



University
of Glasgow

McLean, John (2012) *The investigation of hippocampal and hippocampal subfield volumetry, morphology and metabolites using 3T MRI*. PhD thesis.

<http://theses.gla.ac.uk/3133/>

Copyright and moral rights for this thesis are retained by the Author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

The investigation of hippocampal and hippocampal
subfield volumetry, morphology and metabolites using
3T MRI

John McLean MSci, MSc, DipIPEM(S), Clinical Scientist ¹

January 19, 2012

**Presented as a thesis for the degree of Doctor of
Philosophy (Ph.D)**

Department of Clinical Physics,

School of Medicine,

College of Medical, Veterinary and Life Sciences,

University of Glasgow, University Avenue,

Glasgow, G12 8QQ.

Acknowledgments

My thanks go to my supervisor Professor Barrie Condon who has guided me throughout the duration of this work.

Thanks also to Professor Donald Hadley who reported on all of the MRI scans discussed in this thesis and who was always available for discussions about my research. Thanks also to Dr Celestine Santosh and Dr Joe Bhattacharya for reporting the incomplete hippocampal inversion dataset discussed in chapter 6 and for general advice around the anatomy of the hippocampus.

I am also grateful to Professor Alex Elliott and Professor Dave Wyper who along with Prof Condon gave me the opportunity to work in an excellent department with the freedom to pursue interesting ideas.

I would also like to thank Dr Jim Patterson and Dr David Brennan for useful discussions and technical advice throughout my thesis. Similarly, thanks are due to Brain Patenaude for help with getting the best out of the FIRST algorithm and for providing code to recreate meshes from manually edited volumes. Thanks also to Bill Moorhead from Edinburgh for advice around VBM and in particular the use of the non-stationarity correction.

Thanks to Dr Jonathan Cavanagh of the Sackler Institute of Psychobiological research who facilitated access to the PSOBID and inflammation patients discussed in this thesis. Many thanks also to Dr Rajeev Krishnadas who along with Dr Cavanagh advised on discussions around statistics, inflammatory processes, neurogenesis and stress biology.

A great deal of thanks are also due to Miss Isobel McDonald and the other radiographers at the Institute of Neurological Sciences who co-ordinated and scanned many of the subjects for the PSOBID study.

I am also grateful to Dr Joe Bhattacharya, Professor Henri Duvernoy and Springer Science and Business Media for the use of various images used in this thesis.

Mark Meenan from the University of Glasgow helped me greatly by facilitating my use of the Glasgow University computing science computer cluster grid. This significantly increased the speed at which the relatively large imaging datasets in this thesis could be analysed.

Thanks also to the many volunteers who participated in the various studies presented in this thesis many of whom gave up their free time for no financial reward.

I would also like to thank my girlfriend Paula who has provided support, encouragement and relief from this work over the past few years. Thanks also to all my friends for their support.

Finally and most importantly I would like to thank my Mum and Dad and my brothers David and Marc for providing the platform from which I could achieve my goals. I hope that I have always worked in a manner that would make them proud. I will continue to work in this way and hopefully this thesis is only the beginning of what I can achieve.

Contents

1	The anatomy of the hippocampus	1
1.1	Introduction	1
1.2	The hippocampus within the context of the whole brain	2
1.3	The head, body and tail of the hippocampus	4
1.4	The internal structure of the hippocampus: an introduction to hippocampal subfields	6
1.4.1	The Cornu Ammonis and its Subfields	6
1.4.2	The Gyrus Dentatus	8
1.5	The anatomy surrounding the hippocampus	9
1.6	The function of the hippocampus	10
1.7	Neurogenesis, the Hippocampus and the Dentate gyrus in depression	11
1.7.1	Why the interest in the hippocampus, mood disorders and neurogenesis?	11
1.7.2	What is Neurogenesis?	11
1.7.3	What is the neurogenic hypothesis of MDD?	13
1.8	Aims for the thesis	14

2	Manual and automated hippocampal volume measurements in groups with different socioeconomic status	15
2.1	Abstract	15
2.2	Introduction	16
2.3	Aims	21
2.4	Literature review	21
2.4.1	Hippocampal volumes in normal subjects	21
2.4.2	Hippocampal volumes and reliability at different MRI field strengths	23
2.5	Methodology	26
2.5.1	Ethics	26
2.5.2	Volunteers, materials, methods and study design	26
2.5.2.1	Hippocampal volumes and the relationship to stress and socioeconomic status	28
2.5.2.2	Hippocampal volume and alcohol	30
2.5.3	Imaging hardware	30
2.5.4	Imaging Protocol	30
2.5.5	Manual outlining of the hippocampus	31
2.5.5.1	Manual segmentation strategy	32
2.5.5.2	Anterior border	32
2.5.5.3	Superior borders	33
2.5.5.4	Posterior border	35
2.5.5.5	Inferior border	36
2.5.5.6	Superior-medial border	36

2.5.5.7	Lateral border	37
2.5.5.8	The complete manually segmented hippocampus	37
2.5.6	Automated segmentation of the hippocampus	38
2.5.6.1	Background to automated subcortical brain segmentation	38
2.5.6.2	Choosing an automated segmentation method	38
2.5.6.3	How FIRST works	39
2.5.6.4	Training data	39
2.5.6.5	Linear image registration	39
2.6	Results	41
2.6.1	Introducing the intraclass correlation coefficient (ICC)	41
2.6.2	Introducing the Dice metric	43
2.6.3	Intra operator quality measures: ICC and Dice metric results	44
2.6.4	Inter operator quality measures: ICC and Dice metric results	45
2.6.5	Comparing manual and automated hippocampal volume measurements	46
2.7	Comparing the manual and automated volumes using the Dice metric	51
2.7.1	Displaying differences between the manual and automated segmentation processes	51
2.7.2	Establishing the mean and standard deviation of hippocampal volume measures in the poorer and more affluent groups	55
2.7.2.1	Power calculations for hippocampal volume data acquired using 3T MRI	57
2.7.3	Hippocampal volumes in relation to IL-6 and cortisol	59
2.8	Discussion	61
2.8.1	Assessing the quality of the manually segmented hippocampal volumes	61

2.8.2	Comparing automated and manual measurements made using 3T MRI	63
2.8.3	Hippocampal volumes in the more and least deprived SES groups: discussion of results and limitations of this study	65
2.8.4	Hippocampal volumes and inflammatory markers: discussion of results and limitations of this study	68
2.9	Conclusion	71
3	Voxel Based Morphometry (VBM) in groups with different socioeconomic status	73
3.1	Abstract	73
3.2	Introduction	74
3.3	Aims and hypotheses	75
3.3.1	Primary null VBM hypothesis	75
3.3.2	Secondary null VBM hypothesis	75
3.4	Study methodology	75
3.4.1	Volunteers, materials, methods and study design	76
3.4.2	Data acquisition	76
3.5	Data analysis	76
3.5.1	Data pre-processing	76
3.5.1.1	Manually defining the image origin	76
3.5.1.2	N3 inhomogeneity correction	77
3.5.2	Data processing: 'Optimised' VBM	81
3.5.2.1	Segmentation in native space and removal of non-brain structures	81
3.5.2.2	Normalisation of GM/WM images to template space	83

3.5.2.3	Normalisation of whole brain images to template space . . .	83
3.5.2.4	Segmentation of normalised whole brain image and removal of non-brain structures	84
3.5.2.5	Modulation	84
3.5.2.6	Smoothing	84
3.5.3	The DARTEL-VBM processing pipeline	84
3.5.4	Criticisms of VBM	89
3.5.5	Statistical design, analysis and the correction for multiple comparisons	89
3.5.5.1	Statistical analysis: The General Linear model (GLM) . . .	89
3.5.5.2	The GLM in imaging studies	90
3.5.5.3	Multiple comparison corrections	91
3.5.5.4	The Bonferroni correction and familywise error (FWE) . .	92
3.5.5.5	Random field theory, FWE and the false discovery rate (FDR)	92
3.5.5.6	Non-stationary (NS) statistics	93
3.6	Results	94
3.6.1	Assessing grey matter volume differences between deprived and more affluent groups	94
3.6.2	Regression analysis of inflammatory markers and grey matter volume of the affluent and poorer groups	97
3.6.3	Comparing the results of the 'optimised' VBM approach to DARTEL- VBM	97
3.6.4	Comparing the results of VBM with and without the NS cluster correction	100
3.6.4.1	The utility of NS cluster correction statistics on the DARTEL- VBM	100

3.6.5	Post-hoc analysis of other brain regions of statistically significant GM volume differences	102
3.6.6	Post-hoc analysis of whole brain tissue class volumes	104
3.7	Discussion	105
3.7.1	Discussion of DARTEL-VBM GM volume differences between poorer and more affluent groups	105
3.7.2	Comparing results of optimised VBM and DARTEL-VBM	106
3.7.3	Discussion of post-hoc analysis on DARTEL VBM results across the whole brain	107
3.7.4	Discussion of results of global tissue type volume differences	108
3.8	Conclusion	108
4	Magnetic Resonance Spectroscopy (MRS) measures of the hippocampus in groups of differing socioeconomic status	110
4.1	Abstract	110
4.2	Introduction	111
4.2.1	Overview of hippocampal metabolites and the relationship to SES and stress	112
4.3	Aims and hypotheses	113
4.3.1	Primary null hypotheses	113
4.3.2	Secondary null hypotheses	113
4.3.3	Additional research questions	113
4.4	Literature review and theory	113
4.4.1	Overview of MRS	113
4.4.2	Metabolites of Interest	116

4.4.2.1	NAA	116
4.4.2.2	NAAG	118
4.4.2.3	Creatine (Cr) and phosphocreatine (PCr)	118
4.4.2.4	Choline (Cho), Phosphocholine (PCho) and Glycerophosphorylcholine (GPCho)	119
4.5	Methodology	121
4.5.1	Volunteers, materials, methods and study design	121
4.5.2	Acquiring anatomical MR images on which to plan the MRS voxel	121
4.5.3	MRS acquisition protocol	121
4.5.3.1	Pulse Sequence: PRESS or STEAM	121
4.5.3.2	TE	122
4.5.3.3	TR	122
4.5.3.4	Voxel size and placement	122
4.5.3.5	Number of signal averages	123
4.5.4	Linear Combination (LC) model	123
4.6	Results	127
4.6.1	Comparing NAA/Cr and Cho/Cr between the affluent and poorer groups	128
4.6.2	Assessing correlations between hippocampal metabolites and inflammatory markers	128
4.6.2.1	NAA/Cr vs IL-6	129
4.6.2.2	Cho/Cr vs. IL-6	130
4.6.2.3	NAA/Cr vs. cortisol	130
4.6.2.4	Cho/Cr vs. cortisol	131

4.7	Post-hoc analysis of MRS data	131
4.7.1	Metabolites with age	131
4.7.2	Hippocampal metabolites and hippocampal volumes	132
4.7.2.1	NAA/Cr vs. hippocampal volume	132
4.7.2.2	Cho/Cr vs. hippocampal volume	133
4.8	Discussion	134
4.8.1	Comparing hippocampal metabolites between affluent and poorer subjects	134
4.8.2	Hippocampal metabolites and inflammatory markers	134
4.8.3	Hippocampal metabolite ratios vs. age and hippocampal volume	136
4.8.4	Future technological developments to improve MRS	138
4.9	Conclusion	139
5	Normal adult hippocampal volumes as measured using 3T MRI	140
5.1	Abstract	140
5.2	Introduction	141
5.3	Aims	142
5.4	Methodology	142
5.4.1	Volunteer recruitment and inclusion/ exclusion criteria	142
5.4.2	Imaging hardware	144
5.4.3	Imaging Protocol	144
5.4.3.1	Data acquisition	144
5.4.4	Segmenting the hippocampus	144

5.5	Results	145
5.5.1	Data normality	145
5.5.2	Hippocampal volumes and their relationship to ICV	145
5.5.3	Hippocampal volume with aging	148
5.5.4	Hippocampal volume measures in normal volunteers	149
5.5.5	Hippocampal volumes and sex	150
5.5.6	Hippocampal volume: laterality	151
5.6	Discussion and limitations	152
5.6.1	The relationship between hippocampal volume and ICV	152
5.6.2	The relationship between hippocampal volumes and age within the age range of 19 to 64	152
5.6.3	The normal range for hippocampal volumes in normal subjects between the ages of 19 and 64 using 3T MRI	154
5.6.4	The relationship between hippocampal volume and sex	156
5.6.5	Hippocampal volumes and laterality	157
5.6.6	Further limitations of the normative 3T hippocampal volume data	157
5.6.6.1	Educational and intellectual effects on the volume of the hippocampus	157
5.7	Conclusion	159
6	A quantitative assessment of hippocampal malrotation in normal subjects using 3T MRI	161
6.1	Abstract	161
6.2	Introduction	163
6.3	Literature review	163

