \pi_{i(j)} + \varepsilon_{ijk} \qquad (4)$$

Where Y_{ijk} = the *i*th subject's score for the dependent variable for the *j*th level of Factor A and the *k*th level of Factor B; μ = a constant representing the dependent variable independent of any experimental condition; α_j = the effect of the *j*th level of Factor A; β_k = the effect of the *k*th level of Factor B; $(\alpha\beta)_{jk}$ = the effect of the interaction between the *j*th level of Factor A and the *k*th level of Factor B; $\pi_{i(j)}$ = the random effect of the *i*th subject in the *j*th level of Factor A; and ε_{ijk} = the random error of the *i*th subject in the *j*th level of Factor A and the *k*th level of Factor B (Doncaster and Davey, 2007).

1.4.2 Fixed versus random effects

In the formulations of the ANOVAs described above, the general linear model has the form: data = model + error. Fixed effects involve only fixed levels of factors in the model, which are referred to as 'fixed' because the experimenter has chosen them specifically as the factors of interest and any conclusions drawn from the analysis do not extend beyond them (Rutherford, 2001; Kirk, 2013). In the case of studies in this area of neuroscience, one obvious fixed effect is surgery, where there are two levels or values: BVD versus sham surgery. Animals are randomly allocated to each condition and the researcher has specifically chosen these conditions because of the nature of the research question, i.e. whether normal vestibular function makes a difference to behaviour and hippocampal

function. By contrast, random effects involve only random factors in the model, which are referred to as ‘random’ because they are believed to be only a random sample from a population of experimental conditions. Therefore, conclusions based on their investigation are extrapolated to a wider population of experimental conditions (Rutherford, 2001; Kirk, 2013). Common examples of random effects in the context of this area are the animals themselves, which are a random sample from a population of animals.

Since factorial repeated measures ANOVAs are so commonly used in this area of neuroscience, it is usual to have a ‘mixed’ design which contains both fixed and random factors (Rutherford, 2001). The fixed factor is often surgery and the random factor, animals, over which several repeated measures are made.

1.4.3 General linear model assumptions

General linear model ANOVA designs all make the following assumptions:

1. That the experimental units (animals) have been randomly sampled from a population or from within groups (provided that the lack of independence is represented as a random effect).
2. That the measures within each sample are independent and have uncorrelated model errors.
3. That the variances across the samples are homogeneous.
4. That the model errors are normally distributed (Rutherford, 2001; Doncaster and Davey, 2007; Kirk, 2013).

Study	Statistical Analysis
Behavioural	
Lindemann et al. (2007)	One factor (i.e. vestibular function, with 2 levels, normal and abnormal) non-parametric Kruskal Wallis ANOVAs followed by Mann Whitney U tests. Assumption testing not reported.
Zheng et al. (2008)	Two factor repeated measures ANOVAs followed by Tukey's post-hoc tests. The between group factor was surgery, with 2 levels, bilateral vestibular deafferentation (BVD) and sham lesions; the repeated measure was time. Assumption testing performed.
Goddard et al. (2008)	Two factor repeated measures ANOVAs followed by Tukey's post-hoc tests. The between group factor was surgery, with 2 levels, BVD and sham lesions; the repeated measure was time. Assumption testing performed.
Avni et al. (2009)	t tests. Assumption testing performed.
Shefer et al., (2010)	Two way ANOVAs. No assumption testing reported and inadequate description of statistical analysis, i.e. no statistical analysis section in the Methods description. It was not clear whether the ANOVAs included a repeated measure.
Neo et al. (2012)	Two way repeated measures ANOVAs. The between group factor was surgery, with 2 levels, BVD and sham lesions; the repeated measure was time. Assumption testing performed.
Machado et al. (2012)	Two way repeated measures ANOVAs followed by Student Newman Keuls post-hoc tests. The between group factor was surgery, with 2 levels, BVD and sham lesions; the repeated measure was time. Also one way non-parametric Kruskal Wallis ANOVAs. Assumption testing not reported.

Neurochemical	
Liu et al. (2003)	One factor (i.e. surgery, with 2 levels, BVD and sham lesions) ANOVAs followed by Student Newman Keuls post-hoc tests. Assumption testing not reported.
Besnard et al. (2012)	Two way repeated measures ANOVAs followed by Fisher's Least Significant Difference post-hoc tests. The between group factor was surgery, with 2 levels, BVD and sham lesions; the repeated measure was time. Assumption testing not reported.

Table 1: Statistical analyses used by the relevant behavioural and neurochemical animal studies. Seven out of nine used parametric ANOVAs.

Table 1 shows that of nine studies, only four reported testing the assumptions of ANOVA or using any form of transformation to correct for the violation of these assumptions (Zheng et al., 2008; Goddard et al., 2008; Avni et al., 2009; Neo et al., 2012). Of these four studies, in two (Avni et al., 2009; Neo et al., 2012) the assumption testing was limited to testing the normality assumption. Only one out of nine (Goddard et al., 2008) reported testing all of the assumptions of normality, homogeneity of variance and sphericity (see below). The sample sizes varied from 5 per group (Liu et al., 2003; Lindemann et al., 2007) to 30 per group (Shefer et al., 2010). However, with four exceptions (Zheng et al., 2008; Goddard et al., 2008; Avni et al., 2009; Shefer et al., 2010), the sample sizes were usually less than 15 per group and in six out of nine cases, the sample sizes were not equal. For example, in the study by Lindemann et al. (2007), the sample sizes per group were: 5, 10, 13, 8 and 7. In most cases (6/9), post-hoc testing was used following a significant ANOVA, although the specific tests employed varied from Tukey's tests, to Student Newman Keuls tests, to Fisher's Least Significant Difference tests, in the case of parametric ANOVAs, and the pairwise comparisons were often non-orthogonal and large in number.

1.4.3.1 Normality assumption

It is very common in neuroscience research (see Table 1) to assume that the normality assumption for general linear model t tests and ANOVAs is fulfilled without formally testing it using some form of goodness of fit test such as the Kolmogorov-Smirnov, Shapiro-Wilk or Anderson-Darling tests (Rutherford, 2001; Gamst et al., 2008). This is presumably due to the view that the central limit theorem will protect ANOVAs against the moderate violation of the assumption of normality "when samples sizes are reasonably large and are equal" (Winer et al., 1991, p.101). Unfortunately, the interpretation of "reasonably large" is problematic. Snedecor and Cochran (1989) suggest that while for some populations the sampling distribution of the mean may be normal with sample sizes of 4 or 5, in other cases it may need to be more than 100. The distributions of some variables are inherently non-normal. For example, frequency data have positive integer values in which random variation increases as the mean increases; for example, Poisson distributions (Doncaster and Davey, 2007). On the other hand, Keppel and Wickens (2004) have argued that the normality assumption can be ignored once the sample sizes reach approximately 12 (see also, Rutherford, 2001). However, the symmetry of the distribution is also very important (Winer et al., 1991; Kirk, 2013). In the data sets from the literature that are relevant to this thesis, and in neuroscience more generally, there are two problems with this assumption, even if it is taken at face value. One is that the sample sizes are nearly always less than 12 and are often unequal (e.g., Liu et al., 2003; Lindemann et al., 2007; Besnard et al., 2012; Neo et al., 2012; Machado et al., 2012), which makes the potential violation of the assumption more likely (Motulsky, 1995). The second is that the normality assumption is rarely tested and is assumed to be fulfilled unless there is clear evidence to the contrary (e.g., from a statistical computer program that provides this information automatically when an ANOVA is performed) (e.g., Liu et al., 2003; Lindemann et al., 2007; Shefer et al., 2010; Besnard et al., 2012; Machado et al., 2012). Obviously, if the assumption is not tested, then solutions to the violation of the normality assumption, such as data transformation, cannot be undertaken (Gamst et al., 2008). One issue is that, if the sample size is small, for example, less than or equal to 8 (e.g., Liu et al., 2003; Lindemann et al., 2007; Besnard et al., 2012; Machado et al., 2012), there may not be sufficient information in order to judge whether the data are normally

distributed or not. In this case, some researchers will choose to use a non-parametric statistical analysis (e.g., Avni et al., 2009), although this can be difficult for designs that are not amenable to a one or two-way ANOVA (see Conover and Iman, 1981 for a review).

As an example of this issue in this area of neuroscience, the following is an analysis of normality of CaMKIIα in one of the neurochemical data sets used in this thesis. Both Kolmogorov-Smirnov and Shapiro-Wilk tests indicated that, for the CA1, CA2/3 and dentate gyrus (DG) regions of the hippocampus, the null hypothesis that the residuals are normally distributed, was rejected (Table 2 and Figure 7).

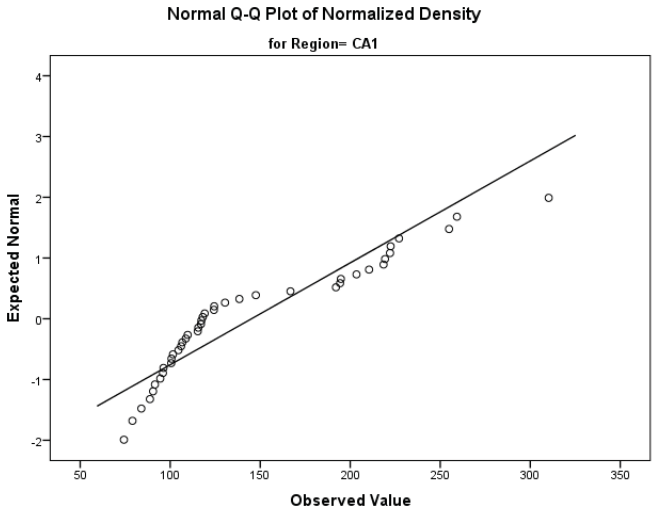
Tests of Normality[a]

	Region	Kolmogorov-Smirnov[b]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Normalized Density	CA1	.231	42	.000	.864	42	.000
	CA23	.249	42	.000	.862	42	.000
	DG	.272	42	.000	.808	42	.000

a. There are no valid cases for Normalized Density when Region = .000. Statistics cannot be computed for this level.

b. Lilliefors Significance Correction

Table 2: Example of Kolmogorov-Smirnov and Shapiro-Wilk tests of normality for hippocampal data from this thesis. Note that in all cases the null hypothesis that the data come from a normal distribution, was rejected, at $P \leq 0.0005$.



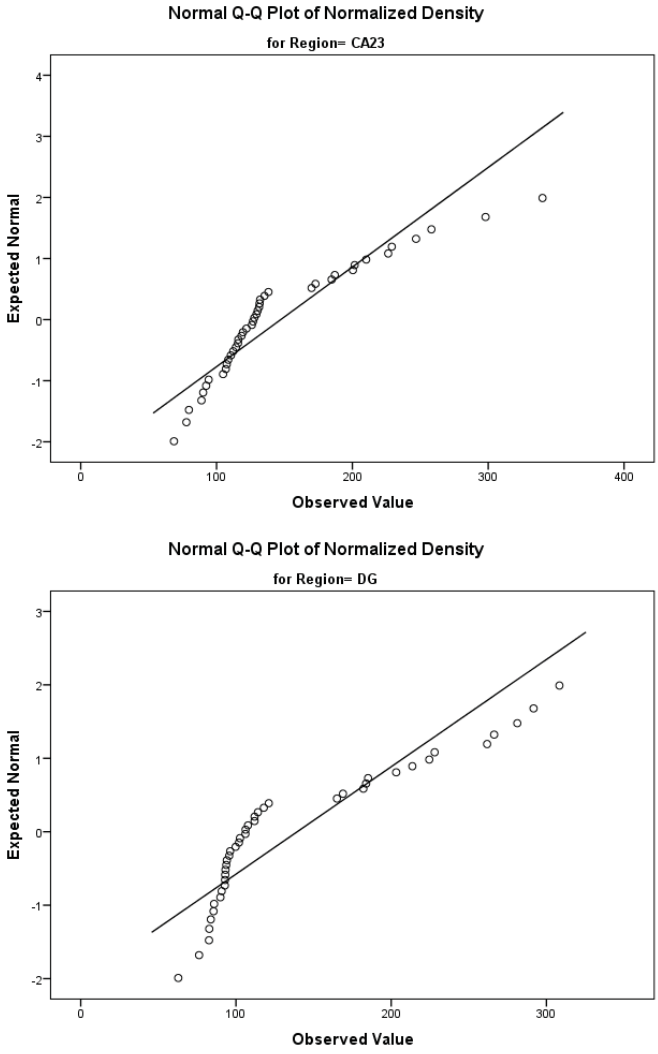


Figure 7: Normal quantile-quantile (Q-Q) plots for the same data, showing the deviation from normality (i.e., the straight line) in the case of all 3 subregions of the hippocampus.

A natural log transformation resolved the problem only for the CA2/3 data for the Shapiro-Wilk test, but not for the Kolmogorov-Smirnov test (Table 3 and Figure 8). The central limit theorem might be expected to protect against any violation of the assumption of normality. However, it does not always apply to random effects, particularly in the case of small sample sizes.

Tests of Normality[a]

	Region	Kolmogorov-Smirnov[b]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Indensity	CA1	.178	42	.002	.920	42	.006
	CA23	.182	42	.001	.953	42	.081
	DG	.210	42	.000	.879	42	.000

a. There are no valid cases for Indensity when Region = .000. Statistics cannot be computed for this level.

b. Lilliefors Significance Correction

Table 3: Normality tests of the data following ln transformation.

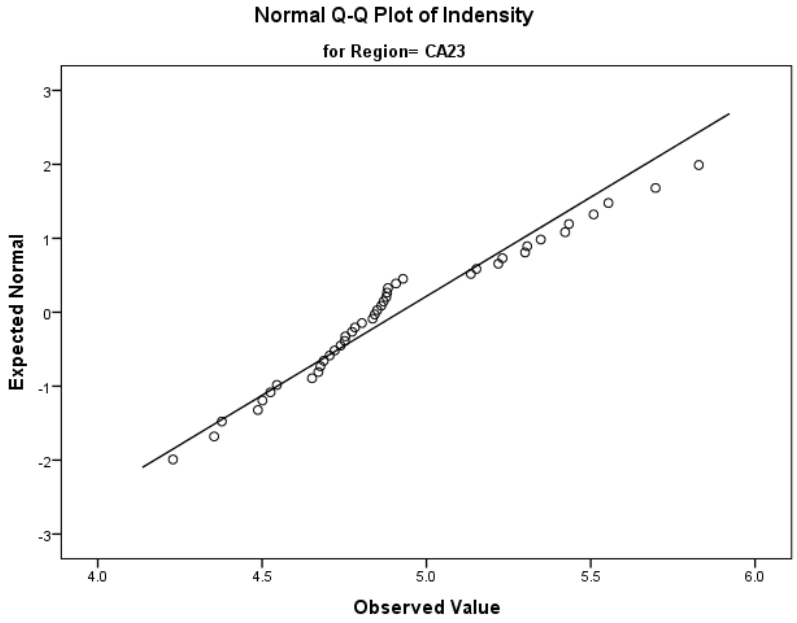


Figure 8: A natural log transformation of the CA2/3 data partially resolves the violation of the normality assumption.

What are the potential consequences of using ANOVA if the normality assumption is violated? Rutherford (2001) suggests that it will affect both the type I error rate and the power of the F test.

However, using bootstrapping in Minitab 16, when 1000 samples of n = 8 were taken with replacement from the CA1 data, the sampling distribution of the mean was normal according to the Anderson-Darling test, suggesting that the central limit theorem held for this sample size (Figures 9 and 10).

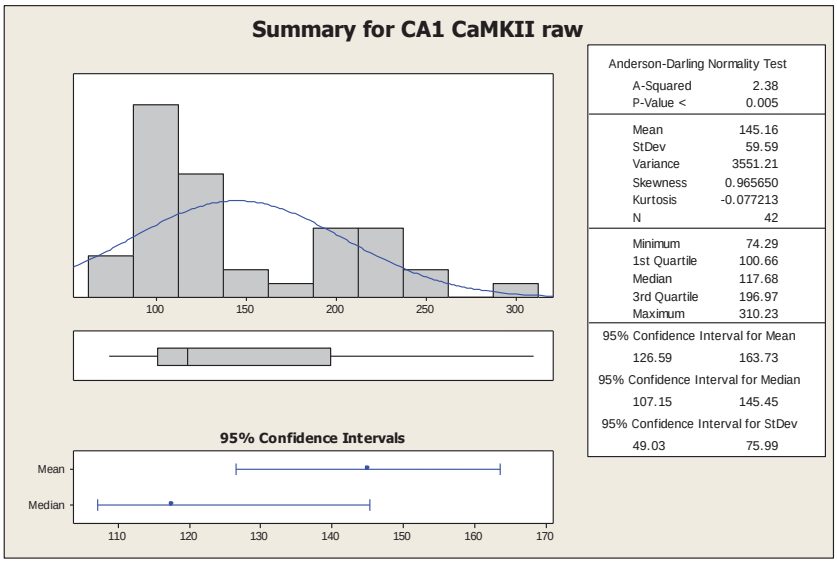


Figure 9: The raw data from CA1 in Figure 7 are not normally distributed (Anderson-Darling test, P ≤ 0.005).

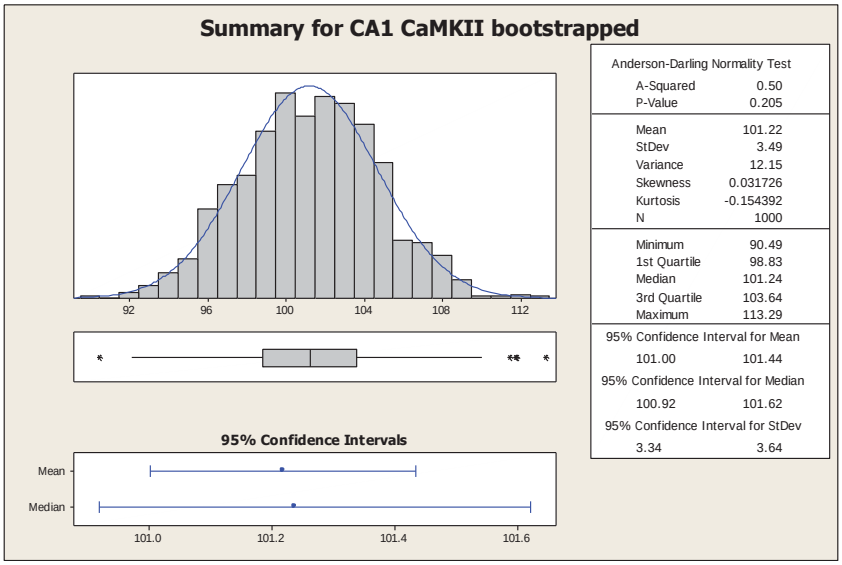


Figure 10: Bootstrapping was used to take 1000 samples of n = 8 from the CA1 data shown in Figure 7. The histogram shows that the bootstrapped data for the sampling distribution of the mean are normally distributed (Anderson-Darling test, P = 0.20).

1.4.3.2 Homogeneity of variance assumption

Another assumption of the general linear model is that the variance associated with experimental error in the treatment populations to be compared, is

homogeneous (the ‘homogeneity of variance’ or ‘homoscedasticity’ assumption) (Winer et al., 1991). Box (1954a; Box, 1954b) demonstrated that the F test is robust against ‘moderate’ violations of this assumption, provided that the sample sizes are equal. However, Box (1954a) showed that small changes in the ratio of the variances between treatment groups can alter the significance level of the F test, in some cases increasing the type I error rate. Wilcox (1987) has suggested that when the homogeneity of variance assumption is violated, the conventional F test should never be used. For this reason, some authors recommend using a more conservative type I error rate (Keppel and Wickens, 2004; Gamst et al., 2008). Winer et al. (1991) suggest that the solution is for experimenters to aim to use large and equal sample sizes and then to use the Box approximation to the F test in situations where this is not possible. The same data transformations that can be used to achieve normality of the data distribution, for example, log or square root transformations, often also result in homogeneity of variance (Winer et al., 1991; Gamst et al., 2008). However, as with the normality assumption, most studies in this area of neuroscience do not report testing the homogeneity of variance assumption using tests such as Levene’s or Bartlett’s tests, before proceeding with an ANOVA (see Table 4).

As an example of this issue, Table 4 below illustrates the results of a Levene’s test of homogeneity of variance on the pre-drug locomotor velocity data that are used later in this thesis. The results show that for this particular small data set (n = 10 sham and n = 8 BVD animals), the null hypothesis that the two variances are equal, was retained. Using the full data set including the drug conditions (none of which had any significant effect on locomotor velocity and therefore the data have been pooled for this purpose), the same result was obtained (Table 5).

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Velocity

F	df1	df2	Sig.
3.901	1	16	.066

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + surgery

Table 4: Levene's test of homogeneity of variance for the pre-drug locomotor velocity data set.

Levene's Test of Equality of Error Variances[a]

Dependent Variable:Velocity

F	df1	df2	Sig.
2.970	1	88	.088

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Surgery

Table 5: Levene's test of homogeneity of variance for the complete locomotor velocity data set.

1.4.3.3 Compound symmetry or sphericity assumption

It is common for behavioural studies in this area of neuroscience to employ repeated measures designs, in which some behavioural variable is measured repeatedly over time (Zheng et al., 2008; Goddard et al., 2008; Shefer et al., 2010; Neo et al., 2012; see Table 1). A two-factor ANOVA with repeated measures on one factor, has the form shown in Equation (5).

One of the assumptions of the F test is that the variance-covariance matrices for the dependent variable for the different treatment groups are equal and symmetrical. This is known as the assumption of 'compound symmetry', which is a special case of 'circularity' (Winer et al., 1991). A less restrictive assumption is 'sphericity', which requires that the variances of the differences between the values of the dependent variable are equal for all pairs of treatments (Quinn and Keough, 2002). Formally stated, for the covariance matrix, Σ, to be spherical:

$$M^{*\prime}\,\Sigma M^{*} = \lambda I \quad (6)$$

Where Σ = the population covariance matrix, M^* = an orthonormal covariance matrix (i.e. a matrix with rows that are a series of orthogonal comparisons), $M^{*\prime}$ = the transpose of M^*, I = an identity matrix, and λ = a scalar number, where $\lambda > 1$ (Kirk, 2013, p. 308). "If Σ is circular, Σ will have all zeros off the diagonal (i.e. have zero covariances) and a constant λ on the main diagonal (i.e., homogeneous variances equal to λ)" (Winer et al., 1991, p. 513). This means that the covariances for all pairs of treatments will be zero and the variances will be equal. Contrary to these assumptions, it is very common in behavioural studies for the data within subjects to be correlated across time, i.e., for the variance for the dependent variable to change systematically with repeated measures over time, and for this to occur with specific changes in the means for the repeated measure. A common example is a recovery phenomenon in which the mean value for a neurological symptom is high initially with a large variance, but then gradually decreases over time as recovery takes place, and with it the variance systematically decreases also (see Figure 11 for an example). The strength of the correlation often decreases as the spatial distance (in time) between observations increases, and in this case an autoregressive order 1 (AR(1)) covariance structure characterizes the relationships across the repeated measure (Brammer, 2003). Data transformations such as log and square root transformations can be used to stabilize the variances in some cases (Quinn and Keough, 2002).

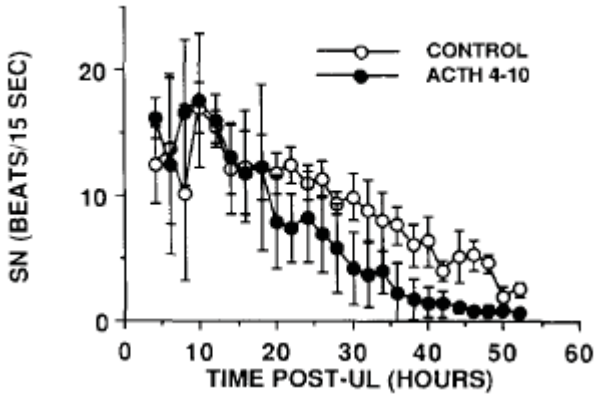


Figure 11: Gradual decrease in a vestibular symptom known as 'spontaneous nystagmus' (SN) following unilateral vestibular lesions in guinea pigs treated with adrenocorticotrophic factor 4-10 (ACTH-(4-10)) or its vehicle (i.e., the solvent). Symbols show means ± 1 standard deviation (SD). Note that as the mean SN decreases, so does the SD. From Gilchrist, D.P.D, Smith, P.F., Darlington, C.L. ACTH (4-10) accelerates ocular motor recovery in the guinea pig following vestibular deafferentation, *Neuroscience Letters.* 118 (1990) 14-16.

This systematic change in the variance for the repeated measure violates the assumption of sphericity and one of the consequences can be an inflation of the type I error rate for the ANOVA (Winer et al., 1991). The degree to which the violation of sphericity occurs can be evaluated using a test such as Mauchly's test of sphericity. However, Mauchly's test has been criticised because it relies on the normality of the data and its sensitivity depends on the sample size; therefore, its routine use is not recommended (Winer et al., 1991; Quinn and Keough, 2002). Quinn and Keough (2002) suggest that it is safer to assume that the sphericity assumption is violated in repeated measures situations. There are various possible solutions, one of which is to use a correction such as the Greenhouse-Geisser or Huynh-Feldt corrections, which make the type I error rate for the F test more conservative for the repeated measure (Winer et al., 1991). However, in practice, very few researchers either test this assumption or employ such corrections (in Table 1 only Goddard et al. (2008) reported testing the assumption of sphericity) (Smith, 2012a).

As an example of this issue, Figure 12 shows the mean escape latencies for rats leaving the open arm of an elevated T maze, that are analysed later in this thesis. For the pre-drug data ('Pre'), the Mauchly's test of sphericity was significant ($P \leq 0.04$), with a test statistic ('Mauchly's W') of 0.65 (Table 6), for which Field (2011) recommends using the Greenhouse-Geisser correction since it is less than 0.75 (Table 7).

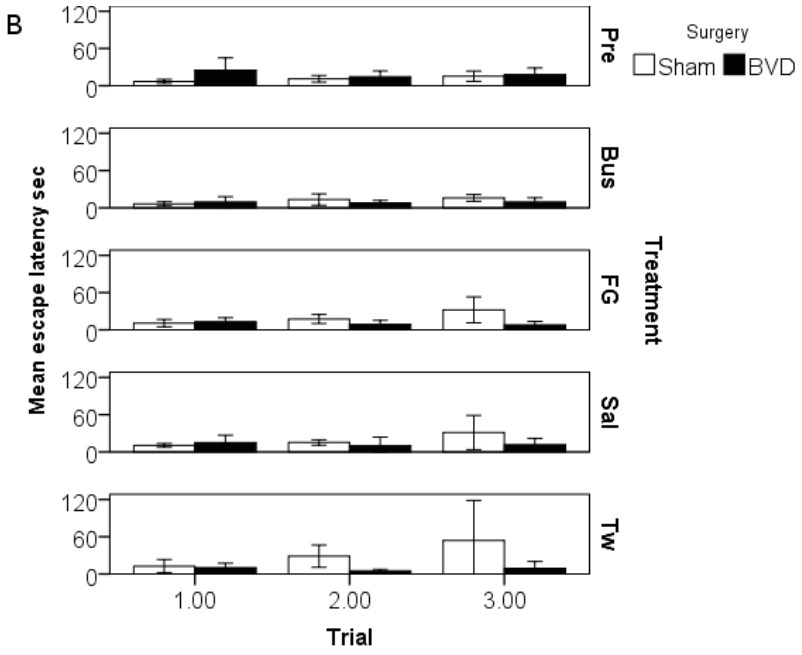


Figure 12: Mean escape latency (sec) in the 3 avoidance trials in the elevated T maze for the BVD and sham groups in the pre-drug condition ('Pre') and in response to buspirone ('Bus'), FG-7142 ('FG'), saline ('Sal') and DW/Tween 20 ('Tw'), ± 95% confidence interval (CI).

Mauchly's Test of Sphericity[b]

Measure:latency

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
trials	.653	6.397	2	.041	.742	.852	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + Surgery
Within Subjects Design: trials

Table 6: Example of Mauchly's test of sphericity for the escape latency data, showing a significant violation of the sphericity assumption, at $P \leq 0.04$.

Tests of Within-Subjects Effects

Measure:latency

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
trials	Sphericity Assumed	143.229	2	71.615	.411	.666
	Greenhouse-Geisser	143.229	1.485	96.479	.411	.607
	Huynh-Feldt	143.229	1.703	84.097	.411	.634
	Lower-bound	143.229	1.000	143.229	.411	.530
trials * Surgery	Sphericity Assumed	675.389	2	337.694	1.940	.160
	Greenhouse-Geisser	675.389	1.485	454.943	1.940	.173
	Huynh-Feldt	675.389	1.703	396.552	1.940	.168
	Lower-bound	675.389	1.000	675.389	1.940	.183
Error(trials)	Sphericity Assumed	5569.499	32	174.047		
	Greenhouse-Geisser	5569.499	23.753	234.477		
	Huynh-Feldt	5569.499	27.250	204.382		
	Lower-bound	5569.499	16.000	348.094		

Table 7: Example of Greenhouse-Geisser and other corrections for the violation of the assumption of sphericity in an ANOVA for the escape latency data.

1.4.3.4 Repeated measures ANOVAs with unbalanced repeated measures designs and missing data

Unequal sample sizes exacerbate the effects of the violation of the assumptions of normality and homogeneity of variance (Box, 1954a; Box, 1954b; Winer et al., 1991). However, general linear model ANOVAs can cope with treatments with unequal sample sizes (Quinn and Keough, 2002). Missing data, on

the other hand, present particular problems for repeated measures ANOVAs. Missing data are common in biology due to experimental subjects dying during an experiment or clinical trial or as a result of a measurement becoming technically or ethically impossible (Quinn and Keough, 2002). In this situation, since the sums of squares for the treatment (SST) are weighted according to the number of observations for the treatments, and the sum of squares for the error (SSE) are weighted according to the number of samples for the experimental subjects, no two mean squares will have equal expected values under the null hypothesis (Kuehl, 2000; Kirk, 2013). Therefore, there can be no exact F test of the null hypothesis (Kuehl, 2000).

Programs such as SPSS 20 delete experimental subjects with missing data when performing a repeated measures ANOVA (Gamst et al., 2008; Field, 2011). Since many experimental designs in neuroscience studies using animals already have small and unequal sample sizes, simply deleting the data from subjects with missing values is difficult to justify (Clark et al., 2012). The reduced sample size will result in less statistical power and it is ethically questionable to make use of animals in research and then not include their data (Smith, 2012b). One alternative is to employ some form of imputation procedure in order to estimate the missing values ('Missing Values Analysis or MVA') (Quinn and Keough, 2002; Gamst et al., 2008). Another is to use a maximum likelihood (ML) and expectation-maximization (EM) approach (a combination of imputation and ML) (Quinn and Keough, 2002). However, the EM algorithm is available only in some programs (e.g., SPSS 20) and the ML and EM methods require that the missing data are 'missing at random' (MAR, i.e. the probability that an observation is missing does not depend on the unobserved missing value but may depend on the group to which it would have belonged) or 'missing completely at random' (MCAR, i.e. the probability that an observation is missing does not depend on the observed or missing values), i.e. there is no bias to the way that data are missing (Quinn and Keough, 2002).

1.4.3.5 Alternatives to repeated measures ANOVAs: Linear mixed model analysis

Since repeated measures data are often correlated and therefore violate the ANOVA assumption of sphericity, an alternative approach is to use a linear mixed

model (LMM) analysis that models the correlation in the data. The term ‘mixed’ derives from the fact that there is a mixture of ‘fixed’ and ‘random’ effects that are to be estimated (Gurka and Edwards, 2011). Many developments in LMM analysis were stimulated by longitudinal data in epidemiological and clinical trial studies, where it is common to have missing data; therefore, LMM analysis can accommodate this scenario (Fitzmaurice et al., 2004; Kutner et al., 2005; Vittinghoff et al., 2005; Brown and Prescott, 2006; West et al., 2007; Gurka and Edwards, 2011). In LMM analysis, parameters are estimated using an iterative maximum likelihood estimation procedure (MLE or restricted maximum likelihood estimation (REML)), an optimization procedure which chooses as the estimates of the parameters the values that give the observed data maximal probability (Miller and Miller, 2004; Fitzmaurice et al., 2004; West et al., 2007; Gurka and Edwards, 2011). Whereas MLEs of the parameters are biased, REMLs are not (Fitzmaurice et al., 2004; Brown and Prescott, 2006; West et al., 2007). Rather than assuming that the data in the repeated measure are independent, or using a correction procedure if they are not, the correlations in the repeated measures data are modelled using various covariance matrix structures (fourteen are available in SPSS 20), for example: unstructured covariance structure, autoregressive (AR, order 1) or autoregressive-moving average (ARMA) covariance structures (Little et al., 2000; Brammer et al., 2003; Clark et al., 2012; see Figure 13). The different covariance matrix structures are described in an accessible way and in the context of pharmacology, in Brammer (2003). LMM analyses have been employed in pharmacological and psychopharmacological/neuroscience studies involving animals (e.g., Brammer, 2003; Stiles et al., 2012; Zheng et al., 2012). Brammer (2003) in particular is an excellent example of the application of LMM analyses to pharmacological data obtained from isolated tissues. It can be difficult to adequately assess the best covariance structure in the case of small sample sizes and the use of the unstructured covariance matrix structure can lead to a loss of statistical power; consequently, adjustments such as the Kenwood Rogers adjustment for small sample sizes have been proposed (Skene and Kenwood, 2010a). Skene and Kenwood (2010a,b) have investigated both a bias-adjusted empirical sandwich estimator and a modified Box correction for use with very small sample sizes and

report that the latter has an acceptable level of power (Skene and Kenwood 2010a,b).

The optimum covariance matrix structure model is usually based on goodness-of-fit assessed using an information criterion such as the Akaike's Information Criterion (AIC, a goodness of fit measure, which in this case reflects how well the covariance matrix structure describes the data; AIC = log(L) – q, where q = the number of covariance parameters), where the smallest value is best (Fitzmaurice et al., 2004; Posada and Buckley, 2004; Brown and Prescott, 2006; West et al., 2007). The AIC and Bayesian Information Criterion (BIC) are often used for this purpose, but some authors (e.g., Fitzmaurice et al., 2004; Gurka and Edwards, 2011) suggest that the BIC results in a higher probability of selecting a model that is too simple for the data. This is due to the fact that the BIC involves greater penalties for models with a large number of parameters (West et al., 2007). Although LMM analysis does not assume a balanced design, sphericity or homogeneity of variance, it still assumes that the sampling is random and that the residuals are normally distributed (Brown and Prescott, 2006). Nonetheless, LMM analysis using a REML has substantial advantages over repeated measures ANOVAs in cases where there are many repeated measures and there are missing data (Smith, 2012b).

Table I. Formulae for the different correlation structures.

Structure	Formula
CS	$\rho_{ij} = \begin{cases} 1 & i = j \\ \rho & \text{otherwise} \end{cases}$
UN	$\rho_{ij} = \begin{cases} 1 & i = j \\ \rho_{ij} & \text{otherwise} \end{cases}$
AR(1) and ARH(1)	$\rho_{ij} = \rho^{\lvert i-j \rvert}$
ARMA(1,1)	$\rho_{ij} = \begin{cases} 1 & i = j \\ \gamma\rho^{\lvert i-j \rvert - 1} & \text{otherwise} \end{cases}$
TOEP and TOEPH	$\rho_{ij} = \begin{cases} 1 & i = j \\ \rho_{\lvert i-j \rvert} & \text{otherwise} \end{cases}$
AR(1)+RE	$\rho_{ij} = \dfrac{\sigma_B^2 + \sigma_W^2 \rho^{\lvert i-j \rvert}}{\sigma_B^2 + \sigma_W^2}$

Figure 13: Examples of covariance matrix structures. From R.J. Brammer, Modelling covariance structure in ascending dose studies of isolated tissues and organs. *Pharmaceutical Statistics*. 2 (2003) 103–112. 'CS': compound symmetry. 'UN': unstructured. 'AR(1)': autoregressive, order 1. 'ARH(1)': heterogeneous AR(1). 'ARMA(1,1)': autoregressive moving average, order 1. 'TOEP': toeplitz. 'TOEPH': toeplitz heterogeneous. 'AR(1) + RE': autoregressive, order 1 + simple random effect.

1.4.4 Multivariate statistical methods

Experimental phenomena in biology in general, and in neuroscience in particular, usually involve the complex, non-linear interaction of multiple variables. Nonetheless, historically, statistical analysis has focussed on the comparison between treatment groups, of one variable at a time. In the context of the behavioural and neurochemical studies that have been conducted following peripheral vestibular lesions, univariate statistical analyses have been used exclusively (see Table 1). Even multivariate ANOVAs (MANOVAs) have not been used. This approach not only tends to inflate the actual type 1 error rate as a

result of large numbers of statistical analyses, but neglects the fact that changes may occur at the level of the interaction within a system of variables that cannot be detected in individual variables (Stevens, 2002; Liu et al., 2010). Consequently, in areas such as the analysis of gene microarray data, protein interaction and medical diagnostics, multivariate statistical analyses and data mining approaches are now being employed in an attempt to understand complex interactions between systems of variables (e.g., Pang et al., 2006; Krafczyk et al., 2006; Ryan et al., 2011; Brandt et al., 2012; Smith, 2012a).

1.4.4.1 MANOVA and linear discriminant analysis

Multivariate analyses of variance (MANOVAs) are an extension of ANOVA to the case in which there is more than one dependent variable. There are various test statistics, including Wilks's λ, Roy's largest root, Pillai's trace statistic (also known as the Pillai-Bartlett trace statistic) and Lawes-Hotelling trace statistic (Manly, 2005).

All four MANOVA statistics appear to be similar in power for small to moderate sample sizes (Field, 2011). According to Seber (1984), simulation studies indicate that Pillai's trace statistic may be more robust against violations of the assumptions of multivariate normality and homogeneity of the covariance matrices, than the other statistics (see below). Field (2011) also draws this conclusion, provided that the sample sizes for the different variables are equal.

Linear discriminant analysis (LDA) is a multivariate statistical method, often used following a MANOVA, in which the membership of two or more groups can be predicted from a linear combination of independent variables (Stevens 2002; Manly, 2005). A linear discriminant function (LDF) has the general form:

$$Z = a_1X_1 + a_2X_2 + ...a_pX_p \quad (7)$$

Where Z refers to the group, X, X_2,...X_p are independent variables, and a_1, a_2,...a_p are coefficients.

LDA is related to canonical correlation analysis, with the set of variables being indicators (0/1) only.

The statistical significance of the LDF can be assessed using statistics such as Wilk's λ. The success of the LDF in separating the groups can be evaluated using cross-validation (e.g., a leave-one-out or 'LOO' procedure), in which the linear equation is used to classify the data, one observation at a time, without knowledge of the actual group membership. It is possible to use a stepwise LDA. However, some authors (Manly, 2005; Field, 2011) suggest that stepwise methods can result in suppressor effects and an increase in type II error.

1.4.4.1.1 Multivariate normality

MANOVA and LDA do make assumptions. The first is that, for formal tests of statistical significance to be valid, the data within groups should have a multivariate normal distribution (Manly, 2005). Unlike univariate statistical analyses such as ANOVA, MANOVA and LDA are quite sensitive to the violation of the assumption of multivariate normality (Marcoulides and Hershberger, 1997; Stevens, 2002; Manly, 2005; Tabachnick and Fidell, 2007). It is difficult to test for multivariate normality, because most programs such as SPSS do not offer such an assumption test (Stevens, 2002). Because univariate normality, i.e. the normality of the individual variables, is necessary but not sufficient for multivariate normality, it is possible for each individual variable to be normally distributed without the multivariate distribution being normally distributed. Stevens (2002) points out that because a multivariate normal distribution entails that all subsets of variables have normal distributions, one way to assess multivariate normality is to determine whether all pairs of variables are bivariate normal. Box's test for the homogeneity of the covariance matrices (see below) is sensitive to violation of multivariate normality; therefore, in order to obtain results from that test that are valid, whether the assumption of multivariate normality is fulfilled, must be of concern (Stevens, 2002). However, there is a multivariate formulation of the central limit theorem and n's of 10-20 per group appear to be sufficient to afford protection against the consequences of violating multivariate normality (Stevens 2002, Field, 2011). It should be noted that LDA may still discriminate between groups even if the assumption of multivariate normality does not hold. On the other hand, multivariate normality does not necessarily mean that LDA will effectively discriminate between the groups.

1.4.4.1.2 Homogeneity of the covariance matrices

The second asumption is that the population covariance matrices are equal for all groups, usually tested using Box's M test (Marcoulides and Hershberger, 1997). If this assumption is violated, a quadratic discriminant analysis (QDA) can be used instead. In a review of several Monte Carlo studies, Stevens (2002) concluded that, provided that the sample sizes are equal, even moderate heterogeneity of the covariances does not substantially affect type I error. Unequal sample sizes, on the other hand, are potentially very problematic if the covariances are unequal.

While Box's M test is often used, its null hypothesis may be rejected only because the multivariate normality assumption is violated (Stevens, 2002). Therefore, it is important to determine whether this is the reason for a significant Box's M test. Box's M test is also very sensitive to departure from homogeneity of the covariances (Field, 2011). Both Stevens (2002) and Field (2011) suggest that even if the Box's M test is significant, the type I error rate will be only slightly affected provided that there are equal sample sizes, although the power may be somewhat reduced.

1.4.4.1.3 Number of variables

One of the common problems in many multivariate statistical analyses is the sample size for each variable, n, relative to the number of variables, p. While unequal sample sizes can be problematic, as described above, when p is greater than n, statistical analyses such as MANOVA and LDA can become invalid. Stevens (2002) and Field (2011) suggest that, unless the n is large, p should be $\leq$ 10. Monte Carlo studies have shown that if the sample size is not large compared to the number of variables, the standardized discriminant function coefficients and correlations obtained in LDA, are unstable (Stevens, 2002). By 'large', Stevens (2002) suggests a ratio of n (total sample size) : p (number of variables) of 20 : 1. He further cautions that a small n : p ratio (i.e., $\leq$ 5) can be problematic for stepwise LDA in particular, because the significance tests are used to determine which variables are included in the solution (Stevens, 2002).

