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.