6.3.1	What is IHI / hippocampal malrotation?	163
6.3.2	What is incomplete hippocampal inversion?	165
6.4	Aims and hypotheses	166
6.4.1	Hippocampal volumes and shape analysis: can a more quantitative analysis add information to the debate on incomplete hippocampal inversion?	166
6.4.2	Volume analysis of the hippocampus and amygdala	166
6.4.3	Morphological analysis of the hippocampus and amygdala	167
6.5	Methodology: defining hippocampal malrotation radiologically	168
6.5.1	Subjects	168
6.5.2	Image acquisition	169
6.5.3	Radiological assessment of malrotation	169
6.6	Quantitative image analysis methodology	170
6.6.1	Pre-processing of MRI data	170
6.6.2	Data processing: automated segmentation of the hippocampus with manual editing	172
6.6.2.1	Vertex analysis	173
6.7	Results of radiological assessment of hippocampal malrotation in normal volunteers	174
6.8	Results of quantitative assessment of hippocampal malrotation	175
6.8.1	Results of hippocampal volume analysis	175
6.8.1.1	Compare malrotated hippocampal volumes against normal hippocampal volumes	175
6.8.1.2	Comparing left malrotated hippocampal volume with right hippocampal volumes in the same subjects	176

6.8.2	Results of amygdala volume analysis	177
6.8.2.1	Comparing amygdala volumes, adjacent to malrotated hippocampi to amygdala volumes adjacent to normal hippocampi	177
6.8.2.2	Compare left-sided amygdala volumes adjacent to malrotated hippocampi against right-sided amygdala volumes adjacent to normal hippocampi	177
6.8.3	Results of hippocampal vertex analysis	178
6.8.3.1	Surface analysis of hippocampal volumes: malrotated vs. normal volumes	178
6.8.4	Results of amygdala vertex analysis	181
6.8.4.1	Surface analysis of amygdala volumes: amygdala volumes adjacent to malrotated hippocampi vs. amygdala volumes adjacent to normal hippocampi	181
6.9	Discussion of radiology outcomes	183
6.10	Discussion of outcomes of the hippocampal volume and surface analysis . . .	185
6.11	Discussion of outcomes of the amygdala volume and surface analysis	186
6.12	Conclusion from radiology outcomes	186
6.13	Conclusions from hippocampal and amygdala volume segmentation	187
7	CA4/dentate hippocampal subfield volumes before and after treatment with an anti-inflammatory drug	189
7.1	Abstract	189
7.2	Introduction	190
7.3	Literature review	191
7.3.1	Review of subfield volume methodology	191
7.3.2	Review of RA, depression and their relevance to the hippocampus . .	194

7.4	Study aim	196
7.4.1	Primary aim	196
7.4.2	Secondary aims	196
7.4.3	Tertiary aim	196
7.4.4	Ethics	196
7.4.5	Subjects	197
7.4.6	Exclusion criteria	197
7.4.7	Image acquisition protocol	197
7.4.8	Analysis protocol for the outlining of the dentate/CA4 hippocampal subfield	198
7.4.8.1	Manual outlining of subfields of the hippocampus	198
7.4.9	Analysis of data by DARTEL-VBM	200
7.4.10	Statistical design and analysis	200
7.5	Results	201
7.5.1	Data quality: Intra-rater variability	201
7.5.2	Data quality: power calculations on identifying CA4/dentate subfield volume differences	202
7.5.3	Establishing the mean and variance of dentate/CA4 volumes before and after a course of anti-inflammatory drug administration	204
7.5.4	Volumes with age	205
7.5.5	Results of DARTEL-VBM to assess hippocampal volume differences	206
7.6	Discussion	210
7.6.1	Intra-operator variability	210
7.6.2	Data quality: Power calculations on CA4/dentate subfield volumes	211

7.6.3	CA4/dentate volumes before and after a course of anti-inflammatory drug treatment	211
7.6.4	CA4/dentate volumes with age	212
7.6.5	Results of DARTEL-VBM comparing hippocampal GM volume before and after anti-inflammatory treatment	213
7.6.6	Comparison with other studies, limitations of the manual outlining procedure and further work	213
7.7	Conclusion	215
8	Hippocampal subfield volumes in normal subjects using 3T MRI	216
8.1	Abstract	216
8.2	Introduction	218
8.3	Literature review	218
8.4	Aim and hypotheses	218
8.5	Methodology	219
8.5.1	Ethics	219
8.5.2	Volunteers	219
8.5.3	Imaging hardware	219
8.5.4	Exclusion criteria	219
8.5.5	Data acquisition	220
8.5.6	Image analysis	220
8.5.6.1	Hippocampus: Manual segmentation of subfield	220
8.5.6.2	Protocol for manually outlining hippocampal subfields	221
8.5.7	Description of hippocampal subfields	221

8.5.7.1	Entorhinal cortex	221
8.5.7.2	The subiculum	221
8.5.7.3	Grey matter of CA1, CA2, and CA3	223
8.5.7.4	The dentate gyrus and CA4	223
8.5.7.5	White matter layer between CA1, CA2, CA3 and the dentate gyrus	224
8.5.7.6	Alveus	225
8.5.7.7	The fimbria	226
8.5.7.8	Order of delineation	226
8.5.7.9	Implementation of manual segmentation protocol	227
8.6	Results of normal subjects hippocampal subfield data	228
8.6.1	Test normality of subfield volumes	228
8.6.2	Normal ranges for raw subfield volumes	228
8.6.3	Assess correlation between raw subfield volumes and ICV	229
8.6.4	Assess correlation between raw subfield volumes and age	231
8.6.5	Left vs. right subfield volumes	232
8.6.6	Male vs. Female	232
8.6.7	Intra-rater variability and the Dice overlap metric	234
8.6.8	Inter-rater variability and Dice overlap metric	235
8.6.9	Power calculations for corrected subfield volumes	236
8.6.9.1	CA4/dentate power calculation	237
8.6.9.2	WM layer power calculation	238
8.6.9.3	CA1/CA2/CA3 power calculation	238

8.6.9.4	Subiculum power calculation	239
8.6.9.5	Fimbria power calculation	239
8.6.9.6	ERC power calculation	239
8.6.9.7	Alveus power calculation	240
8.7	Discussion	240
8.7.1	Mean Volumes and Normal ranges	240
8.7.2	Subfield volumes vs. ICV	245
8.7.3	Subfield volumes vs. age	246
8.7.3.1	Left vs. right subfield volumes	247
8.7.4	Male vs. Female	248
8.7.5	Intra-rater agreement and Dice metric	248
8.7.6	Inter-rater agreement and Dice metric	251
8.7.7	Subfield power and sample size calculations and their implications	254
8.7.8	Further work and future advancements	256
8.8	Conclusion	258
9	Final conclusions	260
9.1	Conclusions and original contributions of this thesis	260
9.2	Limitations of this thesis, methodological issues and future work	268
9.3	Summary of findings and contribution of this thesis	272

List of Figures

1.1	Illustration (1.1a) and dissection (1.1b) showing the hippocampus within the whole brain	2
1.2	3-plane MR image centered on the hippocampus showing the hippocampus in the axial(a), sagittal (b) and coronal planes (c)	3
1.3	Axial MR and illustration showing the major components of the hippocampus and the position of the hippocampus relative to the amygdala	4
1.4	Illustration of the internal structure of the hippocampus showing the CA and the DG	5
1.5	Subfields of the Hippocampus	7
1.6	Subfield cell types illustrated on histological slice	8
2.1	PSOBID study: 3D T1-weighted image displayed in 3-planes using ITKsnap (v1.8), Whole brain and hippocampus, T1-weighted 3D Whole Brain image	31
2.2	PSOBID study: 3D T1-weighted image displayed in 3-planes using ITKsnap (v1.8), Whole brain and hippocampus, 3D T1-weighted view of the hippocampus	32
2.3	Anterior border of the hippocampus	33
2.4	Anterior-superior border of the hippocampus	34
2.5	Medial-superior border of the hippocampus	34
2.6	Posterior-medial border of the hippocampus	35

2.7	Inferior border of the hippocampus	36
2.8	Manual segmentation of the right and left hippocampi as delineated using ITKsnap software	37
2.9	FSL's FIRST, example of hippocampal segmentation, hippocampal mask vol- ume overlaid on the T1-W 3D image and 3D rendering of the hippocampal meshes	40
2.10	Illustration of the Dice metric	44
2.11	Automated and manual volumetry measures of hippocampal volumes for af- fluent and poorer groups	47
2.12	Bland-Altman difference plot to compare agreement between FIRST output and manually segmented hippocampal volumes	48
2.13	Bland-Altman difference plot to compare agreement between FIRST output and manually segmented hippocampal volumes	49
2.14	Bland-Altman difference plot to compare agreement between FIRST output and manually segmented hippocampal volumes	50
2.15	Anterior boundary discrepancies (1)	53
2.16	Anterior boundary discrepancies (2)	54
2.17	Hippocampal head: medial discrepancies	55
2.18	Fornix discrepancies	56
2.19	Power calculation for automatically segmented hippocampal volume difference at 3T	59
2.20	IL6 vs. ICV corrected left and right hippocampal volumes for affluent and poorer groups	60
2.21	Cortisol vs. ICV corrected left and right hippocampal volumes for affluent and poorer groups	60
3.1	T1-weighted 3D FSPGR image pre (image a) and post (image b) N3 inhom- ogeneity correction	80

3.2	'Optimised' VBM image processing pipeline	82
3.3	Global GM and WM volume measures for affluent and poor populations . .	83
3.4	DARTEL-VBM image processing pipeline	88
3.5	SPM map and section of the result of contrasting the affluent and poorer groups and inspecting the left hippocampus	95
3.6	SPM map and section of the result of contrasting the affluent and poorer groups and inspecting the right hippocampus	96
3.7	SPM with statistics and co-ordinates from the results of contrasting the GM volume of the left hippocampus of the affluent and poorer groups	97
3.8	SPM with statistics and co-ordinates from the results of contrasting the GM volume of the right hippocampus of the affluent and poorer groups	98
3.9	Optimised VBM results for affluent vs. poorer groups, GM, t-test	99
3.10	DARTEL-VBM results for affluent vs. poorer groups, GM, t-test	100
3.11	NS corrected statistics for DARTEL-VBM	101
3.12	FDR corrected statistics for DARTEL-VBM	102
3.13	VBM GM result. Affluent vs. Poorer group. NS corrected SPM. Right cerebellum.	103
3.14	VBM GM result. Affluent vs. Poorer group. NS corrected statistics. Right cerebellum.	104
4.1	An example of an MRS spectrum from the hippocampus, PRESS, TR 1500ms, TE 144ms	117
4.2	Chemical Structure of NAA	117
4.3	Chemical Structure of NAAG	118
4.4	Chemical structure of Cr and PCr: methyl (CH ₃) group is circled in red, methylene (CH ₂) group circled in green. Note the Cr + PCr methyl group produces a singlet at 3.0ppm and the methylene group produces a singlet at 3.9ppm.	119

4.5	Chemical structure of choline, PCho, GPCho. Circled in red are the trimethylamine groups with nine protons which contribute to the 'total choline' singlet resonance at 3.20ppm.	120
4.6	PSOBID MRS planning images: coronal, sagittal and axial planes	123
4.7	LCmodel output from the PSOBID study	124
4.8	PSOBID study: Metabolite concentrations as determined by LCmodel in Institutional Units	128
4.9	PSOBID study: Metabolite concentrations as determined by the LCmodel presented relative to Creatine	129
4.10	Left and right hippocampal NAA/Cr vs. IL-6	129
4.11	Left and right hippocampal Cho/Cr vs. IL-6	130
4.12	Left and right hippocampal NAA/Cr vs. cortisol	131
4.13	Left and right hippocampal Cho/Cr vs. cortisol	131
4.14	Left and right hippocampal NAA/Cr vs. age (months)	132
4.15	Left and right hippocampal Cho/Cr vs. age	132
4.16	Left hippocampal NAA/Cr vs. left hippocampal volume and Right hippocampal NAA/Cr vs. right hippocampal volumes	133
4.17	Left hippocampal Cho/Cr vs. left hippocampal volume and Right hippocampal Cho/Cr vs. right hippocampal volumes	133
5.1	Histogram and summary statistics of uncorrected left and right hippocampal volumes	145
5.2	Histogram and summary statistics of ICV corrected left and right hippocampal volumes	146
5.3	Histogram and summary statistics of intracranial volume measurements	146
5.4	Scatterplot illustrating the relationship between ICV and hippocampal volume	147