1.4.4.1.4 Example of LDA in neuroscience

Although LDA has not been used extensively in basic neuroscience to predict categorical membership, an example of its application is the prediction of age on the basis of the concentration of specific neurochemicals in different parts of the brain. The L-arginine metabolic pathway is a biochemical pathway that is critical for neuronal function (see Figure 14) and involves the neurochemicals: agmatine, putrescine, spermidine, spermine, L-arginine, L-ornithine, L-citrulline, glutamate and γ–aminobutyric acid (GABA). Although Figure 14 presents certain causal connections between some of these neurochemical variables, the mechanisms through which they interact with one another are not completely understood and additional pathways, particularly feedback pathways, are possible (Mori and Gotoh, 2004). It is therefore of interest to determine whether the concentrations of one part of this complex neurochemical pathway can be predicted from the other parts.

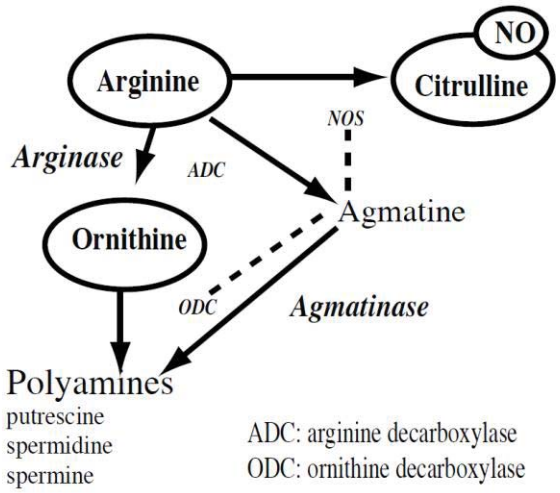


Figure 14. The arginine metabolic pathway showing the conversion of L-arginine to the neurotransmitter, nitric oxide (NO), and L-citrulline, by the enzyme, nitric oxide synthase (NOS), of which there are 3 isoforms (nNOS, eNOS and iNOS); the conversion of L-arginine to agmatine by the enzyme, arginine decarboxylase (ADC), which is then converted to polyamines such as putrescine, spermidine and spermine by agmatinase and ornithine decarboxylase (ODC); and the conversion of L-arginine to L-ornithine by arginase, which is then converted to the same polyamines, which are essential for cell proliferation, differentiation and communication, including neuronal

synaptic plasticity in the brain. The major excitatory neurotransmitter, glutamate, is one of the end products of L-arginine, and glutamate serves as a precursor for the synthesis of the major inhibitory neurotransmitter, GABA.

Liu et al. (2010) examined the concentrations of these neurochemicals in two areas of the rat hindbrain concerned with the control of movement: the brainstem vestibular nucleus complex (VNC) and the cerebellum (CE), in young (4 month old) and aged (24 month old) rats. Using LDA, they found a linear equation that could predict whether the animals were young or old, based on putrescine, spermine, spermidine, L-citrulline, glutamate and GABA concentrations in the VNC. Cross-validation showed that this LDF could discriminate between young and old animals with 100% accuracy and it was also statistically significant ($P \leq 0.0005$, Wilk's λ). The CE results were more surprising. An LDF was found that could predict the animals' age based on only spermine and spermidine. Cross-validation showed that the LDF had 93% accuracy and it was also statistically significant ($P \leq 0.0005$, Wilk's λ). The standardised canonical discriminant function coefficients are shown in Table 8 below. Both the size and the sign of the coefficients have predictive value.

Standardised canonical discriminant function coefficients for the linear discriminant functions for the VNC and CE.

Standardised canonical discriminant function coefficients	Function 1
Zscore(put) vnc	0.734
Zscore(spd) vnc	−2.417
Zscore(spm) vnc	3.458
Zscore(cit) vnc	-0.949
Zscore(glut) vnc	2.107
Zscore(GABA) vnc	−1.800
Zscore(spd) ce	−1.593
Zscore(spm) ce	1.672

Table 8: From Liu, P., Zhang, H., Devaraj, R., Ganesalingam, G., Smith, P.F. A multivariate analysis of the effects of aging on glutamate, GABA and arginine metabolites in the rat vestibular nucleus. *Hearing Research*. 269 (2010) 122-133. 'vnc': vestibular nucleus complex. 'ce': cerebellum. 'put" putrescine. 'spd': spermidine. 'spm': spermine. 'cit': citrulline. 'glut': glutamate. 'GABA': γ–aminobutyric acid.

This method should be applicable to many situations in neuroscience in which multiple variables interact to determine a dependent variable, provided that the sample sizes are sufficient and the cross-validations demonstrate the predictive accuracy of the LDFs. Given that Box's M test of the equality of the covariance matrices assumes multivariate normality, one way to proceed is to determine whether all pairs of variables appear to be bivariate normal. If so, Box's M test can be used as a guide to whether the assumption of the equality of the covariance matrices is fulfilled. However, the cross-validation procedure can be used as the ultimate arbiter of the effectiveness of the LDF.

1.4.4.2 Support vector machines

Support vector machines (SVMs) are an alternative method for classification, which employ 'support vectors', which are observations that form the spatial boundary between different classes (Williams, 2011). These support vectors are then used to determine a hyperplane that defines the boundary between the classes (Williams, 2011). SVMs can employ a variety of functions, such as radial kernel functions (see Figure 15), to remap the data and generate new variables that can separate the different categories (Williams, 2011). The data are usually split into training and test data sets (e.g., 70:30) and the difference between the model based on the support vectors in the training data set, and the test data set, is calculated as a measure of the model's success. As with LDA, classification error matrices can be used to evaluate the success of the classification, as well as receptor operating characteristic (ROC) curves, that quantify the relationship between the true positive rate of classification ('sensitivity') and the false positive rate of classification ('1 – the specificity') (Hastie et al., 2009).

One of the major advantages of SVMs is that they do not make distributional assumptions like MANOVA and LDA, other than that the data are independent and identically distributed. Wilson (2008) suggests that for this reason, even small sample sizes can provide accurate estimates of prediction error when there are a large number of variables.

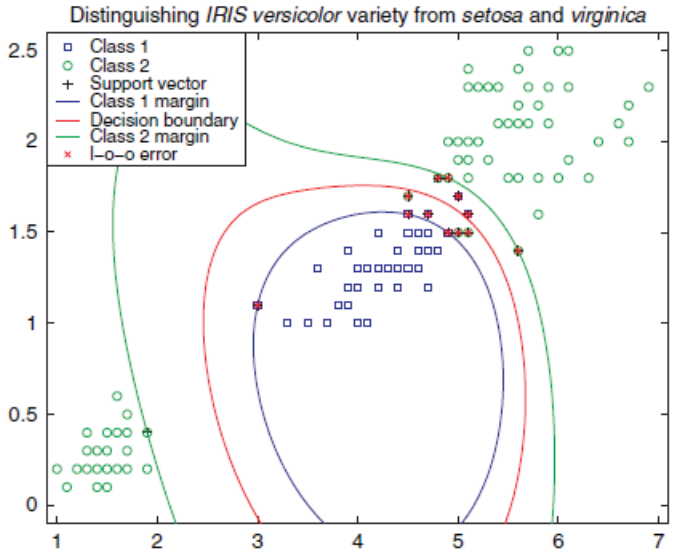


Figure 15: Example of a SVM classification employing a radial basis function to separate iris varieties based on petal width and length. From Wilson, M.D. Support vector machines. In *Encyclopedia of Ecology*. Elsevier, New York. (2008) p. 3436.

1.4.4.3 Cluster analysis

Another multivariate statistical method that has not been used in the context of vestibular neuroscience, is cluster analysis. Cluster analyses (CAs) are another type of non-parametric analysis that is used to explore the natural groupings of variables in a data set (Manly, 2005). Therefore, assumptions such as multivariate normality and equality of the variance-covariance matrices are not required (Marcoulides and Hershberger, 1997; Manly, 2005). Different measurements of the distance between the variables, such as Euclidean or Mahalanobis distance, are used to relate them to one another, and specific algorithms (e.g., Ward Minimal Variance Linkage) are used to determine the clusters (Marcoulides and Hershberger, 1997). The standardized data are usually used in order to avoid bias introduced by differences in scales of measurement. As an example of the application of this method, Figure 16 shows cluster analyses for the data from Liu et al. (2010). Agglomerative CAs, in which each variable is initially considered its own cluster, were used on the correlation coefficient distance. Some algorithms, such as single linkage, are prone to produce long strings of clusters (‘chaining’) (Lattin et al., 2003). Comparisons of the different kinds of CAs for the VNC data suggested that the Complete linkage, McQuitty

linkage, Average linkage and Ward linkage algorithms for determining clusters, all produced similar results. Ward's method, based on the objective of obtaining the smallest within-cluster sum of squares (the 'minimal variance principle'), was used (Lattin et al., 2003).

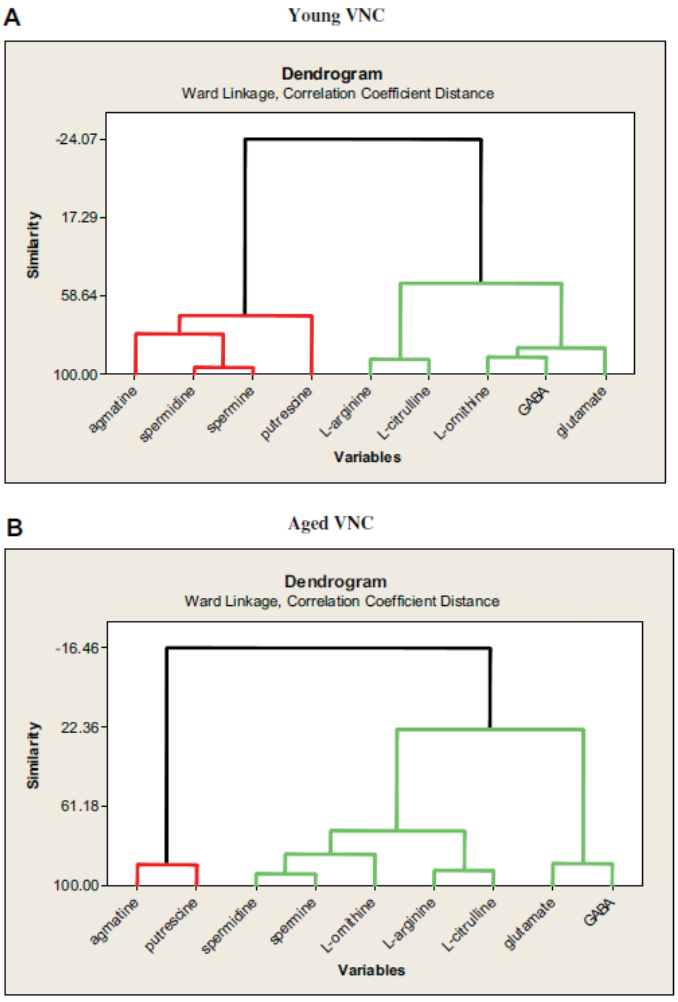


Figure 16: Dendrograms showing the relationship between the expression of the different neurochemical variables in the young and aged VNC. From Liu, P., Zhang, H., Devaraj, R., Ganesalingam, G., Smith, P.F. A multivariate analysis of the effects of aging on glutamate, GABA and arginine metabolites in the rat vestibular nucleus. *Hearing Research*. 269 (2010) 122-133.

1.4.4.4 Multiple linear regression

Yet another statistical method that has been under-employed in neuroscience is multiple linear regression (MLR). Although not strictly a multivariate statistical method, since there is only one dependent variable at a

time, MLR is a part of the general linear model that is useful for determining whether one variable can be predicted from a combination of other variables. Simple linear regression can be expanded to include more than one predictor variable to become MLR.

MLR has the general form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_p X_p + \varepsilon \quad (10)$$

Where Y = the quantitative dependent variable; X_1, X_2,...X_p are independent variables; β_1, β_2... β_p are coefficients; β_0 is the intercept and ε is the error term (Ryan, 2009).

However, formal statistical tests for MLR, like those for simple linear regression, make assumptions regarding the distribution of the data, which cannot always be fulfilled. These assumptions are the same as those for other methods in the general linear model, such as ANOVA and analysis of covariance (ANCOVA): that the residuals are normally distributed, with homogeneity of variance, and that they are independent of one another (e.g. not autocorrelated) (Vittinghoff et al., 2005; see Figure 17 for L-citrulline from Liu et al. (2010) as an example). Furthermore, the predictor variables should be numerical, although indicator variables can be used in order to include nominal variables (e.g., binary coding to represent male and female). The violation of the assumption of normality can sometimes be redressed using data transformation, which may also correct heterogeneity of variance, but other issues such as autocorrelation, are not easily dealt with and may require methods such as time series regression (Ryan, 2009).

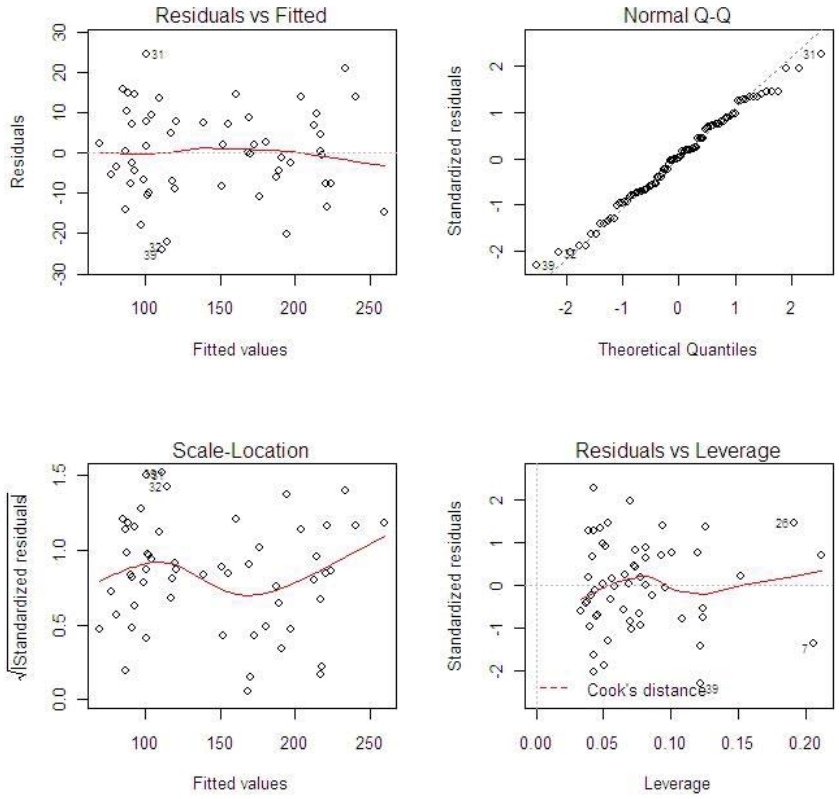


Figure 17: Diagnostic plots for L-citrulline following MLR showing residuals versus fitted values, normal Q-Q, scale location and residuals versus leverage plots. From Liu, P., Zhang, H., Devaraj, R., Ganesalingam, G., Smith, P.F. A multivariate analysis of the effects of aging on glutamate, GABA and arginine metabolites in the rat vestibular nucleus. *Hearing Research*. 269 (2010) 122-133.

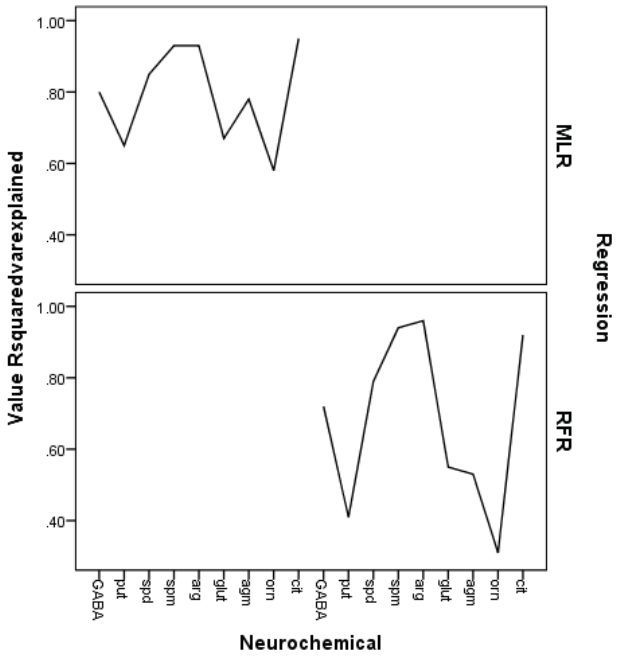
1.4.4.5 Random forest regression and classification

Although modelling using regression trees has been used for over 25 years, its use in the neurosciences has been very limited. In regression tree modelling, a flow-like series of questions is asked about each variable ('recursive partitioning'), subdividing a sample into groups that are as homogeneous as possible by minimising the within-group variance, in order to determine a numerical response variable (Vittinghoff et al., 2005). The predictor variables can be interval or ratio variables also, or they can be ordinal or nominal. By contrast with linear regression, which makes assumptions about the distribution of the data, regression trees make no distributional assumptions. The data are usually split into training and test data sets (e.g., 70:30) and the mean square error (MSE) between the

model based on the training data and the test data, is calculated as a measure of the model's success. Variables are chosen to split the data based on the reduction in the MSE achieved after a split (i.e., the information gained). Unlike linear regression, interactions between different predictor variables are automatically incorporated into the regression tree model and variable selection is unnecessary because irrelevant predictors are excluded from the model. This makes complex, non-linear interactions between variables easier to accommodate than in linear regression modelling (Hastie et al., 2009). Breiman et al. (1984) extended the concept of regression trees by exploiting the power of computers to simultaneously generate hundreds of trees, known as 'random forests', which were based on a random selection of a subset of data from the training set. The various regression tree solutions are averaged in order to predict the target variable with the smallest MSE (Marsland, 2009; Williams, 2011).

Random forests (RFs) can also be used for classification purposes, in which case the solution is based on the number of 'votes' from different trees for a particular category (Williams, 2011). The effect of variable removal on the mean decrease in accuracy, the 'out of bag' (OOB) error, and the overall classification matrix error ('confusion matrix error') are used to evaluate the success of the classification. The 'out of bag' (OOB) error is the error based on the observations that were excluded from the subset of the training data (the 'bag') used to generate the decision tree (Williams, 2011). Unlike LDA, random forest regression and classification make no distributional assumptions and therefore can be applied to situations in which the sample sizes are small relative to the number of variables (Hastie et al., 2009; Williams, 2011).

As a comparison of the application of MLR and random forest regression (RFR) to the data from Liu et al. (2010) described above to illustrate LDA, both forms of regression were applied to the prediction of each of the nine neurochemicals from the other eight. Figure 18 shows the adjusted R^2 (MLR) and variance explained values (RFR), as well as the residual mean square error (RSE) values, for the MLRs and RFRs (respectively).



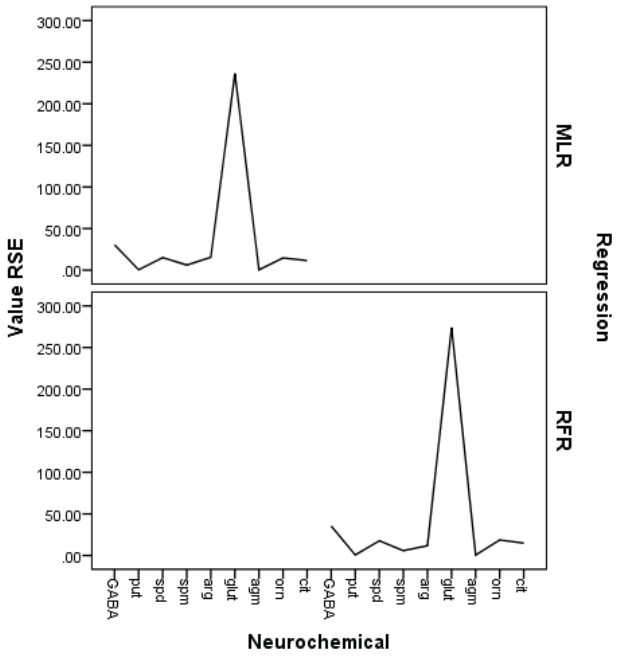


Figure 18: Comparison of the adjusted R^2 values (MLRs) and variance explained values (RFRs), and the RSE values, for the MLRs and RFRs for the neurochemical variables from Liu et al. (2010).

It was apparent that the adjusted R^2 values for the MLRs were generally higher than the variance explained values for the RFRs: 5/9 of them were ≥ 0.80 compared to 3/9 for the variance explained values (Figure 18). The RSE values

were more similar but lower for the MLRs than the RFRs in all but 3 cases (Figure 18). Nonetheless, the general patterns for the adjusted R^2/variance explained values and the RSEs were similar for the MLRs and the RFRs.

Although MLR appeared to be more predictive for this particular data set, both forms of regression could potentially be used to predict behavioural and neurochemical variables in the context of vestibular neuroscience research.

2. Methods

2.1 Behavioural study

2.1.1 Animals

Twenty male Wistar rats (300-370 g at the time of delivery) were obtained from the Hercus Taieri Resource Unit, Dunedin, New Zealand. The animals were randomly allocated to the bilateral vestibular deafferentation (BVD, n=10) or sham surgery conditions (n=10). The behavioural testing began at 1 month following the surgery, in order to allow the animals time to recover from the surgery and anaesthesia and also so that the animals could partially compensate for the symptoms of BVD (Zheng et al., 2006; Zheng et al., 2009). The animals were maintained on a 12:12 h light/dark cycle at 22º C with free access to food and water. During recovery from surgery, the animals were housed individually, and then 2 weeks before the behavioural testing, they were housed in either groups of 2 or 3 within their treatment groups for the remainder of the experiment. Although the groupings of the animals within individual houses could be considered an additional variable, the housing conditions for the animals were otherwise identical and previous studies by this research group have found no effect of the housing arrangements on the behavioural consequences of BVD (Zheng et al., 2006; Zheng et al., 2007; Zheng et al., 2008; Zheng et al., 2009; Baek et al., 2010). All procedures were approved by the University of Otago Animal Ethics Committee (05/10). During the recovery from surgery, 2 BVD animals died from unknown causes. Consequently, 8 BVD and 10 sham animals were used for the behavioural testing.

2.1.2 Surgery

For all surgery, animals were anesthetized with a combination of fentanyl citrate (200 μg/kg, administered subcutaneously (s.c.)) and medetomide hydrochloride (500 μg/kg, s.c.). Atropine (0.05 mg/kg, s.c.) was administered to control respiratory secretion during anaesthesia. Xylocaine (with 1:100,000 adrenaline) was injected around the wound margins. Carprofen (5 mg/kg, s.c) and strepsin (0.1 ml per rat, s.c.) were administered for post-operative analgesia and to prevent infection, respectively.

A complete BVD was performed under microscopic control as described in detail previously (Zheng et al., 2006; Zheng et al., 2007; Zheng et al., 2008;

Zheng et al., 2009). Briefly, the middle ear was exposed so that the tympanic membrane, malleus and incus, could be removed. The stapedial artery was cauterized to prevent bleeding, and then the horizontal and anterior semicircular canal ampullae were drilled open using a high speed dental drill with a fine burr. The canal ampullae, and the utricle and saccule, were aspirated and the temporal bone was sealed with dental cement. Temporal bone histological studies have shown that this surgical procedure results in complete destruction of the peripheral vestibular system with no damage beyond the inner ear (Zheng et al 2006).

For the sham surgical procedure, the temporal bone was exposed in the same way as for the BVD surgery but only the tympanic membrane removed. This served as a partial auditory control, since the sham animals would experience hearing loss without vestibular lesions (Zheng et al 2006).

2.1.3 Drugs

Buspirone hydrochloride (Sigma, USA) was dissolved in methanol and saline (1:40) and administered to animals in a dose of 0.3 mg/kg s.c, 30 min before behavioural testing. Buspirone was chosen specifically because it is a non-benzodiazepine anxiolytic drug and therefore has fewer sedative side effects than benzodiazepines. N-methyl-β-carboline-3-carboxamide (FG-7142; Sigma, USA), was dissolved in Tween 20 and distilled water (1:100) and was administered to animals at a dose of 5 mg/kg s.c, 30 min prior to behavioural testing. These drug doses were chosen based on previous studies (Pellow & File 1986; Bruhwyler et al 1991; Stefanski et al., 1992; Graeff et al., 1998; Siemiątkowski et al., 2000; Poltronieri et al., 2003) Methanol: saline (1:40) and Tween 20: distilled water (1:100) were used as the vehicle controls for buspirone hydrochloride and FG-7142, respectively. These were also administered s.c. 30 min before behavioural testing.

2.1.4 Behavioural testing

Each animal was tested in multiple behavioural assays and in response to the drugs and vehicles, over a period of several months. Therefore, there were multiple repeated measures on each animal. The animals were tested in the spatial T maze (STM) alternation task, to assess spatial learning and memory (Zheng et

al., 2007), and in the open field maze (OFM), elevated plus maze (EPM) and elevated T maze (ETM), to measure their levels of exploratory behaviour and anxiety (Zheng et al., 2008; Goddard et al., 2008). The STM required food deprivation and a period of training; therefore, spatial memory was tested after the completion of the OFM, EPM and ETM. This inevitably meant that there was an inherent order effect in the behavioural tests.

All of the animals were tested once in the OFM, EPM, ETM and STM before any drug or vehicle treatment to obtain the baseline performance for both the sham and BVD groups. After the pre-drug testing, animals receiving buspirone, FG-7142 and their respective vehicles were tested in each task in order to control for the order effects of each treatment and maze exposure and to minimise the number of animals used. Therefore, each behavioural task (OFM, EPM or ETM) was repeated 4 times during which each animal received every drug and vehicle treatment (buspirone, saline/methanol, FG-7142 and DW/Tween20) once, in an order that was counterbalanced (Table 9). Repeated exposure to the same maze has been reported to cause 'one-trial tolerance', in which the effects of anxiolytic drugs did not reduce anxious behaviours on repeated trials (File et al., 1990). This was due to learned behaviour in the initial trial rather than drug tolerance (Dawson et al., 1994). However, it has been reported that a 3 week interval between trials, in combination with re-testing in a novel environment, can prevent 'one-trial tolerance' (Adamec and Shallow, 2000). Therefore, in this study, each trial for one behavioural task was conducted every 3 weeks in a novel environment; therefore, a trial from a different behavioural task was carried out each week. For the STM alternation task, each day of drug testing was followed by a three-day washout period where the animals continued to be tested without any treatment. A three-day washout period was chosen to eliminate any possible carry over effects on the subsequent trials and to ensure that the animals had similar levels of performance before each treatment. The timelines for the behavioural tests are presented in Table 10.

2.1.4.1. Open field maze (OFM)

Apparatus: The test was conducted using a square wooden box of 60 cm (width) x 60 cm (length) x 20 cm (height), 100 cm above the ground. The base was separated into 36 equal-sized boxes. Five mm thick transparent perspex walls

measuring 15 cm in height were fixed onto the existing walls to prevent animals from escaping. The lines on the floor divided the box into 3 zones: outer, middle and inner (see Figure 19). White noise generated from a speaker positioned underneath the box, masked ambient noise. Each trial was recorded using a digital camera mounted above the OFM and Ethovision XT 6.0 tracking software, which tracks and analyzes animal movement and location automatically (Stiles et al., 2012; Zheng et al., 2012).

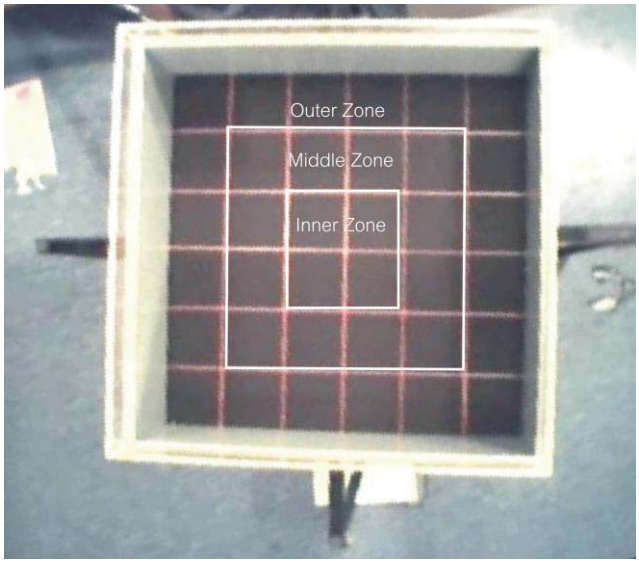


Figure 19. The OFM with its 3 zones.

Procedure: Each animal was placed into one of the corners facing the centre and was allowed to explore the apparatus for a total of 10 min. For each trial the animal was placed into the maze in a different corner. The apparatus was cleaned with mild detergent between each animal in order to remove traces of odour. For analysis, the time spent in the 3 zones was simplified to 2 zones: inner/middle and outer, and this was normalised to the number of grids for each zone. The time in the 3 zones, the distance travelled, in cm, the velocity of locomotion, in m/sec, and the number and the duration of wall-supported and unsupported rearings, an index of exploratory behaviour, were analyzed.

2.1.4.2. Elevated plus maze (EPM)

Apparatus: The EPM consisted of 4 arms perpendicular to each other, 100 cm above the ground. Each arm was 50 cm in length and 10 cm in width. Two of the arms, 180º apart, were enclosed by a 40 cm high wall without ceilings (enclosed arms). The other two arms were fitted with a 1 cm perspex edge to prevent animals from falling (open arms). The 4 arms were separated by the maze centre (10 x 10 cm) (see Figure 20). White noise generated from a speaker

underneath the apparatus, masked ambient noise. Each trial was recorded by a digital camera and Ethovision XT 6.0 tracking software.

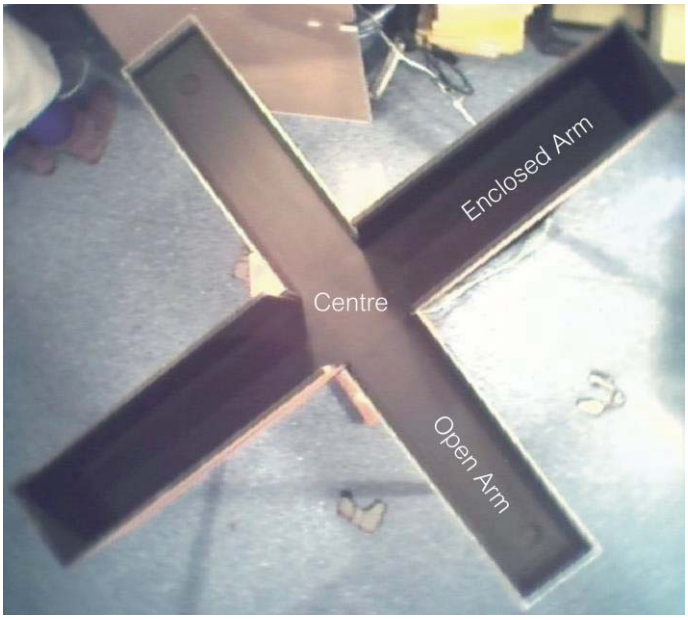


Figure 20. The EPM.

Procedure: The animal was placed at the centre of the maze, facing one of the open arms and left to explore for a total of 10 min. The apparatus was cleaned between each animal with mild detergent and warm water to remove odour traces. The time spent in, and the number of entries into, each arm were recorded. Open arm activity was calculated as the percentage of time in, or the number of entries into, the open arms. The number of total arm entries was the combination of entries into both the open and enclosed arms. The distance travelled in the EPM, in cm, was also analyzed.

2.1.4.3. Elevated T maze (ETM)

Apparatus: The ETM employed the same apparatus as the EPM but with one of the enclosed arms partitioned off. The remaining enclosed arm was enclosed by a 40 cm high wall and was situated perpendicularly to 2 opposite open arms that were fitted with a 1 cm perspex edge (see Figure 21). The apparatus was 100 cm above the floor. White noise generated from a speaker positioned underneath the apparatus, masked ambient noise. Each trial was recorded by Ethovision XT 6.0.

Figure 21. The ETM.

Procedure: To measure inhibitory avoidance, the animal was placed at the distal end of the enclosed arm facing the centre. The baseline latency was measured as the time it took for all 4 paws of the rat to leave the enclosed arm and move into the centre of the maze. This measurement was repeated twice at 30 sec intervals. Following inhibitory avoidance, the animal was placed at the end of one of the open arms, to measure escape latency, as a means of measuring panic. The time for the animal to leave the open arm into the centre of the maze, i.e. escape latency, was measured. This was repeated 2 times at 30 sec intervals. A cut-off time of 300 sec was used for both avoidance and escape latencies, although this was rarely required. The apparatus was cleaned between each animal, with mild detergent and warm water to remove odour traces.

Rat	Surgery	Trial			
		1	2	3	4
1	BVD	Saline/Methanol	Buspirone	FG-7142	DW/Tween20
3	BVD	Buspirone	FG-7142	DW/Tween20	Saline/Methanol
5	BVD	FG-7142	DW/Tween20	Saline/Methanol	Buspirone
6	BVD	DW/Tween20	Saline/Methanol	Buspirone	FG-7142
9	BVD	Saline/Methanol	Buspirone	FG-7142	DW/Tween20
11	BVD	Buspirone	FG-7142	DW/Tween20	Saline/Methanol
13	BVD	FG-7142	DW/Tween20	Saline/Methanol	Buspirone
15	BVD	DW/Tween20	Saline/Methanol	Buspirone	FG-7142
2	Sham	Saline/Methanol	Buspirone	FG-7142	DW/Tween20
4	Sham	Buspirone	FG-7142	DW/Tween20	Saline/Methanol
7	Sham	FG-7142	DW/Tween20	Saline/Methanol	Buspirone
8	Sham	DW/Tween20	Saline/Methanol	Buspirone	FG-7142
10	Sham	Saline/Methanol	Buspirone	FG-7142	DW/Tween20
12	Sham	Buspirone	FG-7142	DW/Tween20	Saline/Methanol
14	Sham	FG-7142	DW/Tween20	Saline/Methanol	Buspirone
16	Sham	DW/Tween20	Saline/Methanol	Buspirone	FG-7142
18	Sham	Saline/Methanol	Buspirone	FG-7142	DW/Tween20
20	Sham	Buspirone	FG-7142	DW/Tween20	Saline/Methanol

Table 9: The order of the drugs administered to each animal during 4 repetitions of each behavioural task.

Weeks	Behavioural Tests
1	Pre-drug OFM
2	Pre-drug EPM
3	Pre-drug ETM
4	OFM Trial 1
5	EPM Trial 1
6	ETM Trial 1
7	OFM Trial 2
8	EPM Trial 2
9	ETM Trial 2
10	OFM Trial 3
11	EPM Trial 3
12	ETM Trial 3
13	OFM Trial 4
14	EPM Trial 4
15	ETM Trial 4
16-19	STM Alteration Task

Table 10: The order of the behavioural tests. Note that the STM was always at the end.

2.1.4.4. Spatial forced alternation in T maze (STM)

Apparatus: The STM had 3 arms, 50 cm in length, 10 cm in width and 20 cm in height, 100 cm above the ground. The starting area was designated as the arm that was perpendicular to the 2 other arms. A removable barrier was placed 25 cm from the end of the starting area to serve as a gate. The other 2 arms contained a small plastic bottle cap at the end, which was used as a food well (see Figure 22). Sucrose tablets (Noyes sucrose reward tablet, 45 mg pellets, Research Diet, Inc., USA) were used as food reward.

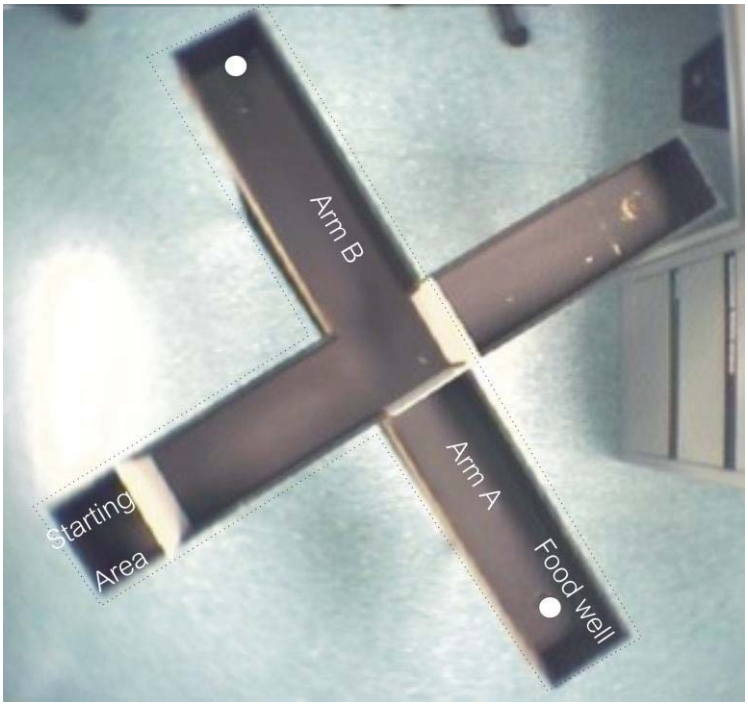


Figure 22. The STM with Arm A blocked off during the sample run.

Procedure: The animals were food-deprived and maintained at 85% of their expected body weight throughout the testing period. During habituation, each food well was baited with a generous number of sucrose pellets and each rat was left to freely explore the maze and retrieve the food for 10 min. After habituation, each animal underwent spatial alternation training for 8 days, with 8 trials per day. Each trial was separated into 2 parts, a sample run followed by a choice run. During the sample run, one of the arms was blocked off with a removable barrier. The arm chosen for each sample run was randomly allocated using a random number generator. The animal was released from the starting area and forced to go into the pre-selected arm and allowed to eat the sucrose pellets.

The animal was then put back into the starting area for 10 sec. During the choice run, the barrier blocking one of the arms was removed and the animal was allowed to enter either of the 2 arms. If the animal chose to go into the opposite arm to the one previously visited in the sample run, this was recorded as a correct response and the animal was allowed to eat the sucrose pellets there before being returned to its cage. If the animal chose to enter the arm previously visited during the sample run, this was recorded as an incorrect response and the animal was confined to that arm for 10 sec as punishment before returning it to its cage. The percentage of correct responses was recorded and the animals were trained until they reached the criterion, i.e., a 90% correct response for 3 consecutive days for sham rats and a stable performance was reached for BVD rats, before the drug testing began.

During the drug testing days, the same testing procedure was used as during training. Each day of drug testing was followed by a 3 day washout period in which the animals continued to be tested. A 3 day washout period was chosen to determine whether the drugs had any lasting effect on spatial memory. There were 4 separate drug testing days, each followed by 3 days of post-drug testing. The number of correct responses was analyzed for each day.

2.1.5. Statistical analyses

Data analysis was performed using SPSS 20, Mintab 16 and R, using the R package, Rattle (Williams, 2009; Williams, 2011). All of the behavioural data were first tested for normality and homogeneity of variance, using the Kolmogorov-Smirnov and Shapiro-Wilk tests (for normality), and the Levene's test (for homogeneity of variance), and then log or square root transformed if necessary and re-tested. Data for the pre-drug conditions were analyzed separately from the drug treatment conditions in order to determine the effects of the surgery independently of drug treatment. The data from the anxiolytic drug, buspirone, were analyzed in relation to its control, methanol/saline. The data from the anxiogenic drug, FG-7142, were analyzed in relation to its control, DW/Tween20. The pre-drug data were not included in the analyses with the drug data because the former were always collected first, whereas the latter conditions were run in a counterbalanced fashion. All significance tests were 2 tailed and the type I error rate was set at 0.05.

2.1.5.1 Sample sizes

The sample sizes of n = 10 for the sham and n = 8 for the BVD groups (originally, n = 10 in each group before 2 animals in the BVD group died) were based on previous studies in which power calculations indicated a 90% power level to detect differences of interest given the variance estimates and an α rate of 5% (Zheng et al., 2006; Goddard et al., 2008; Zheng et al., 2008; Zheng et al., 2009; Baek et al., 2010). Although a total sample size of n = 18 may seem small compared to data sets used in many statistical analyses, in neuroscience studies using animals, such sample sizes are common due to the need to minimise the number of animals used and also because crossover designs are used to maximise statistical power (Lenth, 2001; Festing, 2003; Small et al., 2011; Norma et al., 2012).

2.1.5.2. Open field maze

Animals were tested for locomotor activity in the OFM on 5 separate occasions, first before any drug testing and then 4 times following administration of one of the drugs or their vehicles (see Tables 9 and 10). In all cases, the fixed, between group, variable was surgery, with 2 levels, BVD and sham. Locomotor activity in the OFM was measured in the same animals over different periods of time (e.g., 1, 5 and 10 min for the analysis of time spent in each of the inner/middle and outer zones) and also in response to different drug or vehicle treatments (buspirone, saline/methanol, FG-7142 and DW/Tween20). Therefore, there were 2 within subjects, repeated measures: time, in cases where animals were measured repeatedly over time, e.g. in the zone analysis, with 3 levels, 1, 5 and 10 min; and the drug treatments, since all animals received all treatments. In the case of the first repeated measure, time, it could be considered a fixed factor, since those 3 times were chosen from many possible time periods (Kutner et al., 2005; Gurka & Edwards 2008). In the case of the second repeated measure, involving the comparison of the drugs versus the vehicles, this could be considered a fixed factor, since it was chosen based on the previous literature.

For the analysis of locomotor activity in the OFM over 10 min, distance travelled and velocity were compared between the BVD and sham groups using a 2 sample, independent t test. Comparisons were also made over the 4 drug/vehicle conditions. Because there were few repeated measures, a 2 factor repeated

measures analysis of variance (ANOVA) was used to analyse the OFM tests together, with surgery as the between group factor and drug/vehicle conditions as the repeated measure. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the distance data ($P \leq 0.001$); therefore, the Greenhouse-Geisser correction was used, since sphericity was estimated to be less than 0.75 (Field, 2011).