5.5	Scatterplot of ICV (mm^3) data plotted against age (in months)	147
5.6	Scatterplot of GM, WM and CSF volumes (in Litres) plotted against age (in months)	148
5.7	Left and right hippocampal volumes with age	148
5.8	Normal range ICV corrected left and right hippocampal volumes with age .	150
5.9	Left and right, ICV corrected hippocampal volumes for females and males: Histogram and summary statistics	151
5.10	Left hippocampal volume vs. right hippocampal volume	152
6.1	T1-weighted 3D acquisition with the amygdala and hippocampal head and tail shown	168
6.2	Right and left hippocampi and visualised using 3T MRI and a coronal T2-weighted FSE acquisition	170
6.3	The hippocampus as visualised using 3T MRI and a sagittal T2-weighted FSE acquisition	171
6.4	ICV corrected, left malrotated volumes vs. normal, left hippocampal volumes	176
6.5	Comparing left and right hippocampal volumes in subjects with a malrotated left hippocampus	177
6.6	ICV corrected amygdala volumes, comparing left amygdala volumes adjacent to malrotated hippocampi against normal left sided amygdala volumes . . .	178
6.7	Comparing left and right amygdala volumes in subjects with a malrotated left hippocampus	179
6.8	Results of hippocampal surface analysis comparing malrotated left and normal left hippocampal volumes	180
6.9	FDR corrected results of hippocampal surface analysis comparing malrotated left and normal left hippocampal volumes	181
6.10	Results of amygdala and hippocampal surface analysis comparing volumes from malrotated and normal hippocampal groups	182

6.11	MCC corrected results of amygdala and hippocampal surface analysis comparing malrotated and normal left-sided amygdala and hippocampal volumes	183
7.1	Planning coronal slices along the long axis of the hippocampus	198
7.2	Two, independent manual delineations of the dentate/CA4 region of the hippocampal body	199
7.3	Power calculation estimation for Paired-t-test CA4/dentate subfield volume measurements	202
7.4	Power calculation estimation for 2 sample t-test design of CA4/dentate subfield volume measurements	203
7.5	Results of CA4/dentate volumes (mm^3) before and after anti-inflammatory treatment	204
7.6	Left and right, pre-treatment, ICV corrected CA4/dentate volumes vs. Age (months)	206
7.7	SPM paired t-test design to compare pre and post anti-inflammatory treatment brain by VBM	207
7.8	SPM F-test results comparing pre and post treatment groups	208
7.9	SPM Paired t-test results, post treatment vs. pre treatment	209
7.10	NPM paired t-test (uncorrected) results, post treatment vs. pre treatment	209
7.11	NPM FDR corrected paired t-test results, post treatment vs. per treatment	210
8.1	Delineation of hippocampal subfields at the level of the hippocampal head.	222
8.2	Delineation of hippocampal subfields at the anterior aspect of the hippocampal body.	223
8.3	Delineation of hippocampal subfields at the central portion of the hippocampal body.	224
8.4	Delineation of hippocampal subfields at the hippocampal tail.	225

8.5	Hippocampal substructures: markers delineating subfield boundaries	227
8.6	Normal hippocampal subfield volumes and 95% CI's	229
8.7	Scatterplot of left and right hippocampal subfield volumes vs. ICV	230
8.8	Scatterplot of left and right hippocampal subfield volumes vs. Age	231
8.9	Intraoperator Dice for hippocampal subfield volumes	234
8.10	Inter-rater Dice for left hippocampal subfield volumes	235
8.11	Power calculation for CA4/dentate hippocampal subfield	237
8.12	Power calculation for the CA4/dentate subfield	238
8.13	Power calculation for WM layer hippocampal subfield	239
8.14	Power calculation for CA1, CA2,CA3 hippocampal subfield	240
8.15	Power calculation for subiculum hippocampal subfield	241
8.16	Power calculation for fimbria hippocampal subfield	242
8.17	Power calculation for entorhinal cortex subfield	243
8.18	Power calculation for alveus hippocampal subfield	244
9.1	Permission to use images from The Human Hippocampus	314
9.2	Permission to use images The Human Hippocampus	314
9.3	Poster for recruitment of normal volunteers to the study of the hippocampus 3T MRI	318

List of Tables

2.1	Summary of hippocampal volumes from normal controls in studies assessing hippocampal volumes in MDD patients	23
2.2	Whole hippocampal volume measures of quality: Intra-rater agreement and the Dice metric	45
2.3	Whole hippocampal volume measures of quality: Inter-rater agreement and the Dice metric	45
2.4	Automated and manual, left and right hippocampal volumes for affluent and poorer groups	48
2.5	PSOBID study: Dice similarity index for manual vs. automated hippocampal volume measurements	51
2.6	Results of manually segmented hippocampal volumes for poorer and more affluent groups	56
2.7	PSOBID data: ICV corrected, manually segmented hippocampal volumes presented as a % of ICV	57
7.1	Summary Statistics for repeat measures of right and left CA4/ dentate . . .	202
7.2	Left and right mean CA4/dentate volumes pre and post a course of anti-inflammatory drug. 95% Confidence Intervals (CI) are also shown.	205
8.1	Mean volume, 95% CI and lower and upper CI limits	228
8.2	Comparison of female and male hippocampal subfield volumes	233

8.3	Comparison of female and male hippocampal subfield volumes corrected for ICV	233
8.4	Subfield ICC (agreement) and Dice metric	234
8.5	Inter-rater Dice metric and ICC agreement and consistency measures	236
8.6	Power calculation estimations for hippocampal subfields based on 17 subjects with the anticipation of a 10% volume difference between the groups	255
8.7	Sample size estimation for hippocampal subfields based on the anticipation of a 10% volume difference where a study power of 0.9 is desired	256

Abbreviations

AD: Anderson-Darling Normality test

ANALYZE: a common imaging data format (typically with a .nii or .img/.hdr file extension)

BET: Brain extraction tool

BrdU: bromo-deoxyuridine

CA: Cornu ammonis

CBV: cerebral blood volume

Cho: Choline containing compounds

CTM: Cortical Thickness Measurement

Cr: Creatine containing compounds

DARTEL: Diffeomorphic Anatomical Registration using Exponential Lie Algebra: A fast diffeomorphic image registration algorithm

DBM: Deformation Based Morphometry

DICOM: Digital Imaging and Communications in Medicine (A standardised image data format for medical applications)

DOF: Degrees of Freedom

DWI: Diffusion Weighted Imaging

DTI: Diffusion Tensor Imaging

ERC: Entorhinal cortex

ETL: Echo train length

GE: General Electric

Gln: Glutamine

Glu: Glutamate

GRE: Gradient echo

HIMAL: Hippocampal Malrotation

IHI: Incomplete Hippocampal Inversion

FDR: False discovery rate

FIRST: FMRIB's Integrated Registration and Segmentation Tool

FLAIR: Fluid Attenuation Inversion Recovery sequence

FLIRT: FMRIB's Linear Image Registration Tool

FSE: Fast spin echo

FSL: FMRIB's Software Library

FS: Freesurfer: brain imaging software

FSPGR: Fast Spoiled Gradient Recalled sequence

FMRIB: Functional MRI of the Brain Centre, Oxford

ICC: Intraclass correlation coefficient

ICV: Intracranial volume

IU: Institutional Units (local, arbitrary unit for MR spectroscopy data)

MAP: Maximum a posteriori

mI: myo-inositol

MDD: Major Depressive Disorder

MRI: Magnetic Resonance Imaging

MRS: Magnetic Resonance Spectroscopy

N3: An MR image inhomogeneity correction algorithm

NAA: N-acetylaspartate

NAAG: N-acetylaspartatylglutamate

NEX: Number of excitations (GE terminology)

NIFTI: Neuroimaging Informatics Technology Initiative. A common imaging data format (typically with a .nii or .img/.hdr file extension)

NMR: Nuclear Magnetic Resonance

NPM: Non-parametric mapping

NS: Non-stationary

NSC: Neural Stem Cells

PSOBID: PSOBID study: Psychological, social and biological determinants of ill health

PRESS: Point Resolved Spectroscopy pulse sequence

PROBE: GE's Proton brain examination tool for MRS

RA: Rheumatoid Arthritis

RFT: Random Field Theory

RPV: resels per voxel (resels = resolution element)

SAGE: Spectroscopy Analysis by General Electric

SE: Spin Echo

SES: Socioeconomic status

SNR: Signal to noise ratio

SNPM: Statistical non-parametric mapping

SPM: Statistical Parametric Mapping

SSRI: Selective serotonin reuptake inhibitor

STEAM: Stimulated Echo Spectroscopy pulse sequence

SUSAN: Smallest Univalued Segment Assimilating Nucleus (a non-linear noise reduction algorithm for images)

SVC: Small volume correction

T1: The longitudinal relaxation time

T2: The transverse relaxation time

TE: Echo time

TLE: Temporal Lobe Epilepsy

TR: Repetition time

TBM: Tensor Based Morphometry

VBM: Voxel Based Morphometry

Abstract

A detailed account of the hippocampal anatomy has been provided. This thesis will explore and exploit the use of 3T MRI and the latest developments in image processing techniques to measure hippocampal and hippocampal subfield volumes, hippocampal metabolites and morphology.

In chapter two a protocol for segmenting the hippocampus was created. The protocol was assessed in two groups of subjects with differing socioeconomic status (SES). This was a novel, community based sample in which hippocampal volumes have yet to be assessed in the literature.

Manual and automated hippocampal segmentation measurements were compared on the two distinct SES groups. The mean volumes and also the variance in these measurements were comparable between two methods. The Dice overlapping metric comparing the two methods was 0.81.

In chapter three voxel based morphometry (VBM) was used to compare local volume differences in grey matter volume between the two SES groups. Two approaches to VBM were compared. DARTEL-VBM results were found to be superior to the earlier 'optimised' VBM method. Following a small volume correction, DARTEL-VBM results were suggestive of focal GM volumes reductions in both the right and left hippocampi of the lower SES group.

In chapter four an MR spectroscopy protocol was implemented to assess hippocampal metabolites in the two differing SES groups. Interpretable spectra were obtained in 73% of the 42 subjects. The poorer socioeconomic group were considered to have been exposed to chronic stress and therefore via inflammatory processes it was anticipated that the NAA/Cr metabolite ratio would be reduced in this group when compared to the more affluent group. Both NAA/Cr and Cho/Cr hippocampal metabolite ratios were not significantly different between the two groups.

The aim of chapter 5 was to implement the protocol and methodology developed in chapter 2 to determine a normal range for hippocampal volumes at 3T MRI.

3D T1-weighted IR-FSPGR images were acquired in 39 healthy, normal volunteers in the age range from 19 to 64. Following the automated procedure hippocampal volumes were manually inspected and edited.

The mean and standard deviation of the left and right hippocampal volumes were determined to be: $3421\text{mm}^3 \pm 399\text{mm}^3$ and $3487\text{mm}^3 \pm 431\text{mm}^3$ respectively. After correcting for total ICV the volumes were: $0.22\% \pm 0.03\%$ and $0.23\% \pm 0.03\%$ for the left and right hippocampi respectively.

Thus, a normative database of hippocampal volumes was established. The normative data here will in future act as a baseline on which other methods of determining hippocampal volumes may be compared. The utility of using the normative dataset to compare other groups of subjects will be limited as a result of the lack of a comprehensive assessment of IQ or education level of the normal volunteers which may affect the volume of the hippocampus.

In chapter six Incomplete hippocampal inversion (IHI) was assessed. Few studies have assessed the normal incidence of IHI and of those studies the analysis of IHI extended only to a radiological assessment. Here we present a comprehensive and quantitative assessment of IHI. IHI was found on 31 of the 84 normal subjects assessed (37%). ICV corrected IHI left-sided hippocampal volumes were compared against ICV corrected normal left-sided hippocampal volumes (25 vs. 52 hippocampi). The IHI hippocampal volumes were determined to be smaller than the normal hippocampal volumes ($p \ll 0.05$). However, on further inspection it was observed that the ICV of the IHI was significantly smaller than the ICV of the normal group, confounding the previous result.

In chapter seven a pilot study was performed on patients with Rheumatoid Arthritis (RA). The aim was to exploit the improved image quality offered by the 3T MRI to create a protocol for assessing the CA4/ dentate volume and to compare the volume of this subfield of the hippocampus before and after treatment. Two methodologies were implemented.

In the first method a protocol was produced to manually segment the CA4/dentate region of the hippocampus from coronal T2-weighted FSE images. Given that few studies have assessed hippocampal subfields, an assessment of study power and sample size was conducted to inform future work.

In the second method, the data the DARTEL-VBM image processing pipeline was applied. Statistical nonparametric mapping was applied in the final statistical interpretation of the VBM data. Following an FDR correction, a single GM voxel in the hippocampus was deemed

to be statistically significant, this was suggestive of small GM volume increase following anti-inflammatory treatment.

Finally, in chapter eight, the manual segmentation protocol for the CA4/dentate hippocampal subfield developed in chapter seven was extended to include a complete set of hippocampal subfields. This is one of the first attempts to segment the entire hippocampus into its subfields using 3T MRI and as such, it was important to assess the quality of the measurement procedure. Furthermore, given the subfield volumes and the variability in these measurements, power and sample size calculations were also estimated to inform further work.