For zone analysis, individual variables and interactions with multiple variables were analyzed using a linear mixed model (LMM) analysis, using a REML procedure. The fixed, between group, factor was surgery, with 2 levels, BVD and sham. There were 2 within group, repeated measures: period, with 3 levels, 1, 5 and 10 min; and zone, with 2 levels, inner/middle and outer. Both period and zone were considered fixed factors. In the analysis of drug effects, there was a third, fixed factor, repeated measure, drug treatment, with 2 levels, drug and vehicle (i.e., buspirone and FG-7142 were compared with their respective vehicles). Due to multiple repeated behavioural measures, LMM analyses were considered preferable to repeated measures ANOVAs, as LMM analysis models the covariance structure of the correlated repeated measures (Gurka & Edwards 2008). Additionally, LMM analysis allows missing data and unequal sample sizes. In every case where LMM analysis was used, the smallest Akaike's Information Criterion (AIC) was used to select the best model for the covariance structure from the 14 models offered by SPSS 20, based on the smallest AIC value (Posada & Buckley 2004). Since the time spent in the different zones of the OFM was calculated as a percentage, the natural log of the percentage was used for analysis.

For the frequency and duration of supported and unsupported rearing, the data during the pre-drug trial were analyzed using an independent 2 sample t-test in order to compare the change in supported and unsupported rearing between BVD and sham animals. For the analysis of drug effects, all of the data were analyzed using repeated measures ANOVAs as described above. The fixed, between group, factor was surgery, with 2 levels, BVD and sham; the drug condition, with 2 levels, drug and vehicle, was regarded as fixed factor, repeated measure.

2.1.5.3. Elevated plus maze

For the pre-drug trial, the duration of the total open arm entries, the frequency of open arm entries, and distance travelled, over a 10 min period, were analyzed using an independent 2 sample t-test to compare the BVD and sham groups. During the drug treatment trials, the corresponding data were analyzed using a repeated measures ANOVA, in which surgery was defined as a between group, fixed factor and drug treatment was defined as a within group, repeated measure, fixed factor. Since the time spent in the open arms of the EPM was calculated as a percentage, consideration was given to whether the natural log of the percentage should be used for analysis; however, as with the OFM zone data, this was observed to make no difference to the results.

2.1.5.4. Elevated T-maze

For both avoidance and escape data, there were 3 trials. For the pre-drug data, surgery was defined as a between group, fixed effect, and the avoidance/escape trials were regarded as within group, repeated measures but also a fixed effect. Although a cut-off time of 300 sec was used for the avoidance and escape latencies, it was rarely used and consequently the data were not treated as censored. Therefore, it was decided to use a 2 way repeated measures ANOVA for the pre-drug data. For the post-drug data, drug treatment was an additional within group, repeated measure, fixed effect. Therefore, it was decided to use an LMM analysis for the post-drug data. Post-hoc comparisons were carried out using Bonferroni tests (Gamst et al., 2008).

2.1.5.5. Spatial T-maze alternation task

Since whether the response in the spatial T maze task was correct or incorrect, was a binomial variable, the percentage of correct responses was natural log transformed for all analyses. For the 8 days of training (pre-drug), an LMM analysis was used to analyze the percentage of correct responses. Surgery was treated as a fixed, between group factor and day as a within group, fixed factor.

A three way repeated measures ANOVA was used to analyse the effects of drug treatment on correct responses during the drug injection day and how it compared to the 3 days following drug injection. Surgery was treated as a fixed, between group factor and drug treatment and time as within group, repeated

measures, fixed factors. ANOVA was used in preference to LMM analysis in this case because there were relatively few repeated measures and Mauchly's test of sphericity was not significant, suggesting that the violation of the assumption of sphericity was not a problem for the repeated measures ANOVA.

2.1.5.6 Multivariate statistical and data mining analyses

From the univariate analyses of the behavioural data set which included data on the effects of BVD on performance in the open field maze, the elevated plus and T mazes and the spatial T maze, 3 variables were highly significantly different between the BVD and sham control animals in the pre-drug testing:

1. Zone activity in the open field maze (the amount of time the BVD animals spent in the outer compared to the middle/inner zones was less than for the sham animals)
2. Rearing (the duration of supported and unsupported rearing in the open field maze was less for the BVD than for the sham animals)
3. Percentage of correct responses in the spatial T maze (BVD animals made fewer correct responses than the sham animals)

Therefore, it was decided to further investigate the relationship between these and the non-significant variables (locomotor activity; duration and frequency in the open arms in the EPM; distance travelled in the EPM; avoidance and escape latencies in the ETM) using multivariate statistical and data mining analyses. The effects of the drugs in the other parts of the experiments were negligible, but there were some small but significant effects. In order to avoid possible confounding with drug effects, only the pre-drug data were used.

BVD animals spent more time in the inner/middle zones of the open field maze, which is thought to reflect increased risk taking behaviour, and less time rearing, which may reflect reduced exploration as well as increased postural instability, while poor performance in the spatial T maze is an indication of spatial memory deficits. Therefore, an important question is whether the spatial memory deficits are related to the changes in exploratory and emotional behaviour?

The data were first investigated using a correlation analysis with the Spearman correlation coefficient (since it does not assume normality of the data). Twenty-seven out of 66 pairs of variables were found to be statistically significant

at $P \leq 0.05$. Due to the extent of the correlation amongst some variables (Figures 23 and 24) or because they measured the same underlying concept in slightly different ways, it was decided to exclude variables from the multivariate statistical and data mining analyses that were either highly correlated with others or were otherwise redundant. These were the frequency of unsupported rearing and supported rearing, which were highly correlated with the durations of unsupported and supported rearing; velocity in the OFM, which was highly correlated with distance travelled in the OFM; frequency of open arm entries in the EPM, which was related to the duration of open arm entries; and the distance travelled in the EPM, which was related to the distance travelled in the OFM. The exclusion of these variables also served the purpose of increasing the ratio of the sample size to the number of variables, in order to make analyses such as MANOVA and LDA more reliable (Stevens, 2002). With these 5 variables removed, there were 8 remaining variables to investigate in relation to performance in the STM: surgical group; distance travelled in the OFM; duration of supported rearing; duration of unsupported rearing; zone activity; open arm duration in the EPM; avoidance latency in the ETM; and escape latency in the ETM. In order to obtain measures of zone activity and performance in the ETM that could be compared directly with performance in the STM, data reduction was performed. In the case of zone activity in the OFM, the natural log of the ratio of the time spent in the inner/middle zone to the time spent in the outer zone was used. In the case of avoidance and escape latencies in the ETM, in each case the final latency from the 3 trials was used since this reflected the degree of learning that had taken place (Zheng et al., 2008).

Linear discriminant analysis (LDA), preceded by a multivariate analysis of variance (MANOVA) (Field, 2011), was used to determine whether animals could be identified as BVD or sham based on their measurements for all 8 continuous behavioural variables (i.e., ln % correct in the STM; distance travelled in the OFM; duration of supported rearing; duration of unsupported rearing; zone activity; open arm duration in the EPM; avoidance latency in the ETM; and escape latency in the ETM). For the MANOVA, Pillai's Trace statistic was used because it has been reported to be more robust against violation of assumptions than other MANOVA statistics (Stevens, 2002; Manly, 2005). For the LDA, in the first instance, all variables were entered together. A stepwise linear discriminant

function was then used to determine whether the number of variables in the linear equation could be reduced while still retaining predictive power (Tabachnick and Fidell, 2001; Lattin et al., 2003; Manly, 2005; Marcoulides and Hershberger, 1997). Wilks' λ was used as the test statistic. The success of the linear discriminant function was tested using cross-validation with a leave-one-out procedure. Box's M test was used to assess the homogeneity of the covariance matrices (Tabachnick and Fidell, 2001; Lattin et al., 2003; Manly, 2005; Marcoulides and Hershberger, 1997).

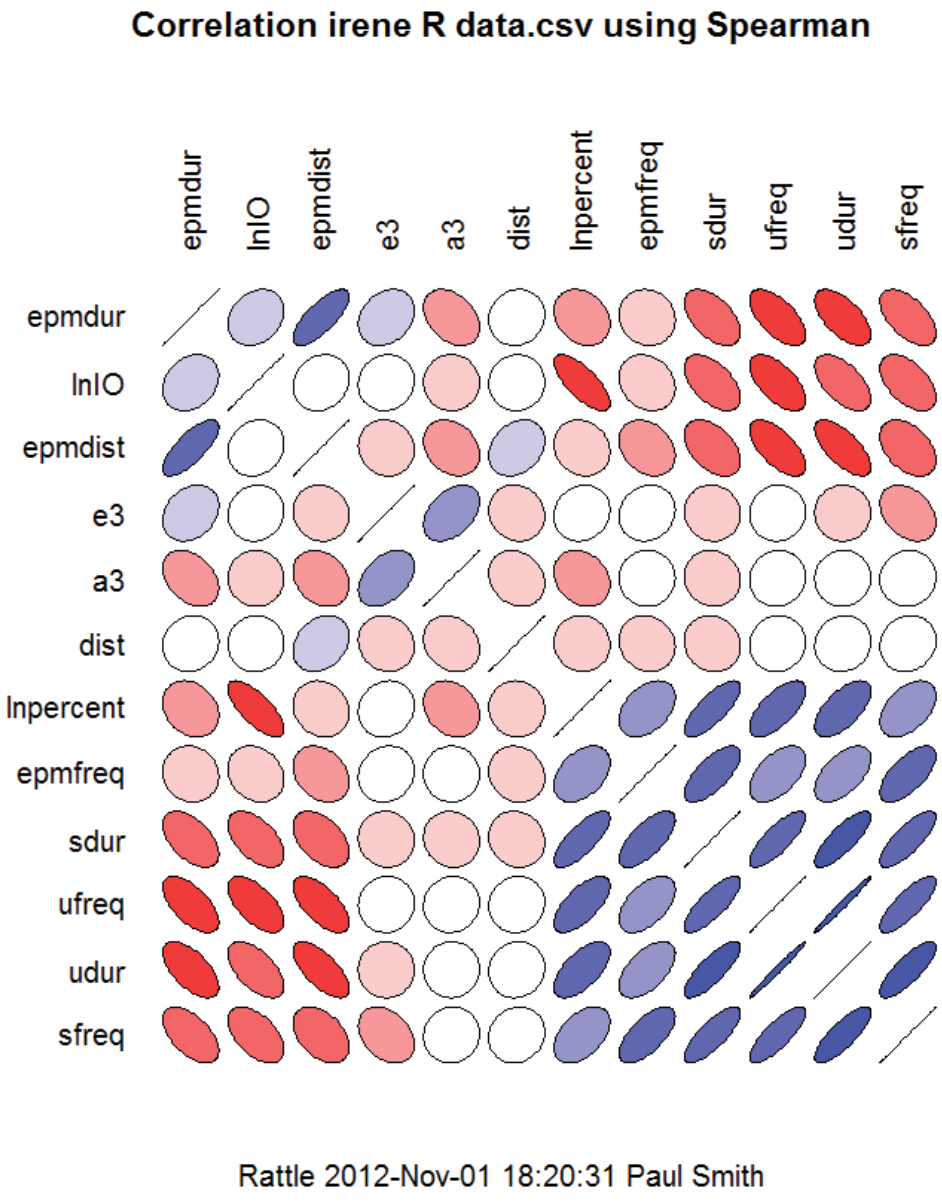


Figure 23. A Spearman correlation matrix for a subset of the behavioural variables. 'epmdur': duration of open arm entries in the EPM. 'lnIO': the ln of the ratio of time spent in the inner/middle to the outer zones of the OFM.

'epmdist': distance travelled in the EPM. 'e3': 3rd escape latency in the ETM. 'a3': 3rd avoidance latency in the ETM. 'dist': distance travelled in the OFM. 'ln percent': ln percent correct performance in the STM. 'epmfreq': frequency of open arm entries in the EPM. 'sdur': duration of supported rearing in the OFM. 'ufreq': frequency of unsupported rearing in the OFM. 'udur': duration of unsupported rearing in the OFM. 'sfreq': frequency of supported rearing in the OFM.

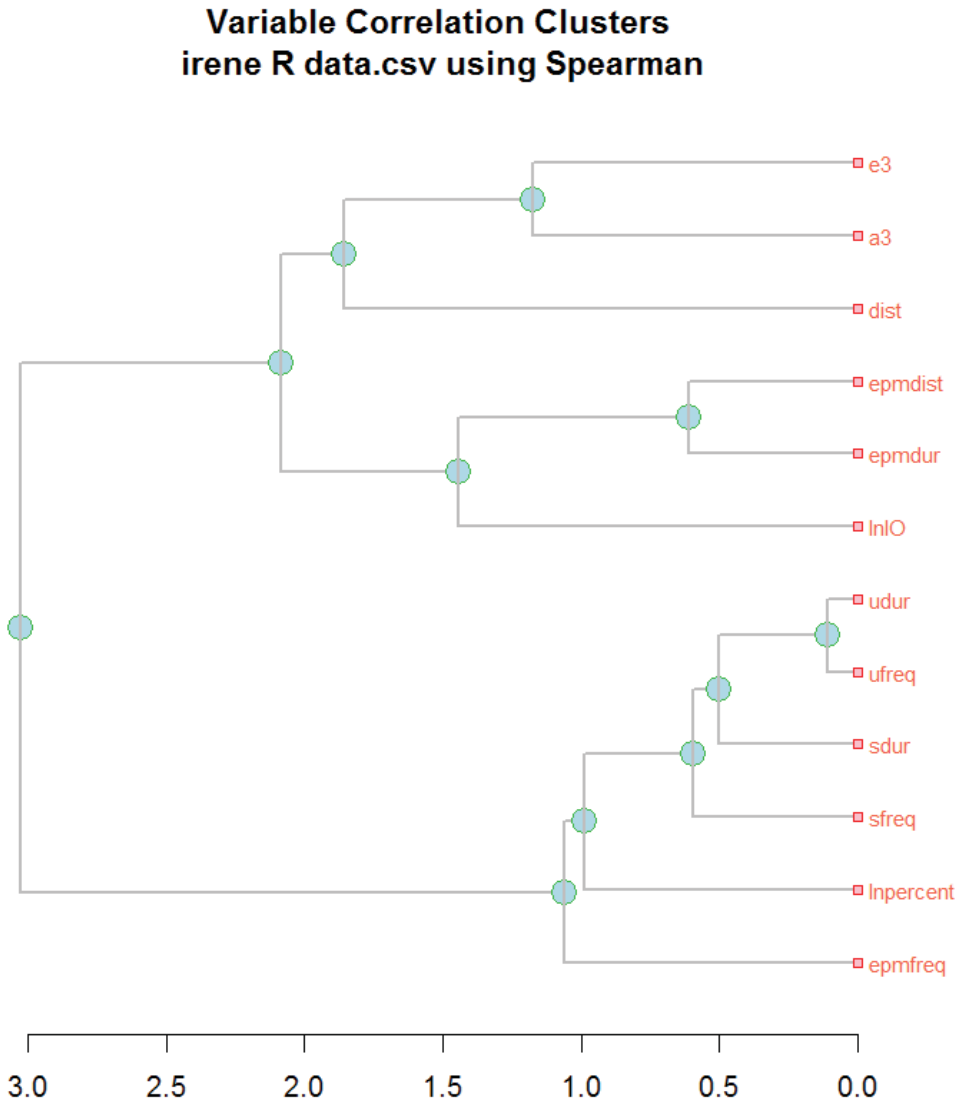


Figure 24. Spearman correlation clusters showing the relationship between each behavioural variable and every other behavioural variable. Abbreviations as in Figure 22.

The random forest (RF) and support vector machines (SVMs) methods were used, as exploratory techniques, using the R package, Rattle (Williams, 2009; Williams, 2011), to explore whether the surgical group of the animals could

be predicted from the other 8 variables ('ln percent correct', 'dist', 'sdur', 'udur', 'ln IO', 'epmdur', 'a3' and 'e3'). In order to perform RFs and SVMs, the data set was divided into training and test sets, using a 70:30 split. For the RFs, since there were 8 predictor variables for the target variable, surgical group (i.e., BVD or sham), it was decided to set m, the number of variables used to determine the decision at a node of the tree, as $\sqrt{p}$, where p is the number of predictor variables (approximately), and use m = 3, with 500 trees (Hastie et al., 2009). The effect of variable removal on the mean decrease in accuracy, the 'out of bag' (OOB) error, and the overall classification matrix error, were used to evaluate the success of the classification. In order to test the effect of kernel function on the SVM analysis, the following kernel functions were used: Gaussian radial basis; polynomial; linear (vanilla); hyperbolic tangent; Laplace; Bessel; Anova RBF; and spline. The classification error was used to evaluate the success of the analysis. Note however that RFs and SVMs are commonly used on larger data sets, so the analysis remains exploratory despite its success.

In order to further explore whether animals could be identified as BVD or sham based on the behavioural variables measured, a cluster analysis using the agglomerative Ward's minimal variance algorithm and squared Euclidean distance, was performed on the individual cases (Tabachnick and Fidell, 2001; Lattin et al., 2003; Manly, 2005; Marcoulides and Hershberger, 1997). All of the data were converted to z scores by variable first. Dendrograms were used to assess the degree of association between animals within and between the sham and BVD groups. Other cluster analysis algorithms were also investigated.

Principal component analysis (PCA) was considered as a possibility for investigating whether the different behavioural variables might form components that represent different aspects of the behavioural effects of BVD. The data were converted to z scores by variable and PCA without rotation on the correlation matrix was performed. However, the Kaiser-Meyer-Olkin Measure of sampling adequacy was < 0.5 (Table 11), which suggested that the sample size was not adequate to provide a stable PCA solution (Field, 2011). Even values between 0.5 and 0.7 are considered to be barely acceptable (Kaiser, 1974). Therefore, it was decided not to pursue PCA any further.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.427
Bartlett's Test of Sphericity	Approx. Chi-Square	54.598
	df	28
	Sig.	.002

Table 11: The results of the Kaiser-Meyer-Olkin Measure of sampling adequacy.

MLR was used in an attempt to determine the extent to which the other variables could be used to predict performance in the STM. Because the dependent variable, the number of correct entries (expressed as a percentage), was a binomial variable, logistic regression was also used but the results were similar (data not shown). Further MLRs were performed to determine whether any of the other variables could be predicted from the remaining variables. Surgical group was included as a binary, indicator variable (i.e., sham or BVD). In general, stepwise regression has received substantial criticism due to the automated nature of the procedure, the arbitrary nature of the criteria for the decision process, the correlation between the different hypothesis tests and the possibility that it can result in the neglect of suppressor effects (Brook and Arnold, 1985; Babayak et al., 2004; Ryan, 2009; Vittinghoff et al., 2005; Kutner et al., 2005; Quinn and Keough, 2002; Field, 2011). Many statisticians prefer the best subsets procedure, since it computes all possible regression models (Brook and Arnold, 1985; Ryan, 2009; Quinn and Keough, 2002). Among the stepwise procedures, backward stepwise regression is preferred by some because negatively confounded variables are less likely to be omitted than in forward stepwise procedures (Vittinghoff et al., 2005). It was therefore decided to compare backward regression and best subsets regression. The success of the regression model was determined using the adjusted R^2; the standard error for the regression; the analysis of variance (ANOVA) for the regression; the t tests for the individual predictor coefficients; variance inflation factors (VIFs) and the Mallows' Cp index as indices of correlation between the variables or 'multicollinearity'. Residual analysis was carried out to determine whether the assumptions of normality and homogeneity

of variance of the residuals ('homoscedasticity') were fulfilled (Brook and Arnold, 1985; Ryan, 2009; Vittinghoff et al., 2005; Kutner et al., 2005; Quinn and Keough, 2002; Field, 2011). The distribution of the residuals for the regression models was tested using the Kolmogorov-Smirnov, Shapiro-Wilk and Anderson-Darling tests.

Random forest regressions (RFRs) were also used as an exploratory technique in order to predict ln % correct performance in the STM, using R. The data set was divided into training and test data sets, using a 70:30 split. Since there were 8 predictor variables (i.e., 'surgical group', 'dist', 'sdur', 'udur', 'ln IO', 'epmdur', 'a3' and 'e3') for the target variable, ln percent correct in the STM, it was decided to set m, the number of variables used to determine the decision at a node of the tree, as p/3 (approximately), where p = the number of predictor variables, and use m = 3, with 500 trees (Hastie et al., 2009). The percentage of variance explained, the mean of the squared error (MSE), and the effect of variable removal on the MSE were used to evaluate the success of the RFR analysis. Note however that RFs are commonly used on larger data sets so, as a general technique, the analysis remains exploratory despite its success.

2.2 Biochemical Study

2.2.1 Animals

At the beginning of the study, male Wistar rats were randomly allocated to BVD or sham surgery conditions at one of 5 time points: 24 h, 72 h, 1 week, 1 month or 6 months post-surgery. For the 24 h, 72 h and 1 week time points, there were 7 animals in each of the sham and BVD groups. For the 1 month time point, there were 6 animals in the sham group and 7 in the BVD group. For the 6 month time point, there were 6 animals in each of the sham groups with and without T maze training (see below), and 7 animals in each of the BVD groups with and without T maze training. This resulted in a total of 81 animals for the entire study. Because the brains had to be harvested at one particular time point, each group was separate. Animals were maintained on a 12:12 h light:dark cycle at 22˚ C and housed in individual cages. All procedures were carried out in accordance with the regulations of the University of Otago Committee on Ethics in the Care and Use of Laboratory Animals.

2.2.2 Surgery

The animals were anaesthetized using a combination of 300 mg/kg fentanyl citrate (administered intraperitoneally (i.p.)) and 300 mg/kg medetomidine hydrochloride (i.p.) and BVD or sham surgery was performed under microscopic control as described in Section 2.1.2 (Zheng et al. 2006; Zheng et al., 2007; Zheng et al., 2008; Zheng et al., 2009). Following surgery, the animals recovered for 24 h, 72 h, 1 week, 1 month or 6 months. By 1 week the animals had recovered from the severe, acute symptoms of BVD and some compensation had occurred, although the vestibulo-ocular and vestibulo-spinal reflexes never return to normal following BVD (see Smith et al., 2005 for a review).

2.2.3 Sample sizes

The sample sizes of n = 6-7 for the sham and BVD groups were based on previous studies in which power calculations indicated a 90% power level to detect differences of interest given the variance estimates and an α rate of 5% (Liu et al., 2003; Zheng et al., 2001; Baek et al., 2011; Stiles et al., 2012). Such sample sizes are common in animal studies in neuroscience due to the need to minimise the number of animals used and also because factorial designs are used to maximise statistical power (Lenth, 2001; Festing, 2003; Small et al., 2011; Norma et al., 2012).

2.2.4 T maze training

Approximately half of the animals in the 6 month group received training in a T maze. The procedure was similar to that described in Zheng et al. (2007). The T-maze was modified from a cross maze by blocking one of the arms with a barrier at the entrance of the arm and the rest of the three arms formed a T-shaped maze (see Figure 22). Each arm was 50 cm long and 12 cm wide with a sunken food well (2 cm in diameter and 0.75 cm deep) located at the end of the arm. The walls of the arms were 20 cm high. The stem was of the same length, width, and height and a starting area was created by inserting a plastic door 25 cm from the end of the stem. The apparatus was elevated 90 cm above the ground. The rats all had one day of habituation (10 min) to the maze with the food wells in all four

arms baited. During this period, the sucrose tablets (Noyes sucrose reward tablet, 45 mg, Research Diet) were continuously replaced so that no arm was found to be empty when first visited. The training was performed according to the procedure described by Zheng et al. (2007). Basically, each trial consisted of two stages. On the first stage (also called the sample run), a barrier was placed at the neck of the T-maze to close off one arm and three sucrose tablets were placed in each food well. The sample run was started by placing the animal in the starting area and removing the barrier. The animal was forced to enter a preselected arm and allowed to eat food there. Then, the animal was picked up and placed in the starting area for a delay of 10 sec (or no delay for sessions 11– 20), during which the barrier was removed and the maze was wiped clean. On the second stage (the choice run), once the door was opened, the animal was allowed a free choice between the two arms of the T-maze. If the animal entered the arm not visited previously on the sample run, it received a reward (i.e., it was allowed to eat the food there). If the animal entered the arm visited previously, it was confined to that arm for about 10 sec and then returned to its cage. The criterion for selecting an arm was that the rat placed a hind foot in one of the arms. The animals received eight trials per session for a total of 20 sessions. For sessions 1–10, the animals were given a delay of 10 sec between the sample run and the choice run, while for sessions 11–20, the animals had choice runs straight after the sample runs with no delays. Six animals at a time were carried into the experimental room in an enclosed carry box with six individual compartments. Each of the six rats had one trial in turn so that the intertrial interval was 3–5 min. An equal number of forced right or left turns was given in a pseudorandom sequence. The number of correct choices was recorded for each session.

2.2.5 Tissue preparation

At the designated time point post-op., the animals were decapitated without anaesthesia, and the hippocampal subregions (CA1, CA2/3 and the dentate gyrus (DG)) were dissected out using the methods described previously (Zheng et al., 2001; Liu et al. 2003), and stored in a -80˚ C freezer until use.

At the time of processing, tissue buffer (containing Complete Proteinase Inhibitor, 50-mM Tris–HCl, pH 7.6) was added to the samples on ice, then the

tissue was homogenised using ultrasonification (Sonifier cell disrupter B-30, Branson Sonic Power Co.) and centrifuged at 12,000 g for 10 min at 4° C. The protein concentration in the supernatant was measured using the Bradford method. The supernatant and a Bio-Rad protein assay dye reagent concentrate were combined in a 96-well plate and analyzed using a Spectramax microplate reader (Zheng et al. 2001; Liu et al., 2003). The protein concentrations in the samples were equalized, then the tissue homogenates were mixed with gel loading buffer (50 mM Tris-HCl, 10% SDS, 10% glycerol, 10% 2-mercaptoethanol, 2 mg/ml bromophenol blue) in a ratio of 1:1 and then boiled for 5 min.

2.2.6 Western blotting

Ten μl of each sample was loaded in each well on a 7.5% SDS-polyacrylamide mini-gel and pre-stained protein markers (10–250 kDa; Bio-Rad, Precision Plus: Dual colour) were used as molecular weight markers on each gel. The samples were electrophoresed with a 90 V variable current (Bio-Rad, PowerPack 3000) until protein flattened at the stacking/resolving interface, and 180 V thereafter. The proteins were transferred to polyvinylidene-difluoride (PVDF) membranes using a transblotting apparatus (2.5 L; Bio-Rad). The transfer was performed overnight in transfer buffer (25% methanol, 1.5% glycine and 0.3% Tris-base) at a 10 V variable current (Bio-Rad PowerPack 3000).

Non-specific IgG binding was blocked by incubation with 5% dried milk protein (Pams) and 0.1% bovine serum albumin (BSA) (Sigma) for 6-7 h at 4˚ C. The membranes were then incubated with affinity-purified polyclonal goat antibodies raised against GluR1, GluR2, GluR3 or GluR 4, and affinity-purified polyclonal rabbit antibodies raised against NMDA ζ1 (NR1), NMDA ε1 (NR2A), NMDA ε2 (NR2B), CaMKIIα or phosphorylated CaMKIIα (pCaMKIIα), overnight at 4° C (see antibody details in Table 12). For the 1 and 6 month time points, only antibodies to the GluR1, GluR2, GluR3, NR1, NR2B, CaMKIIα and pCaMKIIα proteins were used. The secondary antibodies were anti-goat IgG linked to horseradish peroxidase and anti-rabbit IgG linked to horseradish peroxidase (see details in Table 12). Detection was performed using the enhanced chemiluminescence (ECL) system (Amersham Biosciences, NZ). Hyperfilms (Amersham Biosciences, NZ) were analyzed by densitometry to determine the

quantity of protein expressed in each group using a calibrated imaging densitometer (Bio-Rad) and a PowerPC Mac running OS 9.2 and Quantity One software.

Results were expressed as the volume of the band, i.e., optical density × area of the band. An antibody against β-actin (see details in Table 12) was used as a loading control to ensure that the same amount of protein was loaded in each lane, and the density of each target band was then expressed as a percentage of its corresponding loading control. Exploratory regression analyses have shown that any changes in β-actin expression were unlikely to account for changes in the target protein expression (R^2=0.087) (Zheng et al. 2001).

Primary Antibody	Dilution	Secondary Antibody	Dilution
GluR1 (C-19, sc-7609), Santa Cruz	1:1000	donkey-anti-goat IgG, Sigma	1:5000
GluR2 (N-19, sc-7611), Santa Cruz	1:1000	donkey-anti-goat IgG, Sigma	1:5000
GluR3 (N-19, sc-7613), Santa Cruz	1:1000	donkey-anti-goat IgG, Sigma	1:5000
GluR4 (C-20, sc-7614), Santa Cruz	1:1000	donkey-anti-goat IgG, Sigma	1:5000
NR1 (H-54, sc-9056), Santa Cruz	1:1000	goat-anti-rabbit, Sigma	1:1000
NR2A (H-50, sc-9057), Santa Cruz	1:1000	goat-anti-rabbit, Sigma	1:1000
NR2B (H-300, sc-9058), Santa Cruz	1:1000	goat-anti-rabbit, Sigma	1:1000
CaMKII (H-300, sc-13082), Santa Cruz	1:1000	goat-anti-rabbit, Sigma	1:1000
p-CaMKIIα (Thr 286, sc-12886-R), Santa Cruz	1:1000	goat-anti-rabbit, Sigma	1:1000
β-actin (I-19, sc-1616), Santa Cruz	1:5000	donkey-anti-goat IgG, Sigma	1:5000

Table 12: Primary and secondary antibodies and their dilutions used in the western blotting experiments.

2.2.7 Statistical analyses

The data were tested for normality and homogeneity of variance. If necessary, they were natural log transformed and then re-tested. A series of 1 or 2-way multivariate analyses of variance (MANOVAs), using surgery and time point (24 h, 72 h and 1 week data), surgery (1 month data) or surgery and training (6

month data) as between group, fixed factors, were performed in SPSS 20 for each individual hippocampal region, with the 7-9 proteins as dependent variables (Marcoulides and Hershberger, 1997; Manly, 2005). Box's M test was used to assess the homogeneity of the covariance matrices. Pillai's Trace statistic was used as the test statistic because it has been reported to be more robust against violation of assumptions than other MANOVA statistics (Stevens, 2002; Manly, 2005). Since the tissue processing for the early time points (24 h, 72 h and 1 week), the middle time point (1 month) and the late time point (6 months) was done at different times, the data from these 3 conditions were analysed using separate MANOVAs, followed by univariate ANOVAs in the case of a significant MANOVA. For the 1 and 6 month conditions, GluR4 and NR2A were not analysed. In order to further investigate the data, and determine whether combinations of variables were changing in addition to individual variables, a series of further multivariate statistical analyses was performed. Linear discriminant analysis (LDA) was performed, to determine whether surgical group could be predicted from the protein expression profiles, with Wilks' λ as the test statistic and leave one out cross-validation (Manly 2005; Tabachnick and Fidell, 2007). Spearman's correlation coefficient was used to explore the correlations between the different neurochemical variables in the 6 month data set. A RF classification analysis was used on the 6 month data (split 70:30 into training and test samples) as an exploratory technique in order to determine whether surgical group or T maze training, could be predicted, with m = 3 (approximately the $\sqrt{p}$; Hastie et al., 2009) and 500 trees. Finally, cluster analyses were performed on the data expressed as z scores using Ward's minimal variance algorithm and squared Euclidean distance (Marcoulides and Hershberger, 1997; Manly, 2005). The significance level was set at 0.05 for all comparisons.

Consideration was given to whether PCA could be used to investigate the changes in neurochemical variables following BVD, and some preliminary analyses were performed. However, the Kaiser-Meyer-Olkin (KMO) Measure of sampling adequacy was only slightly greater than 0.5 for the sham and BVD data (KMO both 0.52), which is considered barely adequate (< 0.5 being the criterion for inadequacy; Kaiser, 1974). Therefore, PCA was not pursued any further.

3. Results

3.1 Behavioural study: Univariate statistical analyses

3.1.1 Open field maze results

3.1.1.1 Distance travelled and velocity in the OFM

Animals were tested for locomotor activity in the OFM on 5 separate occasions, first before any drug testing and then 4 times following administration of one of the drugs or their vehicles, over the period of time indicated in the Methods section. The distance and velocity data for the pre-drug test were found to be normally distributed with homogeneity of variance (see Tables 13 and 14 for examples of normality and homogeneity of variance tests for the distance data).

Tests of Normality

	surgery	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
distance	BVD	.260	8	.117	.881	8	.195
	Sham	.157	10	.200*	.975	10	.929

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 13: Kolmogorov-Smirnov and Shapiro-Wilk normality tests for the pre-drug locomotor distance data, indicating retention of the null hypothesis that the data in the BVD and sham groups were normally distributed.

Distribution of Data	CI for StDev Ratio	CI for Variance Ratio
Normal	(0.714, 3.214)	(0.510, 10.329)
Continuous	(0.946, 5.447)	(0.894, 29.674)

Tests

Method	DF1	DF2	Test Statistic	P-Value
F Test (normal)	7	9	2.14	0.285
Levene's Test (any continuous)	1	16	3.67	0.074

Table 14: F and Levene's tests of homogeneity of variance for the pre-drug locomotor distance data, indicating retention of the null hypothesis that the data in the BVD and sham groups had equal variance. The confidence

intervals (CIs) show the ratios of the standard deviations and the variances for the BVD to the sham groups.

For the 5 tests, the data were also normally distributed with homogeneity of variance in most cases. With an n of 18, it was decided that the ANOVAs would be robust against violations of these assumptions in a few cases (see Appendix 1 for an example). Although there was no significant difference between the BVD and sham groups in the pre-drug testing ($t(16) = 0.16$, $P \leq 0.88$; Figure 25A; see Appendix 1 for an example), over the drug/vehicle tests there was a large and significant increase in the distance travelled over 10 min by the BVD animals (repeated measures ANOVA: $F(1,16) = 12.57$, $P \leq 0.003$; Figure 25A). Mauchly's test of sphericity was significant ($P \leq 0.001$; see Appendix 1 for an example). Using the Greenhouse-Geisser correction, the repeated measure, drug, representing the different drug/vehicle conditions, was marginally significant ($F(1,79, 28.67) = 4.17$, $P \leq 0.03$; see Appendix 1 for an example). However, the drug x surgery interaction was not significant, indicating that the effects of the drug did not vary according to the surgery the animal received.

When velocity (in m/sec) was analysed over 10 min in the OFM before drug treatment, there was again no significant difference between the BVD and sham animals ($t(16) = 0.09$, $P \leq 093$; Figure 25B). However, when velocity was analysed over the different drug/vehicle conditions, there was a significant increase in velocity in the BVD animals compared to the sham groups $F(1,16) = 12.14$, $P \leq 0.003$; Figure 25B). Mauchly's test of sphericity was significant ($P \leq 0.0005$); however, with or without a Greenhouse-Geisser correction, neither drug nor the interaction between surgery and drug, was significant. These results indicated that the drug treatments had no effect on velocity and that this did not vary as a function of surgery.

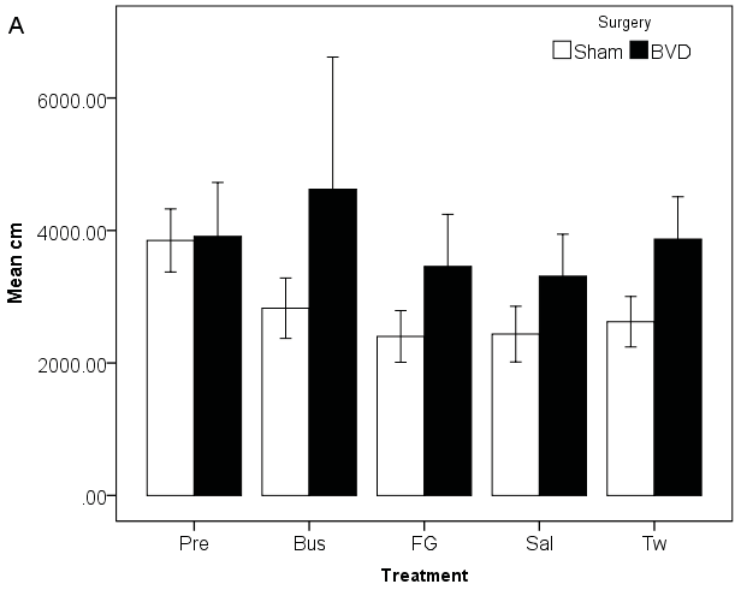


Figure 25A: Mean distance travelled, in cm, in the OFM before ('Pre') and during drug (buspirone 'Bus', FG-7142 'FG') and vehicle treatments (saline 'Sal', Tween 20 'Tw') for the BVD and sham (Sham) animals, ± 95% CI.

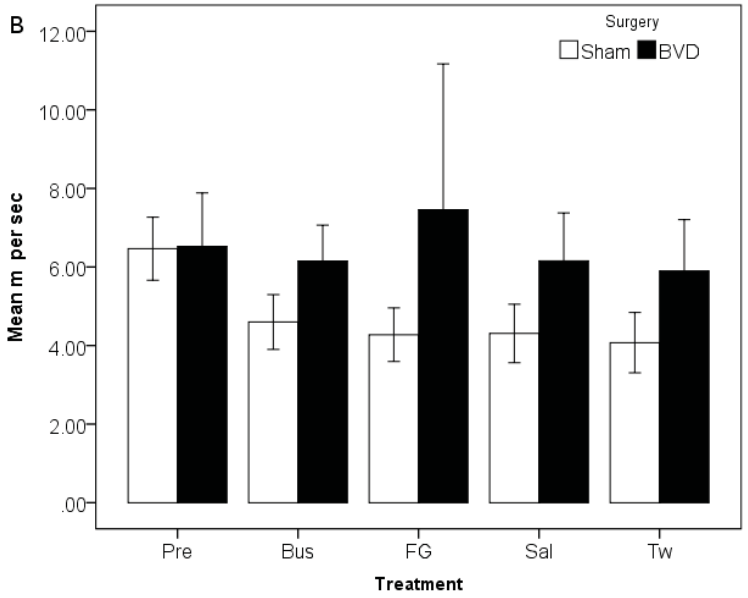
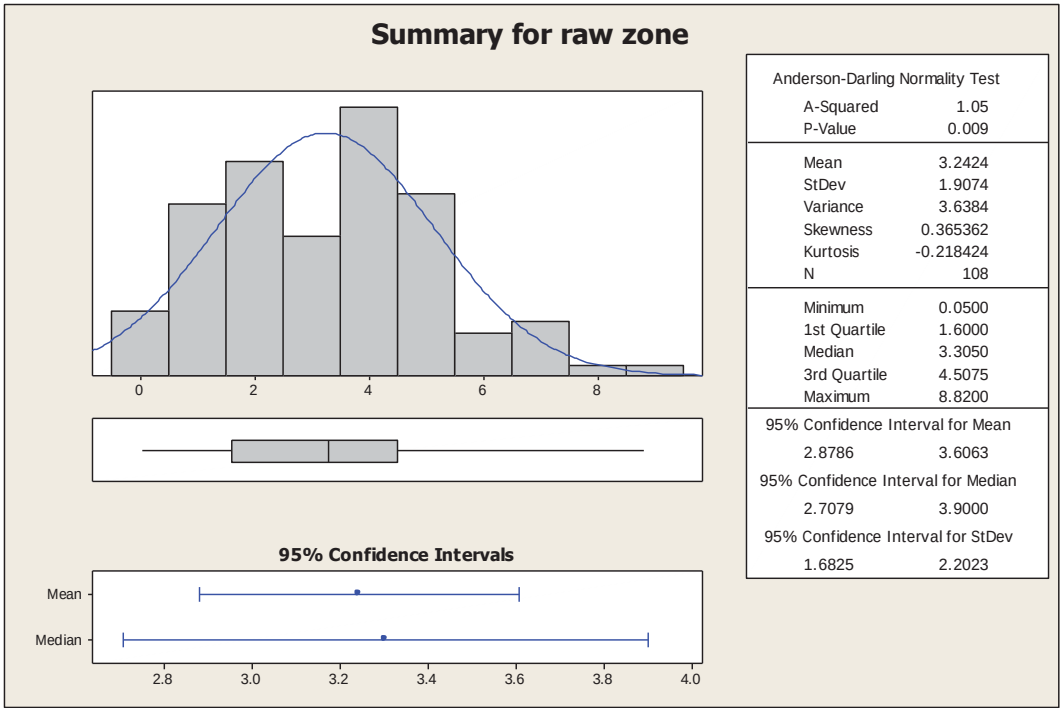


Figure 25B: Mean velocity, in m/sec, travelled in the OFM before ('Pre') and during drug (buspirone 'Bus', FG-7142 'FG') and vehicle treatments (saline 'Sal', Tween 20 'Tw') for the BVD and sham (Sham) animals, ± 95% CI.

3.1.1.2 Zone activity in the OFM

Before drug testing, an LMM analysis was used to compare the ln percentage of time spent in the outer and middle/inner zones of the OFM by BVD and sham animals. The data were extremely non-normal ($P \leq 0.0005$ for both the Kolmogorov-Smirnov and Shapiro-Wilk tests). Transformations did not resolve the problem, although the data were partially normalised by the ln transformation. This was a common problem in all of the data sets. Since the complete sample size for this experiment was 18, based on the bootstrapping performed on the data in Figure 10 with n = 8, the central limit theorem should have provided protection against violation of the normality assumption. In order to test this, 1000 samples of n = 8 were taken with replacement from the zone data and the sampling distribution of the mean compared to the raw zone data. Figure 26 shows that while the raw zone data were not normally distributed, the bootstrapped data for the sampling distribution of the mean were normally distributed. For this reason, in this and the following data sets in which there was evidence of non-normality, the transformed data were used and the central limit theorem was assumed to provide protection against data sets that were not entirely normally distributed.



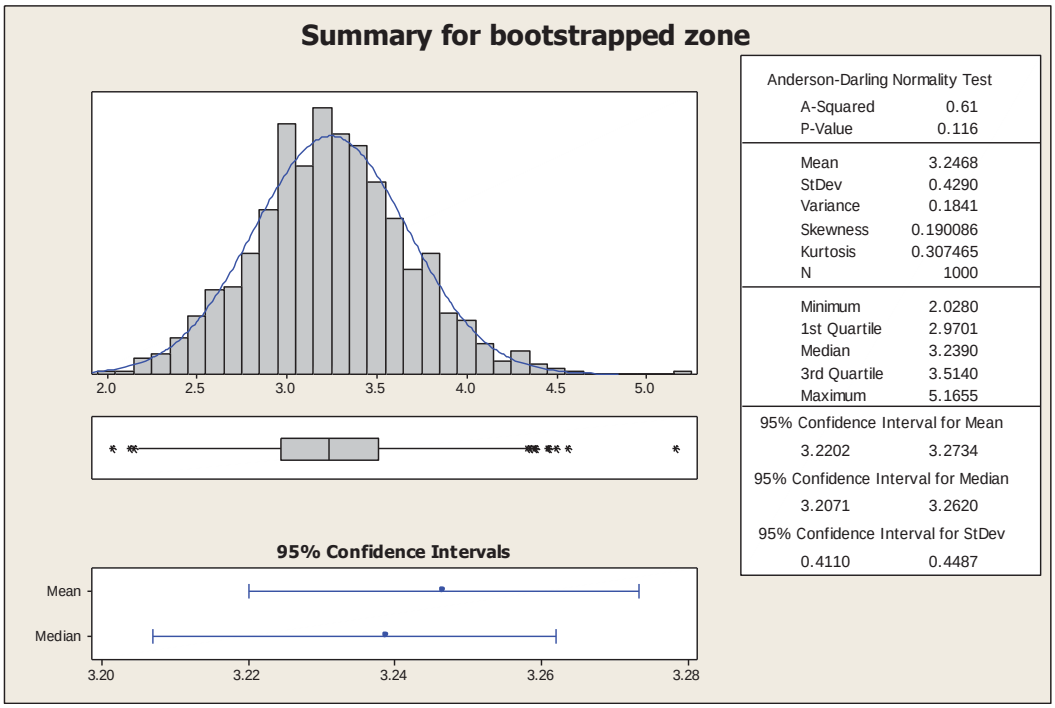


Figure 26. Comparison of the raw zone data with the bootstrapped zone data for the sampling distribution of the mean. Note that while the Anderson-Darling statistic is significant for the raw data, it is not for the bootstrapped data.

An Unstructured covariance matrix structure resulted in the smallest Akaike's Information Criterion (AIC), so an LMM analysis with this covariance matrix structure was used (see Appendix 1 for an example). There was a large and significant surgery effect ($F(1,16) = 99.31$, $P \leq 0.0005$), with BVD animals spending more time in the inner/middle zone and less time in the outer zone (significant surgery x zone interaction, $F(1,16) = 85.15$, $P \leq 0.0005$). Zone was not significant (Figure 27A).

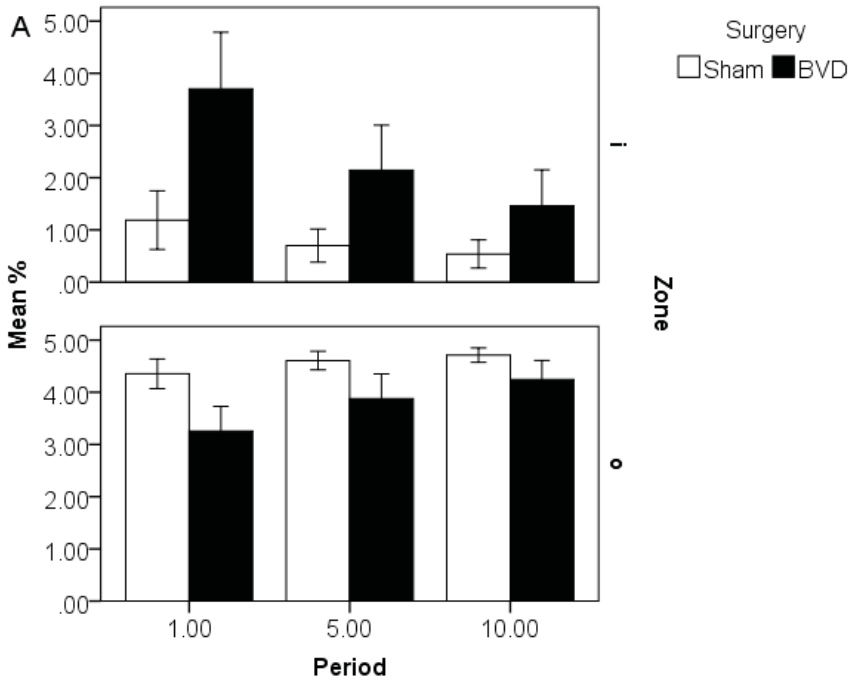


Figure 27A: Mean ln % time spent in the inner/middle (i) and outer (o) zones of the OFM for BVD and sham animals over a 1, 5 and 10 min period, ± 95% CI. The percentages were normalised to the number of grids (area) for each zone.

3.1.1.2.1 Effects of buspirone on zone activity

From the 14 covariance structure models offered by SPSS 20 for a LMM analysis, the Toeplitz model resulted in the smallest AIC, therefore this model was used. Surgery was still significant ($F(1,14.71) = 30.57$, $P \leq 0.0005$), as was zone ($F(1,57.91) = 114.63$, $P \leq 0.0005$), and the surgery x zone ($F(1,62.40) = 14.86$, $P \leq 0.0005$) and surgery x zone x drug interactions ($F(1,48.81) = 5.34$, $P \leq 0.03$; Figure 27B). The latter suggested that buspirone slightly reduced the amount of time the BVD animals spent in the inner/middle zone and increased the amount of time they spent in the outer zone. There was no significant main effect of drug and no other significant interactions.

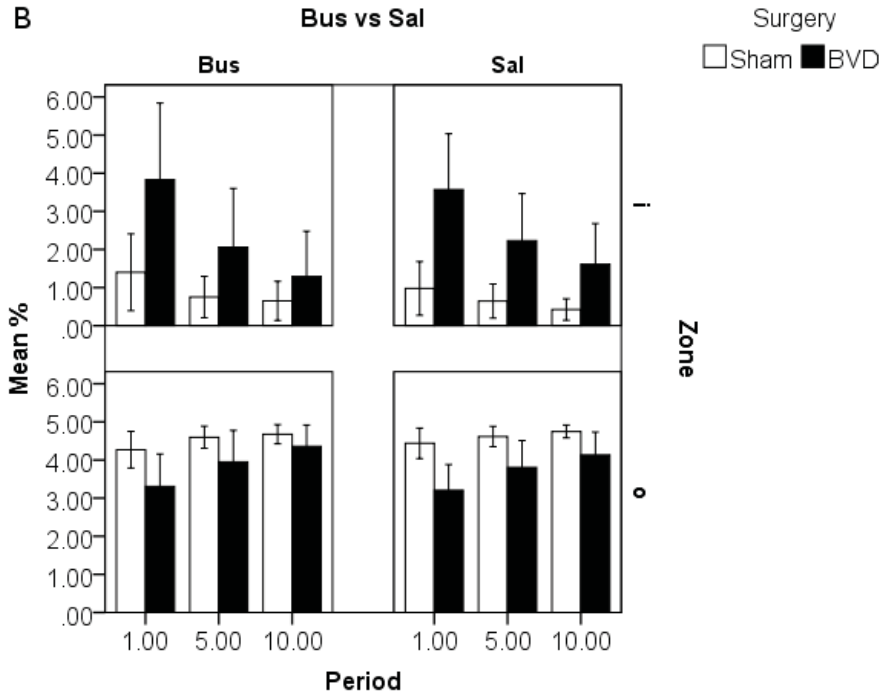


Figure 27B: Mean ln % time spent in the inner/middle (i) and outer (o) zones of the OFM for BVD and sham animals treated with buspirone (Bus) or vehicle (saline, Sal) over a 1, 5 and 10 min period, ± 95% CI. The percentages were normalised to the number of grids (area) for each zone.

3.1.1.2.2 Effects of FG-7142 on zone activity

The smallest AIC was provided by a Heterogeneous Compound Symmetry matrix structure model, and therefore an LMM analysis was carried out using this covariance structure. Surgery was still significant ($F(1,29.40) = 24.24$, $P \leq 0.0005$), as was zone ($F(1,81.43) = 526.47$, $P \leq 0.0005$), and the surgery x zone ($F(1,81.39) = 51.55$, $P \leq 0.0005$) and zone x drug ($F(1,77.55) = 7.84$, $P \leq 0.006$; Figure 27C) interactions. The significant zone x drug interaction appeared to be due to the FG-7142 changing the amount of time the BVD and sham animals spent in the inner/middle zone over the 3 periods. There was no significant main effect of drug and no other significant interactions.

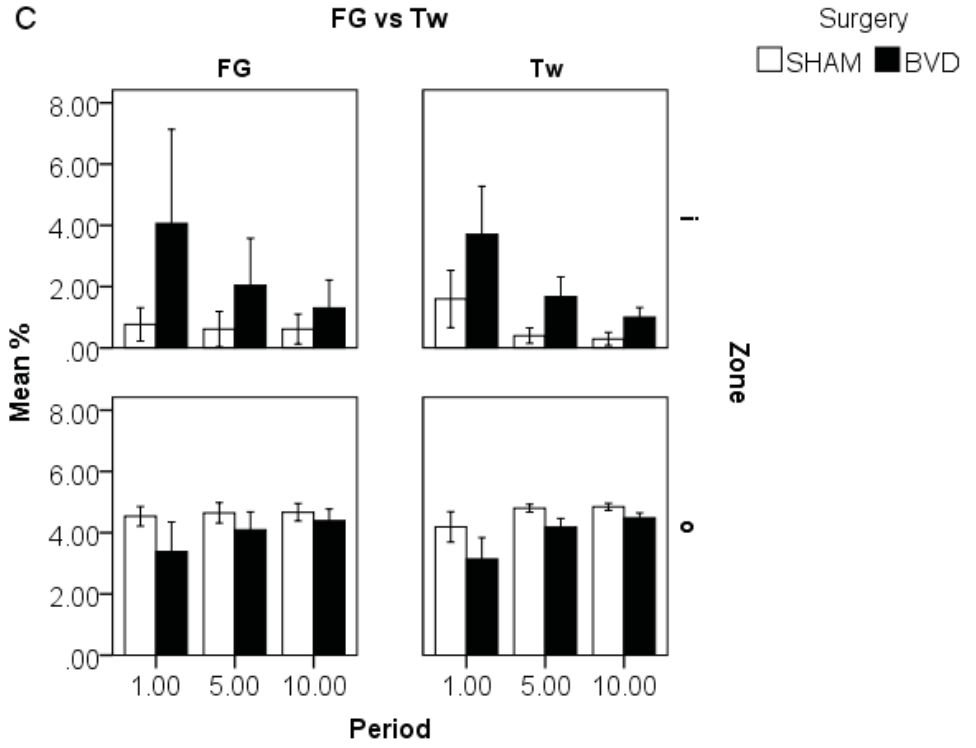


Figure 27C: Mean ln % time spent in the inner/middle (i) and outer (o) zones of the OFM for BVD and sham animals treated with FG-7142 ('FG') or vehicle (tween 80, 'Tw') over a 1, 5 and 10 min period, ± 95% CI. The percentages were normalised to the number of grids (area) for each zone.

3.1.1.3 Pre-drug rearing in the OFM

The pre-drug data fulfilled the assumptions of normality and homogeneity of variance; therefore, it was not transformed and was analysed using a 2 sample independent t test. There was no significant difference in the frequency of supported rearing between the BVD and sham groups (Figure 28); however, the frequency of unsupported rearing (Figure 29) was significantly less in the BVD group ($t(16) = -3.72$, $P \leq 0.002$) and the durations of supported and unsupported rearing were significantly lower in the BVD group ($t(16) = -6.65$, $P \leq 0.0005$ and $t(16) = -4.59$, $P \leq 0.0005$, respectively; Figure 30A and B).

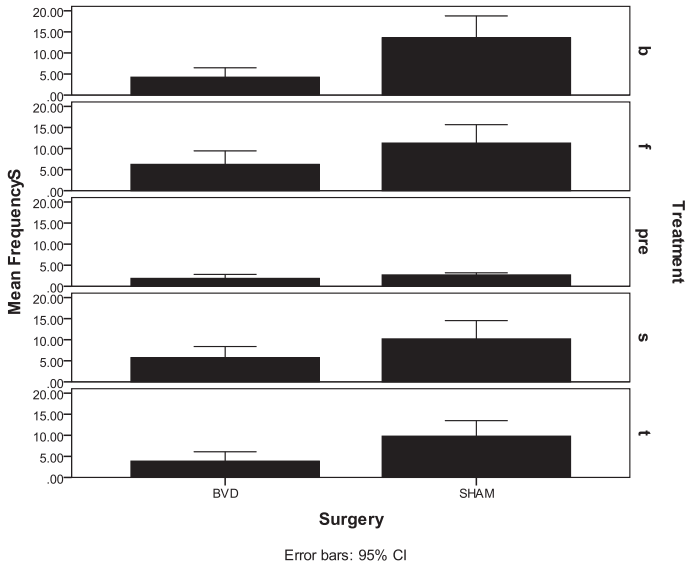


Figure 28: Mean frequency of supported rearing ('FrequencyS') in the OFM in BVD and sham groups in the pre-drug ('pre'), buspirone ('b'), FG-7142 ('f'), saline ('s') and Tween-20 ('t') conditions, ± 95% CI.

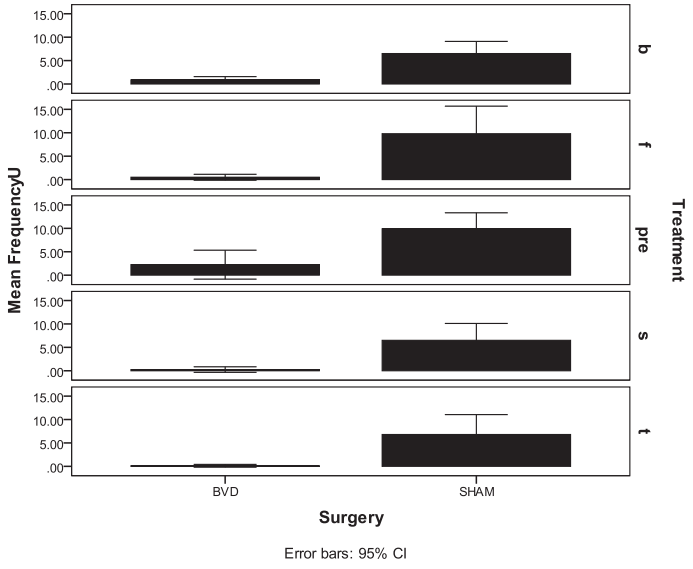


Figure 29: Mean frequency of unsupported rearing ('FrequencyU') in the OFM in BVD and sham groups in the pre-drug ('pre'), buspirone ('b'), FG-7142 ('f'), saline ('s') and Tween-20 ('t') conditions, ± 95% CI.

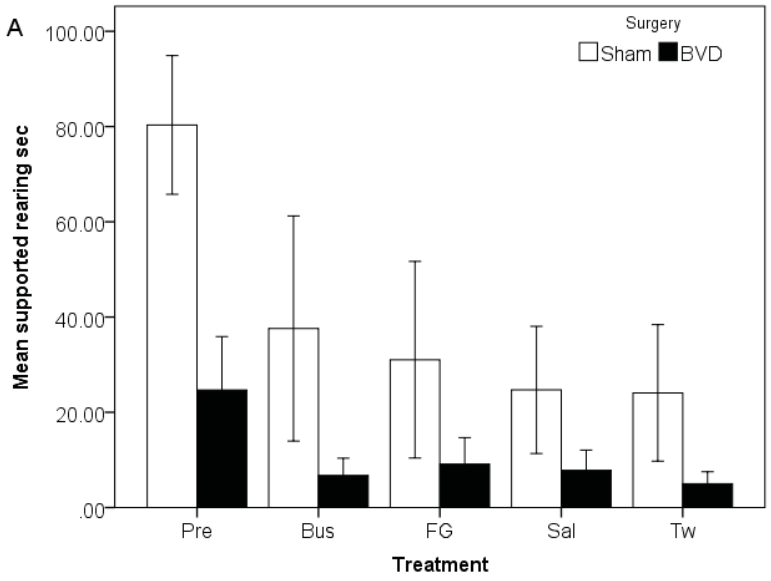


Figure 30A: Mean duration of supported rearing ('DurationS') in the OFM in BVD and sham groups in the pre-drug ('Pre'), buspirone ('Bus'), FG-7142 ('FG'), saline ('Sal') and Tween-20 ('Tw') conditions, ± 95% CI.

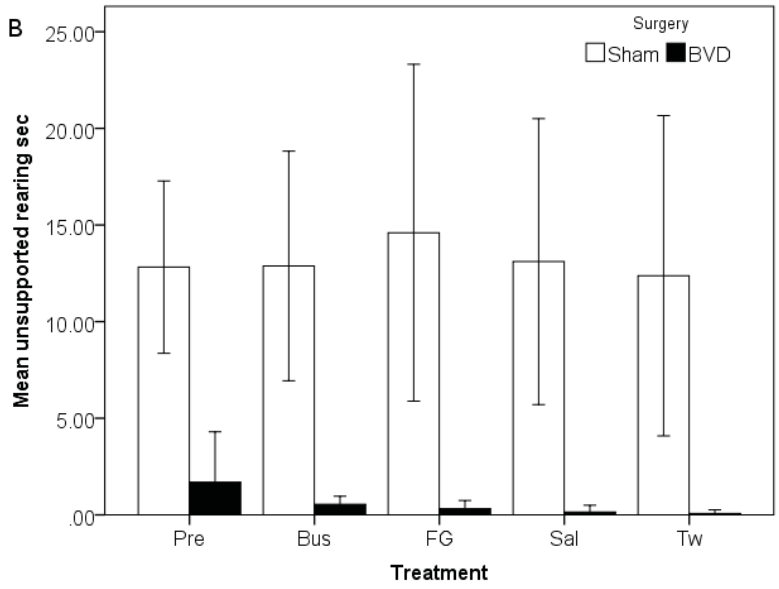


Figure 30B: Mean duration of unsupported rearing ('DurationU') in the OFM in BVD and sham groups in the pre-drug ('Pre'), buspirone ('Bus'), FG-7142 ('FG'), saline ('Sal') and Tween-20 ('Tw') conditions, ± 95% CI.

3.1.1.3.1 Effects of buspirone on rearing

The effects of buspirone and its vehicle were analysed using a repeated measures ANOVA. When treated with buspirone, surgery still had a significant effect on the frequency of supported rearing, with a large decrease in the BVD group compared to the sham group (repeated measures ANOVA: $F(1,16) = 8.88$, $P \leq 0.009$; Figure 28). Buspirone itself had no significant effect, but the interaction between surgery and drug was marginally significant due to the drug increasing the frequency of supported rearing in the sham group ($F(1,16) = 5.47$, $P \leq 0.03$).

Surgery had a significant effect on the frequency of unsupported rearing, with a large decrease in the BVD group compared to the sham group ($F(1,16) = 25.92$, $P \leq 0.0005$; Figure 29). However, there was no significant drug effect and no significant interactions.

For the duration of supported rearing, surgery was significant, with a large decrease in the BVD group compared to the sham group ($F(1,16) = 7.38$, $P \leq 0.02$; Figure 30A). However, there was no significant drug effect and no significant interactions.

For the duration of unsupported rearing, surgery was significant, with a large decrease in the BVD group compared to the sham group ($F(1,16) = 32.76$, $P \leq 0.0005$; Figure 30B). Drug treatment was not significant, nor was the interaction between surgery and drug.

3.1.1.3.2 Effects of FG-7142 on rearing

When treated with FG-7142, surgery still had a significant effect on the frequency of supported rearing, with a large decrease in the BVD group compared to the sham group (repeated measures ANOVA: $F(1,16) = 7.42$, $P \leq 0.02$; Figure 28). FG-7142 had no significant effect and there was no significant interaction between surgery and drug. Surgery had a significant effect on the frequency of unsupported rearing, with a large decrease in the BVD group compared to the sham group ($F(1,16) = 10.36$, $P \leq 0.005$; Figure 29). FG-7142 also had a significant effect on the frequency of unsupported rearing ($F(1,16) = 7.24$, $P \leq 0.02$) and there was a significant interaction between drug and surgery ($F(1,16) = 4.38$, $P \leq 0.05$), due to the drug increasing the frequency of rearing in the BVD animals compared to the controls.

For the duration of supported rearing, surgery was significant, with a large decrease in the BVD group compared to the sham group ($F(1,16) = 6.06$, $P \leq 0.03$; Figure 30A). However, there was no significant drug effect and no significant interactions.

For the duration of unsupported rearing, surgery was significant, with a large decrease in the BVD group compared to the sham group ($F(1,16) = 11.61$, $P \leq 0.004$; Figure 30B). Drug treatment was not significant and nor was the interaction between surgery and drug.

3.1.2 Elevated plus maze

Performance in the elevated plus maze (EPM) was evaluated before any drug treatment and then after treatment with buspirone, FG-7142 or their vehicles. The percentage of time spent in the open arms relative to the total time (10 mins), the frequency of open arm entries divided by the total number of entries, and the distance travelled, during 10 min, were analysed for the BVD and sham controls before drug treatment using a 2 sample independent t test. There were no significant differences in either the percentage of time spent in the open arms or the frequency of open arm entries (Figure 31A).

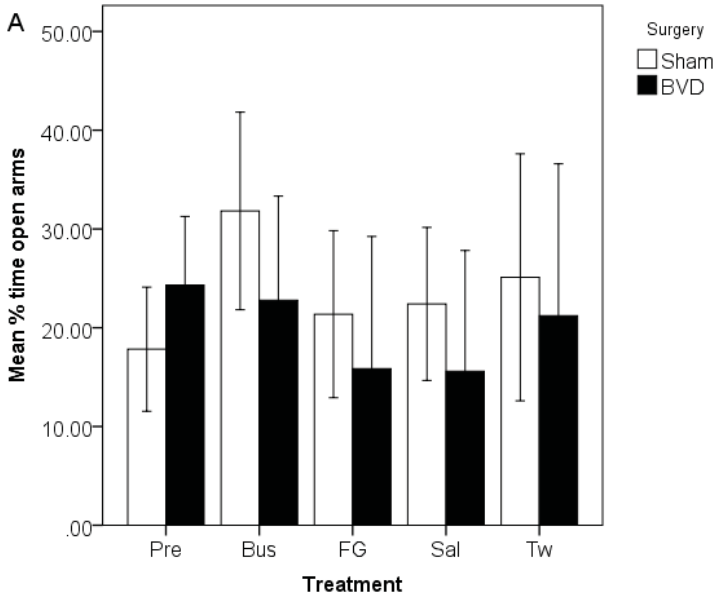


Figure 31A: Mean % time in the open arms for the BVD and sham groups pre-drug ('Pre') and in response to buspirone ('Bus'), FG-7142 ('FG'), saline ('Sal') and Tween 20 ('Tw'), ± 95% CI..

However, BVD animals did travel significantly further in the EPM compared to sham controls ($t(16) = 3.44$, $P \leq 0.003$; Figure 31B).

In the buspirone and FG-7142 conditions, neither surgery, drug nor their interactions, had any significant effect on the frequency or duration of open arm entries, or the distance travelled (Figures 31A and B).

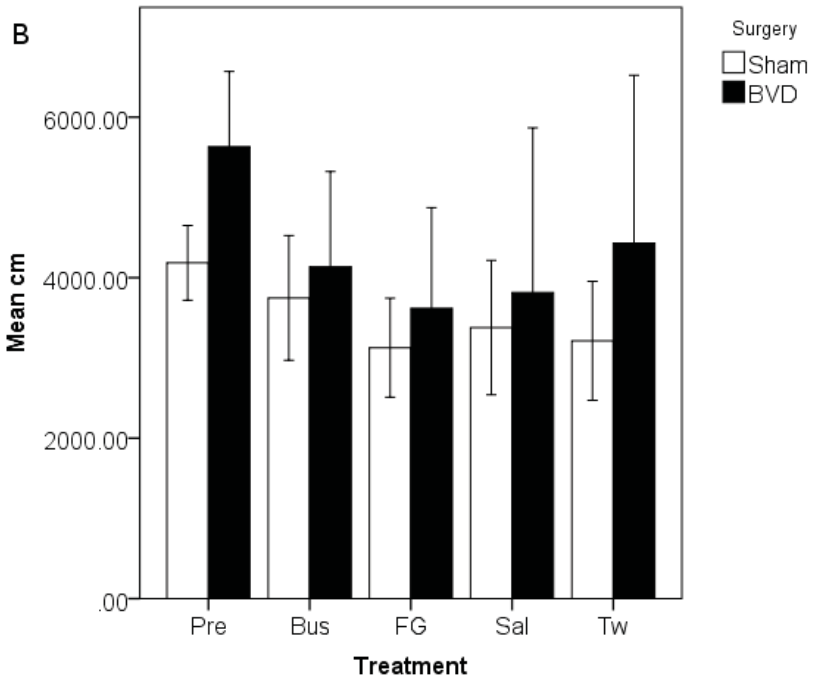


Figure 31B: Mean distance travelled in the EPM for the BVD and sham groups pre-drug ('Pre') and in response to buspirone ('Bus'), FG-7142 ('FG'), saline ('Sal') and Tween 20 ('Tw'), ± 95% CI.

3.1.3 Elevated T maze

3.1.3.1 Pre-drug performance

The data were analysed using a 2 way repeated measures ANOVA since there were only 3 repeated measures (i.e. the 3 avoidance trials). Mauchly's test of sphericity indicated that the assumption of sphericity was violated; therefore, the Greenhouse-Geisser correction was used for the repeated measure and its interactions. When avoidance learning was examined in the ETM before drug treatment, there was no significant difference between the BVD and sham groups in the latency to leave the enclosed arm. Trial was significant (repeated measures

ANOVA: $F(1.77,28.32) = 10.15$, $P \leq 0.001$), indicating that the animals increased their latency to leave the enclosed arm over the 3 trials, i.e. they learned to avoid the open arm (Figure 32A). However, the trial x surgery interaction was not significant. When escape learning was examined in the ETM before drug treatment, there was a significant difference between the BVD and sham groups in the latency to escape the open arm, with BVD animals exhibiting a longer latency ($F(1,16) = 5.56$, $P \leq 0.03$; Figure 32B). Neither trial nor the interaction between surgery and trial was significant.

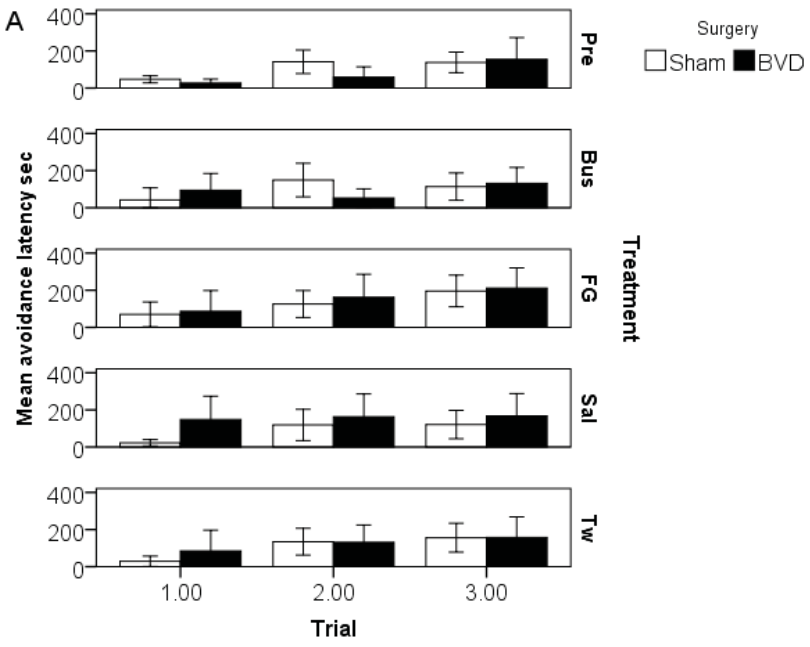


Figure 32A: Mean avoidance latency ('Alatency') in sec in the 3 avoidance trials in the ETM for the BVD and sham groups in the pre-drug condition ('Pre') and in response to buspirone ('Bus'), FG-7142 ('FG'), saline ('Sal') and Tween 20 ('Tw'), ± 95% CI.

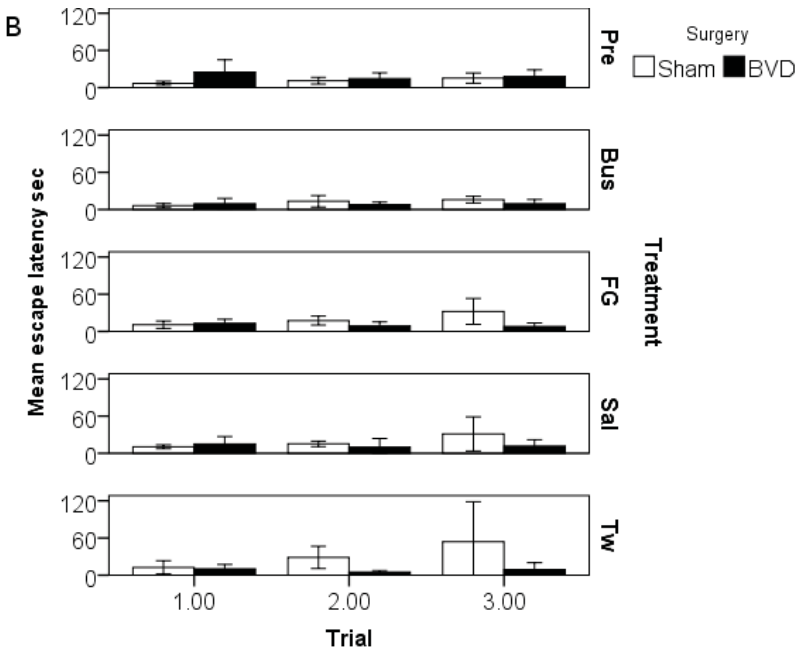


Figure 32B: Mean escape latency ('Elatency') in sec in the 3 avoidance trials in the ETM for the BVD and sham groups in the pre-drug condition ('Pre') and in response to buspirone ('Bus'), FG-7142 ('FG'), saline ('Sal') and Tween 20 ('Tw'), ± 95% CI.

3.1.3.2 Effects of buspirone on avoidance and escape

For avoidance latency in the buspirone condition, an ARMA (1,1) covariance matrix structure had the smallest AIC. Surgery was not significant and neither was drug. However, there were significant trial (LMM: $F(2,68.72) = 5.75$, $P \leq 0.005$; Figure 32A), surgery x drug ($F(1,26.98) = 4.69$, $P \leq 0.04$), and surgery x trial effects ($F(2,62.72) = 6.10$, $P \leq 0.004$). The significant surgery x drug interaction appeared to be due to buspirone reducing the avoidance latency. Bonferroni post-hoc comparisons showed that the significant trial effect was due mainly to differences between trial 1 and trial 2 ($P \leq 0.03$) and trial 1 and trial 3 ($P \leq 0.006$; see Appendix 1 for an example). For the escape latency, surgery was not significant and there were no significant interactions; however, buspirone did have a marginally significant effect, reducing the escape latency compared to the saline control condition ($F(1,19.09) = 5.42$, $P \leq 0.03$; Figure 32B).

3.1.3.3 Effects of FG-7142 on avoidance and escape

For avoidance latency in the FG-7142 condition, a Compound Symmetry covariance matrix structure had the smallest AIC. Neither surgery, drug nor their interaction, was significant. However, there was a significant trial effect (LMM: $F(2,79.08) = 15.65$, $P \leq 0.0005$; Figure 32A), suggesting that BVD and sham rats showed similar levels of avoidance behavour. Bonferroni post-hoc comparisons showed that this was due mainly to differences between trial 1 and trial 2 ($P \leq 0.003$) and trial 1 and trial 3 ($P \leq 0.0005$). For the escape latency, there were no significant effects except for a significant surgery x trial interaction ($F(2,16.20) = 4.07$, $P \leq 0.04$; Figure 32B).

3.1.4 Spatial T Maze Results

3.1.4.1 Pre-drug performance

The animals were tested in the spatial T maze for 8 days prior to drug testing, so an LMM analysis was used in this case. For the ln percentage of correct responses, the AR(1) covariance matrix structure was found to have the smallest AIC; therefore this model was used. The BVD animals performed significantly worse than the sham animals in terms of the mean percentage of correct responses ($F(1, 18.10) = 48.25$, $P \leq 0.0005$; Figure 33A).

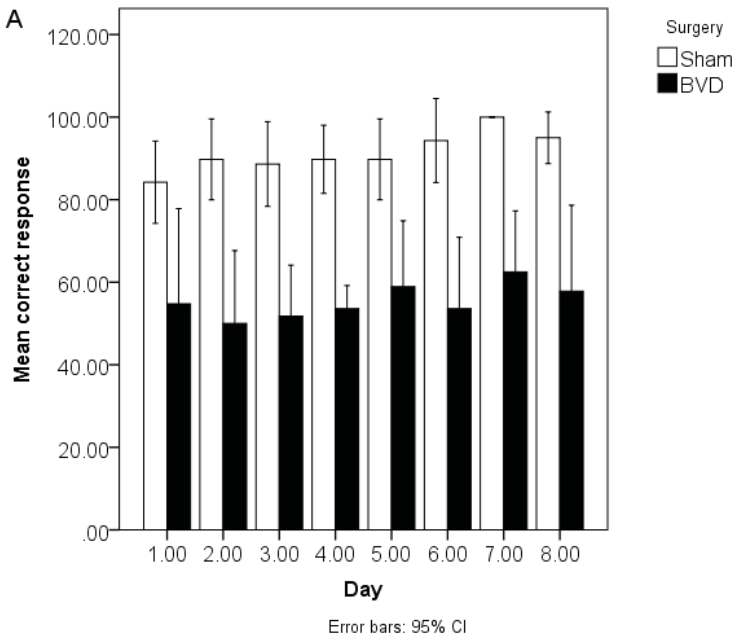


Figure 33A: Mean % correct responses in the spatial T maze task before any drug treatment for the BVD and sham (Sham) animals, ± 95% CI.

3.1.4.2 Effects of buspirone and FG-7142 on spatial memory

The effects of buspirone or FG-7142 on the percentage of correct responses in the BVD and sham animals were determined by comparing performance at 30 min following the drug or vehicle injection and 3 days post the drug or vehicle treatment. A repeated measures ANOVA was used in preference to LMM analysis in this case because there were relatively few repeated measures and Mauchly's test of sphericity was not significant, indicating that the violation of the assumption of sphericity was not a problem for the repeated measures ANOVA.

Neither the buspirone, FG-7142 nor time had any significant effect on the mean percentage of correct responses. However, the effect of surgery was significant for both the buspirone and FG-7142 data, with BVD animals performing significantly worse than sham animals ($F(1,16) = 30.58$, $P \leq 0.0005$ and $F(1,16) = 35.33$, $P \leq 0.0005$, respectively; Figure 33B). There were no significant interactions between surgery and drug, drug and time, time and surgery or drug x time x surgery.

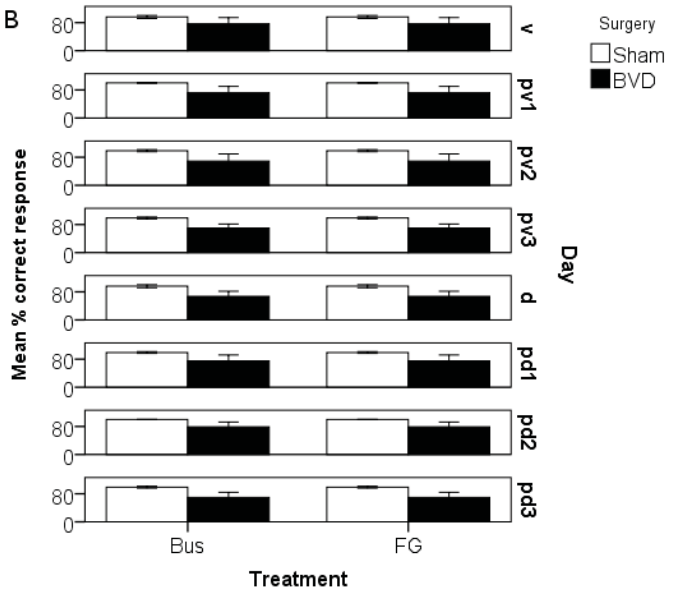


Figure 33B: Effect of buspirone ('Bus') or FG-7142 ('FG') on mean % correct responses on the day of treatment (vehicle 'v' or drug 'd') and the 3 following days (pv1, pv2, pv3, pd1, pd2 and pd3) in the BVD and sham (Sham) groups, ± 95% CI.

3.2 Behavioural study: Multivariate statistical and data mining analyses

3.2.1 MANOVA

In order to further investigate the data, a MANOVA was performed prior to linear discriminant analysis, with the independent variable, surgery, and the behavioural measures ('ln % correct', 'dist', 'sdur', 'udur', 'ln IO', 'epmdur', 'a3' and 'e3') as the dependent variables. Pillai's trace statistic was used since it is the least susceptible to violation of assumptions such as multivariate normality (Stevens, 2002; Manly, 2005). Surgery was significant ($F(8,9) = 27.33$, $P \leq 0.0005$; see Appendix 2 for an example). Levene's test of the homogeneity of the error variances indicated that only 3 of the 8 variables had significantly different variances (see Appendix 2 for an example).

3.2.2 Linear discriminant analysis

Linear discriminant analysis was used to determine whether animals could be identified as BVD or sham based on their measurements for the 8 behavioural variables ('ln % correct', 'dist', 'sdur', 'udur', 'ln IO', 'epmdur', 'a3' and 'e3'). In the first instance, all variables were entered together. A linear discriminant function was identified that could significantly predict the animals' group membership (Wilks' $\lambda = 0.04$, $P \leq 0.0005$). Box's M test for the homogeneity of the covariance matrices, was significant ($P \leq 0.007$; see Table 15). However, because the sample sizes were equal ($n = 18$ per group), it was assumed that the LDA would be robust against violation of this assumption, as argued by Stevens (2002) and Field (2011). The standardized canonical discriminant function coefficients are shown in Table 16 below. Furthermore, cross-validation showed that this discriminant function was 100% successful in classifying the animals as BVD or sham (Table 17).

A stepwise linear discriminant analysis was used to determine whether the number of variables in the linear equation could be reduced while still retaining predictive power. A linear function was identified that could significantly predict the animals' group membership (Wilks' $\lambda = 0.06$, $P \leq 0.0005$; see Appendix 2 for an example). The standardized canonical discriminant function coefficients are shown in Table 18. Cross-validation showed that this discriminant function, which was based only on the duration of unsupported rearing, ln zone activity, and ln %

correct in the STM, was also 100% successful in classifying the animals as BVD or sham (Table 19).

Test Results

Box's M		22.405
F	Approx.	2.945
	df1	6
	df2	1582.708
	Sig.	.007

Tests null hypothesis of equal population covariance matrices.

Table 15: Box's M test for the equality of the covariance matrices, which was significant at P ≤ 0.007.

Standardized Canonical Discriminant Function Coefficients

	Function
	1
sduration	.266
uduration	.844
distance	-.374
epmdur	-.254
atrial3	.245
etrial3	.135
lnIO	-.659
lnpercent	1.085

Table 16: The standardized canonical discriminant function coefficients for the LDA with all variables entered. 'sduration': duration of supported rearing. 'uduration': duration of unsupported rearing. 'distance': distance

travelled in the OFM. 'epmdur': duration of open arm entries in the EPM. 'atrial3': 3rd avoidance latency trial in the ETM. 'etrial3': 3rd escape latency trial in the ETM. 'lnIO': ln of the ratio of time spent in the inner to the outer zones in the OFM. 'lnpercent': ln percent correct in the STM.

Classification Results[a,c]

		Group	Predicted Group Membership		Total
			.00	1.00	
Original	Count	.00	10	0	10
		1.00	0	8	8
	%	.00	100.0	.0	100.0
		1.00	.0	100.0	100.0
Cross-validated[b]	Count	.00	10	0	10
		1.00	0	8	8
	%	.00	100.0	.0	100.0
		1.00	.0	100.0	100.0

a. 100.0% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 100.0% of cross-validated grouped cases correctly classified.

Table 17: The classification matrix for the LDA with all variables entered.

Standardized Canonical Discriminant Function Coefficients

	Function
	1
uduration	.810
lnIO	-.655
lnpercent	.986

Table 18: The standardized canonical discriminant function coefficients for the stepwise LDA. Abbreviations as in Table 15.

Classification Results[a,c]

		Group	Predicted Group Membership		Total
			.00	1.00	
Original	Count	.00	10	0	10
		1.00	0	8	8
	%	.00	100.0	.0	100.0
		1.00	.0	100.0	100.0
Cross-validated[b]	Count	.00	10	0	10
		1.00	0	8	8
	%	.00	100.0	.0	100.0
		1.00	.0	100.0	100.0

a. 100.0% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 100.0% of cross-validated grouped cases correctly classified.

Table 19: The classification matrix for the stepwise LDA.

3.2.3 Random forests to predict group membership

The RF method was also used as an exploratory technique in order to predict the surgical status of the animals from the 8 behavioural variables ('ln % correct', 'dist', 'sdur', 'udur', 'ln IO', 'epmdur', 'a3' and 'e3'). Five hundred trees were generated using 3 variables at each split. The 'out of bag' (OOB) estimate of the error rate was 0% and the confusion matrix is shown in Table 20 below.

OOB estimate of error rate: 0%

Confusion matrix:

	BVD	SHAM	class.error
BVD	5	0	0
SHAM	0	7	0

Table 20: The OOB error for the RF

Despite this, the error matrix for the RF model showed an overall error of 0.17 (Table 21).

Error matrix for the Random Forest model on irene R data.csv [test] (counts):

	Predicted	
Actual	BVD	SHAM
BVD	2	1
SHAM	0	3

Overall error: 0.1666667

Table 21: Overall error for the RF model.

Figure 34 shows the variables in order of importance for the model. Consistent with the results of the LDA, the ln % of correct responses in the STM and the duration of unsupported rearing, were important for the prediction of the surgical status of the animals. Figure 35 shows the OOB, BVD and sham group error rates as a function of the number of trees.