Seventeen healthy volunteers were scanned using 3T MRI. A detailed manual segmentation protocol was created to guide two independent operators to measure the hippocampal subfield volumes. Repeat measures were made by a single operator for intra-operator variability and inter-operator variability was also assessed. The results of the intra-operator comparison proved reasonably successful where values compared well but were typically slightly poorer than similar attempts in the literature. This was likely to be the result of the additional complication of trying to segment subfields in the head and tail of the hippocampus where previous studies have focused only on the body of the hippocampus. Inter-rater agreement measures for subfield volumes were generally poorer than would be acceptable if full exchangeability of the data between the raters was necessary. This would indicate that further refinements to the manual segmentation protocol are necessary. Future work should seek to improve the methodology to reduce the variability and improve the reproducibility in these measures.

Chapter 1

The anatomy of the hippocampus

1.1 Introduction

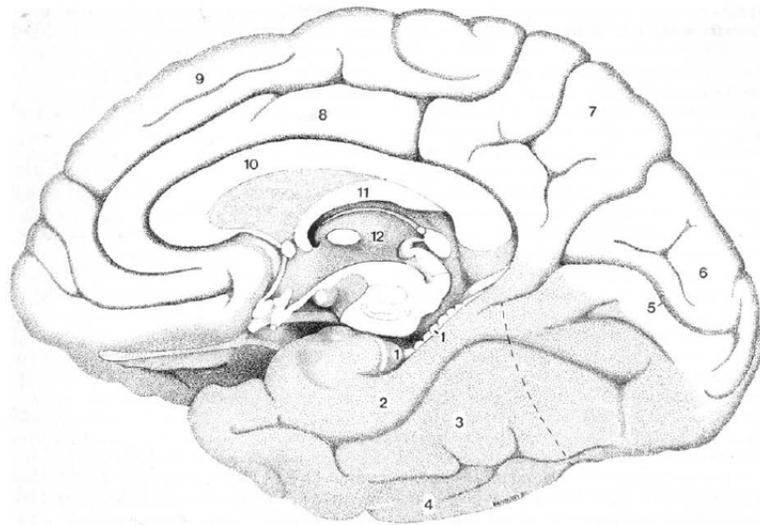
Throughout this thesis various magnetic resonance (MR) based measurement techniques will be applied to the brain and in particular to the hippocampus. Given the importance of the hippocampus to this work it is appropriate to begin with a comprehensive description of the anatomy of the hippocampus, its internal structure, and the role this structure plays in brain and human functioning. A description of the neighbouring anatomy of the hippocampus, while not exhaustive, will also be introduced in this chapter as it is important to appreciate that the hippocampus is neither anatomically nor functionally isolated. Throughout this thesis, in aiming to define robust measurement protocols of the hippocampus, a clear description of the neighbouring anatomy proved equally as important as the description of the hippocampus itself.

The human hippocampus is a structure of great complexity. The word hippocampus comes from the Latin for seahorse which conjures up images of a curling, intricate object which is indeed the reality of this structure. A major authority on the anatomy of the hippocampus is the book: *The Human Hippocampus* by Henri M Duvernoy (1998) and this was used as the primary anatomical reference text throughout this chapter [114]. This work has been cited many times in studies where outlining of the hippocampus from MRI data has been performed, for example, in work assessing human hippocampal volumes in depression [237, 360, 204].

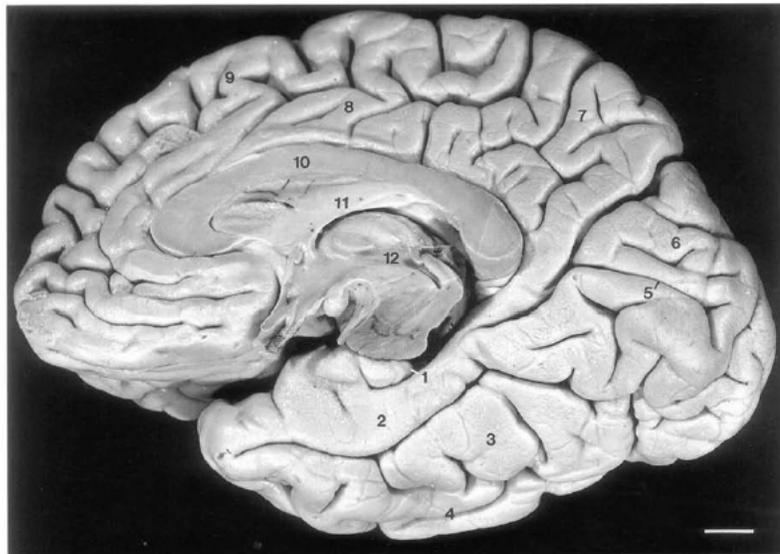
Finally, the concept of neurogenesis will be introduced along with a discussion of the importance of the role of the hippocampus and dentate gyrus in this process.

1.2 The hippocampus within the context of the whole brain

The hippocampus lies within the medial temporal lobe of the brain as illustrated in figure 1.1 below. Each person has two hippocampi, with each one located within the temporal lobe, being roughly symmetrical across the midline of the brain when viewed in the axial or coronal planes.



1.1a



1.1b

Figure 1.1: Illustration (1.1a) and dissection (1.1b) showing the hippocampus within the whole brain

Figure 1.1 illustrates the location of the hippocampus within the brain. The hippocampus is only partially visible on this midline sagittal slice and is denoted by the number 1. The other numbered regions are 2, the parahippocampal gyrus; 3, fusiform gyrus; 4, the inferior temporal lobe; 5, the calcarine sulcus ; 6, the occipital lobe, 7; the parietal lobe; 8, cingulate gyrus; 9, the frontal lobe; 10, the corpus collusum; 11, the fornix; 12, the 3rd ventricle. These images were reproduced from: *The Human Hippocampus* by Henri M Duvernoy (1998) with kind permission from Springer Science and Business Media and Henri M Duvernoy [114].

When viewing the hippocampus in the sagittal plane as shown in figures 1.1a and 1.1b, the hippocampus can be seen to lie at a relatively steep angle, running parallel to the parahippocampal gyrus. The long axis and body of the hippocampus runs anteriorly to posteriorly, with the head of the hippocampus sitting at the anterior aspect of the structure within the temporal lobe. The tail of the hippocampus curls medially from its superior position, and when viewed from the axial plane, this would represent the tail of the seahorse.

The hippocampus can be clearly seen using T1-weighted MR image and figure 1.2 below shows the hippocampus in 3-planes located within the whole brain.

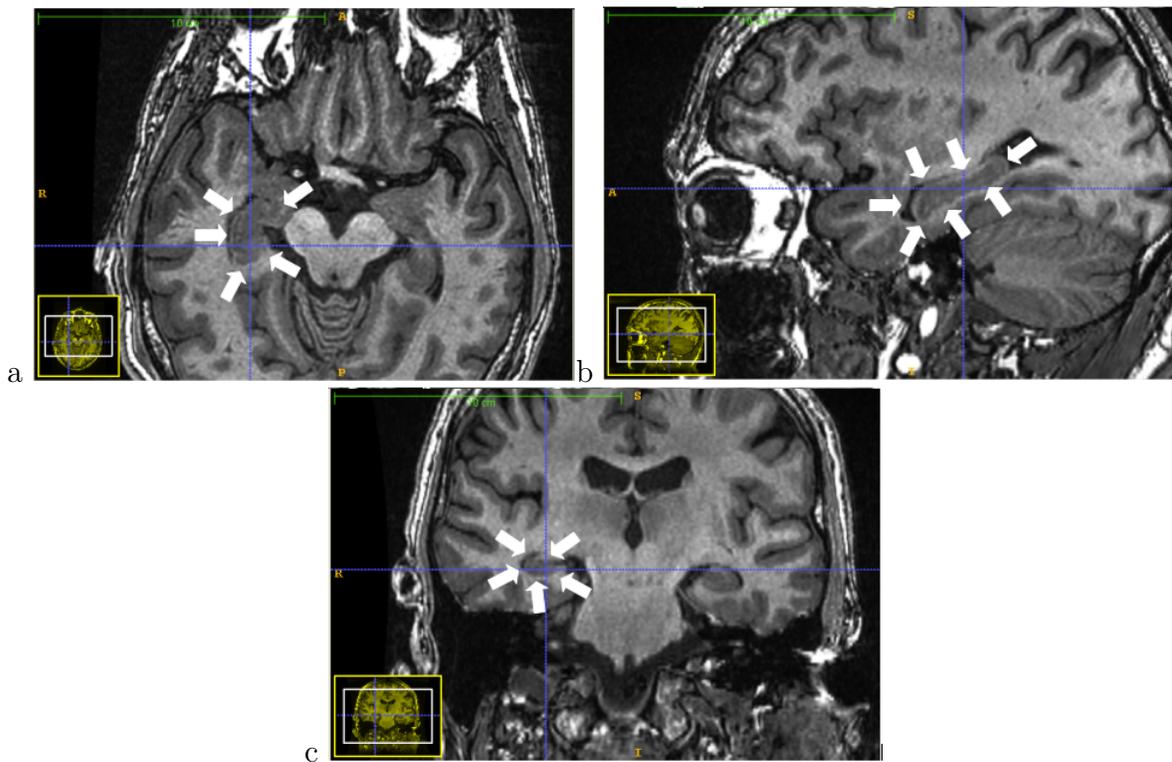


Figure 1.2: 3-plane MR image centered on the hippocampus showing the hippocampus in the axial(a), sagittal (b) and coronal planes (c)

Figure 1.2 shows the right hippocampus in 3-planes centred on the body of the hippocampus. The long axis of the sausage-like hippocampal structure is clearly visible on the sagittal image. On this image the grey matter of the hippocampus (on this slice) is almost entirely surrounded by the white matter of the alveus at the anterior and superior boundaries of the hippocampus, as well as the white and grey matter of the parahippocampal gyrus below the hippocampus. The coronal image shows clearly the position of both the right and left hippocampi within the temporal lobe.

1.3 The head, body and tail of the hippocampus

The hippocampus can be further subdivided into its head, body and tail. The bulb-like head of the hippocampus is distinct in that it protrudes medially into the vascular space towards the brain stem. The hippocampal head also has a characteristic undulation on its surface as a result of the developmental folding of the structure. Throughout the head, body and tail, the hippocampus comprises two cortical layers rolled up inside one another. These are the cornu ammonis (CA) and the gyrus dentatus (dentate gyrus (DG)). These layers are illustrated in figure 1.3 below.

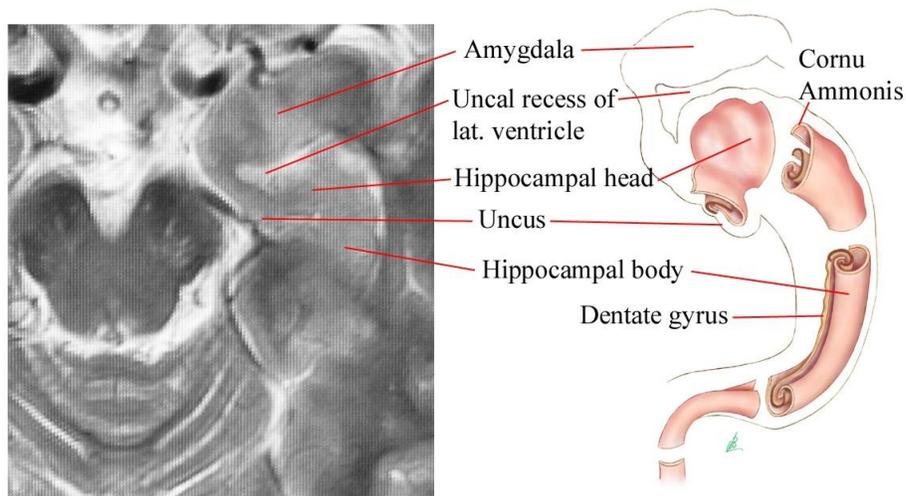


Figure 1.3: Axial MR and illustration showing the major components of the hippocampus and the position of the hippocampus relative to the amygdala

From figure 1.3 the body of the hippocampus can be seen to constitute the majority of the hippocampal volume, defining its long axis. This image is replicated here from Thammaroj et al, 2005, with permission [344]. Figure 1.4 below illustrates in greater detail the internal structure of the hippocampal body and surrounding anatomy. This image is also replicated here from Thammaroj et al, 2005, with permission [344].