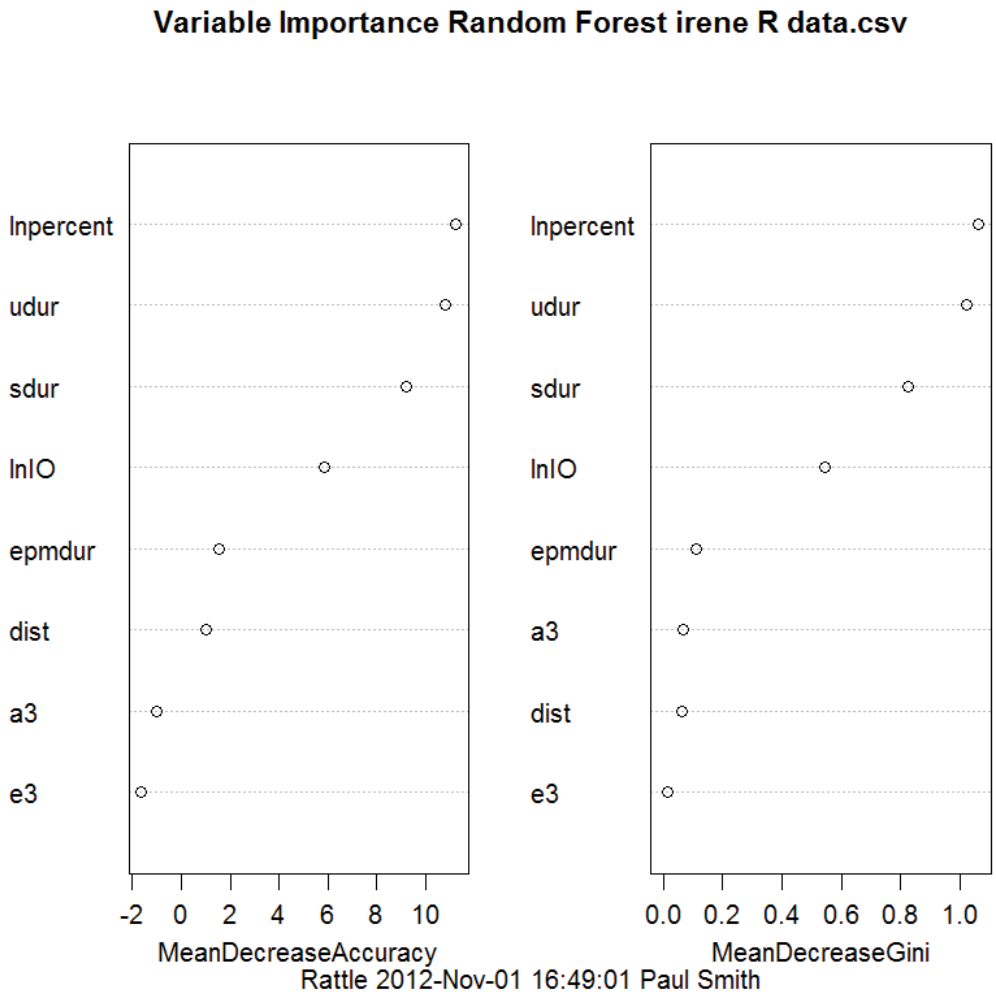


Figure 34: The variables in order of importance for the RF model.

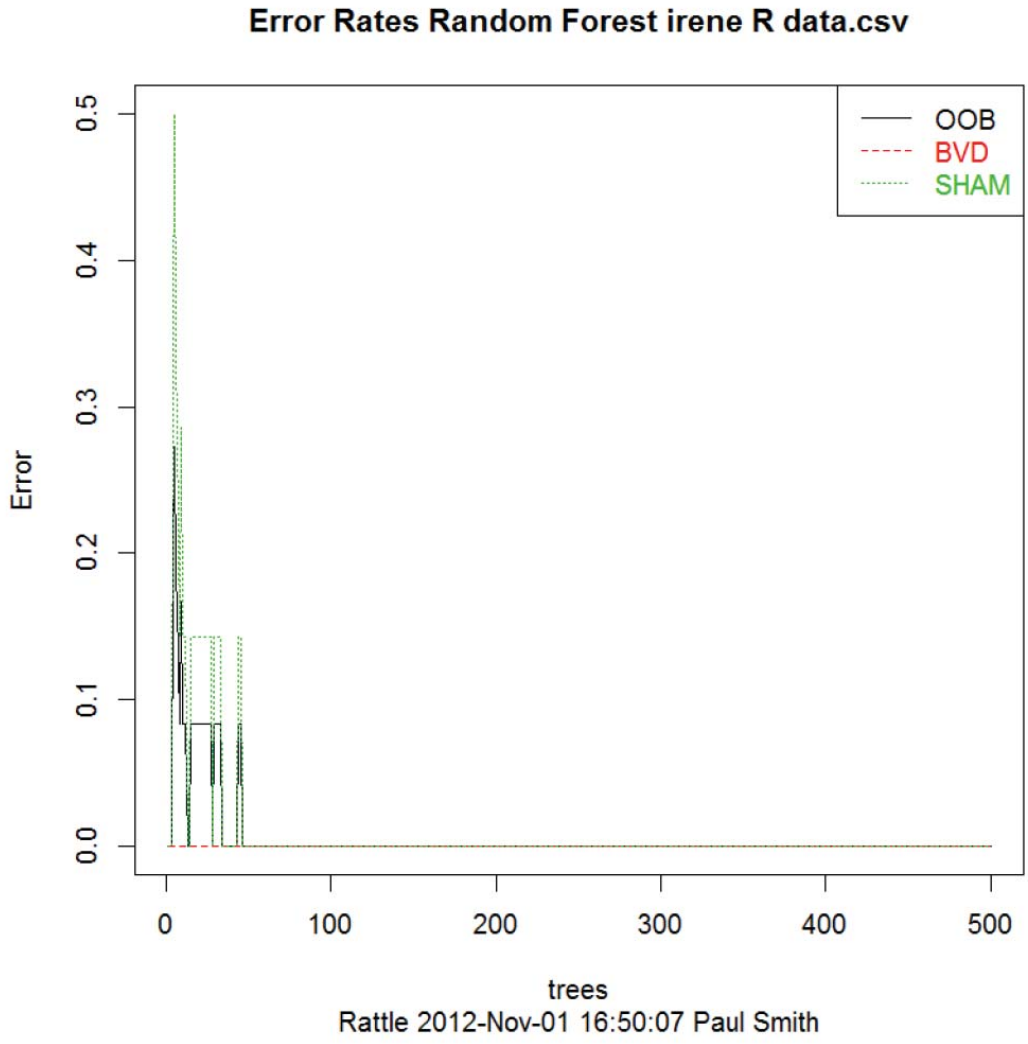


Figure 35: The OOB, BVD and sham group error rates for the RF model as a function of the number of trees.

3.2.4 Support vector machines to predict group membership

Support Vector Machines (SVMs) were also used as an exploratory technique in order to predict the group membership of the animals from the behavioural variables (i.e., 'ln % correct', 'dist', 'sdur', 'udur', 'ln IO', 'epmdur', 'a3' and 'e3') . The following kernels were used in order to test the effect of kernel function: Gaussian radial basis; polynomial; linear (vanilla); hyperbolic tangent; Laplace; Bessel; ANOVA RBF; and spline. Most of these functions (Gaussian radial basis, polynomial, linear, hyperbolic tangent, and Bessel) resulted in 0% error. The results for the Gaussian radial basis kernel function are shown below in Table 22. The exceptions were the Laplace kernel function (equivalent to an exponential kernel function; overall error rate: 50%), the Anova RBF (another radial basis kernel function; overall error rate: 17%) and the spline kernel function (overall error rate: 33%). The latter results indicate that these particular non-linear kernel functions were not successful in using the behavioural data to separate the BVD from the sham animals in higher dimensional space (Williams, 2011).

Support Vector Machine object of class "ksvm"
SV type: C-svc (classification)
parameter : cost C = 1
Gaussian Radial Basis kernel function.
Hyperparameter : sigma = 0.0619525415836334
Number of Support Vectors : 9
Objective Function Value : -3.6246
Training error : 0
Probability model included.

Error matrix for the SVM model on irene R data.csv [test] (counts):

	Predicted	
Actual	BVD	SHAM
BVD	3	0
SHAM	0	3

Overall error: 0

Table 22: Results of the SVM analysis using the Gaussian radial basis kernel function.

3.2.5 Cluster analyses

The dendrogram in Figure 36 shows that, using the combination of 8 behavioural variables, it was possible to almost, but not completely, distinguish between the BVD and sham animals, using the Ward's minimal variance algorithm, consistent with the results of the LDA. Other methods of cluster analysis, such as between group (average) linkage, within group (average) linkage, nearest neighbour (single linkage), furthest neighbour (complete linkage) centroid clustering and median clustering, were also investigated, but yielded similar results in terms of classifiying BVD and sham animals (see Appendix 3).

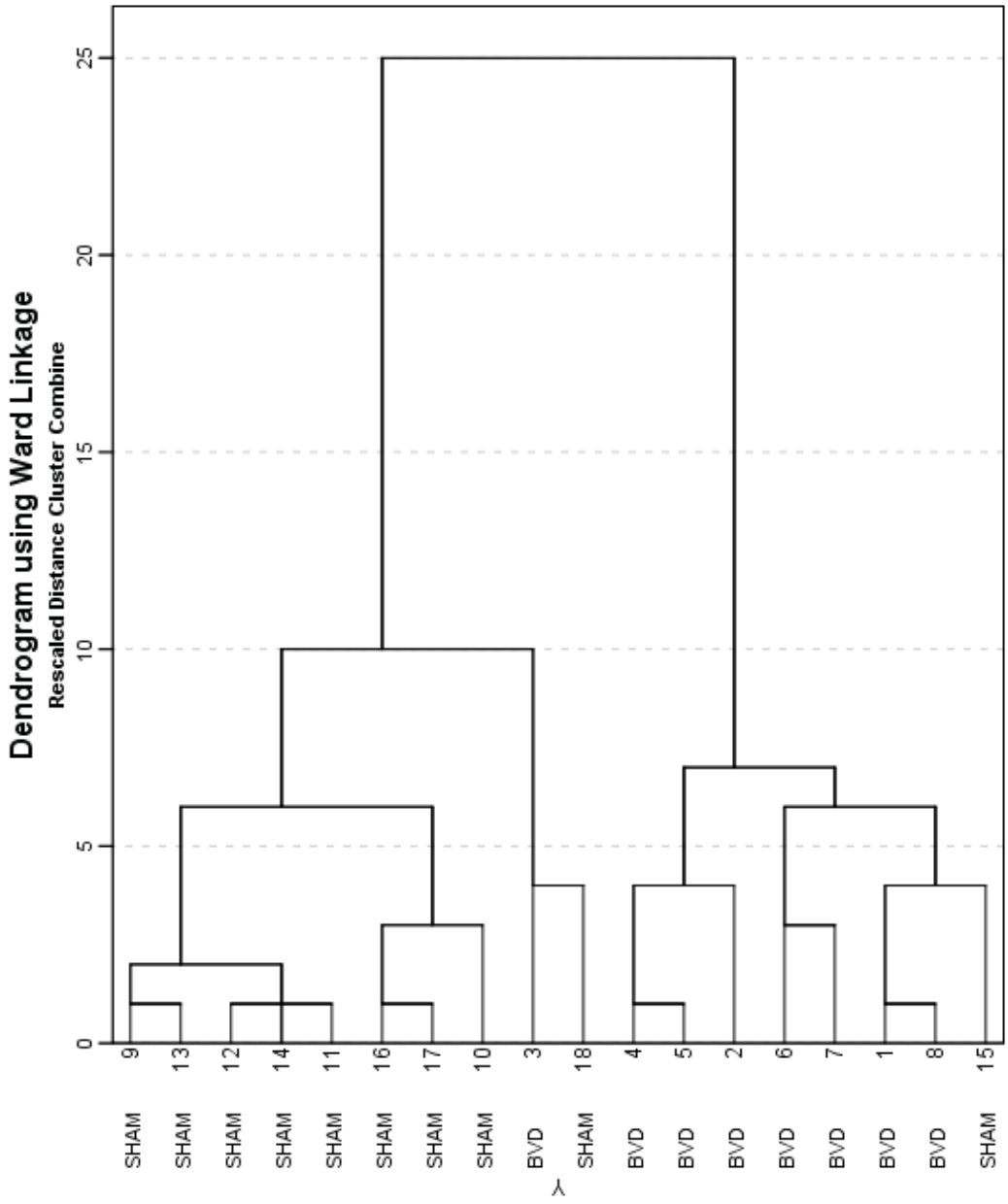


Figure 36: Cluster analysis of the behavioural data.

Behavioural Tests	Dependent Variables
OMF	Distance travelled (‘dist’), Duration Supported Rearing (‘sdur’), Duration Unsupported Rearing (‘udur’), Zone activity (‘lnIO’)
EPM	Duration Open Arm Entries (‘epmdur’)
ETM	Avoidance Latency (‘a3’), Escape Latency (‘e3’)
STM	% Correct (‘lnpercent’)

Table 23: Behavioural variables used in the MLRs.

3.2.6 Multiple linear regression

For the dependent variable, ln STM performance, with backward regression (with a probability of F to remove of ≥ 0.10), the highest adjusted R^2 of 0.90 was obtained for the predictors, distance, duration of unsupported rearing, surgical group and avoidance and escape latencies (Tables 23 and 24), which also had the smallest standard error for the regression of 0.09. The overall ANOVA was significant ($F(5,12) = 31.59$, $P \leq 0.0005$; Table 25). However, in the t tests for the individual coefficients, only surgical group ($t = -7.78$, $P \leq 0.0005$) and avoidance latency (i.e., ‘atrial3’; $t = -2.29$, $P \leq 0.04$) were significant, suggesting that the other variables were redundant. For this reason, the regression model with only surgical group and avoidance latency was considered the best, which had an adjusted R^2 of 0.88, with a standard error of 0.10 (ANOVA: $F(2,15) = 64.73$, $P \leq 0.0005$; $t = -2.34$, $P \leq 0.03$ for avoidance latency; $t = -10.92$, $P \leq 0.0005$ for group; Table 25). For the model with avoidance latency and group as predictors, the VIFs were both 1.01. Since the VIF values were < 10, multicollinearity was not considered to be a problem (Field, 2011). Finally, the distribution of the residuals for the regression model was approximately normal, since both the Kolmogorov-Smirnov and Shapiro-Wilk tests were non-significant ($P \leq 0.200$ and $P \leq 0.865$, respectively; Figure 37). Since the dependent variable was a binomial one that had been converted to a percentage, logistic regression was also used, but the results were similar (data not shown).

Model Summary[h]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.967[a]	.935	.877	.09930	.935	16.180	8	9	.000
2	.967[b]	.935	.889	.09422	.000	.004	1	9	.954
3	.965[c]	.932	.895	.09191	-.003	.468	1	10	.509
4	.964[d]	.929	.900	.08962	-.003	.409	1	11	.536
5	.961[e]	.923	.900	.08971	-.006	1.027	1	12	.331
6	.953[f]	.908	.889	.09457	-.015	2.556	1	13	.134
7	.947[g]	.896	.882	.09720	-.012	1.848	1	14	.196

a. Predictors: (Constant), lnIO, atrial3, epmdur, distance, etrial3, uduration, sduration, Group

b. Predictors: (Constant), atrial3, epmdur, distance, etrial3, uduration, sduration, Group

c. Predictors: (Constant), atrial3, distance, etrial3, uduration, sduration, Group

d. Predictors: (Constant), atrial3, distance, etrial3, uduration, Group

e. Predictors: (Constant), atrial3, distance, etrial3, Group

f. Predictors: (Constant), atrial3, etrial3, Group

g. Predictors: (Constant), atrial3, Group

h. Dependent Variable: lnpercent

Table 24: Backward regression results for the MLR.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.276	8	.160	16.180	.000[b]
	Residual	.089	9	.010		
	Total	1.365	17			
2	Regression	1.276	7	.182	20.537	.000[c]
	Residual	.089	10	.009		
	Total	1.365	17			
3	Regression	1.272	6	.212	25.095	.000[d]
	Residual	.093	11	.008		
	Total	1.365	17			
4	Regression	1.269	5	.254	31.589	.000[e]
	Residual	.096	12	.008		
	Total	1.365	17			
5	Regression	1.260	4	.315	39.149	.000[f]
	Residual	.105	13	.008		
	Total	1.365	17			
6	Regression	1.240	3	.413	46.211	.000[g]
	Residual	.125	14	.009		
	Total	1.365	17			
7	Regression	1.223	2	.612	64.733	.000[h]
	Residual	.142	15	.009		
	Total	1.365	17			

a. Dependent Variable: lnpercent

b. Predictors: (Constant), lnIO, atrial3, epmdur, distance, etrial3, uduration, sduration, Group

c. Predictors: (Constant), atrial3, epmdur, distance, etrial3, uduration, sduration, Group

d. Predictors: (Constant), atrial3, distance, etrial3, uduration, sduration, Group

e. Predictors: (Constant), atrial3, distance, etrial3, uduration, Group

f. Predictors: (Constant), atrial3, distance, etrial3, Group

g. Predictors: (Constant), atrial3, etrial3, Group

h. Predictors: (Constant), atrial3, Group

Table 25: ANOVA table for the backward regression MLR.

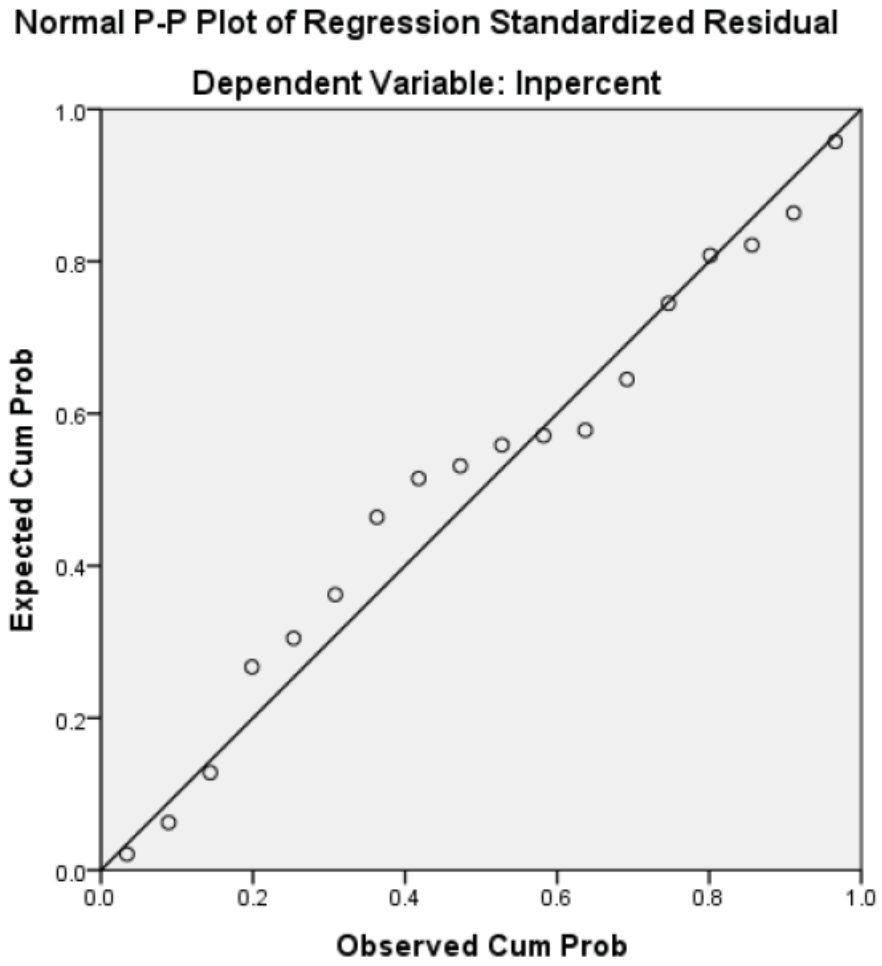


Figure 37: Normality plot for the residuals of the backward MLR analysis.

When MLR was used to determine whether any of the other 8 variables could be predicted from the remaining variables, the adjusted R^2 values were generally low (i.e. < 0.5), except for the durations of supported (adjusted R^2 = 0.71) and unsupported rearing (adjusted R^2 = 0.69) and the ln ratio of time spent in the inner to the outer zones of the OFM (adjusted R^2 = 0.62). Since these R^2 values were not particularly high and were based on 6, 5 and 4 predictor variables (respectively), they were not investigated any further.

When a best subsets regression was carried out for ln STM performance as the dependent variable, the highest adjusted R^2 was 0.90 for the predictors, duration of unsupported rearing, distance travelled, surgical group, avoidance and escape latencies (Table 26). However, the Mallows' Cp index was less than the number of variables, suggesting that multicollinearity was a problem (Brook and Arnold, 1985; Ryan, 2009; Vittinghoff et al., 2005; Kutner et al., 2005; Quinn and

Keough, 2002; Field, 2011). The most parsimonious regression model with the best Mallows' Cp index (approximately 1.0 more than the number of variables) and the smallest standard error was that with surgical group, duration of unsupported rearing and avoidance latency (adjusted $R^2 = 0.88$, standard error = 0.10) (Table 26). Finally, the distribution of the residuals for the regression model was approximately normal, since the Anderson-Darling test was non-significant ($P \leq 0.778$; Figure 38).

The results of both the backward regression and best subsets regression indicated that the overwhelming predictor of correct performance in the STM was whether the animals had BVD surgery or not, with avoidance latency (backward and best subsets regression) and duration of unsupported rearing (best subsets regression) also contributing but less important.

Best Subsets Regression: ln% versus sdur, udur, ...

Response is ln%

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	sdur	udur	dius	Groups	epmdur	a3	e3	lnIO
1	85.7	84.8	6.0	0.11012				X				
1	59.1	56.5	43.2	0.18634	X							
1	43.9	40.4	64.4	0.21810								X
1	37.4	33.4	73.6	0.23048		X						
1	11.7	6.1	109.5	0.27371					X			
2	89.5	88.1	2.7	0.097361				X		X		
2	87.4	85.8	5.5	0.10655		X		X				
2	86.8	85.0	6.5	0.10929				X			X	
2	86.1	84.2	7.5	0.11231			X	X				
2	86.0	84.2	7.5	0.11236				X				X
3	90.8	88.8	2.9	0.094589				X		X	X	
3	90.4	88.3	3.4	0.096524			X	X		X		
3	90.0	87.8	4.0	0.098581		X		X		X		
3	89.9	87.8	4.1	0.098751				X		X		X
3	89.5	87.3	4.6	0.10065				X	X	X		
4	92.4	90.0	2.7	0.089342			X	X		X	X	
4	91.7	89.1	3.6	0.093105		X	X	X		X		
4	90.9	88.1	4.8	0.097629				X		X	X	X
4	90.8	88.0	4.8	0.097746		X		X		X	X	
4	90.8	88.0	4.8	0.097772	X			X		X	X	
5	92.9	90.0	3.9	0.089450		X	X	X		X	X	
5	92.5	89.4	4.5	0.092029	X		X	X		X	X	
5	92.5	89.3	4.5	0.092273			X	X	X	X	X	
5	92.4	89.2	4.7	0.092960			X	X		X	X	X
5	92.3	89.0	4.8	0.093581		X	X	X		X		X
6	93.2	89.6	5.4	0.091309	X	X	X	X		X	X	
6	93.0	89.3	5.7	0.092606		X	X	X	X	X	X	
6	93.0	89.2	5.8	0.093042		X	X	X		X	X	X
6	92.8	88.8	6.1	0.094545	X		X	X	X	X	X	
6	92.5	88.4	6.5	0.096054	X		X	X		X	X	X
7	93.6	89.1	7.0	0.093467	X	X	X	X	X	X	X	
7	93.3	88.6	7.4	0.095347	X	X	X	X		X	X	X
7	93.1	88.2	7.7	0.097016		X	X	X	X	X	X	X
7	92.8	87.8	8.0	0.098500	X		X	X	X	X	X	X
7	92.4	87.1	8.6	0.10129	X	X	X	X	X	X		X
8	93.6	87.8	9.0	0.098514	X	X	X	X	X	X	X	X

Table 26: Best subsets regression results.

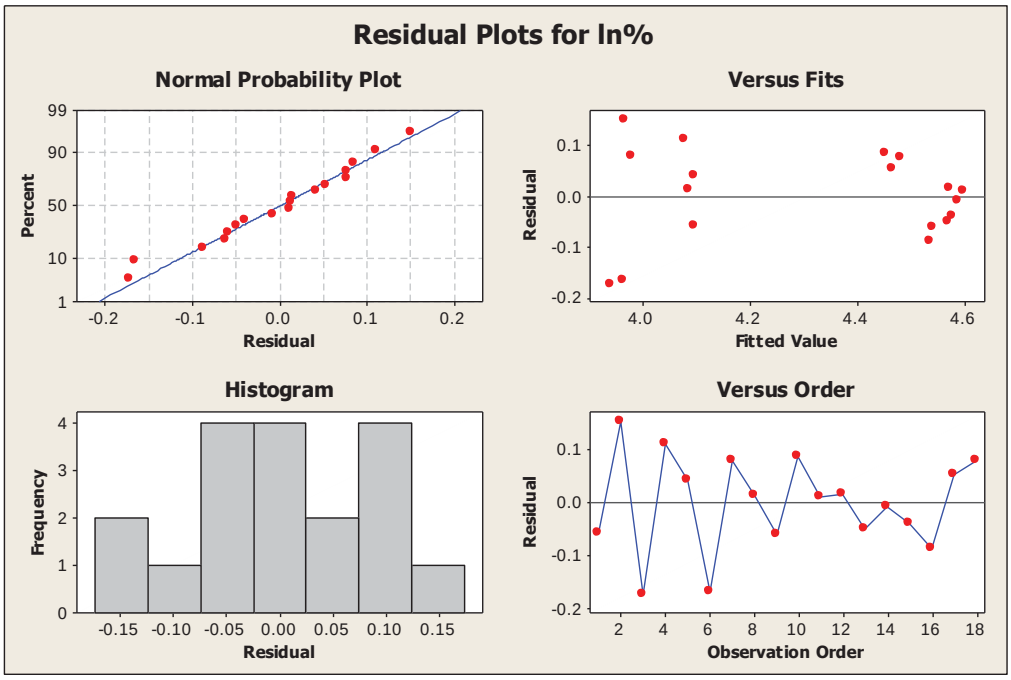


Figure 38: Residual analysis for the best subsets regression model.

3.2.7 Random forest regression

Random Forest Regression (RFR) was also used as an exploratory technique in order to determine if the ln % of correct responses in the STM could be predicted from the other 8 variables ('surgical group', 'dist', 'sdur', 'udur', 'ln IO', 'epmdur', 'a3' and 'e3'). With the data split into 70% for the training sample and 30% for the test sample, 500 trees were built using 3 variables at each split (approximately p/3, where p is the number of variables; Hastie et al., 2009). The percentage of the variance explained was 70% with a mean of the squared residuals of 0.02. Figure 39 shows the variables in order of importance for the regression. Similar to the results of the MLR, surgical group ('group') was an important variable, although duration of supported rearing ('sdur') was more important. Figure 40 shows that the error rate for the regression stabilizes after only about 50 trees. The predicted versus the observed values for the ln % of correct responses is shown in Figure 41. The 70:30 split for the training and test data samples meant that there were only 6 observations in the latter group for the ln % of correct responses, and the validity of the test results for the RFR must be considered with this in mind. However, the pseudo-R^2 value was 0.65 and RFRs

with different training: test sample splits, e.g. 50:50, yielded similar results (data not shown).

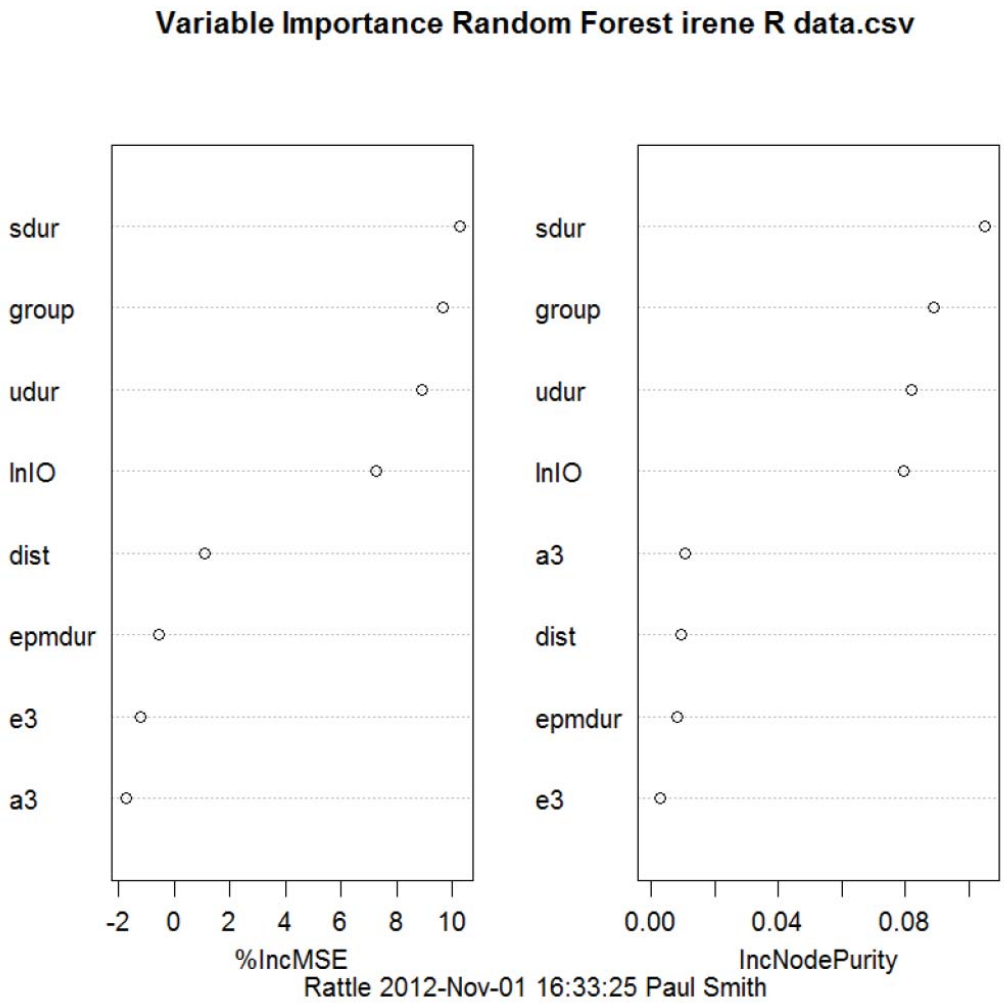


Figure 39: Variables in order of importance for the RFR.

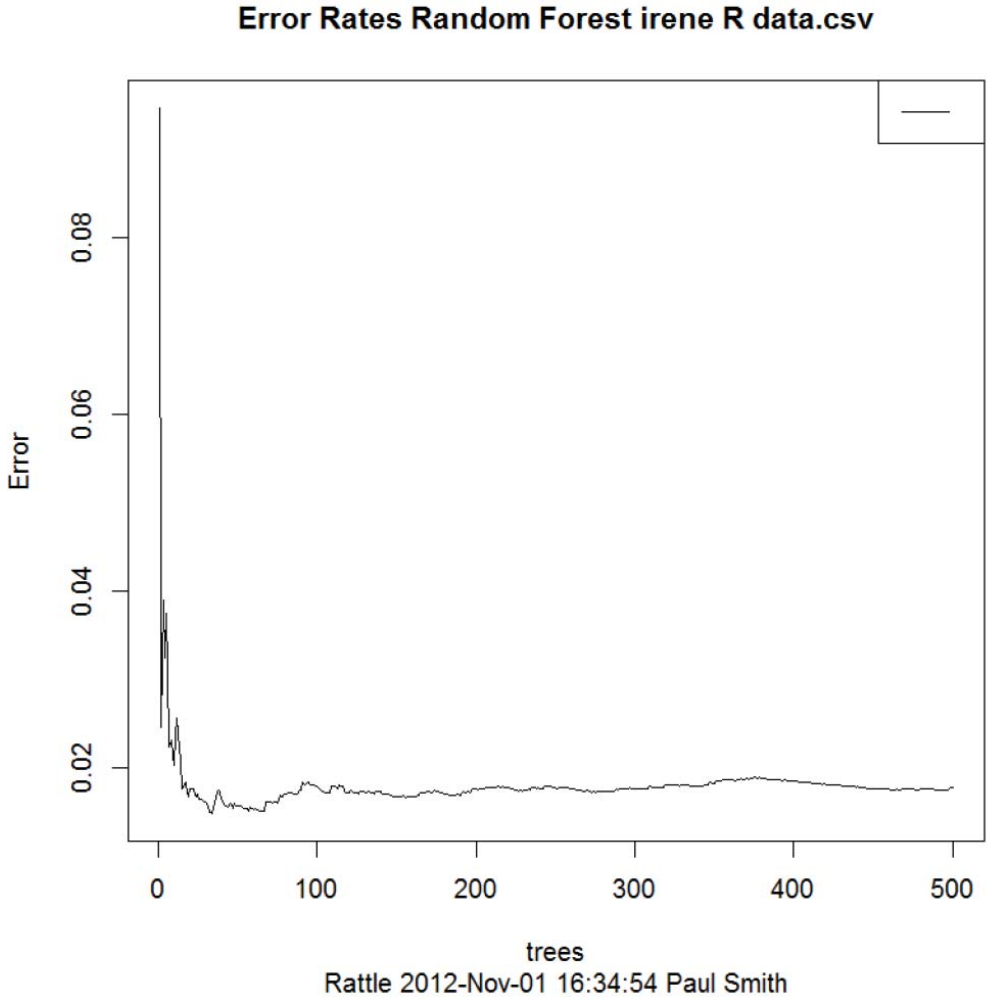


Figure 40: Error for the RFR as a function of the number of trees.

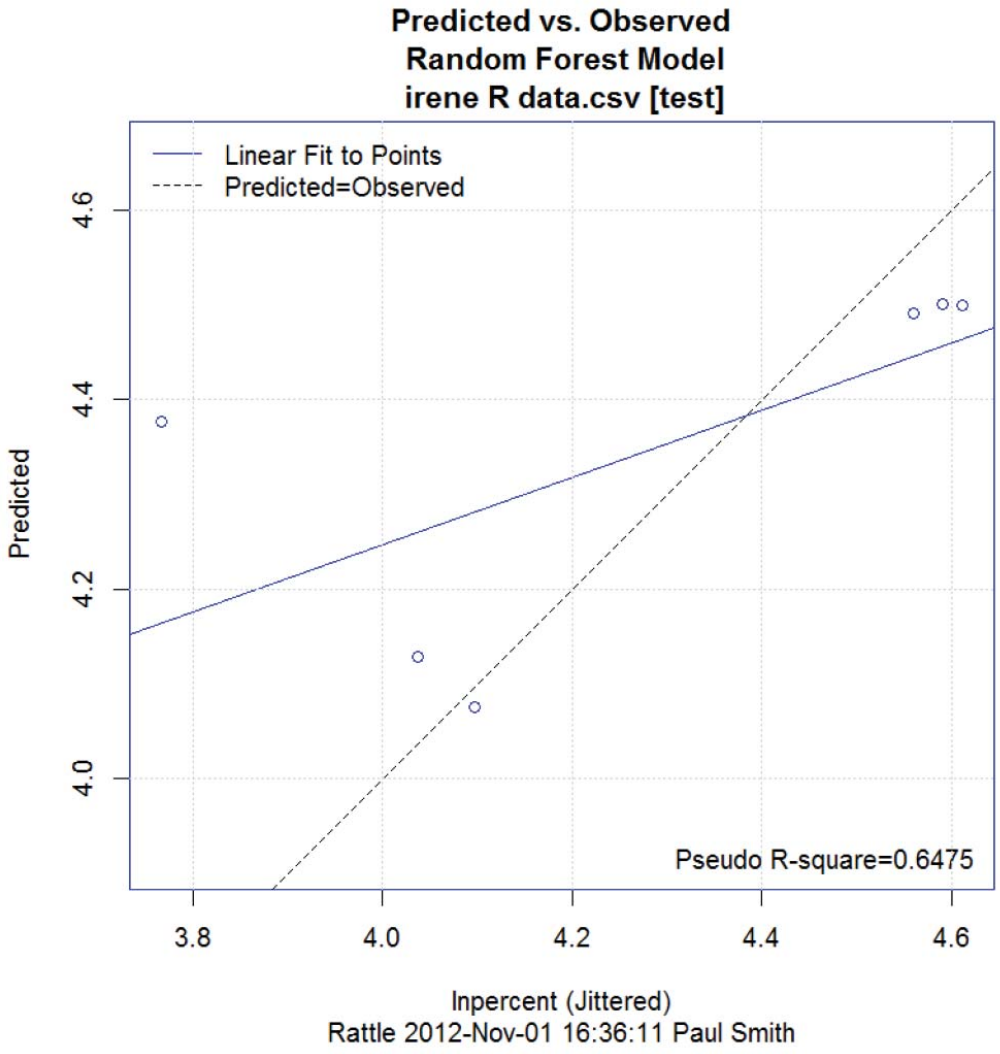


Figure 41: Predicted versus observed values for ln % correct for the test data in the RFR.

3.3 Biochemical study: Multivariate statistical analyses

The data were not normally distributed and a square root transformation was performed, which partially resolved the problem. However, with 42 measurements for each protein in each brain region, the sample size was likely to be large enough that the central limit theorem provided protection against violation of the assumption of normality (see boostrapping example in Figure 10 based on n = 8). Because of differences in the assay conditions at the different time points, the data for the 24 h, 72 h and 1 week time points were analysed separately from those for the 1 and 6 month time points. For the 6 month time point, there were separate groups of animals that received T maze training or no T maze training; therefore, the 1 and 6 month post-op. data were also analysed separately.

BVD resulted in a number of postural and locomotor behavioural symptoms that are characteristic of bilateral vestibular loss. These included: gait ataxia, marked hyperactivity, head-dorsiflexion, head-weaving, and circling (Zheng et al., 2006; Zheng et al., 2008; Goddard et al., 2008; Zheng et al., 2009; Baek et al., 2010; Machado et al., 2012; Stiles et al., 2012). In no case were these behaviours seen in the sham-operated animals. In addition, the BVD animals in the 6 month condition were demonstrated to have significant memory impairment in the spatial forced alternation T maze task, as reported in Zheng et al. (2007) (see Figure 42).

3.3.1 Protein expression in the hippocampus at 24 hours, 72 hours and 1 week post-BVD

There was no significant effect of surgery on the protein expression for any hippocampal subregion, and no significant interaction between surgery and time point, for the 24 h, 72 h and 1 week conditions (Figure 43). However, there was a significant effect of time point on the protein expression in all cases (CA1: $F(18,58) = 29.87$, $P \leq 0.0005$; CA2/3: $F(18,58) = 35.38$, $P \leq 0.0005$; DG: $F(18,58) = 57.30$, $P \leq 0.0005$; Figure 43). Using a LDA on the CA1, CA2/3 or DG data, no linear discriminant function could be identified that significantly predicted whether the brain tissue came from a BVD or a sham animal.

3.3.2 Protein expression in the hippocampus at 1 month post-BVD

The results were similar for the 1 month data set: there were no significant differences between BVD or sham animals either in the MANOVA or univariate ANOVAs for any hippocampal subregion (data not shown). Using a LDA, no linear discriminant function could be identified that significantly predicted whether the brain tissue came from a BVD or a sham animal.

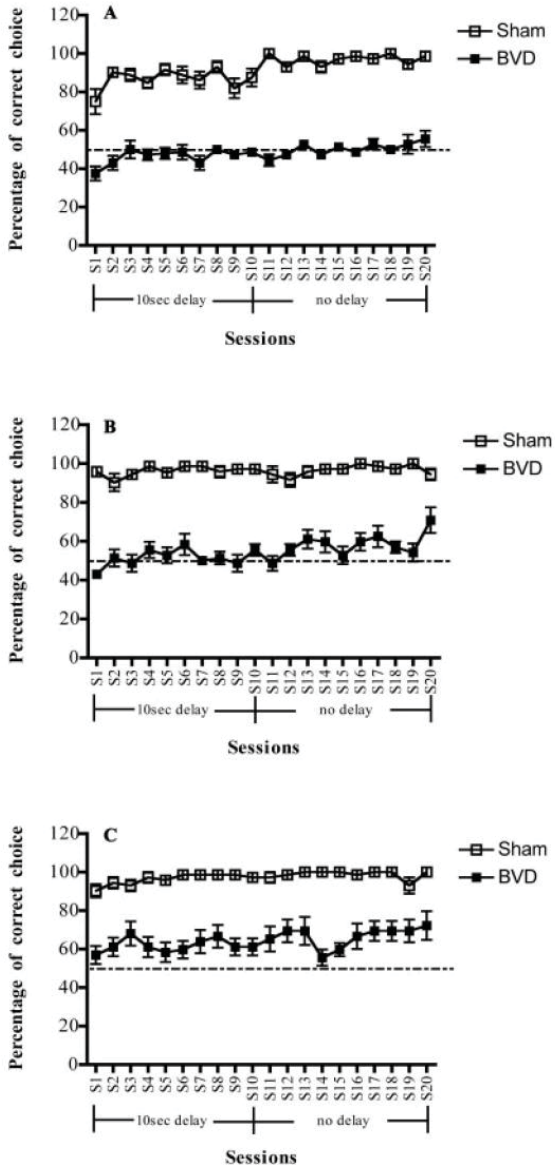


Figure 42: Percentage of correct choices in the T maze task for bilateral vestibular deafferentation (BVD) and sham surgery controls at 3 weeks (A), 3 months (B) and 5 months (C) post-op. Symbols represent means and bars ± 1 SE. From Zheng, Y., Goddard, M., Darlington, C.L., Smith, P.F. Bilateral

vestibular deafferentation impairs performance in a spatial forced alternation task in rats. *Hippocampus.* 17 (2007) 253-256.