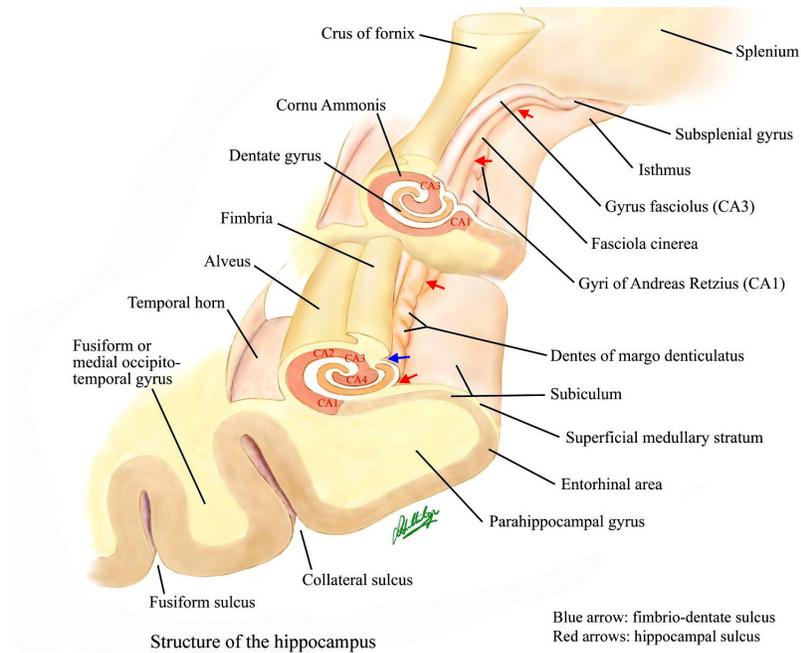


Figure 1.4: Illustration of the internal structure of the hippocampus showing the CA and the DG

Figure 1.4 illustrates, based on the work by Duvernoy [114], the folds of both the DG and CA. These layers can be seen to run from the head of the hippocampus throughout its body to its tail. Other pertinent features include the dentes of the margo denticulatus. These digitations are part of what makes the hippocampus a particularly intricate and complex structure and are what makes it a challenge to accurately delineate from MR images. From figure 1.4, the main body of the hippocampus, consisting of grey matter cellular regions within the white matter of the alveus and fimbria can be seen. Moving posteriorly along the long axis of the hippocampus, the white matter bundle of the fimbria can be observed on the superior-medial aspect of the hippocampus before receding towards the major white matter fibre bundle known as the fornix. The fornix connects the right and left hippocampi, where the fibre bundles merge at the midline of the brain.

Having described the position of the hippocampus relative to the rest of the brain, and having

introduced its major components, a more detailed description of the internal structure of the hippocampus at a cellular level will now follow.

1.4 The internal structure of the hippocampus: an introduction to hippocampal subfields

The two major cellular subfields within the hippocampus are the cornu ammonis and the gyrus dentatus. These subfields consist of various cellular regions with separate cellular subtypes. An overview of these cellular subfields will now follow.

1.4.1 The Cornu Ammonis and its Subfields

The CA can be further sub-divided into four cellular regions or subfields (CA1-CA4). The regions CA1-CA4 are defined by cell type.

CA1 cellular subfield continues from the subiculum. It consists of mostly triangular cells known as pyramidal somata which are generally small and scattered [102]. The stratum pyramidalae layer of the human CA1 is large in contrast to its narrow and dense appearance observed in rats [337]. The two sublayers of the stratum pyramidalae have been identified in the human CA1: the stratum profundum and the stratum superficiale [52]. The stratum profundum lies in contact with the stratum oriens at the outer layer of CA1 and has few pyramidal neurons while the stratum superficiale contains large pyramidal neurons.

CA2 is composed of large, ovoid, densely packed somata making the stratum pyramidale dense and narrow in contrast to CA1 [53].

CA3 refers to the inwardly curving aspect of the cornu ammonis as it enters the gyrus dentatus. Like CA2, CA3 consists of pyramidal somata but their population is less dense. A feature of CA3 is the presence of fine, non-myelinated fibers that arise from the gyrus dentatus. These fibers surround the pyramidal somata and are compressed between the strata radiatum and pyramidale forming an additional layer known as the stratum lucidum that is characteristic of the CA3 subfield.

CA4 is situated within the cavity of the gyrus dentatus which distinguishes it from CA3. Somata in this region are large, ovoid, sparse, and scattered amongst large mossy interweaving myelinated fibers characteristic of this subfield.

An illustration of the the cellular subfields are shown below in figure 1.5.

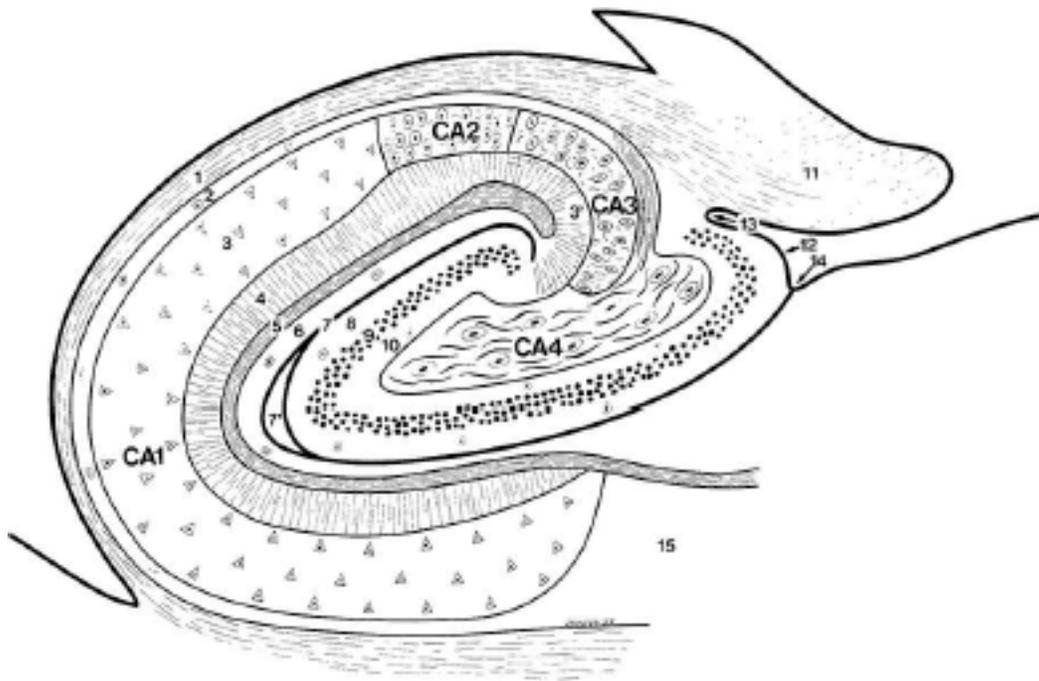


Figure 1.5: Subfields of the Hippocampus

Figure 1.5 illustrates the subfields of the hippocampus as viewed in the coronal plane. Subfields of the cornu ammonis (CA1-CA4): 1, alveus; 2, stratum oriens; 3, stratum pyramidale; 3', stratum lucidum; 4, stratum radiatum; 5 stratum lacunosum; 6, stratum moleculare; 7, vestigial hippocampal sulcus and residual cavity (7'). Subfields of the gyrus dentatus: 8, stratum moleculare; 9, stratum granulosum; 10, polymorphic layer; 11, fimbria; 12, margo denticulatus; 13, fimbriodentate sulcus; 14, superficial hippocampal sulcus; 15, subiculum. This image was reproduced from: *The Human Hippocampus* by Henri M Duvernoy (1998) with kind permission from Springer Science and Business Media and Henri M Duvernoy [114].

Thus we have seen that the cellular regions differ in a number of aspects, namely: cell type, size, shape and packing density. As a result of the varying degrees of cellular density it is possible to view some of the cellular subfields on histological slices following the injection of India ink. This allows variation in vascular density throughout the hippocampal subfields to be highlighted.



Figure 1.6: Subfield cell types illustrated on histological slice

Figure 1.6 above shows a coronal section of the hippocampal body after intravascular injection of India ink. The layers of the hippocampus may be differentiated due to differences in their vascular density. The stratum moleculare of the cornu ammonis (8) and that of the gyrus dentatus (9) are separated by the vestigial hippocampal sulcus (10). The subiculum (11) can be seen to have a relatively high vascular density relative to the adjacent stratum radiatum of CA1 (7). The subfield layers of the cornu ammonis are: 5, the alveus; 6, stratum pyramidale, 7; strata radiatum and lacunosum; 8, stratum moleculare. The subfield layers of the gyrus dentatus are: 9, stratum moleculare; 10, vestigial hippocampal sulcus; 11, subiculum; 12, margo denticulatus; 13, superficial hippocampal sulcus; 14, fimbriodentate sulcus; 15, fimbria; 16 choroid plexuses; 17, tail of the caudate nucleus; 18 inferior temporal horn of the lateral ventricle. This image was reproduced from: *The Human Hippocampus* by Henri M Duvernoy (1998) with kind permission from Springer Science and Business Media and Henri M Duvernoy [114].

1.4.2 The Gyrus Dentatus

The gyrus dentatus is most easily visualised on coronal sections of the hippocampal body. The DG is a narrow, curved tissue that envelops the CA4 region of the CA. The gyrus

dentatus is separated from the CA1-CA3 regions by the hippocampal sulcus that disappears early in the developmental process and becomes redundant. Thus the cornu ammonis and the gyrus dentatus are fused together, separated only by the vestigial hippocampal sulcus. At the far end of this sulcus the layers of the cornu ammonis and the gyrus dentatus become so close together it becomes impossible to distinguish between them.

A narrow segment of the gyrus dentatus, the margo denticulatus is visible on the temporal lobe surface. In the human brain the margo denticulatus has a distinctive toothed appearance. It overlaps the subiculum separated by it from the superficial hippocampal sulcus and from the fimbria by the fimbriodentate sulcus.

Anatomically the gyrus dentatus is simpler than the cornu ammonis. The three layers of the allocortex: the strata moleculare, stratum granulosum and the polymorphic layer are clearly visible on histological slices under the microscope. The stratum granulosum consists of somata of granular neurons. These are small and densely packed making this layer easily distinguishable on histology. Their axons are mossy and traverse the polymorphic layer to CA3 and CA4. A single dendrite escapes from the basal pole of each granular soma and projects into the stratum moleculare. The stratum moleculare is thick and separated from the stratum moleculare of the CA by the vestigial hippocampal sulcus. Its outer two-thirds receive fibers from the perforant pathway (i.e. the connective pathway from the entorhinal cortex into the hippocampus) while the inner third, in contact with the stratum granulosum, are occupied by commissural and septal fibers [76]. The polymorphic (or plexiform) layer merges the granular layer to CA4 and is crossed by axons of granular neurons. There are few interneurons in the molecular and polymorphic layers.

The internal cellular structure of the hippocampus and its associated terminology is clearly sophisticated. Considerations beyond the anatomy of the hippocampus of both the biological and electro-chemical dynamism only serve to compound its complexity. Measurements of the hippocampal subfields made later in this thesis will focus on volume measurements of the anatomy and in doing so will refer to some of anatomical landmarks and boundaries discussed above (Chapters 7 and 8).

1.5 The anatomy surrounding the hippocampus

The nearest neighbour to the hippocampus is the amygdala. The amygdala lies anteriorly and superiorly to the hippocampus. The uncus recess of the temporal horn of the lateral ventricle, illustrated in figure 1.3, denotes the border separating the hippocampal head from

the amygdala [344]. Early MR studies assessing the temporal lobe sought to investigate the amygdala and the hippocampus, in these earlier works these structures were not differentiable and so the amygdala-hippocampal complex was assessed as a single entity. Further developments in MR technology has meant that the hippocampus and amygdala are now more easily separated and with the advent of 3T MRI and higher field strength magnets it is now conceivable to assess in greater detail the internal structure of the hippocampus which in large part was the motivation behind this thesis.

Anteriorly and inferiorly to the hippocampal head lies the parahippocampal gyrus. Medially to the hippocampal head, the hippocampus borders the temporal horn of the lateral ventricle in close proximity to the brainstem. Along the hippocampal body it continues to be surrounded to its lateral aspects by the lateral ventricles with white matter lying superiorly and the parahippocampal gyrus running parallel to the hippocampus along the base of its body. As the tail of the hippocampus curves medially it begins to become thinner and is bordered laterally by the posterior lateral ventricle, posterior-medially by cerebral white matter, before joining the white matter bundle known as the fornix. The tail of the hippocampus lies inferiorly and adjacent to the posterior aspects of the thalamus with these two structures being separated by the lateral ventricle.

As mentioned in the introduction to this chapter a clear description of the anatomy surrounding the hippocampus is as important as a description of the anatomy of the hippocampus and its subfields. The brief description of the surrounding anatomy above serves as an introduction. It is almost impossible, and it would certainly be considered suboptimal, to try and accurately describe anatomy without the use of images. Therefore, a complete description of the anatomy surrounding the hippocampus is not given in this introductory chapter but a more detailed description with supporting images will be presented in the appropriate chapters elsewhere in this thesis where measurement protocols are being defined (Chapters 2, 4, 7 and 8).

1.6 The function of the hippocampus

The function of the hippocampus is too broad a topic to cover within a single chapter of a thesis. Where relationships to functional aspects of the hippocampus are assessed in this thesis, the particular function of interest will be introduced and discussed in the appropriate chapter. However, a brief statement highlighting the key functional processes of the hippocampus shall be made here.

While the layered and interweaving anatomy of the hippocampus introduced earlier in this chapter is complex, the way in which these cellular regions interact to enable function adds yet more complexity. Thus, striving to derive any detailed understanding of how the hippocampus functions poses an appreciable challenge. The Hippocampus itself plays a crucial role in a number of key functional processes in the brain such as short and long term memory [42], emotional processing [67], learning [207] and motor control [7]. Despite its importance to all of the aforementioned functional processes it is the role of the hippocampus in mood disorders that is of particular interest in this study [37] and this will be elaborated on in the next section.

1.7 Neurogenesis, the Hippocampus and the Dentate gyrus in depression

1.7.1 Why the interest in the hippocampus, mood disorders and neurogenesis?

While there has long been interest in assessing the hippocampus in mood disorders this has recently been reignited as a result of the demonstration of neurogenesis occurring within the hippocampus of the adult human brain [116]. One of the primary reasons for the continued investigation into neurogenesis is that it may provide a better understanding of diseases such as Major Depressive Disorder (MDD). MDD is a debilitating and complex psychiatric disease. It is a widespread problem with severe depression affecting approximately 2-5% of the population (USA data), with mood disorders impacting on 7% of the world's population [261]. The disorder involves a number of neural circuits with the origin of the disease believed to occur as a result of both genetic and non-genetic risk factors. In trying to unravel the various neurobiological processes involved in MDD, neurogenesis of the hippocampus has emerged as a candidate substrate that may go some way to explaining both the aetiology and treatment of MDD. In the following sections the evidence pertaining to neurogenesis from both animal and human studies will be summarised with a particular focus on how these studies relate to depression.

1.7.2 What is Neurogenesis?

Neurogenesis refers to the formation of new cells and is the process by which neurons are

created. Neurogenesis is most prevalent during pre-natal development with neurogenesis being responsible for populating the developing brain. Neurogenesis is important as it is the process by which self-renewing, pluripotent, neural stem cells (NSC) become neurons. Neurons are the functional components of the nervous system and are responsible for information processing and transmission.

While embryonic neurogenesis is obviously critical to the initial development of the brain, the majority of the interest in the domain of mood disorders and in particular anti-depressant treatment is with adult neurogenesis. It was the seminal work by Altman and Das in 1965 which demonstrated that the adult hippocampus is host to the birth and integration of newborn dentate granule cells in the dentate gyrus [8]. In the rat an estimated 9000 new cells are produced each day in the dentate gyrus. Approximately 50% of these go on to express neuron specific markers. At these proliferation rates the number of new granule neurons born each month is equal to 6% of the mature granule cell population [71].

There are few studies that have quantified the degree of neurogenesis in the adult human brain. The first study to achieve this involved cancer patients injected with a mitotic marker, the thymidine analogue, bromo-deoxyuridine (BrdU) before death [116]. From the post-mortem human sections neurogenesis was observed. The number of BrdU-labelled neurons entering the neuronal lineage was lower than that observed in marmoset [192] and rodent studies. However, given the advanced years of the patients in the human study and the fact that neurogenesis has been shown to decline with age, this may account for the discrepancy [189, 296]. One must also consider the differences between the lab environments which the animals experienced and the external environment which humans are exposed to [142, 178]. Such differences could lead to differences in the degree of neurogenesis in the lab animal as opposed to the human. Thus considering the uncertainty about the impact factors such as age and environment, it is uncertain as to whether the relatively low levels of human neurogenesis observed, compared to what has been observed in other animals, relates to a genuine species difference.

There are only two studies that have sought to try and detect neurogenesis in the in-vivo human brain. Both studies took different approaches based on MR technology. The first of these studies developed a magnetic resonance spectroscopy (MRS) strategy to monitor the number of neural progenitor cells in both live animals and humans. The study confirmed the presence of cell proliferation in the hippocampus and its absence in the cortex [215]. While the methodology of this study appeared comprehensive there was still some ambiguity over the nature of the underlying MRS metabolite marker being measured. Indeed, this work has been broadly criticised casting doubt over the findings [162, 170, 129].

The second study of this type aimed to exploit the differences in cerebral blood volume (CBV) during exercise and at rest. Exercise has been shown to increase neurogenesis in adult mice [355]. Moreover, CBV can be measured in-vivo using MRI making it a viable biomarker to be assessed. This was first tested in rats and subsequently in humans. The study assumes a correlation between angiogenesis and neurogenesis [272, 197]. Hippocampal subregions were defined and regional changes of CBV were recorded in the dentate gyrus. In addition, exercise induced CBV increased as a result of increased angiogenesis and this correlated with the degree of neurogenesis as determined by post-mortem measurements on the animal brains. As in the case of the animals, when the same study was performed in humans the same correlation between CBV in the dentate was found, as well as improvements in cognitive and cardiopulmonary function. Thus it appears that exercise induces neurogenesis in the dentate of humans and that there is a correlation between CBV and the degree of the neurogenesis which occurs [281].

There has been much written about the role of neurogenesis and in particular neurogenesis in depression but ultimately the proposed mechanism can be reduced to a single idea, discussed in the next section.

1.7.3 What is the neurogenic hypothesis of MDD?

The neurogenic hypothesis posits that a decrease in the production of newborn dentate granule cells in the hippocampus causally relates to the pathogenesis and pathophysiology of MDD and that enhanced neurogenesis is necessary for treatment of depression (Duman et al. 2000; Jacobs, Praag, & Gage 2000) [111, 169]. The basis for the hypothesis when first posited was a result of several converging areas of evidence. Firstly, stress, a major causal factor in MDD, is known to suppress neurogenesis [141]. Secondly, most anti-depressant treatments increase neurogenesis [205, 209]. Thirdly, an imbalance in the serotonin system affects hippocampal neurogenesis [277]. Fourthly, neurogenesis is dependent upon chronic and not acute selective serotonin re-uptake inhibitor (SSRI) treatment, mirroring the time course for the therapeutic action of antidepressants [101]. Finally, the therapeutic lag in the response to SSRI's in patients with MDD parallels that of the timeline of maturation and integration of newborn dentate granule cells [208]. Therefore, given the evidence, it is reasonable to postulate that neurogenesis within the adult dentate gyrus is a potential substrate for antidepressant response.

1.8 Aims for the thesis

A detailed account of the hippocampal anatomy has been provided. Supporting evidence for the importance of the hippocampus and in particular the involvement of neurogenesis in the dentate gyrus has also been examined. This thesis will explore and exploit the use of 3T MRI and the latest developments in image processing techniques to measure hippocampal and hippocampal subfield volumes, hippocampal metabolites and morphology.

Automated and manual hippocampal volume measurements in normal healthy subjects of two different socioeconomic groups will be compared (Chapter 2). Further volumetric analysis, in the form of voxel-based morphometry (VBM), will also be performed on these groups as another means of investigating volumetric changes in the hippocampus while also allowing other brain regions to be assessed (Chapter 3). MRS metabolites will be measured in the two differing socioeconomic group to assess markers for neuronal integrity and turnover (Chapter 4). Hippocampal volumes in healthy normal volunteers will be assessed in Chapter 5 while in Chapter 6 the phenomenon of hippocampal malrotation will be assessed using novel hippocampal volumetric and morphology methods. A combination of volumetric methods will be used to assess the hippocampal subfields of a small, pilot group of patients with chronic inflammatory disease. The patients will be assessed before and after treatment with an anti-inflammatory drug which may have benefits to mood (Chapter 7). Finally, in Chapter 8 an extended protocol for measuring hippocampal subfields will be presented and measurements will be made in normal healthy volunteers.

The brief description of each chapter above will be elaborated upon within each chapter where the individual aims and hypotheses will be detailed. As the reader progresses through this thesis it is hoped that the common themes will emerge, these being, the development of reproducible acquisition and analysis protocols using 3T MRI and the latest image processing methods to establish accurate methods for making measurements of the hippocampus. Efforts will focus on assessing the quality of the measurements, comparing methodologies, establishing normal ranges, and reviewing intra- and inter-operator variability. Furthermore, this preliminary work may be used as a base for future work including an assessment of the variance of the measurements and subsequently the study power, something which is often overlooked but is of great importance in developing good study design.

Chapter 2

Manual and automated hippocampal volume measurements in groups with different socioeconomic status

2.1 Abstract

In a pilot imaging study to assess hippocampal volumes using 3T MRI in two distinct groups of differing socioeconomic status (SES), manual and automated methods of assessing the hippocampus were developed and implemented. The author developed a protocol for segmenting the hippocampus based mainly on observations recorded in Konrad et al around a review of differing hippocampal volumetry protocols [185]. Performing manual segmentation of the hippocampus is time consuming and challenging where even experienced operators may take between 30 and 45 minutes to complete a single segmentation. Therefore, the author applied an automated method of hippocampal segmentation where we aimed to determine whether or not it was possible to replace the manual method. Intra and Inter rater quality measures including the Intraclass correlation coefficient for agreement and the Dice metric were determined for repeat measures of the manually delineated hippocampal volumes. Both Intra and Inter rater measures were typically of a good standard (>0.85). This provided confidence in the clarity and comprehensiveness of the manual segmentation protocol and the skill of the rater performing the measurements. The mean volumes for the manually segmented hippocampal volumes and automatically segmented volumes and the variance in these measurements were then found to be comparable. The dice overlapping metric comparing the two

methods was 0.81. This is comparable to what has already been observed in the literature for automated segmentation methods of the hippocampus [123, 278]. Despite the reasonable results that were obtained by running the automated segmentation algorithm, in some circumstances the automated method had clear, gross errors. Therefore it is recommended that if automated segmentation algorithms are employed in future work that they are manually inspected and that volumes are corrected wherever necessary. It is hoped this approach will provide the best balance between operator analytical time and volumetric accuracy.

Given the mean hippocampal volume and the standard deviation of the two SES groups, a sample size calculation was performed to determine how many subjects would be necessary to achieve a study power of 0.90. It was found that approximately 57 subjects would be required in each group to achieve this level of study power when assessing hippocampal volume differences as a result of deprivation. It is hoped that this calculation will inform future work which may be performed to assess if significant volume differences exist between differing SES groups.

2.2 Introduction

A persons socioeconomic status (SES) and the reasons why such disparities exist is a topic that is of political as well as academic interest. Socioeconomic status (SES) is increasingly recognised as an important determinant of ill health. Growing up in an family with low SES is associated with substantially poorer health and impaired psychological well-being, impaired cognition and emotional development throughout the lifespan [2, 54, 64, 91, 118, 233].

Animal models can offer important insights into the effects of SES on brain development. While animals may not be considered to have an SES of their own per se, models can be established to assess various components and correlates of SES which can be used to investigate aspects of SES. This includes, assessing pre-natal factors, post-natal parental behaviour and cognitive stimulation [150]. The benefit of animal models is that they allow a greater degree of control of various aspects of the experimental conditions that is not possible, desirable or perhaps even ethical in human studies. There are various candidate mechanisms by which the effects of SES may be exerted. SES influences the quality of the physical and psychosocial environment throughout development [119]. Factors such as the exposure to cognitive stimulation, toxins, nutrition, prenatal drug exposure, and stress, including prenatal stress and its associated effects on parenting may all play a mediating role in the effects of SES on the brain [149, 150].

Animal research supports a relationship between prenatal chronic stress and neuroplasticity. In rodents, pre and peri-natal glucocorticoid administration to pregnant females reduces brain weight at birth, inhibits neurogenesis and delays neuronal maturation, myelination, gliogenesis and synapse formation [320]. In addition, maternal stress during pregnancy decreases spine density in multiple brain areas that are related to emotion regulation, including the hippocampus, anterior cingulate and orbitofrontal cortex [260], and increases behavioural and hormonal responses to stress in the offspring in adulthood [234, 320, 201, 367, 136]. Similarly, it has been shown that levels of pro-inflammatory cytokines (small cell-signalling proteins) such as Interleukin-6 (IL-6) can affect neurogenesis in the hippocampus in adult mice [352]. Mechanistically, changes in a variety of factors such as glucocorticoid levels and immune markers (IL-6) are implicated in linking chronic stress to brain changes. In a further animal study in Rhesus monkeys, fetal exposure to elevated glucocorticoid levels reduces hippocampal volumes in adulthood [349]. The offspring of female Rhesus monkeys that were stressed during pregnancy exhibit decreased birthweight, impaired neuromotor development, attention deficits and emotional dysregulation across the lifespan [314]. Furthermore, there is evidence from rodent studies that prenatal influences the limbic-hypothalamus-pituitary-adrenal (LHPA) axis activity and that this can be transmitted across generations in an epigenetic manner [219].