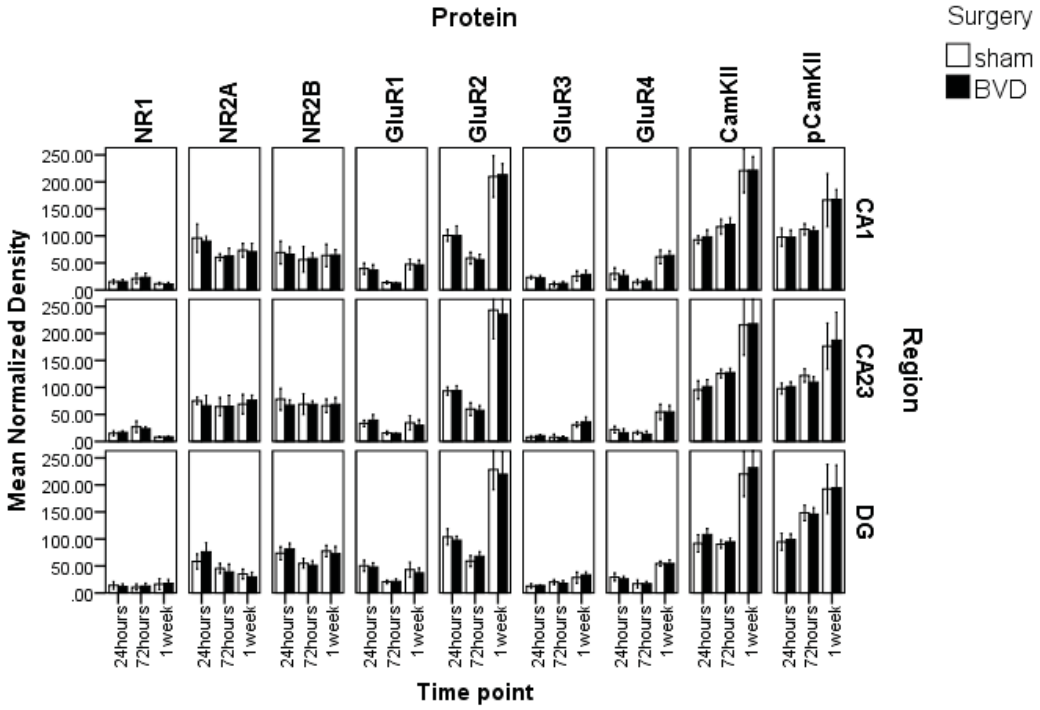


Figure 43: Mean protein expression for NR1, NR2A, NR2B, GluR1, GluR2, GluR3 and GluR4 receptors and CamKIIα and pCamKIIα in the CA1, CA2/3 and DG regions of the hippocampus at 24 hs, 72 hs or 1 week following BVD and sham surgery, ± 95% CI.

3.3.3 Protein expression in the hippocampus at 6 months post-BVD

The results for the 6 month data set were more complicated. For the MANOVA for CA1, surgery was still non-significant, as was the interaction between surgery and training. However, T maze training had a significant effect on the protein expression ($F(7,13) = 16.60$, $P \leq 0.0005$; see Appendix 4 for an example). Univariate ANOVAs indicated that spatial training significantly increased protein expression for CaMKIIα ($F(1,19) = 43.22$, $P \leq 0.0005$), NR1 ($F(1,19) = 8.02$, $P \leq 0.01$) and NR2B ($F(1,19) = 7.89$, $P \leq 0.01$), whereas GluR1 expression ($F(1,19) = 5.45$, $P \leq 0.03$) was significantly decreased in both sham

and BVD animals (Figure 44; see Appendix 4 for an example). There were no significant surgery and training interactions in the univariate ANOVAs.

For the MANOVA for CA2/3, surgery and the interaction between surgery and training were again non-significant. However, training was significant ($F(7,9) = 4.92$, $P \leq 0.02$), with significant univariate ANOVAs for CaMKIIα ($F(1,15) = 20.04$, $P \leq 0.0005$), pCaMKIIα ($F(1,15) = 21.83$, $P \leq 0.0005$), GluR1 ($F(1,15) = 15.31$, $P \leq 0.001$), GluR 2 ($F(1,15) = 20.64$, $P \leq 0.0005$) and GluR3 ($F(1,15) = 17.55$, , $P \leq 0.001$; Figures 44 and 45). There were no significant interactions in the univariate ANOVAs.

The results were similar for the DG: the MANOVA was significant only for training ($F(7,12) = 6.15$, $P \leq 0.003$), with significant univariate ANOVAs only for CaMKIIα ($F(1,18) = 47.29$, $P \leq 0.0005$), pCaMKIIα ($F(1,18) = 33.73$, $P \leq 0.0005$), GluR1 ($F(1,18) = 9.68$, $P \leq 0.006$) and GluR3 ($F(1,18) = 5.88$, $P \leq 0.03$; Figure 44). There were no significant interactions in the univariate ANOVAs.

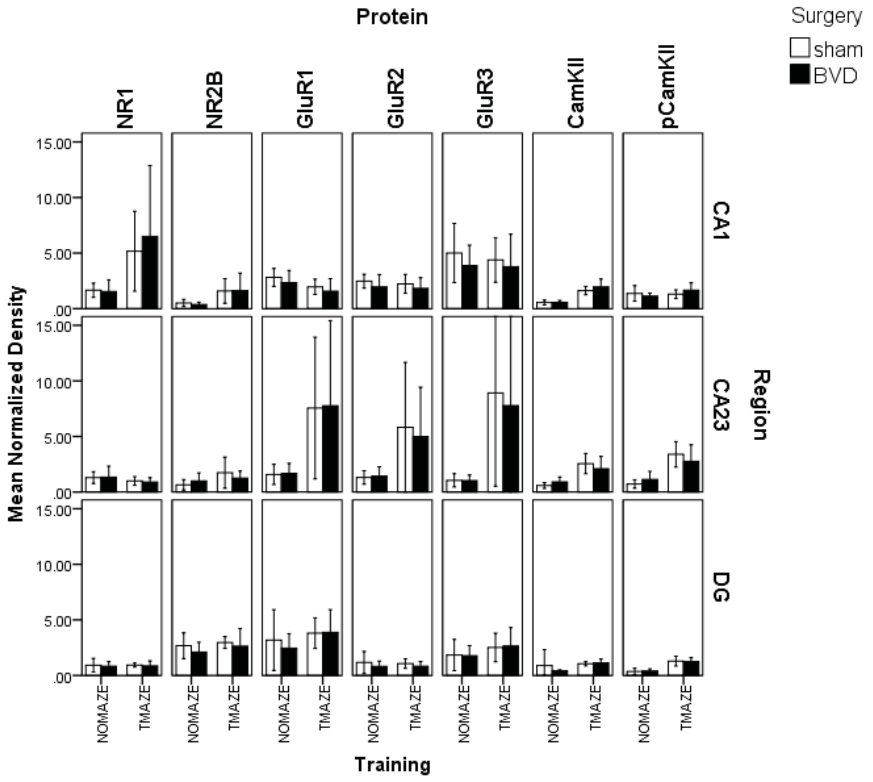


Figure 44: Mean normalized density of expression of NR1, NR2B, GluR1, GluR2, GluR3, CaMKII and pCaMKIIα in the CA1, CA2/3 and DG regions of the hippocampus at 6 months following BVD or sham surgery for animals

trained in a T maze or not trained in a T maze. Error bars represent 95% confidence intervals for the mean.

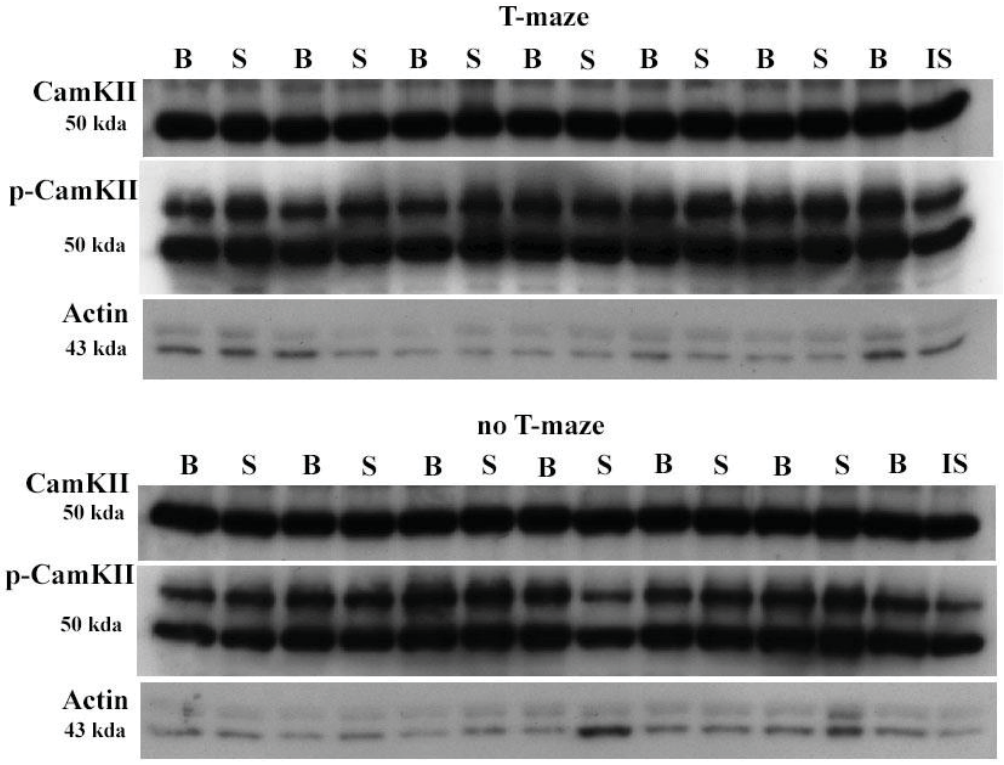


Figure 45: Example of western blots for CamKIIα and pCaMKIIα in CA2/3 for the BVD ('B') and sham ('S') animals that received T maze training or no T maze training at 6 months post-op. 'IS' is the internal standard and β–actin ('Actin') is also shown.

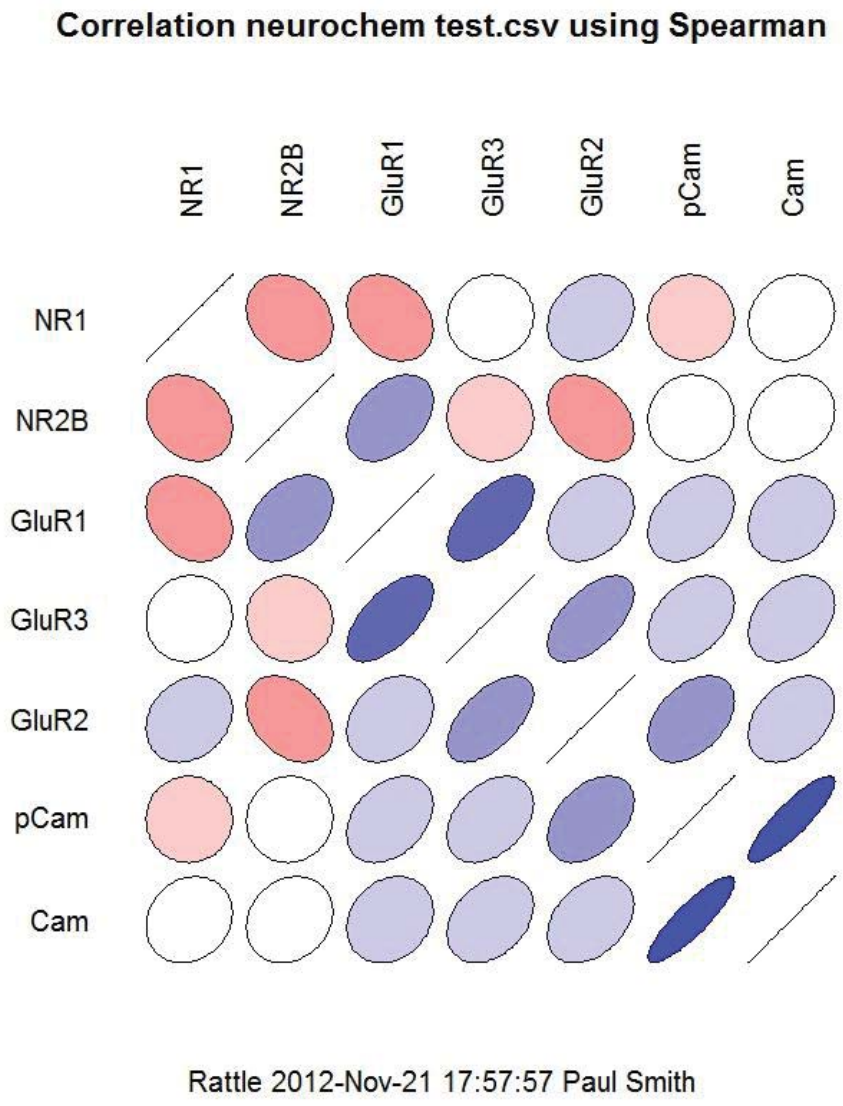


Figure 46: Spearman correlation analysis for the different glutamate receptor subunits, calmodulin kinase IIα ('Cam') and phosphorylated calmodulin kinase IIα ('pCam') in the 3 hippocampal subregions at 6 months post-op., with the BVD and sham animals and T maze-trained and non-T maze-trained animals' data pooled. Note the strong correlation between the expression of Cam and pCam.

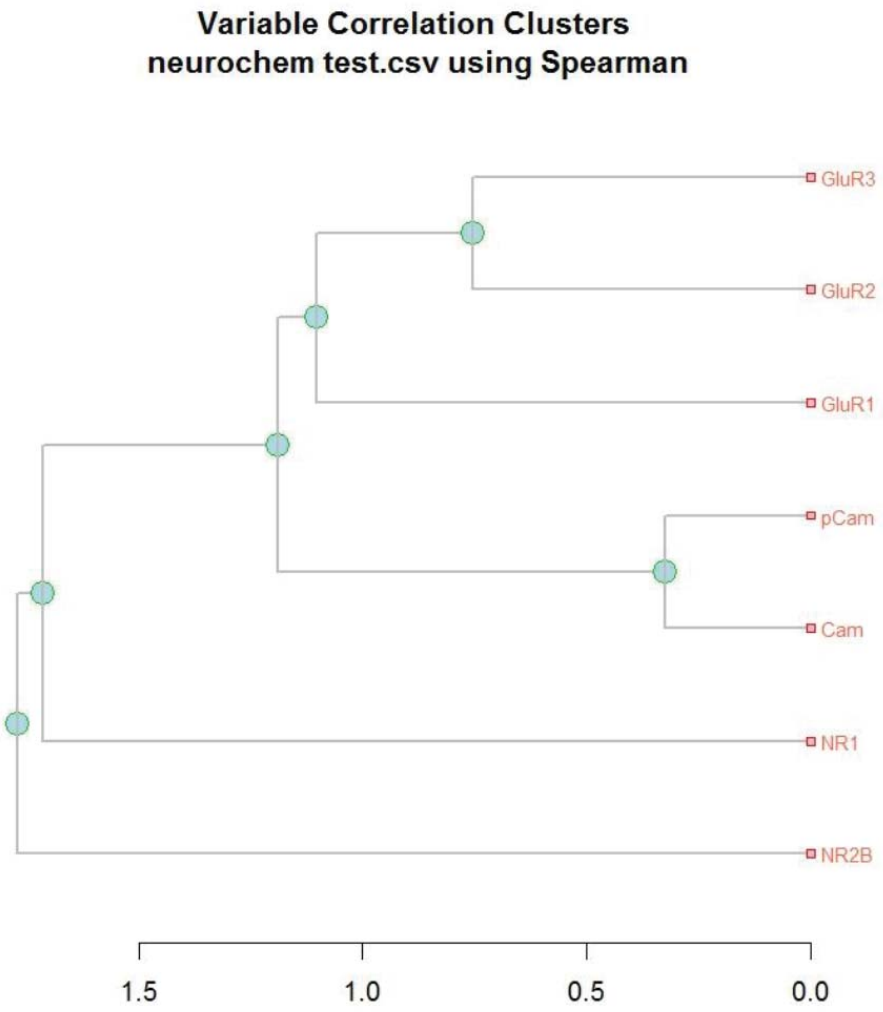


Figure 47: Spearman correlation cluster analysis for the different glutamate receptor subunits, calmodulin kinase II ('Cam') and phosphorylated calmodulin kinase II ('pCam') in the 3 hippocampal subregions at 6 months post-op., with the BVD and sham animals and T maze-trained and non-T maze-trained animals' data pooled. Note the correlation between the expression of the AMPA receptor subunits, GluR1-GluR3, the NMDA receptor subunits, NR1 and NR2B, and Cam and pCam.

When a Spearman correlation analysis was carried out on the data for the different glutamate receptor subunits and CaMKIIα and pCaMKIIα in the 3 hippocampal subregions together at 6 months post-op., with the BVD and sham animals and T maze-trained and non-T maze-trained animals' data pooled, there was an obvious relationship between the expression of the AMPA receptor

subunits, GluR1-GluR3, the NMDA receptor subunits, NR1 and NR2B, and CaMKIIα and pCaMKIIα (see Figures 46 and 47). However, using a LDA on the CA1, CA2/3 or DG data, no linear discriminant function could be identified that significantly predicted whether the brain tissue came from a BVD or a sham animal. A RF classification analysis for surgical group, with the proteins as predictor variables and using m = 3 and 500 trees, resulted in a large OOB error rate of 52.83%, with an even higher classification error rate on the test data of 63.16%. Figure 48 shows the variables in order of importance for predicting surgical group, and Figure 49 shows that the error rate actually increases with the increasing number of trees. When a RF analysis was used on the same data to predict whether animals had received T maze training or not, the OOB error rate was 13.21%, with a classification error rate on the test data of 10.5% (see Table 27).

OOB estimate of error rate: 13.21%

Confusion matrix:

	NOMAZE	TMAZE	class.error
NOMAZE	18	5	0.21739130
TMAZE	2	28	0.06666667

Error matrix for the Random Forest model on neurochem test.csv [test] (counts):

		Predicted	
Actual	NOMAZE	TMAZE	
	NOMAZE	11	1
	TMAZE	1	6

Overall error: 0.1052632

Table 27: OOB and classification error rates for the RF analysis of the pooled 6 month neurochemical data for the prediction of training in a T maze.

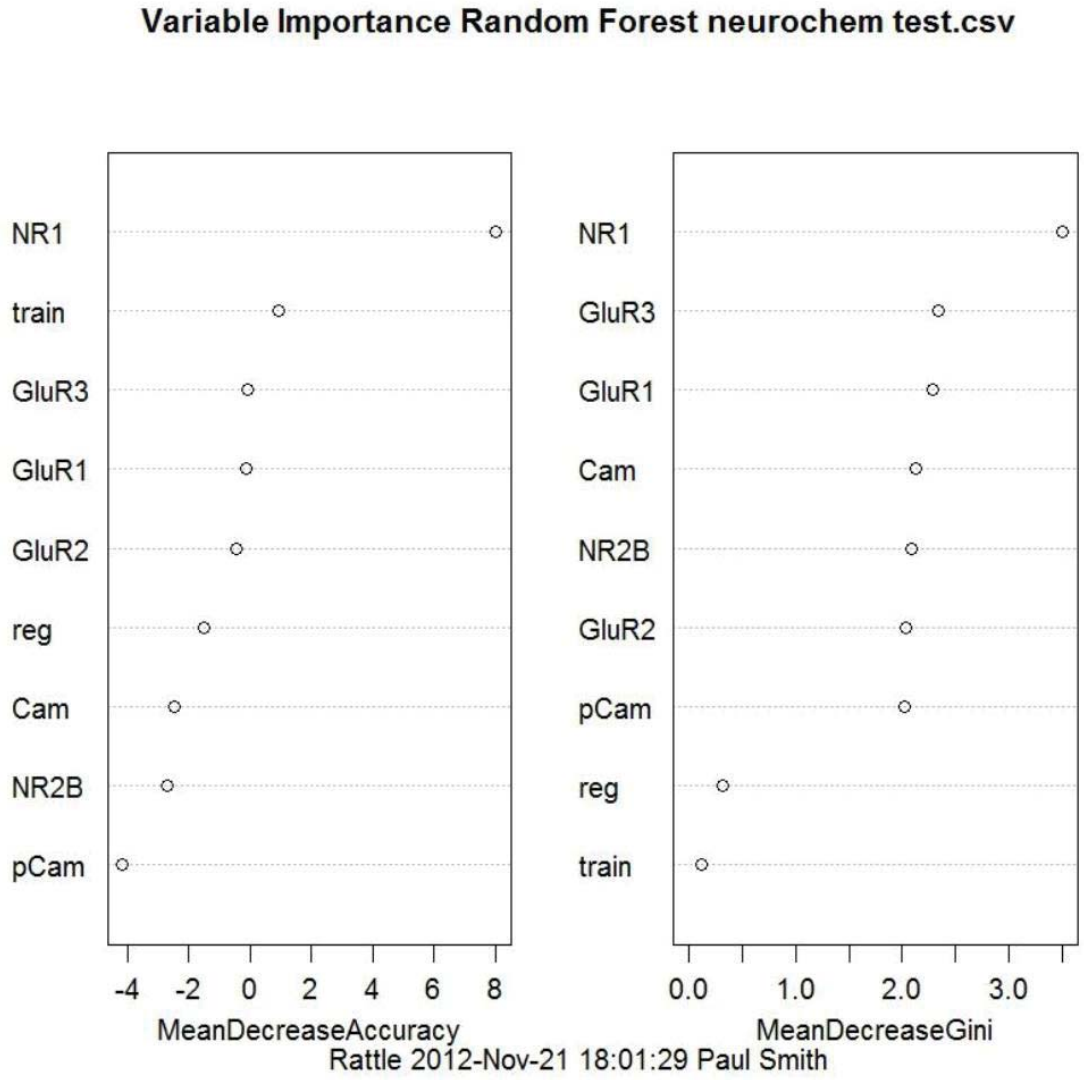


Figure 48: Variables in order of importance for the prediction of surgical group using a RF classification analysis.

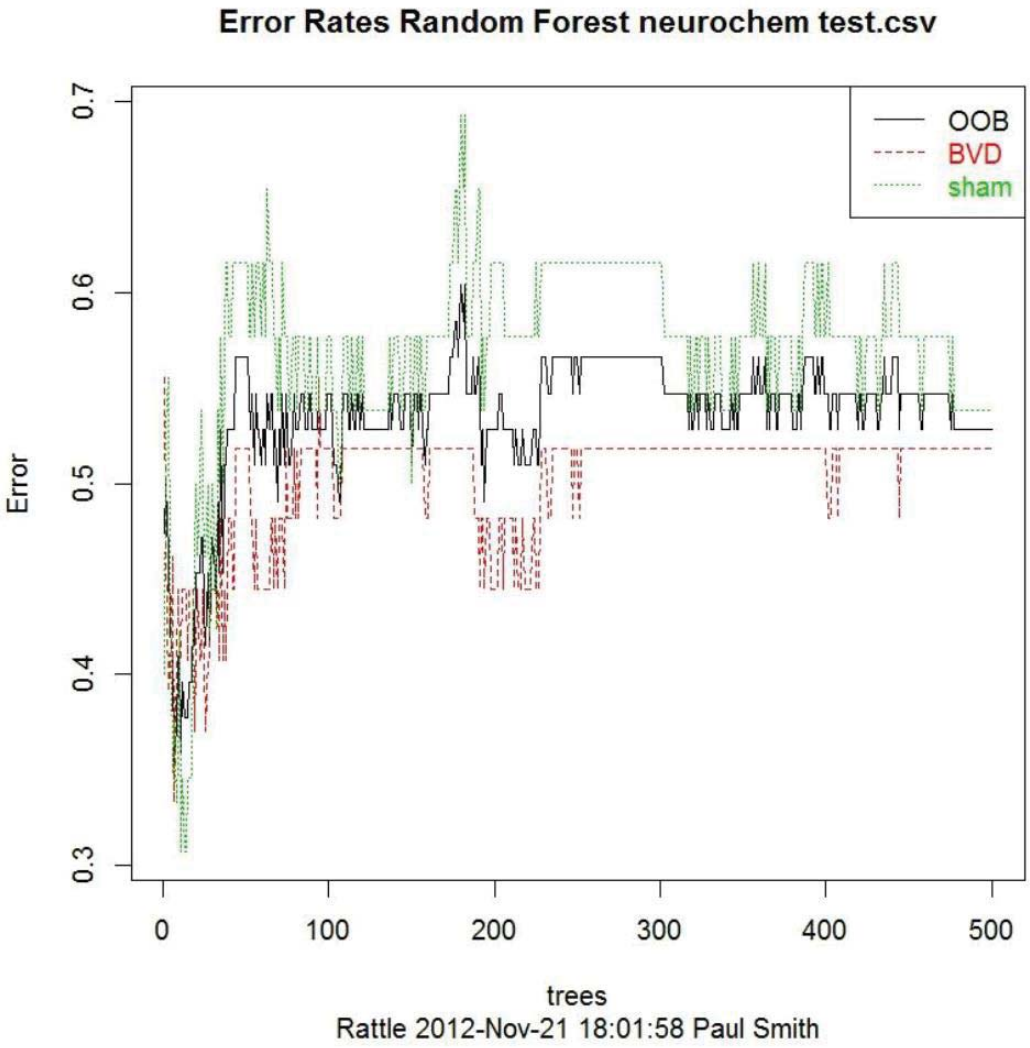


Figure 49: OOB, BVD and sham classification error for the prediction of surgical group using a RF classification analysis.

Figures 50 and 51 shows the variables in order of importance for the prediction of T maze training and the OOB, T maze training and no T maze training error rates as a function of the number of trees, respectively.

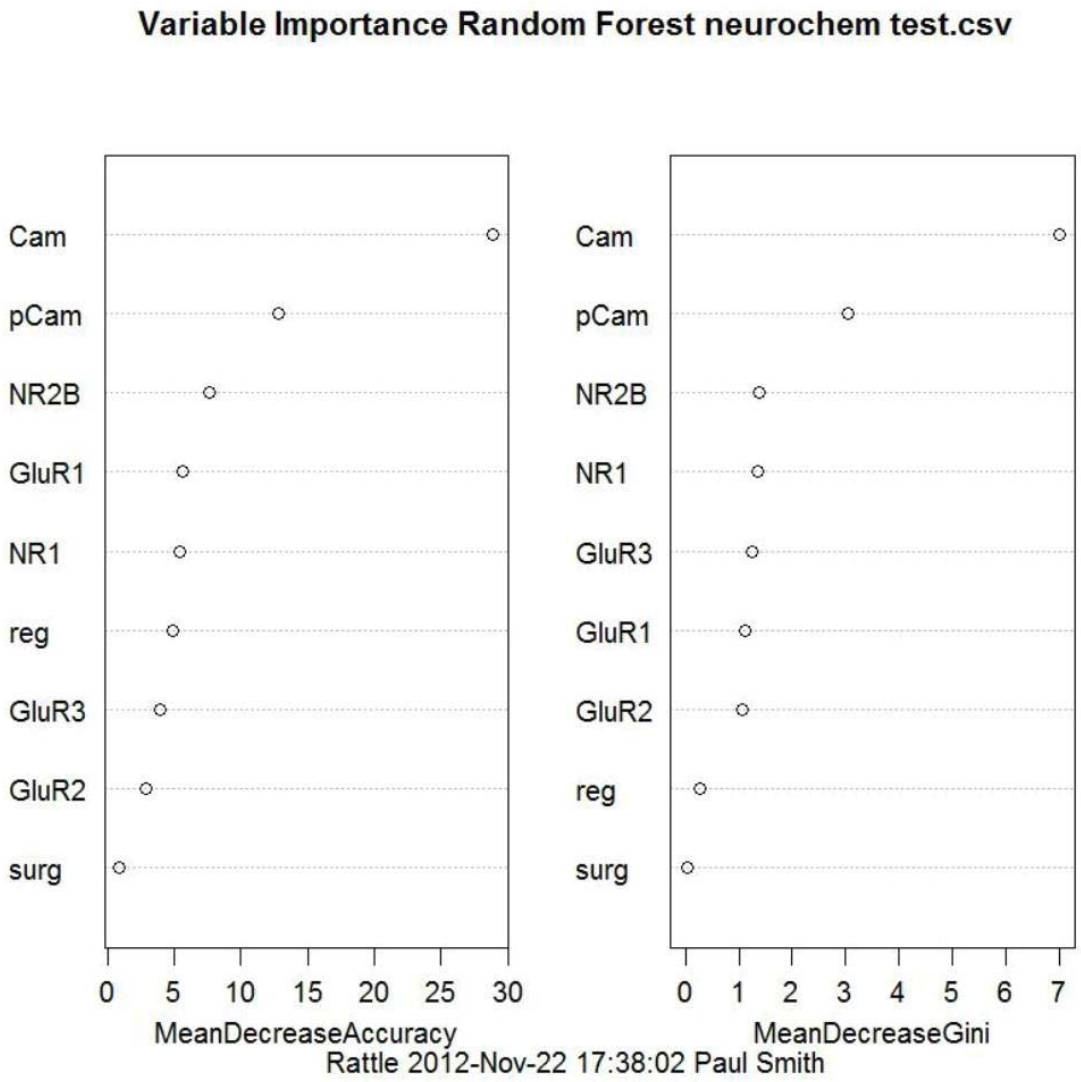


Figure 50: Variables in order of importance for the RF classifcation for T maze training and no T maze training.

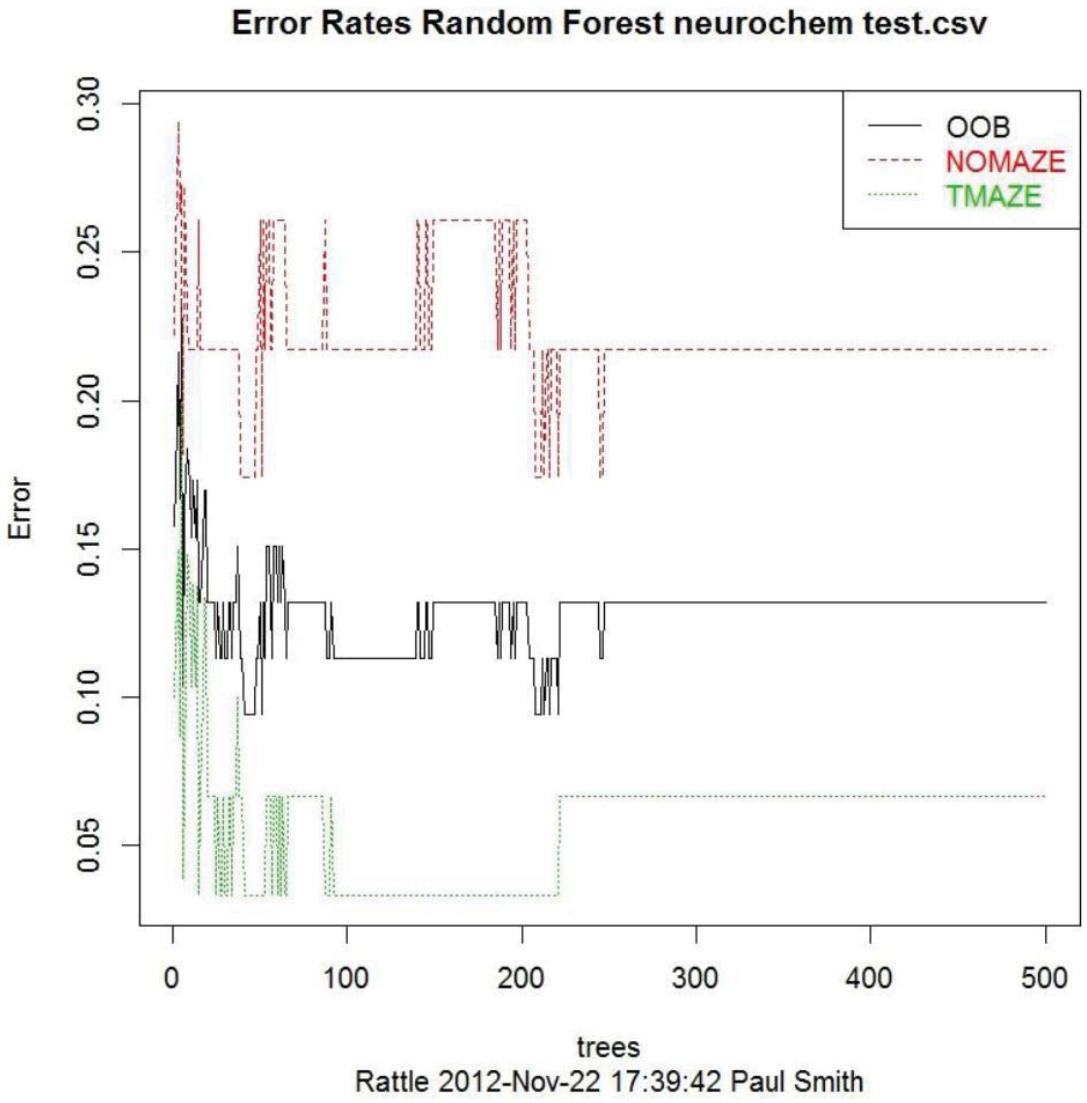
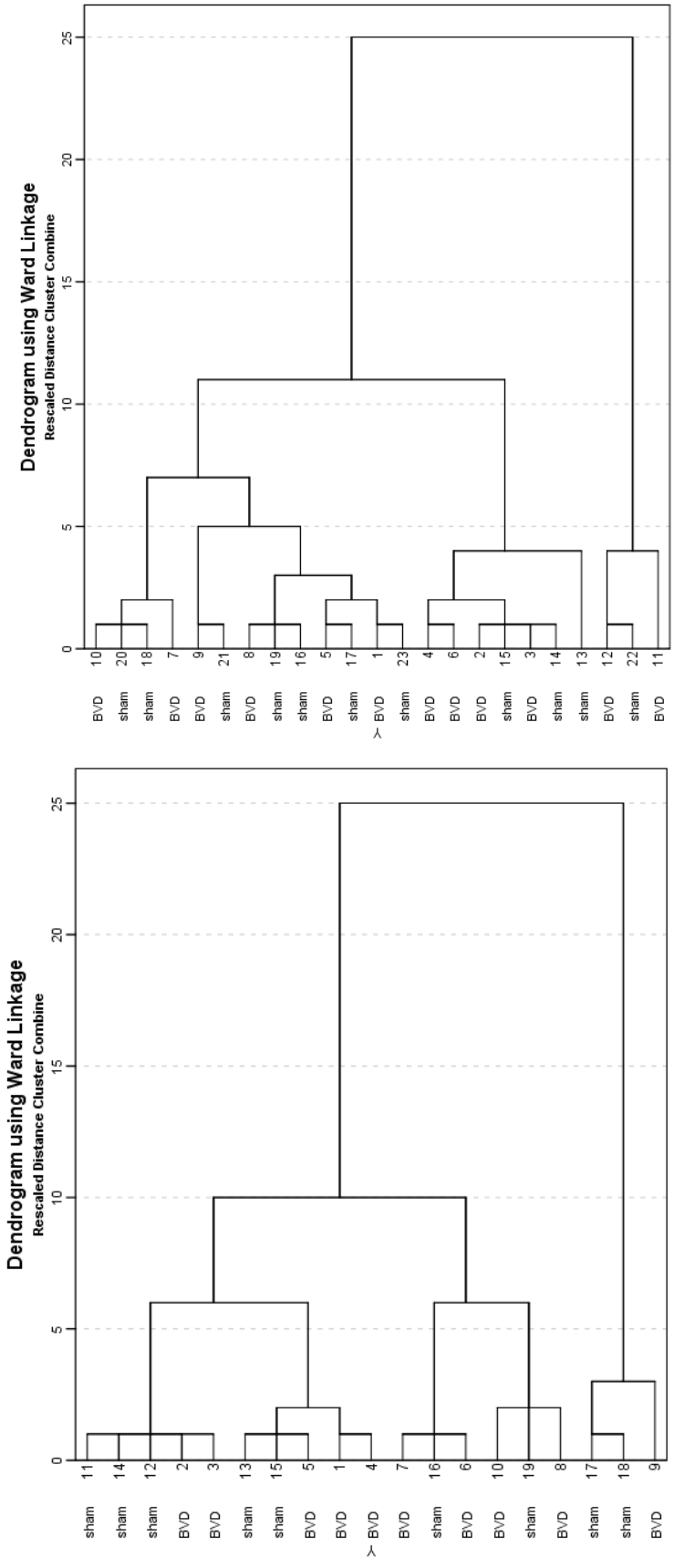


Figure 51: OOB, T maze training and no T maze training error rates as a function of the number of trees.

Cluster analyses also confirmed that surgical group could not be predicted from the neurochemical variables (see Figure 52). Training could not be predicted from the neurochemical variables for CA1 and the DG; however, a cluster analysis of the CA2/3 data revealed that the T maze - and non -T maze-trained

animals could be accurately separated based on the neurochemical variables (Figure 53).



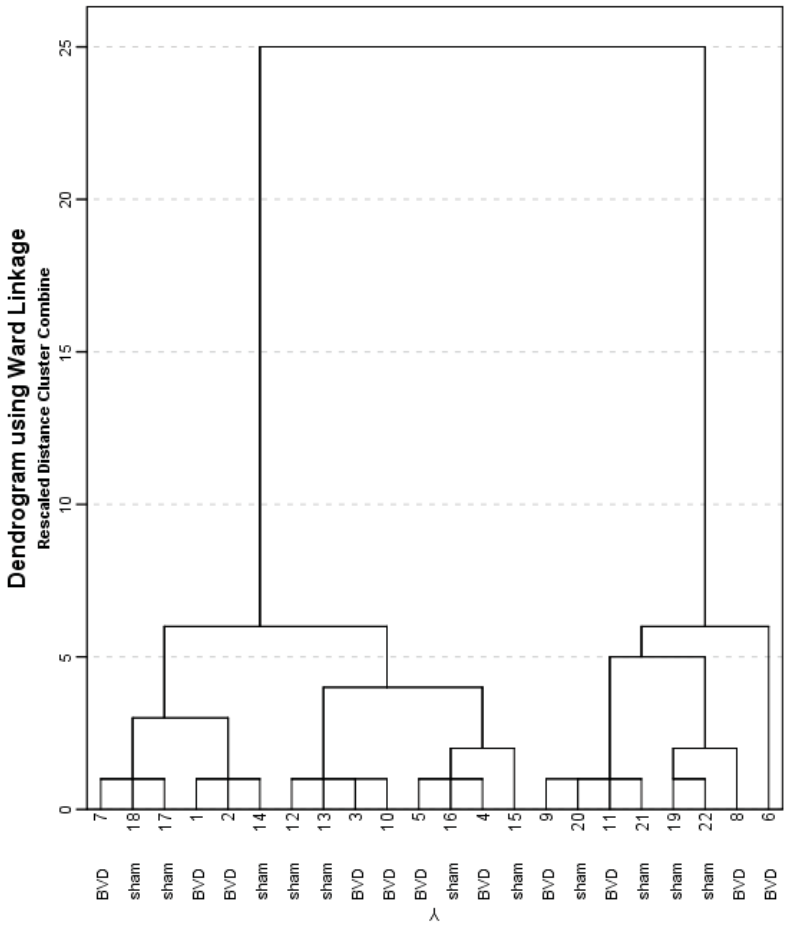
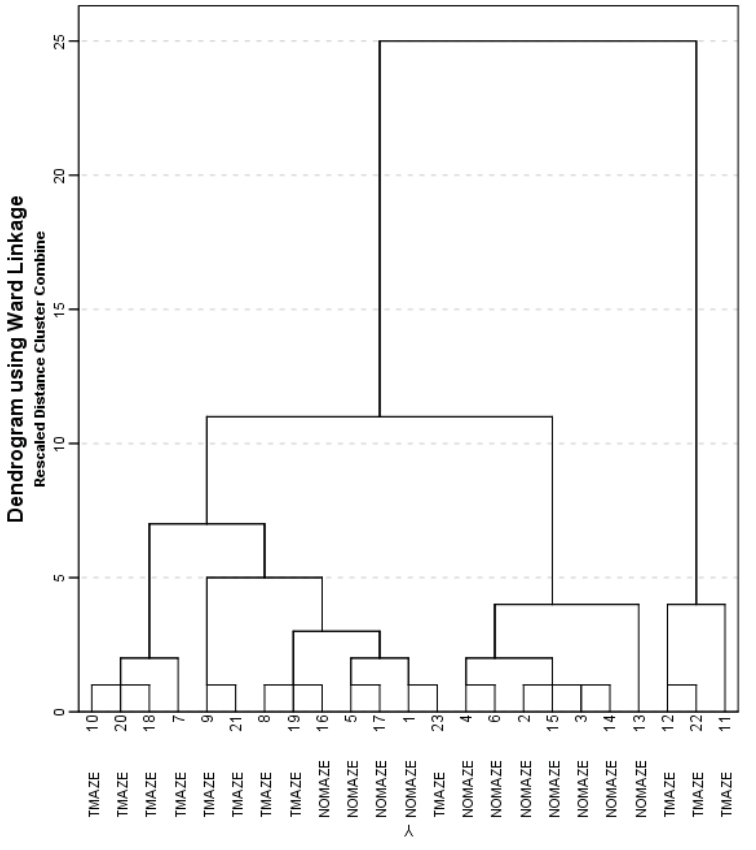
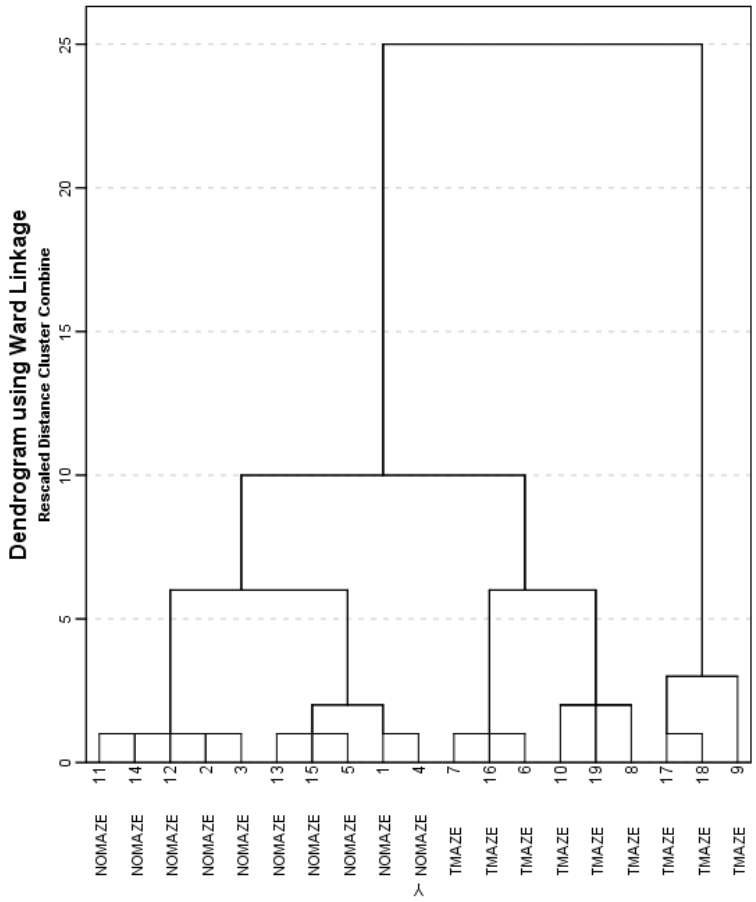


Figure 52: Cluster analysis, using squared Euclidean distance and Ward's minimal variance algorithm, on all of the neurochemical data in CA1 (top), CA2/3 (middle) and the DG (bottom), showing individual animals according to surgery (BVD or sham).