Similarly, when considering prenatal influences in humans, low SES of pregnant women increases the likelihood of premature birth and impaired fetal growth [331], both of which predict increased rates of childhood mental illness and poor performance at school [171, 341, 234, 48, 313]. Low SES is also been associated with higher levels of stress, higher rates of infection and poor nutrition during pregnancy. All of these factors can increase levels of corticotropin releasing factors and glucocorticoids in both the mother and the fetus [78, 229, 234, 320] and can restrain fetal growth [234, 320] and trigger prematurity [78]. Furthermore, glucocorticoid administration during pregnancy is associated with increased externalising behaviour, shyness, distractibility and lower IQ in children [374]. In addition, lower birthweight is linked to smaller hippocampal volumes adults [69]. All of the above suggest that conditions which are associated with low SES compromise fetal growth and subsequently neurodevelopment which affects neural function that persist into adulthood.

Parental care has also been shown to play an important role in establishing SES. Studies on rodents and non-human primates have revealed evidence for direct stress on the quality of the mother and infant interaction, on gene expression and on neurodevelopment. For example, in a study of macaque monkeys, restricted access to food has been observed to be a stressor that greatly impairs mother and infant interactions, this leads to increased stress reactivity in the

offspring, reflecting a long standing effect of parental care [92]. Similarly, in a further study on rodents the frequency of licking and grooming of pups by the mother was diminished by chronic stress induced during pregnancy [80, 304]. Cross-fostering studies in rats have also revealed direct effects of maternal care, independent of genomic influences, on hippocampal physiology and on the response to stress in the adult offspring [128, 70]. Effects of chronic stress in rats has also been observed to be transmitted across generations [79].

These animal studies mirror that which has been observed in human studies of SES and parenting where stressful environments alter the quality of parenting and therefore developmental outcomes. For example, in humans low SES has been associated with greater irritability and depressed and anxious moods in parents which compromises parent and child interactions [145]. Quality of parenting, in particular in relation to early life, predicts children's emotional behaviour and behavioural patterns in later years [267, 298, 86, 146]. Interestingly, high quality parent and child interactions are associated with resilience among children who live in stressful, impoverished environments [218]. Furthermore, clinical programmes aimed at improving parental practices in poor, high-risk, families improve behavioural and cognitive outcomes in children [353, 268, 125]. This demonstrates that to some degree, high quality parenting mediates some of the effects typically observed in subjects from low SES backgrounds and importantly offers a window of opportunity by which an intervention to low SES could be applied.

The developing human brain also appears to be particularly sensitive to early life stress and relative social standing. For example, childhood SES influences neural development of language and executive function as reflected by school performance [247, 300, 113, 44, 127]. The latter appears especially important in achieving positive life outcomes, despite adversity, in low SES in children and adolescents [68, 191]. Furthermore, the neurobiological impact of poverty appears to be greater if experienced in earlier rather than later in life. This is supported by cross-fostering study data and by twin studies suggesting cognitive ability is almost entirely predicted by environmental factors at lower SES [150].

Though the area of cognitive neuroscience is comparatively small it is a rapidly growing field and much of the work performed in this area has focused on the effects observed in the developmental human brain. While assessing the developing brain is clearly essential in elucidating the early life changes that occur as a result of SES, what of the adults who find themselves in these differing SES groups? Little work has been performed in this area, for example, in assessing differences in brain volume. One of the few studies in adults focusing on stress related brain areas and assessing neurocorrelates of SES found that SES was positively correlated to the size of the perigenual ACC [133]. Similarly, further work by Gianaros

et al reported chronic life stress predicts decreased grey matter (GM) volume in the right hippocampus [134].

Given the animal evidence and the preliminary data human studies discussed above, low SES may be considered as a chronic stressor [149] and as such for those people of a lower SES may be considered to be exposed to chronic stress through an inflammatory pathway [226]. This has led to substantial research on putative stress-related pathways that might explain the SES-illness gradient [324]. In this context, the brain is the central organ mediating stress reactivity and recovery processes, insofar as the brain determines what is threatening and thus stressful. The brain jointly regulates peripheral patterns of stress physiology that ultimately impact on it through several feedback mechanisms [227]. By extension, low SES can be construed as a state of significant, ongoing threat processed by, and ultimately affecting, the brain. An anatomical and functional network of neural circuitry including hippocampus, amygdala and prefrontal cortex is central to the coordination of cognition, experience and behaviour with neuroendocrine, immune and autonomic functions in the service of adaptively coping to effectively meet environmental and psychosocial challenges [227]. Closely allied to this is the construct of allostatic load, which refers to the ‘wear and tear’ the body experiences when repeated allostatic responses are activated during stressful situations [224]. This framework is useful in defining physiological markers of chronic stress and phenotyping the adversity experienced by subjects of lower SES.

Thus there is a need to further study the adult brain in groups of people of different SES. Determining whether global or localised volumetric differences exist between fully developed brains in groups of people of different SES may be informative as to assessing which regions are affected in the longer term. As previous work has implicated volume changes in the hippocampus [134], this region will be of particular interest although, by the nature of the VBM analysis (See Chapter 3), other regions may well be considered if statistically significant differences are found in other areas of the brain. As stated above, there are not many studies that have assessed brain volumetric differences in groups of differing socioeconomic status and there are fewer still which have included the multiple biological, psychological and social measures which may be used to explain the resultant differences. Thus we are well placed to perform a pilot assessment of the brain differences between two groups of differing SES to establish the methodology which will inform future work.

In assessing brain volume differences between a low SES group or more deprived group with a less deprived group, potentially via mechanisms involving stress one may well question the robustness of this model. That is, it may be perceived that we are suggesting and assuming that more affluent subjects are not subject to stressors. This is not the case. People with

higher SES may of course be subjected to what in layman's terms would be described as stressful situations. These types of environmental stressors, which often lead to the person feeling 'stressed', while significant at an individual, with potential consequences for the individual's health, are likely to only form a small part of the myriad of stress a given individual is exposed to throughout their life. Indeed, environmental stressors for example, from a stressful working environment or from stressful life events such as bereavements are likely to be apparent on both low SES and higher SES subjects. However, as has been discussed above, there is a considerable body of evidence from animal work and from developmental human studies assessing the impact of biological and environmental stress and the profound impact of low SES on health and cognitive outcomes. These effects may be particularly potent on the developing brain or to put it another way, the fetal/child brain is more vulnerable to stress exerted by biological and environmental factors where stressful environmental life events acting on an adult are likely to form only a small part of the overall life stress or 'allostatic load' as it is referred to in the peer reviewed literature [222].

Thus, the stress discussed in the neuropsychological literature can come from a range of sources, most often of which the person (or animal) has no knowledge, recognition or awareness of, such sources of stress also permeate throughout all stages of a subject's lifespan, and arguably, prior to conception via the transfer of genetic traits, through prenatal, antenatal and childhood care and then later into adulthood i.e. the adult stress factors may only account for a fraction of the biological and environmental vehicles for stress which a particular person is likely to have been exposed to throughout their life.

Therefore, in the PSOBID study, two groups of subjects were recruited, a more deprived (MD) and least deprived (LD) group, where the assumption is that those subjects in the more deprived groups are likely to have been exposed to a greater degree of life stress, this being a result of the range of factors discussed above which are likely to have affected brain development and cognition. Thus, this study is a snapshot taken in adulthood, across two distinct SES groups, to assess the effect on the adult brain as a result of the cumulative effects of lifespan stressors.

We wish to consider the implications of low SES i.e. the long term exposure to deprivation, potentially mediated by stress, on the brain and in particular, on the hippocampus. SES differences are present in all aspects of society throughout the world. If the mechanisms by which subjects of lower SES are more prone to ill health than more affluent people may be understood then it may be possible to try and develop increasingly more powerful schemes to mediate its effects. In order to address this issue in a robust manner, given as we had access to a well categorized sample a pilot study was performed to establish methodologies

for manually and automatically segmenting the hippocampus. These two methods were compared and following assessment of the methodology, estimation or power and sample size were performed to inform future work which may be used to establish whether hippocampal volume differences exist between less deprived and more deprived subjects.

2.3 Aims

The aims of the pilot imaging study aspect of the PSOBID study was to establish methodologies for assessing hippocampal volumes, morphology and metabolites. Given the methodology, the mean and variances of two differing SES groups will be assessed with a view to estimating the necessary power and sample sizes necessary for subsequent work that may be used to determine if significant differences exist between the two groups. Hippocampal volumes are discussed in this chapter, focal GM volume differences are investigated in chapter 3 and 3T MRS metabolite levels in the hippocampus are discussed in chapter 4.

The aim in this chapter was to establish a methodology employed for manually delineating the hippocampus and also to perform a comparison between automated and manual hippocampal volumetry measurements. Furthermore, intra and inter rater variability and other hippocampal volume quality metrics such as the intraclass correlation coefficient (ICC) and Dice overlap metric were assessed. As few studies have provided a comprehensive comparison between automated and manual hippocampal volumes from 3T MRI data, it is hoped the work presented here will add to the current body of literature and thus help with planning further studies

2.4 Literature review

2.4.1 Hippocampal volumes in normal subjects

In order to gain an estimation of what the normal hippocampal volume should be, a review of the literature was conducted. While there are large numbers of studies in which hippocampal volumes have been assessed, two key factors make it difficult to compare volumes across studies. Firstly, protocols for delineation of the hippocampus can vary, this issue itself has been studied with a view to highlighting best practice [185]. Typical variations in protocols between studies include whether or not the WM components of the hippocampus are included in the total hippocampal volume. Discrepancies in locating various anatomical boundaries may also lead to volume differences. A further complication in comparisons across studies

of hippocampal volume is how the hippocampal volume is corrected relative to head size. It has been known for some time that MRI can be used for measuring the size of subcortical structures and that it is beneficial to correct for intracranial cavity volume (ICV) when making these measurements. This adjustment can improve comparisons of tissue volumes for subjects of different ages, where a decrease in volume of 1.6% per decade has been observed in male subjects in the age range 20-60 [90]. Given the variation in methodologies for correcting hippocampal volumes, it was simplest in the first instance to compare uncorrected hippocampal volumes from normal controls and to accept that some of the inherent variation in this measurement will be the result of the differing subject head size and age.

As there are not very many studies which solely assess volumes of normal subjects, the hippocampal volumes from control subjects from a meta-analysis of studies on depressed subjects were used. McKinnon et al [232], provides a comprehensive review and meta-analysis of hippocampal volumes in patients with major depressive disorder (MDD). Within this analysis McKinnon et al list estimated uncorrected hippocampal volumes taken from 32 MRI studies of MDD patients and associated normal controls. Studies were performed between 1999 and 2008 and when combined in this way hippocampal volumes are presented for approximately 1000 normal control subjects. Of the 32 studies most were performed using 1.5T MRI, a couple of studies used 1T [196, 266] and a single study used 3T MRI [262]. To supplement the work performed by McKinnon et al, for the purposes of this review, the meta-analysis was extended to include more recent publications of hippocampal volumes from normal controls in studies assessing MDD. Four further studies published between 2008 and 2010 were added, two which used 1.5T MR scanners [213, 210] and two which used 3T MR scanners [176, 292]. The results of the extended meta-analysis are shown in table 2.1 on the following page.

Therefore, having extended the meta-analysis the average hippocampal volume for normal controls was $3079\text{mm}^3 \pm 377\text{mm}^3$ and $3159\text{mm}^3 \pm 381\text{mm}^3$ for the left and right hippocampus respectively. Although it is difficult to make a definitive comparison between 1.5T and 3T from the meta-analysis as significantly fewer studies have been performed at 3T, there does appear to be a tendency for volumes acquired at 3T to be larger than volumes acquired at 1.5T. It remains to be proven whether this observation is explicitly the result of improved image quality at the higher field strength or perhaps as 3T has come along later, the delineation protocols are better as a result of the experience of earlier studies.