Dendrogram using Ward Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25
10 20 18 7 9 21 8 19 16 5 17 1 23 4 6 2 15 3 14 13 12 22 11
TMAZE TMAZE TMAZE TMAZE TMAZE TMAZE TMAZE TMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE TMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE TMAZE TMAZE TMAZE


Dendrogram using Ward Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25
11 14 12 2 3 13 15 5 1 4 7 16 6 10 19 8 17 18 9
NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE NOMAZE TMAZE TMAZE TMAZE TMAZE TMAZE TMAZE TMAZE TMAZE TMAZE

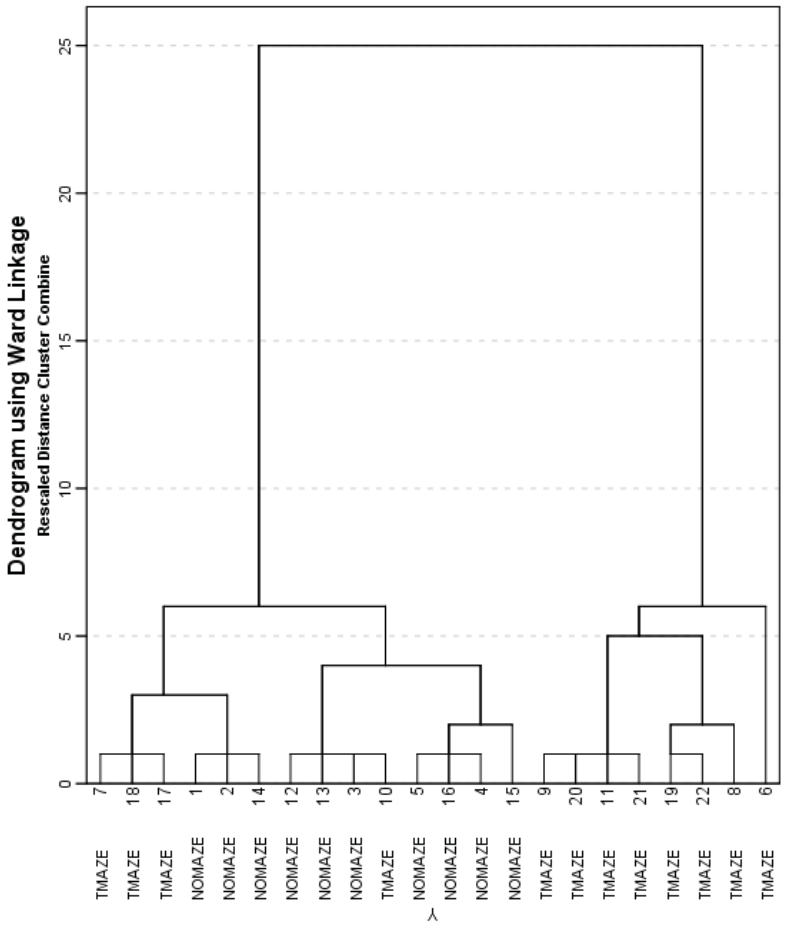


Figure 53: Cluster analysis, using squared Euclidean distance and Ward's minimal variance algorithm, on the neurochemical data at 6 months post-op. in CA1 (top), CA2/3 (middle) and the DG (bottom), showing individual animals according to training (T maze or no T maze). Note that the neurochemical data for CA2/3 (middle) can completely distinguish the trained from the non-trained animals, in contrast to CA1 (top) and the DG (bottom).

4. Discussion

4.1 Scientific interpretation

4.1.1 Behavioural data

The results of this study replicate some but not all of the findings from previous studies of rats with BVD. Consistent with previous studies (Zheng et al., 2007; Zheng et al., 2008; Goddard et al., 2008; Avni et al., 2009; Zheng et al., 2009; Stiles et al., 2012; Besnard et al., 2012; Neo et al., 2012; Machado et al., 2012), BVD rats were found to be generally hyperactive, exhibited reduced thigmotaxis and rearing, and performed very poorly in the spatial T maze alternation test. However, in contrast to previous studies (Zheng et al., 2007; Neo et al., 2012), there was no significant effect of BVD on performance in the EPM, no effect on avoidance latency in the ETM and a small effect on escape latency in the ETM. Despite the fact that the doses of buspirone and FG-7142 were based on previous studies (Pellow & File 1986; Bruhwyler et al 1991; Stefanski et al 1992; Graeff et al 1998; Siemiątkowski et al 2000; Poltronieri et al., 2003) as well as pilot studies, relatively few effects of these drugs in sham or BVD animals were observed and some of the effects were paradoxical.

The largest discrepancy between the current results and those of previous studies concerns the EPM. In a previous study (Zheng et al., 2008), it was found that BVD rats spent significantly more time in the open arms of the EPM, suggesting that they experienced *reduced anxiety* compared to sham controls. These results appeared to be consistent with those reported by Lindemann et al. (2007) using the *ci2* mutant rat with cochlear and vestibular deficits. In the current study, there was no significant effect of BVD on time spent in the open arms of the EPM, and drug treatment with buspirone or FG-7142 made no difference to this. However, other studies have reported evidence of *increased anxiety* in *Hdb* mice using the EPM (Shefer et al., 2010). Furthermore, Machado et al. (2012) have recently reported that rats with bilateral chemical vestibular lesions exhibited evidence of increased anxiety in the black and white box test of anxiety. The results of the current study were also inconsistent for the ETM. Previously, Zheng et al. (2008) found that BVD rats exhibited a significantly shorter latency to leave the enclosed arm, and a longer latency to escape the open arm, compared to sham controls. However, in the current study, there was no significant difference in avoidance latency between BVD and sham animals, although BVD animals

exhibited a slightly longer escape latency. The ETM test did seem to be working properly, since the animals increased their latency to leave the enclosed arms of the ETM over 3 trials, suggesting that they were learning to avoid the open arm. However, another study has recently been completed which also did not find any significant differences in avoidance or escape latencies in the ETM in BVD rats (Neo et al., 2012); therefore, any effect of BVD on performance in the ETM may be difficult to replicate.

The difficulty in replicating results in the EPM and ETM in BVD rats, raises the question of how useful these tests might be to measure anxiety in animals with vestibular dysfunction, given that the same animals also exhibit deficits in learning and memory (Kalueff and Murphy, 2007; Machado et al., 2012). It is conceivable that any result in the EPM and ETM could be influenced by cognitive deficits. Other results from the current study are relevant to anxiety; however, they are difficult to interpret. Rearing was reduced in BVD animals, which has been interpreted as indicating reduced anxiety (Neo et al., 2012). However, rats with no vestibular function may exhibit less rearing due to their abnormal postural control; therefore, it is difficult to interpret this symptom purely in terms of anxiety. As has been reported in previous studies (Goddard et al., 2008; Neo et al., 2012; Stiles et al., 2012), thigmotaxis was reduced in BVD rats, which may suggest increased risk taking behaviour in the OFM. However, rats with BVD have no vestibulo-ocular reflex function, and therefore they will experience blurred vision (‘oscillopsia’) even during head movement that is as small as that produced by the pulse beat. Therefore, it is possible that the changes in their exploratory behaviour are partly due to visual instability. The fact that the time spent in inner/middle zone of the open field by the BVD animals gradually decreased over the 10 min period, suggests that the reduced thigmotaxis in BVD rats might be a result of disorientation which could be improved by further exploration of the environment.

The results from the STM are entirely consistent with those from previous studies (Zheng et al., 2007; Neo et al., 2012; Besnard et al., 2012; Machado et al., 2012) in demonstrating that BVD causes a profound deficit in spatial alternation, which is not affected by drug treatment. This is consistent with the evidence that BVD results in hippocampal dysfunction that is long-lasting (Brandt et al., 2005; Zheng et al., 2006; Zheng et al., 2007; Zheng et al., 2009; Baek et al., 2010; Neo

et al., 2012; Besnard et al., 2012; Machado et al., 2012). Given that the performance of the same animals in the EPM and ETM did not suggest any difference in anxiety levels compared to sham controls, this result may suggest that the cognitive deficits caused by BVD are independent of anxiety. This hypothesis has received recent support by Machado et al. (2012), who observed that anxiolytic doses of diazepam had no significant effect on the deficits exhibited by BVD rats in the 8 arm radial maze.

The main significant drug effects were a small surgery x zone x drug interaction from buspirone, suggesting that buspirone slightly reduced the amount of time that the BVD animals spent in the inner/middle zone of the OFM; a zone x drug interaction from FG-7142, suggesting that it increased the amount of time animals spent in the inner/middle zone; a surgery x drug interaction for buspirone for the frequency of supported rearing; a main effect and surgery x drug interaction for FG-7142 for the frequency of unsupported rearing, in the OFM (both drugs increasing the frequencies in the sham group); and a main effect of buspirone on escape latency in the ETM (buspirone reducing the escape latency in sham rats compared to the saline control condition) as well as a surgery x drug interaction for avoidance latency (buspirone reducing the avoidance latency in BVD rats). The doses of buspirone and FG-7142 that were used in the current study were clearly ineffective for the most part and therefore this weakens this study as a test of the influence of anxiety on performance in the various behavioural tests. From the significant drug effects listed above, the only expected result was that buspirone significantly reduced avoidance latency in the ETM, suggesting an anxiolytic effect. However, that there were significant drug effects on performance in the OFM and ETM, made it impossible to exclude drug treatment as a variable. Doses of buspirone and FG-7142 were chosen, based on previous studies (Pellow & File 1986; Bruhwyler et al 1991; Stefanski et al 1992; Graeff et al 1998; Siemiątkowski et al 2000; Poltronieri et al., 2003), which should have had anxiolytic and anxiogenic effects (respectively), without producing adverse side effects. However, the effects of these drugs at these doses are variable (Pellow & File 1986; Bruhwyler et al 1991; Stefanski et al 1992; Graeff et al 1998; Siemiątkowski et al 2000; Poltronieri et al., 2003) and the doses may have been too low for the conditions of this experiment.

The results of the multivariate statistical and data mining analyses suggested that: 1) BVD animals could be clearly distinguished from sham animals by their behavioural syndrome, in particular their decreased duration of unsupported rearing, reduced thigmotaxis (decreased time spent in the outer zone of the OFM), and spatial memory deficits; 2) the best predictors of performance in the STM were whether the animals had received a BVD and the duration of supported rearing. These results held consistently for a variety of different statistical analyses with some differences that are discussed below, with links to substantive physiological findings.

The results of the LDA, RF, SVM and cluster analyses demonstrated that the performance of the animals in the various behavioural tests could be used to classify rats as BVD or sham animals with nearly complete accuracy. An LDA using all of the variables could classify animals as BVD or sham with 100% accuracy. However, this could also be achieved using a linear discriminant function including only the duration of unsupported rearing, the time spent in the outer zone of the OFM, and performance in the STM. Similar results were obtained using RFs and SVMs. The cluster analysis based on all of the behavioural data separated the BVD from the sham animals except for 2 cases. The results of these multivariate statistical and data mining analyses are consistent with those of the univariate statistical analyses, that demonstrated significant differences in the duration of unsupported rearing, time spent in the outer zone of the OFM, and performance in the STM, between BVD and sham animals.

Because of the complexity of the behavioural syndrome that occurs following BVD, it has always been a concern that what appear to be spatial memory deficits could be attributable to locomotor hyperactivity or anxiety-related behavior (Zheng et al., 2009). Previous studies have used anxiolytic drugs, such as diazepam, in an attempt to dissect the influence of anxiety on performance in spatial memory tasks (Machado et al., 2012). The results of these studies have suggested that anxiety-related behavior cannot account for poor performance in such tasks, suggesting that the performance of BVD animals in spatial memory tasks may be attributable to spatial memory deficits themselves. Likewise, previous regression analyses have suggested that changes in locomotor activity also cannot account for the poor performance of BVD animals in spatial memory tasks (Baek et al., 2010). The results of the regression analyses in this study are

consistent with those studies in suggesting that the most important predictors of performance in the STM are whether the animals had received BVD or sham lesions and the duration of supported rearing, which may represent an index of postural instability associated with vestibular dysfunction. While surgical group was clearly the most important predictor of STM performance in the MLRs, the duration of supported rearing was the most important in the RFR. It is difficult to compare the MLRs and RFRs directly, since the latter were developed on a training data set and then tested on a test data set. Nonetheless, the variance explained value for the RFR, 70.11%, was substantially less than the adjusted R^2 values for the backward and best subsets regressions (88%), suggesting that the MLRs may have provided a superior model of the data. The MLRs suggested that whether or not the animals had intact vestibular systems was the most important predictor of spatial memory performance in the STM. Therefore, the impairment in performance in spatial memory tasks that has been documented in so many studies of BVD animals, may actually be due to spatial memory deficits caused by loss of vestibular function rather than merely an indirect consequence of changes in locomotor activity or anxiety. This is consistent with evidence from human patients with BVD, who have demonstrated spatial memory impairment even when the amount of movement required to perform a task was minimal and they have had 5-10 years to compensate (Brandt et al., 2005). In both the backward and best subsets regression, avoidance latency in the ETM, while only marginally significant, also appeared as a significant effect in the regression models with the highest adjusted R^2 values. Since avoidance of the open arms in the ETM is also a form of learning, it is not surprising that latency to leave the open arm might be related to performance in the STM.

The most likely explanation for the spatial memory deficits observed in this study and in other studies is that the hippocampus and other areas of the medial temporal lobe become dysfunctional following BVD. Previous studies have shown that BVD results in the loss of selective firing in hippocampal place cells (Stackman et al., 2002; Russell et al., 2003) and abnormal theta rhythm (Russell et al., 2006; Neo et al., 2012; Tai et al., 2012). However, partial restoration of theta rhythm in the hippocampus using electrical stimulation of the septum, did not improve the performance in a T maze task, suggesting that abnormalities in theta rhythm alone may not be responsible for the spatial memory

deficits (Neo et al., 2012). Despite this evidence for hippocampal dysfunction following BVD, long-term potentiation (LTP) appears to remain intact (Zheng et al., 2010), suggesting that the hippocampus is still capable of synaptic plasticity. One possibility is that the loss of vestibular input, which causes other sensory information to become ambiguous, leads to the encoding of incorrect spatial memories of the environment.

In conclusion, the results of this study confirm those from previous studies that BVD results in locomotor hyperactivity, reduced thigmotaxis and rearing, and spatial memory deficits, which, for the most part, were not modulated by anxiolytic or anxiogenic drugs. By contrast with some previous studies, there was little evidence of changes in anxiety in the EPM or ETM, suggesting the possibility that the spatial memory deficits observed in the STM are independent of anxiety, consistent with recent evidence (Machado et al., 2012).

4.1.2 Neurochemical data

The results of this study show that, using western blotting, the expression of AMPA and NMDA glutamate receptor subunits, and CaMKIIα, in the hippocampus is not significantly different in BVD compared to sham animals at 24 h, 72 h, 1 week, 1 month or 6 months post-op. Spatial training in a T maze, however, had a significant effect on the expression of CaMKIIα, NR1, NR2B and GluR1 in CA1, on CaMKIIα, pCaMKIIα, GluR1, GluR2 and GluR3 in CA2/3, and on CaMKIIα, pCaMKIIα, GluR1, and GluR3 in the DG. However, this effect occurred independently of surgery. The results of the LDAs showed that no linear discriminant function could be found that significantly discriminated the BVD from the sham animals on the basis of the neurochemical data. RF classification analysis was similarly unsuccessful in separating the BVD from the sham animals; however, it could separate the T maze-trained from the non-T-maze trained animals in the 6 month post-op. group, with a moderate error rate (OOB error rate of 13.21%).

In a previous study, a significant decrease in NR1 expression was observed in the ipsilateral CA2/3 region at 2 weeks following UVD (Liu et al., 2003). Besnard et al (2012), who performed sequential UVD's several weeks apart using intratympanic sodium arsanilate injections, observed a significant up-regulation of

NMDA receptors in the hippocampus, with reduced affinity, using receptor autoradiography. These findings appear to be in disagreement with the current results. However, there are several differences between the studies that probably account for the apparent discrepancy. First and most importantly, UVD results in an imbalance in the vestibulo-ocular (VOR) and vestibulo-spinal reflexes (VSR), causing symptoms such as spontaneous ocular nystagmus (SN, with quick phase toward the intact side) and postural asymmetry toward the lesioned side (see Smith and Curthoys, 1989 for a review). These symptoms, which are a result of an imbalance between the left and right central vestibular systems, are so severe initially that animals such as rats and guinea pigs have difficulty standing immediately after recovery from anaesthesia. Gradually, over a period of 2-3 days, the SN and postural asymmetry decrease in severity in a process known as 'vestibular compensation' (see Smith and Curthoys, 1989 for a review). If a UVD is then performed on the contralateral side after compensation has occurred for the first UVD, this generates SN and postural asymmetry in the opposite direction to the original symptoms, in a phenomenon known as Bechterew's syndrome (see Smith and Curthoys, 1989 for a review). Following BVD, in which one labyrinth is lesioned after the other under anesthesia, SN and postural asymmetry do not occur, because there is no imbalance in activity between the two labyrinths following recovery from the anaesthetic. Rather, BVD results in a complete loss of the VORs and VSRs. Therefore, the behavioural symptoms which follow UVD, or two UVD procedures performed sequentially, are quite different from those that follow a simultaneous BVD under anaesthesia. The most likely explanation for the difference between the current results for the NR1, NR2A and NR2B subunits of the NMDA receptor and Besnard et al.'s (2012) results for the NMDA receptor, is the different temporal sequence of the lesions. It must also be considered that whereas the current study used surgical lesions of the labyrinth, Besnard et al. (2012) used intratympanic injections of the ototoxin, sodium arsanilate. Besnard et al (2012) conducted their analyses of the hippocampus at 2 months after the second lesion, whereas the current time points were 24 h, 72 h, 1 week, 1 month and 6 months post-BVD. Finally, the current study used western blotting to analyse NMDA receptor subunit expression in different hippocampal subregions, whereas Besnard et al. (2012) used receptor autoradiography to measure the NMDA receptor number and affinity in the whole hippocampus. One

possibility is that the increase in hippocampal NMDA receptor expression observed by Besnard et al. (2012) was a response to the sequence of UVD behavioural syndromes. However, this seems too simplistic an explanation since a decrease in NR1 expression was observed in the ipsilateral CA2/3 region at 2 weeks following UVD (Liu et al., 2003). It is possible that the up-regulation occurs in response to the change in vestibular input to the hippocampus as the second UVD occurs. Whatever the explanation, since both sequential and simultaneous vestibular lesions occur clinically in humans, both paradigms are of interest in terms of their effects on spatial memory and the hippocampus.

It was very surprising to find no significant change in the expression of the different NMDA and AMPA receptor subunits, and CaMKIIα, in the hippocampal subregions following BVD. Given the evidence that hippocampal place cell firing and theta rhythm are dysfunctional following BVD (Stackman et al., 2002; Russell et al., 2003b; Russell et al., 2006; Neo et al., 2012; Tai et al., 2012) and hippocampal field potentials are reduced in hippocampal slices from rats that had received a UVD several months previously (Zheng et al., 2003), significant changes in glutamate receptor subunit expression were predicted. One possibility is that the receptor changes that underlie the physiological abnormalities in the hippocampus that are caused by BVD, are too subtle to be detected using western blotting and that they can only be detected using receptor autoradiography or immunohistochemistry. However, it is conceivable that many of the neurophysiological changes that take place in the hippocampus following BVD do not require changes in the expression of glutamate receptors, but occur as a result of changes in receptor affinity or efficacy, or perhaps do not require receptor plasticity at all. Interestingly, when field potentials were analysed in anesthetised or alert behaving animals following BVD, there were no significant differences in baseline field potentials or in the induction or maintenance of long-term potentiation (Zheng et al., 2010).

It could be argued that the lack of a significant difference between sham and BVD animals was merely due to experimental error. However, there were significant effects of T maze training in all hippocampal subregions at 6 months post-op. In CA1, CaMKIIα, NR1 and NR2B expression were significantly increased, and the expression of GluR1 was significantly decreased. In CA2/3,

CaMKIIα and pCaMKIIα expression were significantly increased, as was the expression of GluR1-3. In the DG, CaMKIIα and pCaMKIIα expression were also significantly increased, as was the expression of GluR1 and GluR3. These results are consistent with previous studies in showing that experience can alter the expression of glutamate receptor subunits in the hippocampus (e.g., Ghafari et al., 2012; Heo et al., 2012). It was particularly interesting that, using cluster analysis, the expression of the neurochemical variables in CA2/3 could reliably distinguish between the animals that received T maze training and those that did not. These results also demonstrate that significant changes in protein expression could be detected using these assays, and that the lack of effect of BVD was unlikely to be due to methodological problems.

It was surprising to see that spatial training resulted in an increased protein expression of glutamate receptors and CaMKIIα in the hippocampus in the same BVD rats that were impaired in spatial alternation (Zheng et al., 2007). It has been shown that performance in T-maze spatial alternation is impaired by the NMDA receptor antagonist, D-(-)-2-Amino-5-phosphonopentanoic acid (D-AP5) and in GluR1 knockout mice (Bannerman et al., 2003; McHugh et al., 2008), which suggests that NMDA and AMPA receptors are important for spatial alternation. However, in the present study, spatial training produced the same degree of increase in protein expression in both sham and BVD rats when compared to the untrained rats, regardless of their spatial alternation performance. This, together with the previous finding that LTP is intact in BVD rats (Zheng et al, 2010), suggests that learning and memory impairment in BVD animals cannot be explained simply by altered receptor plasticity.

Overall, the results of these experiments suggest that BVD is not associated with large changes in glutamate receptor subunit or CaMKIIα expression in the rat hippocampus, but that the neurophysiological changes that occur are more likely to be due to smaller, more subtle alterations in receptor affinity and/or efficacy, or perhaps occur without receptor plasticity.

4.2 Statistical discussion

As mentioned in the Introduction, univariate statistical analyses have been used almost exclusively as a means of analysing experimental results in this area of neuroscience. Probably more than 98% of experimental papers would have employed nothing more than t tests or ANOVAs followed by post-hoc tests (see Table 1). The main objective of this thesis was to compare the use of univariate and multivariate statistical and data mining procedures in the analysis of data in this area of research, using two new data sets as examples. A critical question, therefore, is, what can the multivariate statistical and data mining analyses reveal that the univariate statistical analyses cannot?

4.2.1. Behavioural data set

From the univariate statistical analyses of the behavioural data set which included data on the effects of BVD on performance in the OFM, EPM, ETM and STM, 3 variables were highly significantly different between the BVD and sham control animals in the pre-drug testing: 1) zone activity in the OFM (the amount of time the BVD animals spent in the outer compared to the middle/inner zones was less than for the sham animals); 2) rearing (the duration of supported and unsupported rearing in the OFM was less for the BVD than for the sham animals); 3) the percentage of correct responses in the STM (BVD animals made fewer correct responses than the sham animals). There were some small but significant differences in the ETM data but the effects of the drugs used were negligible for the most part. These significant univariate statistical results are largely consistent with those from previous studies (Stackman et al., 2002; Russell et al., 2003; Zheng et al., 2006; Zheng et al., 2007; Goddard et al., 2008; Zheng et al., 2009; Baek et al., 2010; Besnard et al., 2012; Neo et al., 2012; Machado et al., 2012). The nature of this data set is such that it could have been analysed solely with unvariate statistical procedures. However, none of the previous studies has been able to dissect the relationship between the different behavioural variables that are affected by BVD. For example, in most studies, animals that perform poorly in spatial memory tasks also exhibit locomotor hyperactivity, reduced rearing, a tendency to avoid the outer zone of the OFM, and changes in behavour in anxiety tasks (the current study did not detect the latter to a significant level, however).

Therefore, the question of, to what extent, these other behavioural symptoms account for the apparent spatial memory deficits, needs to be addressed.

In this thesis, it was decided to further investigate the relationship between the different behavioural variables using multivariate statistical and data mining analyses. Because the effects of the drugs used in the experiments were negligible, but there were some significant effects, only the pre-drug data were used in order to avoid potential confounds.

Due to the extent of the correlation amongst some of the variables (see Figure 23), or because they measured the same underlying concept in slightly different ways, it was decided to exclude some variables from the multivariate statistical and data mining analyses. With these 5 variables removed, there were 8 remaining variables to investigate in relation to performance in the STM: surgical group; distance travelled in the OFM; duration of supported rearing; duration of unsupported rearing; zone activity; open arm duration in the EPM; avoidance latency in the ETM; and escape latency in the ETM. Some data reduction was performed in order to make the different variables comparable, as described in Section 2.1.5.6.

A MANOVA was performed prior to LDA, with the independent variable, surgery, and the behavioural measures as the dependent variables. Surgery was significant, which was not surprising given the results of the univariate ANOVAs and LMM analyses. LDA was then used to determine whether animals could be identified as BVD or sham based on their measurements for the 8 behavioural variables. In the first instance, all variables were entered together. A linear discriminant function was identified that could significantly predict the animals' group membership and cross-validation showed that this discriminant function was 100% successful in classifying the animals as BVD or sham. A stepwise LDA was used to determine whether the number of variables in the linear equation could be reduced while still retaining predictive power. A linear function was identified that could significantly predict the animals' group membership based only on the duration of unsupported rearing, zone activity, and % correct in the STM, and this was also 100% successful in classifying the animals as BVD or sham. The assumptions of multivariate normality and homogeneity of the covariance matrices were not necessarily fulfilled and therefore consideration was given to the question of whether the MANOVA and LDA results were valid.

Multivariate normality was not tested but clearly the assumption of univariate normality was not fulfilled in all cases and for this reason data transformations were undertaken. Box's M test suggested that the assumption of homogeneity of the covariance matrices was violated. However, MANOVA and LDA are considered to be reasonably robust against violation of both of these assumptions, provided that the sample sizes are equal (Stevens, 2002; Field, 2011), which they were in this case. Furthermore, the cross-validation results of the LDA strongly suggested that the analysis was valid.

Similar results were obtained using RFs and SVMs for classification, which have the advantage that they do not make distributional assumptions about the data (Wilson, 2008; Hastie et al., 2009; Williams, 2011). Interestingly, it has been suggested that these methods are ideal for cases in which the n:p ratio is small (Wilson, 2008; Hastie et al., 2009). These results again supported the idea that the behavioural profile of BVD animals is quite distinct from that of sham animals and that the spatial memory impairment, as indicated by the decrease in the % correct responses in the STM, is a major component of the behavioural syndrome. Cluster analysis on cases, using the combination of the 8 behavioural variables, showed that it was possible to almost, but not completely, distinguish between the BVD and sham animals, consistent with the results of the LDA, RF and SVM analyses.

Although MLR is not, strictly speaking, a multivariate statistical method, because there is only one dependent variable at a time, in this case it was compared with multivariate statistical methods due to its ability to investigate the relationship between different independent variables. The results of both the backward regression and best subsets regression indicated that the overwhelming predictor of correct performance in the STM, as an index of spatial memory, was whether the animals had BVD surgery or not, with avoidance latency (backward and best subsets regression) and duration of unsupported rearing (best subsets regression) also contributing but less important. The RFR indicated that the duration of supported rearing, rather than whether the animals had BVD surgery, was the most important predictor variable. However, the variance explained by the RFR (70%) was considerably less than the adjusted R^2 for the MLR (88%), suggesting that the MLR was the superior model. It is difficult to compare error estimates for MLR and RFR, since in the latter case the MSE is calculated from

the difference between the model based on the training data set, and the test data set. Nonetheless, the MLR appeared to provide greater explanatory power. These results suggest very strongly that the spatial memory deficits associated with BVD are a direct result of the loss of vestibular information rather than an indirect consequence of the other behavioural effects of the surgery. This is a conclusion that could not have been drawn from the univariate statistical analyses because they did not take into consideration the interactions between the different behavioural variables. Further support for the importance of these MLR results came from the fact that when MLR was used to determine whether any of the other 7 variables could be predicted from the remaining variables, the adjusted R^2 values were generally low (i.e. < 0.5), except for the durations of supported (adjusted $R^2 = 0.71$) and unsupported rearing (adjusted $R^2 = 0.69$) and the ln ratio of the time spent in the inner to the outer zones of the OFM (adjusted $R^2 = 0.62$).

Overall, the results of the multivariate statistical and data mining analyses of the behavioural data set demonstrated that they could answer questions related to the interaction between the different behavioural variables that simple univariate statistical analyses could not.

4.2.2. Neurochemical data set

Unlike the behavioural data set, the neurochemical data set was not analysed first using univariate statistical methods. Because of the complex design of the experiments and the fact that there were 8 target proteins in 3 brain regions at 5 time points, it was decided to use MANOVAs initially to analyse the data and then follow this with univariate ANOVAs, LDAs, RF and cluster analyses.

The results showed that surgery had no significant effect on the expression of the target proteins, but that training in a T maze did have a significant effect on the expression of many of these proteins in the different subregions of the hippocampus, as indicated by many significant univariate ANOVAs.

LDA was then used in an attempt to predict whether animals were BVD or sham animals based on their protein expression profiles. This was uniformly unsuccessful, as was RF analysis, consistent with the results of the MANOVAs. Likewise, whether the animals had received T maze training or not, could not be predicted using LDA. However, RF analyses did show that the trained and the

non-trained animals could be separated from one another to a certain extent, and cluster analyses yielded similar results for CA2/3.

Overall, the results of the multivariate statistical analyses of the neurochemical data set were somewhat disappointing. Rather than multivariate statistical methods being an alternative to univariate methods, it could be argued that this data set was appropriate only for multivariate statistical analysis. However, the effects of T maze training that were found to be significant using the MANOVA, would have been found significant using univariate ANOVAs as well, and the results of the other multivariate statistical tests (LDA, RF and cluster analysis) were unremarkable. Nonetheless, the fact that whether the animals had intact vestibular function or not, could not be predicted from their protein expression profiles, was consistent with the MANOVA results.

4.3. General conclusion

Experimental phenomena in neuroscience usually involve the complex interaction of multiple variables, and yet historically, statistical analyses have focussed on comparisons between treatment groups, of one variable at a time, using simple univariate statistical methods such as t tests and ANOVAs. This approach not only tends to inflate the type 1 error rate as a result of large numbers of statistical analyses, but neglects the fact that changes may occur at the level of the interaction within a system of variables that cannot be detected in individual variables (Stevens, 2002; Liu et al., 2010). In areas such as the analysis of gene microarray and proteomic data, and medical diagnostics, multivariate statistical and data mining analyses are now being employed in an attempt to understand complex interactions between systems of variables (e.g., Pang et al., 2006; Krafczyk et al., 2006; Ryan et al., 2011; Brandt et al., 2012), and there is no reason that they cannot be exploited in neuroscience as well (Smith, 2012b). The relatively small sample sizes that are often used in factorial experimental designs in neuroscience experiments involving animals, does inevitably place some limitations on the analyses that can be done (for example, PCA was not advisable for the behavioural and neurochemical data sets in this case due to the small sample sizes). Nonetheless, the sample sizes for the data sets used in this thesis were adequate for most of the multivariate statistical and data mining analyses performed and the assumption and diagnostic tests suggested that the results were

valid. In order to make the most of the data obtained in neuroscience research, multivariate statistical and data mining methods need to be employed alongside univariate statistical methods, wherever they are appropriate and can offer additional information. This is important not just for the scientific value that is obtained from the research but also for the ethical use of experimental animals (Festing, 2003; Smith, 2012b). One implication of this conclusion is that experiments in this area need to be designed so that multiple variables can be measured under the one set of conditions. This may be easier to do for behavioural experiments; however, in the context of methods such as western blot studies for protein expression, the experimental design would have to include controls for differences in assay conditions between gels (i.e., internal standards), since there are practical limitations to how many proteins can be run on a single gel.

5. References

Adamec, R., Shallow, T. (2000) Effects of baseline anxiety on response to kindling of the right medial amygdala. *Physiology and Behavior.* 70; 67-80.

Agrawal, Y., Carey, J.P., Della Santina, C.C., Schubert, M.C., Minor, L.B. (2009) Disorders of balance and vestibular function in US adults: data from the National Health and Nutrition Examination Survey, 2001-2004. *Archives of Internal Medicine.* 169(10); 938-944.

Avni, R., Elkan, T., Dror, A.A., Shefer, S., Eilam, D., Avraham, K.B., Mintz, M. (2009) Mice with vestibular deficiency display hyperactivity, disorientation, and signs of anxiety. *Behavioural Brain Research.* 202(2); 210-217.

Babayak, M.A. (2004) What you see may not be what you get: A brief, non-technical introduction to over-fitting in regression-type models. *Psychosomatic Medicine.* 66; 411-421.

Baek, J-H., Zheng, Y., Darlington, C.L., Smith, P.F. (2011) Cannabinoid CB1 receptor expression and affinity in the rat hippocampus following bilateral vestibular deafferentation. *Neuroscience Letters.* 487; 330-334.

Baek, J.H., Zheng, Y., Darlington, C.L., Smith, P.F. (2010) Evidence that spatial memory deficits following bilateral vestibular deafferentation in rats are probably permanent. *Neurobiology of Learning and Memory.* 94; 402-413.

Balaban, C.D. (2002) Neural substrates linking balance control and anxiety. *Physiology and Behavior*. 77; 469-475.

Balaban, C.D., Thayer, J.F. (2001) Neurological bases for balance-anxiety links. *Journal of. Anxiety Disorders*. 15(1–2); 53–79.

Bannerman, D.M., Deacon, R.M., Seeburg, P.H., Rawlins, J.N. (2003) GluR-A-deficient mice display normal acquisition of a hippocampus-dependent spatial reference memory task but are impaired during spatial reversal. *Behavioral Neuroscience.* 117; 866-870.

Barria, A., Derkach, V., Griffith, L., Muller, D., Soderling, T. (1997) Regulatory phosphorylation of AMPA-type glutamate receptors by CaMKII during long-term potentiation. *Science* 276; 2042-2045.

Besnard, S., Machado, M.L., Vignaux, G., Boulouard, M., Coquerel, A., Bouet, V., Freret, T., Denise, P., Lelong-Boulouard, V. (2012) Influence of vestibular input on spatial and nonspatial memory and on hippocampal NMDA receptors. *Hippocampus.* 22(4); 814-826.

Best. C., Eckhardt-Henn, A., Diener, G., Bense, S., Breuer, P., Dieterich, M. (2006) Interaction of somatoform and vestibular disorders. *Journal of Neurology, Neurosurgery and Psychiatry*.77(5); 658–664.

Black, F.O., Pesznecker, S., Stallings, V. (2004) Permanent gentamicin vestibulotoxicity. *Otology and Neurotology.* 25(4); 559-569.

Bolmont, B., Gangloff, P., Vouriot, A., Perrin, P.P. (2002) Mood states and anxiety influence abilities to maintain balance control in healthy human subjects. *Neuroscience Letters.* 329(1); 96–100.

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. *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 a two-way classification. *Annals of Mathematical Statistics.* 25; 484-498.

Brammer, R. J. (2003) Modelling covariance structure in ascending dose studies of isolated tissues and organs. *Pharmaceutical Statistics*. 2; 103–112.

Brandt, T., Schautzer, F., Hamilton, D.A., Bruning, R., Markowitsch, H., Kalla, R., Darlington, C.L., Smith, P.F., Strupp. M. (2005) Vestibular loss causes hippocampal atrophy and impaired spatial memory in humans. *Brain.* 128; 2732-2741.

Brandt, T., Strupp, M., Novozhilov, S., Krafczyk, S. (2012) Artificial neural network posturography detects the transition of vestibular neuritis to phobic postural vertigo. *Journal of Neurology.* 259; 182-184.

Breiman, L., Friedman, J., Stone, C.J., Olshen. R.A. (1984) *Classification and Regression Trees*. 1st Edition, CRC Press, Boca Raton.

Brook, R.J., Arnold, G.C. (1985) *Applied Regression Analysis and Experimental Design*. Chapman and Hall/CRC, Boca Raton.

Brown, H., Prescott, R. (2006) *Applied Mixed Models in Medicine. 2nd Edition.* Wiley, Chichester, England.

Bruhwyler, J., Chleide, E., Houbeau, G., Mercier, M. (1991) Stimulant effect of the β–carboline FG 7142 in the open-field test. *European Journal of Pharmacology* 200; 183-185.

Clark, R.A., Shoaib, M., Hewitt, K.N., Stanford, S.C., Bate, ST. (2012) A comparison of InVivoStat with other statistical software packages for analysis of data generated from animal experiments. *Journal of Psychopharmacology*. 26(8); 1136-1142.

Conover, W.J., Iman, R.L. (1981) Rank transformations as a bridge between parametric and non-parametric statistics. *American Statistician.* 35(3); 124-129.

Craig A, Blackstone C, Huganir R, Banker G. (1993) The distribution of glutamate receptors in cultured rat hippocampal neurons: postsynaptic clustering of AMPA selective subunits. *Neuron* 10; 1055-1068.

Cullen, K.E. (2012) The vestibular system: Multimodal integration and encoding of self-motion for motor control. *Trends in Neuroscience*. 35; 185-196.

Dawson, G.R., Crawford, S.P., Stanhope, K.J, Iversen, S.D., Tricklebank, M.D. (1994) One-trial tolerance to the effects of chlordiazepoxide on the elevated plus maze may be due to locomotor habituation, not repeated drug exposure. *Psychopharmacology.* 113; 570-572.

Doncaster, C.P., Davey, A.J.H. (2007) *Analysis of Variance and Covariance.* Cambridge University Press, Cambridge.

Festing, M.F.W. (2003) Principles: The need for better experimental design. *Trends in Pharmacological Sciences.* 24; 341-345.

Field, A. (2011) *Discovering Statistics Using SPSS.* Sage, Los Angeles.

File, S.E., Mabbutt, P.S., Hitchcott, P.K. (1990) Characterisation of the phenomenon of "one-trial tolerance" to the anxiolytic effect of chlordiazepoxide in the elevated plus-maze. *Psychopharmacology.* 102; 98-101.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H. (2004) *Applied Longitudinal Analysis.* Wiley, Hoboken, NJ.

Furman, J.M., Jacob, R.G. (2001) A clinical taxonomy of dizziness and anxiety in the otoneurological setting. *Journal of. Anxiety Disorders*. 15(1–2); 9–26.

Furman, J.M., Redfern, M.S., Jacob, R.G. (2006) Vestibulo-ocular function in anxiety disorders. *Journal of Vestibular Research*. 16(4–5); 209–215.

Gamst, G., Meyers, L.S., Guarino, A.J. (2008) *Analysis of Variance Designs. A Conceptual and Computational Approach With SPSS And SAS.* Cambridge University Press, New York.

Geiger, J., Melcher, T., Koh, D., Sakmann, B., Seeburg, P., Jonas, P., Monyer, H. (1995) Relative abundance of subunit mRNAs determines gating and Ca^{2+} permeability of AMPA receptors in principal neurons and interneurons in rat CNS. *Neuron* 15; 193-204.

Ghafari, M., Falsafi, S.K., Hoeger, H., Lubec, G. (2012) Hippocampal levels of GluR1 and GluR2 complexes are modulated by training in the multiple T-Maze in C57BL/6J mice. *Brain Structure and Function.* 217(2); 353-362.

Giese, K., Fedorov, N., Filipkowski, R., Silva, A. (1998) Autophosphorylation at Thr286 of the calcium-calmodulin kinase II in LTP and learning. *Science* 279; 870-873.

Gilchrist, D.P.D, Smith, P.F., Darlington, C.L. (1990) ACTH (4-10) accelerates ocular motor recovery in the guinea pig following vestibular deafferentation, *Neuroscience Letters.* 118; 14-16.

Goddard, M., Zheng, Y., Darlington, C.L., Smith, P.F.. (2008) Locomotor and exploratory behavior in the rat following bilateral vestibular deafferentation. *Behavioral Neuroscience.* 122; 448-459.

Graeff, F.G., Ferreira Netto, C., Zangrossi, Jr H. (1998) The elevated T-maze as an experimental model of anxiety. *Neuroscience and Biobehavioral Reviews* 23; 237-246.

Gurka, M.J., Edwards, L.J. (2011) Mixed Models. In: *Essential Statistical Methods for Medical Statistics*, Rao, C.R., Miller, J.P., Rao, D.C. (Eds), Elsevier, Amsterdam, pp. 146-173.

Halberstadt, A.L., Balaban, C.D. (2006) Serotonergic and nonserotonergic neurons in the dorsal raphe nucleus send collateralized projections to both the vestibular nuclei and the central amygdaloid nucleus. *Neuroscience*. 140(3); 1067–1077.

Hastie, T., Tibshirani, R., Friedman, J. (2009) *Elements of Statistical Learning: Data Mining, Inference and Prediction.* 2nd Edition. Springer Verlag, Heidelberg.

Heo S, Jung G, Beuk T, Höger H, Lubec G. (2012) Hippocampal glutamate transporter 1 (GLT-1) complex levels are paralleling memory training in the Multiple T-maze in C57BL/6J mice. *Brain Structure and Function.* 217(2); 363-378.

Hüfner, K., Hamilton, D.A., Kalla, R., Stephan, T., Glasauer, S., Ma, J., Brüning, R., Markowitsch, H.J., Labudda, K., Schichor, C., Strupp, M., Brandt, T. (2007) Spatial memory and hippocampal volume in humans with unilateral vestibular deafferentation. *Hippocampus.* 17(6); 471-485.

Ishii, T., Moriyoshi, K., Sugihara, H., Sakurada, K., Kadotani, H., Yokoi, M., Akazawa, C., Shigemoto, R., Mizuno, N., Masu, M. (1993) Molecular characterization of the family of the N-methyl-D-aspartate receptor subunits. *Journal of Biological Chemistry* 268; 2836-2843.

Jauregui-Renaud, K., Ramos-Toledo, V., Aguilar-Bolanos, M., Montano-Velazquez, B., Pliego-Maldonado, A. (2008a) Symptoms of detachment from the self or from the environment in patients with an acquired deficiency of the special senses. *Journal of Vestibular Research*. 18; 129-137.

Jauregui-Renaud, K., Sang, F.Y., Gresty, M.A., Green, D.A., Bronstein, A.M. (2008b) Depersonalisation/derealisation symptoms and updating orientation in patients with vestibular disease. *Journal of Neurology, Neurosurgery and Psychiatry*. 79; 276–283.

Kaiser, H.F. (1974) An index of factorial simplicity. *Psychometrika.* 39; 31-36.

Kalueff, A.V., Ishikawa, K., Griffith, A.J. (2008) Anxiety and otovestibular disorders: linking behavioral phenotypes in men and mice. *Behavioural Brain Research.* 186; 1–11.

Kalueff, A.V., Murphy, D.L. (2007) The importance of cognitive phenotypes in experimental modeling of animal anxiety and depression. *Neural Plasticity*. 2007; 52087.

Keppel, G., Wickens, T.D. (2004) *Design and Analysis: A Researcher's Handbook. 4th Edition.* Pearson Prentice Hall, Saddle River, NJ.

Kirk, R.E. (2013) *Experimental Design. Procedures for the Behavioral Sciences. 4th Edition.* Sage, LA.

Krafczyk, S., Tietze, S., Swoboda, W., Valkovic, P., Brandt, T. (2006) Artificial neural network: A new diagnostic posturographic tool for disorders of stance. *Clinical Neurophysiology.* 117; 1692-1698.

Kuehl, R.O. (2000) *Design of Experiments: Statistical Principles of Research Design and Analysis. 2nd Edition.* Duxbury Press, Pacific Grove, CA.

Krupp, J., Vissel, B., Heinemann, S., Westbrook, G. (1998) N-terminal domains in the NR2 subunit control desensitization of NMDA receptors. *Neuron* 20; 317-327.

Kutner, M.H., Nachtsheimm C.J., Neter, J., Li, W. (2005) *Applied Linear Statistical Models.* McGraw-Hill Irwin, Boston.

Kutsuwada, T., Kashiwabuchi, N., Mori, H., Sakimura, K., Kushiya, E., Araki, K., Meguro, H., Masaki, H., Kumanishi, T., Arakawa, M. (1992) Molecular diversity of the NMDA receptor channel. *Nature.* 358; 36-41.

Lattin, J., Carroll, J.D., Green, P.E. (2003) *Analyzing Multivariate Data.* Duxbury, Pacific Grove, CA.

Leonard, A., Lim, I., Hemsworth, D., Horne, M., Hell, J. (1999) Calcium/calmodulin-dependent protein kinase II is associated with the N-methyl-D-aspartate receptor. *Proceedings of the National Academy of Sciences of the United States of America*. 96; 3239-3244.

Lenth, R.V. (2001) Some practical guidelines for effective sample size determination. *American Statistician.* 55; 187-193.

Lepicard, E.M., Venault, P., Perez-Diaz, F., Joubert, C., Berthoz, A., Chapouthier, G. (2000) Balance control and posture differences in the anxious BALB/cByJ mice compared to the non-anxious C57BL/6J mice. *Behavioural Brain Research.* 117(12); 185–195.

Lindemann, S., Gernert, M., Bennay, M., Koch, M., Loscher, W. (2007) Comparative analysis of anxiety-like behaviors and sensorimotor functions in two rat mutants, ci2 and ci3, with lateralized rotational behavior. *Physiology and. Behavior*. 91(5): 551–560.

Little, R.C., Pendergast, J., Natarajan, R. (2000) Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine.*19; 1793-1819.

Liu, P., Zhang, H., Devaraj, R., Ganesalingam, G., Smith, P.F. (2010) A multivariate analysis of the effects of aging on glutamate, GABA and arginine metabolites in the rat vestibular nucleus. *Hearing Research*. 269; 122-133.

Liu, P., King, J., Zheng, Y., Darlington, C.L., Smith, P.F. (2003) Long-term changes in hippocampal NMDA receptor subunits following peripheral vestibular damage. *Neuroscience.* 117; 965-970.

Machado, M.L., Lelong-Boulouard, V., Smith, P.F., Freret, T., Philoxene, B., Denise, P., Besnard, S. (2012) Influence of anxiety in spatial memory impairments related to the loss of vestibular function in rat**.** *Neuroscience.* 218; 161-169.

Manly, B.F.J. (2005) *Multivariate Statistical Analyses. A Primer.* 3rd Edition. Chapman and Hall/CRC, London.

Marcoulides, G.A., Hershberger, S.L. (1997) *Multivariate Statistical Methods. A First Course.* Lawrence Erlbaum Assoc., Mahwah, New Jersey.

Marsland, S. (2009) *Machine Learning. An Algorithmic Perspective.* CRC Press, Boca Raton.

McHugh, S.B., Niewoehner, B., Rawlins, J.N., Bannerman, D.M. (2008) Dorsal hippocampal N-methyl-D-aspartate receptors underlie spatial working memory performance during non-matching to place testing on the T-maze. *Behavioural Brain Research*. 186; 41-47.

Miller, I., Miller, M. (2004) *John E. Freund's Mathematical Statistics with Applications.* Pearson Prentice Hall, NJ.

Mori, M., Gotoh, T. (2004) Arginine metabolic enzymes, nitric oxide and infection. *Journal of Nutrition.* 134; 2820S-2028S.

Motulsky, H. (1995) *Intuitive Biostatistics.* Oxford University Press, New York.

Neo, P., Carter, D., Zheng, Y., Smith, P.F., Darlington, C.L., McNaughton, N. (2012) Septal elicitation of hippocampal theta rhythm did not repair the cognitive and emotional deficits resulting from vestibular lesions. *Hippocampus.* 22: 1176-1187.

Norma, G., Monteiro, S. and Salama, S. (2012) Sample size calculations: Should the emperor's clothes be off the peg or made to measure? *British Medical Journal*. 345; 19-21.

Pang, H., Lin, A., Holford, M., Enerson, B.E., Lu, B., Lawton, M.P., Floyd, E., Zhao, H. (2006) Pathway analysis using random forests classification and regression. *Bioinformatics.* 22; 2028-2036.

Pellow, S., File, S.E. (1986) Anxiolytic and anxiogenic drug effects on exploratory activity in an elevated plus-maze: a novel test of anxiety in the rat. *Pharmacology Biochemistry and Behavior*. 24; 525-529.

Petralia, R., Yokotani, N., Wenthold, R. (1994) Light and electron microscope distribution of the NMDA receptor subunit NMDAR1 in the rat nervous

system using a selective anti-peptide antibody. *Journal of Neuroscience* 14; 667-696.

Poltronieri, S.C., Zangrossi, H., de Barros Viana, M. (2003) Antipanic-like effect of serotonin reuptake inhibitors in the elevated T-maze. *Behavioural Brain Research.* 147; 185-192.

Posada, D., Buckley, T.R. (2004) Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology.* 53(5); 793-808.

Quinn, G.P., Keough, M.J. (2002) *Experimental Design and Data Analysis for Biologists.* Cambridge University Press, Cambridge.

Redfern, M.S., Talkowski, M.E., Jennings, J.R., Furman, J.M. (2004) Cognitive influences in postural control of patients with unilateral vestibular loss. *Gait and Posture*. 19; 105-114.

Russell, N., Horii, A., Smith, P.F., Darlington C.L., Bilkey. (2003a) Long-term effects of permanent vestibular lesions on hippocampal spatial firing. *Journal of Neuroscience* 23; 6490-6498.

Russell, N., Horii, A., Smith, P.F., Darlington, C.L., Bilkey, D. (2003b) Effects of bilateral vestibular deafferentation on radial arm maze performance. *Journal of Vestibular Research.* 13; 9-16.

Russell, N., Horii, A., Smith P.F., Darlington C.L., Bilkey, D. (2006) Lesions of the vestibular system disrupt hippocampal theta rhythm in the rat. *Journal of Neurophysiology* 96; 4-14.

Rutherford, A. (2001) *Introducing ANOVA and ANCOVA. A GLM Approach.* Sage Publications, London.

Ryan, T.P. (2009) *Modern Regression Methods.* Wiley, New Jersey.

Ryan, M., Mason-Parker, S.E., Tate, W.P., Abraham, W.C., Williams, J.M. (2011) Rapidly induced gene networks following induction of long term potentiation at perforant synapses *in vivo*. *Hippocampus.* 21; 541-553.

Seber, G.A.F. (1984) *Multivariate Observations.* Wiley, New York.

Shefer, S., Gordon, C.R., Avraham, K.B., Mintz, M. (2010) Progressive vestibular mutation leads to elevated anxiety. *Brain Research.* 1317; 157-164.

Siemiątkowski, M., Sienkiewicz-Jarosz, H., Członkowska, A., Bidziński, A., Płaźnik, A. (2000) Effects of buspirone, diazepam, and zolpidem on open

field behavior, and brain [^{3}H] muscimol binding after buspirone pretreatment. *Pharmacology Biochemistry and Behavior*. 66; 645-651.

Simon, N.M., Parker, S.W., Wernick-Robinson, M., Oppenheimer, J.E., Hoge, E.A., Worthington, J.J., Korbly, N.B., Pollack, M.H. (2005) Fluoxetine for vestibular dysfunction and anxiety: a prospective pilot study. *Psychosomatics* 46(4); 334–339.

Skene, S. S., Kenward, M. G. (2010a) The analysis of very small samples of repeated measurements I: An adjusted sandwich estimator. *Statististics in Medicine.* 29; 2825–2837.

Skene, S. S., Kenward, M. G. (2010b) The analysis of very small samples of repeated measurements II: A modified Box correction. *Statististics in Medicine.* 29; 2838–2856.

Small, D.S., Volp, K.G., Rosenbaum, P.R. (2011) Structured testing of 2 x 2 factorial effects: An analytic plan requiring fewer observations. *American Statistician.* 65(1); 11-15.

Smith, P.F. (2012a) A note on the advantages of using linear mixed model analysis with maximal likelihood estimation over repeated measures ANOVAs in psychopharmacology: Comment on Clark et al. (2012). *Journal of Psychopharmacology*. 26(12); 1605-1607.

Smith, P.F. (2012b) Statistical analysis in pharmacology is not always BO. *Trends in Pharmacological Sciences.* 33(11); 565-566.

Smith, P.F., Curthoys, I.S. (1989) Mechanisms of recovery from unilateral labyrinthectomy: A review. *Brain Research Reviews.* 14; 155-180.

Smith, P.F., Darlington, C.L., Zheng, Y. (2010) Move it or lose it: Is stimulation of the vestibular system necessary for normal spatial memory? *Hippocampus.* 20; 36-43.

Smith, P.F., Horii, A., Russell, N., Bilkey, D.K., Zheng, Y., Liu, P., Kerr, D.S., Darlington, C.L. (2005) The effects of vestibular lesions on hippocampal function in rats. *Progress in Neurobiology*. 75; 391-405.

Smith, P.F., Zheng, Y., Horii, A., Darlington, C.L. (2005) Does vestibular damage cause cognitive dysfunction in humans? *Journal of Vestibular Research.* 15; 1-9.

Snedecor, G.W., Cochran, W.G. (1989) *Statistical Methods. 8th Edition.* Iowa State University Press, Ames.

Staab, J.P. (2006) Chronic dizziness: the interface between psychiatry and neuro-otology. *Current Opinion in Neurology.* 19(1); 41–48.

Staab, J.P., Ruckenstein, M.J., Solomon, D., Shepard, N.T. (2002) Serotonin reuptake inhibitors for dizziness with psychiatric symptoms. *Archives of Otolaryngology and Head Neck Surgery.* 128(5); 554–560.

Stackman, R., Clark, A., Taube, J. (2002) Hippocampal spatial representations require vestibular input. *Hippocampus* 12; 291-303.

Stefanski, R., Kostowski, W., Palznik, A. (1992) The comparison of benzodiazepine derivatives and serotonergic agonists and antagonists in two animal models of anxiety. *Neuropharmacology.* 31; 1251-1258.

Stevens, J.P. (2002) *Applied Multivariate Statistics for the Social Sciences. 4th Edition.*, Hillsdale NJ, Lawrence Erlbaum.

Stiles, L., Zheng, Y., Darlington, C.L., Smith, P.F. (2012) The D_2 dopamine receptor and locomotor hyperactivity following bilateral vestibular deafferentation in the rat. *Behavioural Brain Research.* 227; 150-158.

Tabachnick, B.G., Fidell, L.S. (2007) *Using Multivariate Statistics. 5th Edition.* Pearson Education Inc. Boston.

Tai, S.K., Ma, J., Ossenkopp, K.P., Leung, L.S. (2012) Activation of immobility-related hippocampal theta by cholinergic septohippocampal neurons during vestibular stimulation. *Hippocampus.* 22(4); 914-925.

Venault, P., Rudrauf, D., Lepicard, E.M., Berthoz, A., Jouvent, R., Chapouthier, G. (2001) Balance control and posture in anxious mice improved by SSRI treatment. *NeuroReport.* 12(14); 3091–3094.

Vittinghoff, E., Glidden, D.V., Shiboski, S.C., McCulloch, C.E. (2005) *Regression Methods in Statistics: Linear, Logistic, Survival and Repeated Measures Models.* Springer, New York.

Wenthold, R., Petralia, R., Blahos, J., Niedzielski, A. (1996) Evidence for multiple AMPA receptor complexes in hippocampal CA1/CA2 neurons. *Journal of Neuroscience* 16; 1982-1989.

West, B.T., Welch, K.B., Galecki, A.T. (2007) *Linear Mixed Models. A Practical Guide Using Statistical Software.* Chapman and Hall/CRC, Boca Raton.

Wilcox, R.R. (1987) New designs in analysis of variance. In: Rosenweig, M.R. and Porter, L.W. (Eds.), *Annual Review Of Psychology.* Annual Reviews Inc., Palo Alto, CA., pp. 29-60.

Williams, G.J. (2009) Rattle: A data mining GUI for R. *The R Journal.* 1/2; 45-55.

Williams, G.J. (2011) *Data Mining With Rattle And R.* Springer, New York.

Williams, K., Zappia, A., Pritchett, D., Shen, Y., Molinoff, P. (1994) Sensitivity of the N-methyl-D-aspartate receptor to polyamines is controlled by NR2 subunits. *Molecular Pharmacology* 45; 803-809.

Wilson, M.D. (2008) Support vector machines. In *Encyclopedia of Ecology.* Elsevier, New York. pp. 3431-3437.

Winer, B.J., Brown, D.R., Michels, K.M. (1991) *Statistical Principles in Experimental Design. 3rd Edition.* McGraw-Hill, New York.

Zheng, Y., Darlington, C.L., Smith, P.F. (2006) Impairment and recovery on a food foraging task following unilateral vestibular deafferentation in rats. *Hippocampus.* 16; 368-378.

Zheng, Y., Goddard, M., Darlington, C.L., Smith, P.F. (2007) Bilateral vestibular deafferentation impairs performance in a spatial forced alternation task in rats. *Hippocampus.* 17; 253-256.

Zheng, Y., Goddard, M., Darlington, C.L., Smith, P.F. (2008) Effects of bilateral vestibular deafferentation on anxiety-related behaviours in Wistar rats. *Behavioural Brain Research.* 193; 55-62.

Zheng, Y., Goddard, M., Darlington, C.L., Smith, P.F. (2009) Long-term deficits on a foraging task after bilateral vestibular deafferentation in rats. *Hippocampus.* 19; 480-486.

Zheng, Y., Kerr, D.S., Darlington, C.L., Smith, P.F. (2003) Peripheral vestibular damage causes a lasting decrease in the electrical excitability of CA1 in hippocampal slices *in vitro*. *Hippocampus* 13; 873-878.

Zheng, Y., Horii, A., Appleton, I., Darlington, C.L., Smith, P.F. (2001) Damage to the vestibular inner ear causes long-term changes in neuronal nitric oxide synthase expression in the rat hippocampus. *Neuroscience.* 105; 1-5**.**

Zheng, Y., Mason-Parker, S.E., Logan, B., Darlington, C.L., Smith, P.F., Abraham, W.C. (2010) Hippocampal synaptic transmission and LTP *in vivo* are intact following bilateral vestibular deafferentation in the rat. *Hippocampus.* 20; 461-468.

Zheng, Y., Vagal, S., McNamara, E., Darlington, C.L., Smith, P.F. (2012) A dose-response analysis of the effects of L-baclofen on chronic tinnitus induced by acoustic-trauma in rats. *Neuropharmacology.* 62; 940-946.

6. Appendices

6.1 Appendix 1

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
test1	.097	18	.200*	.976	18	.900
test2	.235	18	.010	.716	18	.000
test3	.126	18	.200*	.930	18	.197
test4	.112	18	.200*	.942	18	.310
test5	.098	18	.200*	.974	18	.864

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 28: Example of the results of the normality tests for the 5 data sets for the locomotor distance data analysed in Section 3.1.1.1. Note that only the data for test 2 demonstrated a departure from the assumption of normality.

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
test1	4.163	1	16	.058
test2	3.346	1	16	.086
test3	1.280	1	16	.275
test4	.706	1	16	.413
test5	.629	1	16	.439

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + surgery

Within Subjects Design: time

Table 29: Example of the results of the homogeneity of variance tests for the 5 data sets for the locomotor distance data analysed in Section 3.1.1.1. Note none of the data sets demonstrated a departure from the assumption of homogeneity of variance.

Independent Samples Test

		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig. (2-tailed)
test1	Equal variances assumed	4.163	.058	.156	16	.878
	Equal variances not assumed			.149	11.914	.884

Table 30: Example of the results of a 2 sample independent t test for the pre-drug locomotor distance data analysed in Section 3.1.1.1, with (top) and without (bottom) homogeneity of variance assumed. There was no significant difference between the BVD and sham groups in either case.

Mauchly's Test of Sphericity[a]

Measure: dist

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.333	16.167	5	.007	.597	.710	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + surgery

Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Table 31: Example of Mauchly's test of sphericity for the locomotor distance data analysed in Section 3.1.1.1. Note that the test demonstrated a departure from the assumption of sphericity at $P \leq 0.007$.

Tests of Between-Subjects Effects

Measure: dist

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	725094188.081	1	725094188.081	331.560	.000
surgery	27489595.769	1	27489595.769	12.570	.003
Error	34990638.537	16	2186914.909		

Table 32: Example of a repeated measures ANOVA, in this case, for the locomotor distance data analysed in Section 3.1.1.1. The between group factor, surgery, was significant, $F(1,16) = 12.57$, $P \leq 0.003$.

Tests of Within-Subjects Effects

Measure: dist

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	8119414.093	3	2706471.364	4.167	.011
	Greenhouse-Geisser	8119414.093	1.792	4530540.769	4.167	.029
	Huynh-Feldt	8119414.093	2.130	3812434.941	4.167	.022
	Lower-bound	8119414.093	1.000	8119414.093	4.167	.058
time * surgery	Sphericity Assumed	2120028.750	3	706676.250	1.088	.363
	Greenhouse-Geisser	2120028.750	1.792	1182951.944	1.088	.344
	Huynh-Feldt	2120028.750	2.130	995450.114	1.088	.352
	Lower-bound	2120028.750	1.000	2120028.750	1.088	.312
Error(time)	Sphericity Assumed	31179766.075	48	649578.460		
	Greenhouse-Geisser	31179766.075	28.674	1087372.190		
	Huynh-Feldt	31179766.075	34.075	915020.070		
	Lower-bound	31179766.075	16.000	1948735.380		

Table 33: Example of a repeated measures ANOVA, in this case, for the locomotor distance data analysed in Section 3.1.1.1. The repeated measure, time, representing the different drug conditions, was significant, using the Greenhouse-Geisser correction ($F(1.79, 28.67) = 4.17$, $P \leq 0.03$); however, there was no significant interaction between time and surgery.

Information Criteria[a]

-2 Restricted Log Likelihood	101.522
Akaike's Information Criterion (AIC)	145.522
Hurvich and Tsai's Criterion (AICC)	158.016
Bozdogan's Criterion (CAIC)	225.699
Schwarz's Bayesian Criterion (BIC)	203.699

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Time.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	14.936	2821.632	.000
Surgery	1	16.000	99.311	.000
Zone	1	16.011	.870	.365
Surgery * Zone	1	15.999	85.154	.000

a. Dependent Variable: Time.

Table 34: Example of an LMM analysis, using an Unstructured covariance matrix structure, for the zone data analysed in Section 3.1.1.2. The surgery condition was significant, $F(1,16) = 99.31$, $P \leq 0.0005$), as was the surgery x zone interaction $F(1,16) = 85.15$, $P \leq 0.0005$); however, zone itself was not.

Pairwise Comparisons[b]

(I) Trial	(J) Trial	Mean Difference (I-J)	Std. Error	df	Sig.[a]	95% Confidence Interval for Difference[a]	
						Lower Bound	Upper Bound
1.00	2.00	-44.453*	16.577	63.822	.028	-85.212	-3.694
	3.00	-56.977*	17.846	83.414	.006	-100.580	-13.375
2.00	1.00	44.453*	16.577	63.822	.028	3.694	85.212
	3.00	-12.524	16.577	63.822	1.000	-53.283	28.235
3.00	1.00	56.977*	17.846	83.414	.006	13.375	100.580
	2.00	12.524	16.577	63.822	1.000	-28.235	53.283

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

b. Dependent Variable: Alatency.

Table 35: Example of Bonferroni post-hoc tests on avoidance latency, as reported in Section 3.1.3.2. There were significant differences between trials 1 and 2 ($P \leq 0.03$) and 2 and 3 ($P \leq 0.006$).

6.2 Appendix 2

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	1.000	3809.828[b]	8.000	9.000	.000
	Wilks' Lambda	.000	3809.828[b]	8.000	9.000	.000
	Hotelling's Trace	3386.514	3809.828[b]	8.000	9.000	.000
	Roy's Largest Root	3386.514	3809.828[b]	8.000	9.000	.000
surgery	Pillai's Trace	.960	27.326[b]	8.000	9.000	.000
	Wilks' Lambda	.040	27.326[b]	8.000	9.000	.000
	Hotelling's Trace	24.290	27.326[b]	8.000	9.000	.000
	Roy's Largest Root	24.290	27.326[b]	8.000	9.000	.000

a. Design: Intercept + surgery

b. Exact statistic

Table 36: Example of the MANOVA performed on the behavioural data, as reported in Section 3.2.1. Using Pillai's Trace statistic, there was a significant surgery effect ($F(8,9) = 27.33$, $P \leq 0.0005$).

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
sduration	3.203	1	16	.092
uduration	5.156	1	16	.037
distance	4.163	1	16	.058
epmdur	.181	1	16	.676
atrial3	19.239	1	16	.000
etrial3	.786	1	16	.388
lnIO	1.192	1	16	.291
lnpercent	8.077	1	16	.012

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + surgery

Table 37: Example of the homogeneity of variance tests performed on the behavioural data, as reported in Section 3.2.1. Three of the 8 variables (uduration, atrial 3 and ln percent) had significantly heterogeneous variances.

Stepwise Statistics

Summary of Canonical Discriminant Functions

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.056	41.878	3	.000

Standardized Canonical Discriminant Function Coefficients

	Function
	1
uduration	.810
lnIO	-.655
lnpercent	.986

Classification Results[a,c]

		treatment	Predicted Group Membership		Total
			.00	1.00	
Original	Count	.00	10	0	10
		1.00	0	8	8
	%	.00	100.0	.0	100.0
		1.00	.0	100.0	100.0
Cross-validated[b]	Count	.00	10	0	10
		1.00	0	8	8
	%	.00	100.0	.0	100.0
		1.00	.0	100.0	100.0

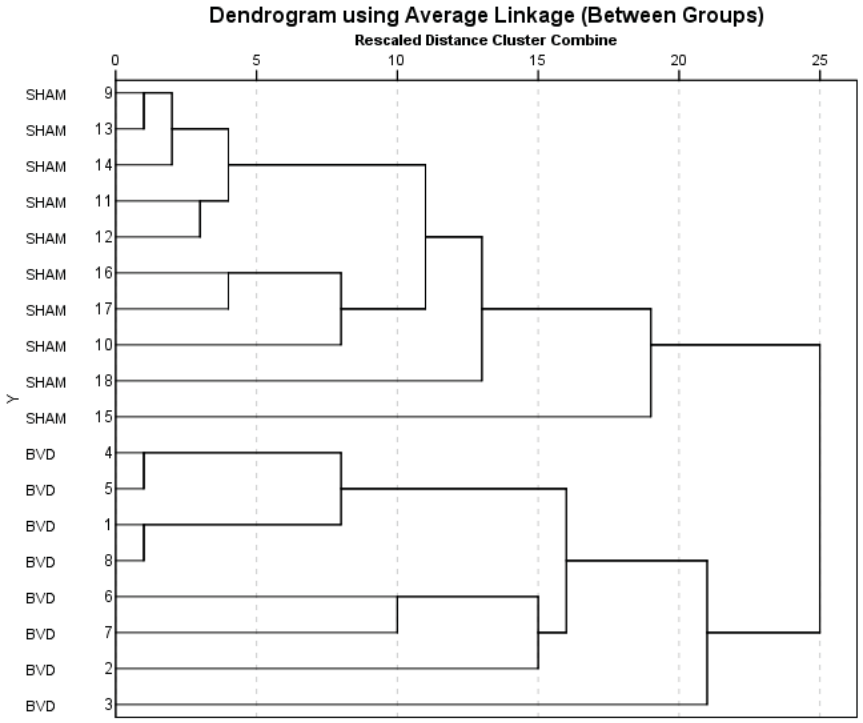
a. 100.0% of original grouped cases correctly classified.

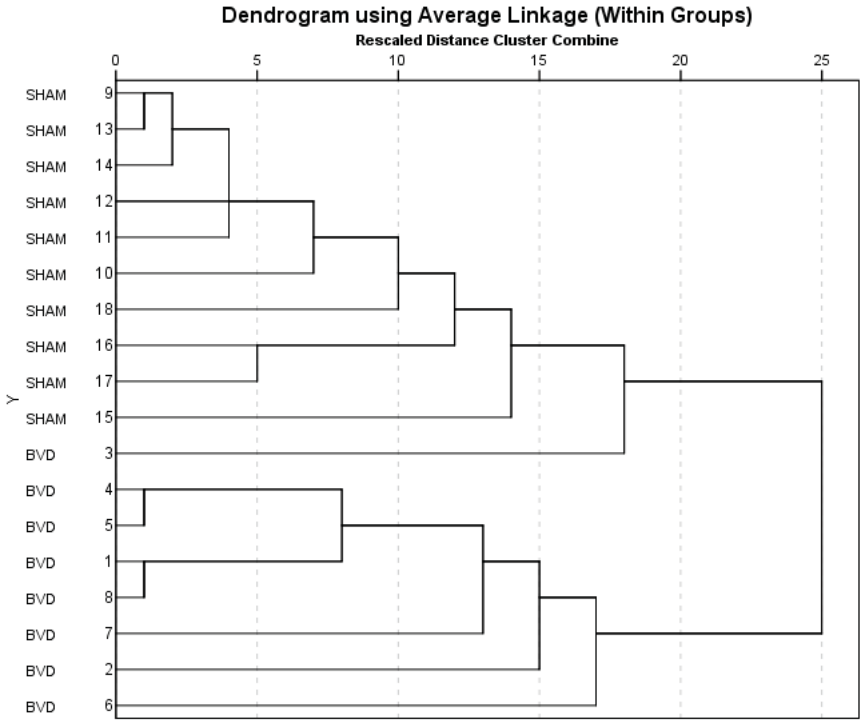
b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

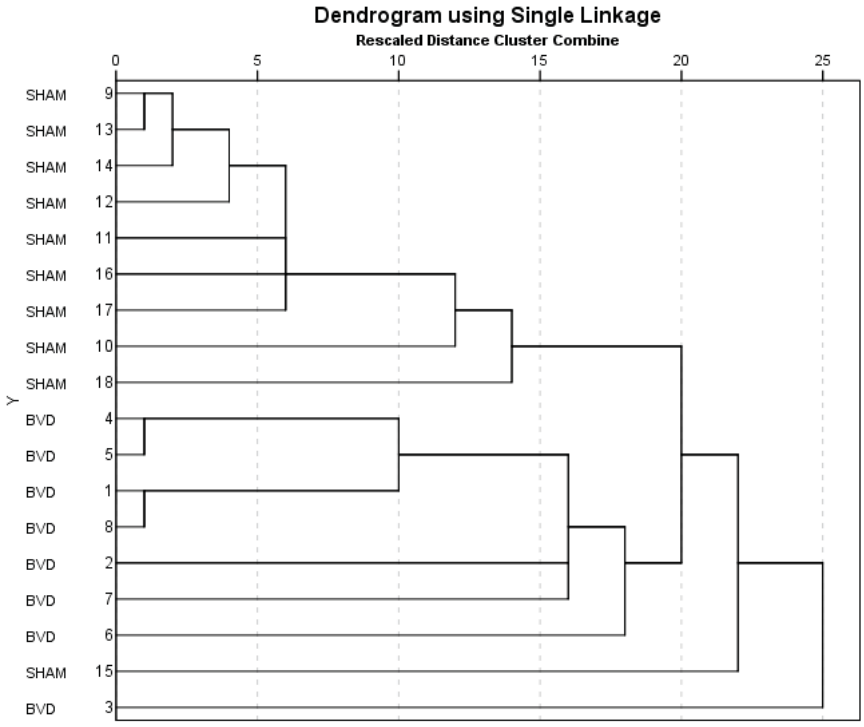
c. 100.0% of cross-validated grouped cases correctly classified.

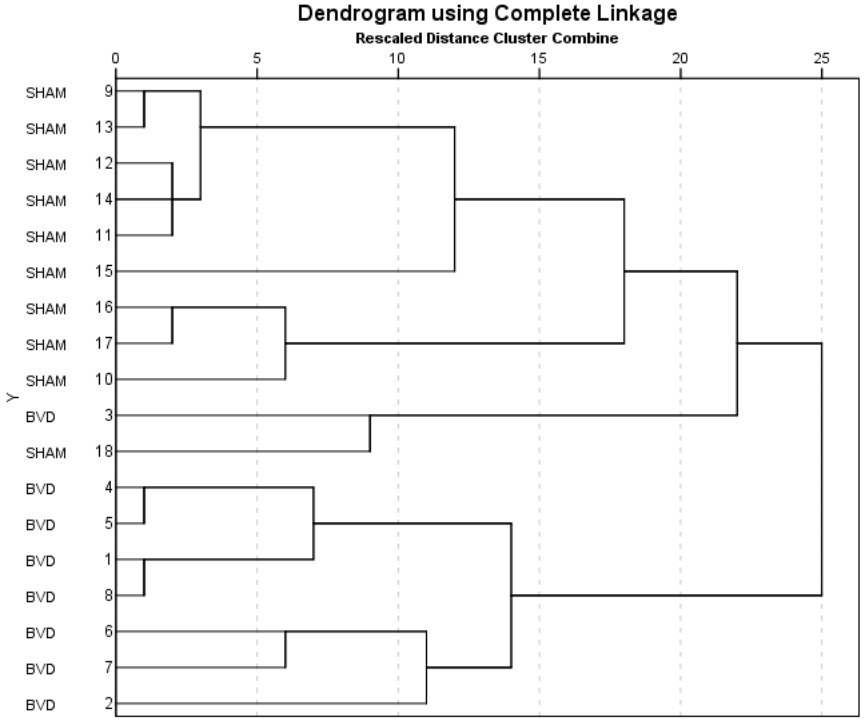
Tables 38-42: Example of the stepwise LDA reported in Section 3.2.2.

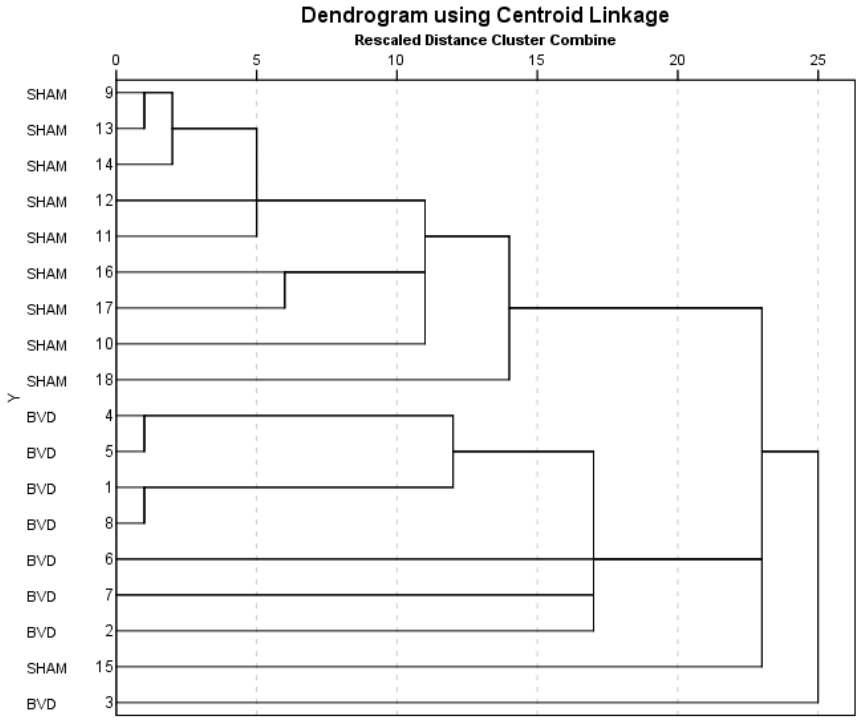
6.3 Appendix 3

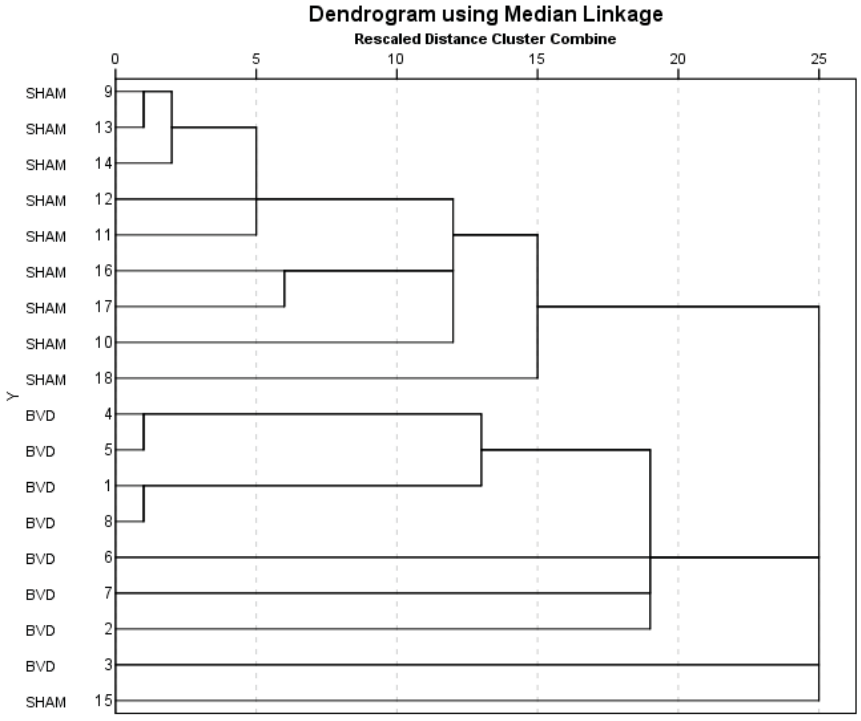





Dendrogram using Single Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25
Y
SHAM 9
SHAM 13
SHAM 14
SHAM 12
SHAM 11
SHAM 16
SHAM 17
SHAM 10
SHAM 18
BVD 4
BVD 5
BVD 1
BVD 8
BVD 2
BVD 7
BVD 6
SHAM 15
BVD 3


Dendrogram using Complete Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25
Y
SHAM 9
SHAM 13
SHAM 12
SHAM 14
SHAM 11
SHAM 15
SHAM 16
SHAM 17
SHAM 10
BVD 3
SHAM 18
BVD 4
BVD 5
BVD 1
BVD 8
BVD 6
BVD 7
BVD 2





Figures 54-59: Other methods of cluster analysis, such as between group (average) linkage, within group (average) linkage, nearest neighbour (single linkage), furthest neighbour (complete linkage) centroid clustering and median clustering, for the 8 variables from the behavioural data.

6.4 Appendix 4

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.970	60.647[b]	7.000	13.000	.000
	Wilks' Lambda	.030	60.647[b]	7.000	13.000	.000
	Hotelling's Trace	32.656	60.647[b]	7.000	13.000	.000
	Roy's Largest Root	32.656	60.647[b]	7.000	13.000	.000
surgery	Pillai's Trace	.191	.439[b]	7.000	13.000	.861
	Wilks' Lambda	.809	.439[b]	7.000	13.000	.861
	Hotelling's Trace	.236	.439[b]	7.000	13.000	.861
	Roy's Largest Root	.236	.439[b]	7.000	13.000	.861
training	Pillai's Trace	.899	16.597[b]	7.000	13.000	.000
	Wilks' Lambda	.101	16.597[b]	7.000	13.000	.000
	Hotelling's Trace	8.937	16.597[b]	7.000	13.000	.000
	Roy's Largest Root	8.937	16.597[b]	7.000	13.000	.000
surgery * training	Pillai's Trace	.230	.555[b]	7.000	13.000	.779
	Wilks' Lambda	.770	.555[b]	7.000	13.000	.779
	Hotelling's Trace	.299	.555[b]	7.000	13.000	.779
	Roy's Largest Root	.299	.555[b]	7.000	13.000	.779

a. Design: Intercept + surgery + training + surgery * training

b. Exact statistic

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	CamKII	9.787[a]	3	3.262	15.386	.000
	pCamKii	1.286[b]	3	.429	1.299	.304
	NR1	144.599[c]	3	48.200	3.047	.054
	NR2B	8.988[d]	3	2.996	2.795	.068
	GluR1	3.532[e]	3	1.177	1.889	.166
	GluR2	.213[f]	3	.071	.116	.950
	GluR3	.164[g]	3	.055	.012	.998
Intercept	CamKII	33.437	1	33.437	157.697	.000
	pCamKii	43.258	1	43.258	131.065	.000
	NR1	368.516	1	368.516	23.294	.000
	NR2B	29.283	1	29.283	27.319	.000
	GluR1	119.585	1	119.585	191.893	.000
	GluR2	116.684	1	116.684	190.270	.000
	GluR3	442.575	1	442.575	99.568	.000

	CamKII	.313	1	.313	1.475	.239
	pCamKii	.041	1	.041	.123	.730
	NR1	10.392	1	10.392	.657	.428
surgery	NR2B	.024	1	.024	.022	.883
	GluR1	9.837E-005	1	9.837E-005	.000	.990
	GluR2	.064	1	.064	.105	.749
	GluR3	.105	1	.105	.024	.880
	CamKII	9.165	1	9.165	43.224	.000
	pCamKii	.474	1	.474	1.435	.246
	NR1	126.945	1	126.945	8.024	.011
training	NR2B	8.457	1	8.457	7.890	.011
	GluR1	3.396	1	3.396	5.449	.031
	GluR2	.156	1	.156	.255	.619
	GluR3	.002	1	.002	.000	.984
	CamKII	.256	1	.256	1.207	.286
	pCamKii	.710	1	.710	2.152	.159
	NR1	6.563	1	6.563	.415	.527
surgery * training	NR2B	.345	1	.345	.322	.577
	GluR1	.076	1	.076	.122	.731
	GluR2	3.359E-005	1	3.359E-005	.000	.994
	GluR3	.063	1	.063	.014	.907
	CamKII	4.029	19	.212		
	pCamKii	6.271	19	.330		
	NR1	300.582	19	15.820		
Error	NR2B	20.366	19	1.072		
	GluR1	11.841	19	.623		
	GluR2	11.652	19	.613		
	GluR3	84.454	19	4.445		
	CamKII	49.038	23			
	pCamKii	51.112	23			
	NR1	836.308	23			
Total	NR2B	60.004	23			
	GluR1	134.201	23			
	GluR2	128.671	23			
	GluR3	530.938	23			
	CamKII	13.816	22			
	pCamKii	7.557	22			
	NR1	445.181	22			
Corrected Total	NR2B	29.354	22			
	GluR1	15.373	22			
	GluR2	11.865	22			
	GluR3	84.618	22			

a. R Squared = .708 (Adjusted R Squared = .662)

b. R Squared = .170 (Adjusted R Squared = .039)

c. R Squared = .325 (Adjusted R Squared = .218)

d. R Squared = .306 (Adjusted R Squared = .197)

e. R Squared = .230 (Adjusted R Squared = .108)

f. R Squared = .018 (Adjusted R Squared = -.137)

g. R Squared = .002 (Adjusted R Squared = -.156)

Tables 43 and 44: Example of the MANOVA performed on the neurochemical data for CA1 at 6 months post-op, as reported in Section 3.3.3. Pillai's Trace statistic was significant for training ($F(7,13) = 16.60$, $P \leq 0.0005$), with significant univariate ANOVAs for training for CamKIIα, NR1, NR2B and GluR1. However, there was no significant surgery effect or significant interaction.