Hippocampal volumes of control subjects in studies assessing Hippocampal volumes in MDD
 Table is updated from the meta-analysis of McKinnon et al, 2009.
 All volumes are presented in cubic mm

Paper author (et al)	year	N	Mean Age	Age range	MRI field strength	Left Hippo vol	Std dev	Right Hippo vol	Std dev	Normalisation method
Sheline	1999	24	53	18	1.5	2482	305	2468	309	uncorrected
Mervaala	2000	17	42	15	1.5	3441	436	3700	467	uncorrected
Steffens	2000	18	67	5	1.5	3170	440	3300	440	uncorrected
Vakili	2000	20	40	10	1.5	2460	380	2600	510	uncorrected
Von Gunten	2000	14	58	ng	1.5	2644	410	2700	322	uncorrected
Rusch	2001	15	37	14	1.5	2130	270	2200	240	uncorrected
Frodl	2002	30	41	13	1.5	3772	397	3763	411	uncorrected
Vythilingham	2002	14	ng	ng	1.5	3179	460	3037	501	uncorrected
Macmillan	2003	23	14	2	1.5	3240	440	3260	400	uncorrected
Macqueen	2003	20	36	12	1.5	2761	368	2784	342	uncorrected
Posener	2003	42	33	11	1.5	2475	359	2994	414	uncorrected
Sheline	2003	38	51	17	1.5	2421	318	2429	326	uncorrected
Caetano	2004	31	37	11	1.5	3370	420	3320	430	uncorrected
Frodl	2004	30	46	13	1.5	3820	340	3930	350	uncorrected
Janssen	2004	41	52	11	1.5	3200	520	3120	450	uncorrected
Lange	2004	17	32	6	1.5	2990	460	3190	370	uncorrected
Lloyd	2004	39	73	7	1	2800	400	3000	400	uncorrected
MacMaster	2004	17	16	2	1.5	3050	110	2880	110	uncorrected
O'Brien	2004	40	73	7	1	2820	420	3000	410	uncorrected
Vythilingham	2004	33	34	10	1.5	3334	390	3235	407	uncorrected
Xia	2004	13	35	9	1.5	3352	46	3710	37	uncorrected
Hickie	2005	20	56	10	1.5	3300	500	3300	600	uncorrected
Neumeister	2005	57	38	11	3	3576	342	3679	351	uncorrected
Taylor	2005	83	69	6	1.5	2960	450	3120	440	Corrected for total cerebral volume
Frodl	2006	34	47	13	1.5	3060	300	3140	300	uncorrected
Saylam	2006	24	30	6	1.5	2787	249	2806	257	uncorrected
Weniger	2006	23	32	7	1.5	3000	500	3200	400	uncorrected
Frodl	2007	60	44	12	1.5	3886	412	3970	434	uncorrected
Hickie	2007	16	56	10	1.5	3190	280	3290	470	uncorrected
MacMaster	2008	35	15	3	1.5	3150	460	3160	420	uncorrected
Keller	2008	22	32	12	3	3780	490	3890	460	Corrected for age and brain volume
Qiu	2009	31	>60	ng	3	3520	481	3607	542	uncorrected
Maller	2010	76	41	35	1.5	2773	409	2629	403	uncorrected
Malykhin	2010	34	34	8	1.5	2798	257	2982	244	uncorrected
		Total N				Mean volume	Mean Std dev	Mean volume	Mean Std dev	
ng = data not given		1051				3079	377	3159	381	

Table 2.1: Summary of hippocampal volumes from normal controls in studies assessing hippocampal volumes in MDD patients

2.4.2 Hippocampal volumes and reliability at different MRI field strengths

The overwhelming majority of studies assessing hippocampal volume have been performed using 1.5T MRI. A review of hippocampal volumes in MDD alone lists almost 30 studies at 1.5T [232]. More recently, a small number of studies have used 3T, 4T and 7T.

In a study comparing hippocampal volumes, scanned twice at both 1.5T and 3T in eight healthy controls it was found that there was no significant difference between the volume estimates from these two systems, though the repeatability of the measurements was better

at 3T [60]. In a similarly designed study, in 10 epilepsy patients, the patients were again scanned at both 1.5T and 3T field strengths. A high linear relationship was found between the two different field strengths (Pearson r correlation coefficient = 0.975) with the authors concluding that there was no significant difference in the same measurement done at the two different field strengths [319]. Importantly the same raters applied the same protocol, in this case, with only hippocampal GM being manually outlined. One might argue that any large variability observed between measures on different systems is more likely to be a result of differing raters working to differing protocols than any affect of field strength.

In a study comparing normal volunteers on 1.5T and 4T system it was found that manually traced hippocampal volumes were on average 16% smaller when measured compared to the volume acquired at the higher field strength [Levy-Reis, Proc. Intl. Soc. Mag. Reson. Med, 2000, 8, 15] . In a study (N=20) where the reliability of assessing hippocampal volumes at 3T was assessed it was found that uncorrected hippocampal volumes were $3185\text{mm}^3 \pm 411\text{mm}^3$ and $3302\text{mm}^3 \pm 411\text{mm}^3$ for the left and right hippocampus respectively [174]. Intraclass-correlation-coefficients (ICC) of 0.86 (left) and 0.86 (right) were determined which were comparable to what had already been observed in the literature, where ICC values range from 0.73 to 0.95 [202, 330, 274]. Importantly Jeukens et al also made the point that in comparing the reliability of volumes that some measure to assess the degree the volumes overlap or intersect should be included as well as the ICC and inter reader volume difference [174]. This reflects the fact that two operators could theoretically delineate the same volume but in different spatial locations, thus the ICC alone may give a false impression of conformance which is not illuminated until a measure of overlap is considered. Overlap measures such as the Dice metric are an important and useful factor to define in delineating volumes but these values are reported less often in the literature. When calculating the Dice metric values greater than 0.8 are typically considered satisfactory, though this is a somewhat arbitrary assignment. The Dice metric will be discussed in more detail later in this chapter, see section 2.6.2.

A further study using 3T MRI aimed to compare hippocampal volumes of heavy and light users of alcohol. For the 8 lighter users of alcohol the hippocampal volumes were $3600\text{mm}^3 \pm 590\text{mm}^3$ and $2890\text{mm}^3 \pm 500\text{mm}^3$ for the left and right hippocampus respectively. The larger standard deviations observed here may be a result of the smaller numbers of subjects involved in this study [39]. Another study performed using data acquired at 3T was the study by Hanamiya et al [155]. The mean hippocampal volume from the 15 normal controls in this study was found to be $3156 \pm 250\text{mm}^3$, the method used in this study was different to the majority of more recent studies in that the hippocampus was outlined on a T2-weighted FSE

images in the coronal plane, this is in contrast to the 3D T1-weighted images often used. However, the volumes compare reasonably well to what has been observed in other studies.

A single study using 4.1T MRI [273] and two studies using 4.7T MRI [214, 359] have been performed to assess the hippocampus. These studies focused on the internal structure of the hippocampus and not the hippocampal volume as a whole. Thus, whole hippocampal volume measurements were not made. Similarly, a handful of studies have been performed using 7T MRI [59, 290, 346, 84, 369, 345], with a couple of further studies having been performed at even higher field strengths such as the work by Chakeres et al using an 8T magnet [77] and also the work by Fatterpeker using a 9.4T magnet [121]. It should be noted, to avoid any confusion, that the early study by Wieshmann et al using a 7T magnet [369] and the more recent studies at field strengths greater than 7T [77, 121] were ex-vivo studies where the hippocampus had been excised from the brain. Of the studies performed more recently using 7T MRI, the majority have provided a purely radiological assessment of the in-vivo hippocampus [346, 59, 345, 290], thus these studies lack any volumetry or other quantitative measures. However, these studies do impressively show, albeit by a qualitative assessment, the improved resolution and image quality at 7T, with in-plane voxel sizes down to the order of around $230\mu\text{m} \times 230\mu\text{m}$ [345, 290]. The only study at 7T to provide detailed, in-vivo hippocampal volumes, to-date, to the authors knowledge, is the study by Cho et al [84]. This study acquired 3D T1-weighted images of the hippocampus with a voxel size of $350\mu\text{m} \times 350\mu\text{m} \times 350\mu\text{m}$, the study also compared volumes acquired from 1.5T and 7T. The volumes, obtained by manual tracing, were found to be larger at 7T when compared to the same volume measured at 1.5T. This likely reflects the benefit of the increased resolution, providing more detail in the images, and for example, allowing the internal digitations of the hippocampal surface to be more reliably defined. Reliability and sensitivity of determining the hippocampal volumes at the two field strengths were also compared. Volumes obtained from the 7T were consistently found to be more reliable and more sensitive than volumes obtained from images using 1.5T MRI. Mean volumes for the left and right hippocampus from images acquired at 1.5T were $2884 \pm 297\text{mm}^3$ and $3112 \pm 332\text{mm}^3$ respectively whereas mean volumes for the left and right hippocampus from images acquired at 7T were $3232 \pm 232\text{mm}^3$ and $3340 \pm 323\text{mm}^3$ [84].

2.5 Methodology

2.5.1 Ethics

The study was approved by the local NHS research ethics committee.

2.5.2 Volunteers, materials, methods and study design

A large study of the psychological, social and biological factors was conducted in Glasgow, this was known as the PSOBID study [356]. The aim of this study was to investigate factors which may contribute to ill health in disadvantaged communities where such communities have been shown to have increased incidence of, for example, heart disease, diabetes and some cancers [327]. The author of this thesis had no involvement with the recruitment or screening of the subjects selected for the imaging component of the PSOBID study, however, a description of how participants were selected is clearly of relevance to chapters in this thesis related to the PSOBID study and is therefore described below.

The Health Information and Technology (HIT) section of Greater Glasgow Health board were responsible for sample selection of the full PSOBID study. Subjects were recruited based upon the Scottish Index for Multiple Deprivation (SIMD) [<http://www.scotland.gov.uk/Topics/Statistics/SIMD/>]. The SIMD ranks small areas on the basis of multiple deprivation indicators across six domains, namely: income; employment; health; education, skills and training; geographic access; telecommunications and housing. The SIMD was used to identify the least and most deprived areas of Glasgow. Five GP practices with the highest percentage of patients aged 35-64 year olds living in areas classified as being in the bottom 5% of the SIMD were approached and agreed to participate in the study. These subjects would form the most deprived group (MD). A further 5 GP practices with the highest percentage of 35-64 year olds living in areas as classified as being in the top 20% of SIMD, these subjects would form the least deprived group (LD).

The HIT generated a target population of 21,672 people from the GP lists of the ten practices who agreed to participate in the study. From this target population 12 groups of 300 subjects were selected stratified by a combination of SIMD category, age and sex. GP's were excluded patients who had recently died or had a terminal illness. Due to the nature of the psychological questionnaires and cognitive assessment, only those who understood and spoke English were invited to participate in the study. The eligibility of subjects was checked by GP's and practice managers. Following GP practice approval letters were sent to the

potential participants inviting them to participate in the study. Invitation letters to selected subjects were sent in batches of 150 every two weeks. Accompanying the letter was a form for the subject to return (in a reply paid envelope) recording their contact details and indicating their willingness to consider participation. Subjects who agreed to receive further information about the study were sent the PSOBID participant information booklet. Following this initial recruitment period 700 subjects were selected for the study. Of the 700 subjects recruited, 327 of the subjects were male. From the 327 male participants, 140 volunteered for the neuroimaging component of the study. From the 140 subjects a subset of 42 subjects were randomly selected for MRI to assess brain volumetric differences between the least deprived and more deprived groups and to assess correlations between brain volume differences and inflammatory measures. MRS measures of hippocampal metabolites were also acquired and these are discussed in Chapter 4 of this thesis.

The 42 volunteers chosen for imaging were a subset of the full PSOBID study, with 21 subjects each in one of either a deprived or affluent group. The age range of the deprived group was 38 - 65 years old, with a mean age of 52; the age range of the affluent group was 40 - 67, with a mean age of 53. The groups were therefore considered to be well matched in terms of their age.

Subjects completed the National Adult Reading Test (NART-II) which provides an index of peak intellectual function [96]. They also completed the General Health Questionnaire (GHQ) [137]. The GHQ is a questionnaire to assess psychological well-being. A subset of the questions from the GHQ, the GHQ-12, may be used to assess the likelihood of subjects having some form of psychosis [165, 270]. Subjects were not screened prior to imaging. A retrospective analysis of the GHQ-12 confirmed there was no significant difference in the psychological well-being between the more and least deprived subject groups.

The exclusion criteria for this neuroimaging study included: history of head injury, stroke or neurological disorder. Subjects were also chosen such that they were suitable for MRI scanning and had no medical implants that were contraindicated with the MRI environment.

The data acquired for the PSOBID study which forms the basis of chapters 2, 3 and 4 of this thesis were collected concurrently i.e. the same subjects provided both the volumetric and MRS data. Moreover, the PSOBID study also ran concurrently with the study of normal subjects detailed in chapter 5.

A full treatment of all the factors which may potentially mediate or moderate differences brain volume differences in differing SES groups is beyond the scope of this thesis and the

knowledge of the author. However, discussed below are several factors where evidence exists for the variables to have an appreciable impact on hippocampal volume. This discussion will include how these factors were controlled for in subsequent analysis.

2.5.2.1 Hippocampal volumes and the relationship to stress and socioeconomic status

SES is a multifaceted construct that includes measures of economic resources in addition to social factors such as power, prestige and hierarchical social status [3, 149]. The measurement of SES is both complex and controversial, the most common indicators used to measure SES are income, education and occupation or a combination of these measures [188, 56]. Although these factors often correlate, there are enough discrepancies that they should not be used interchangeably as they reflect related but distinct components of SES [188, 56]. SES operates on multiple levels, this can be at the individual level, at the household or community level and often different factors which comprise SES effect outcomes in divergent ways [188, 56].

As has been detailed in the previous section, the SIMD multi-faceted criteria was used to separate the individuals in the PSOBID study into two separate groups with differing SES. The construct of SES is a complex one with which many factors may contribute. Measures for SES can vary widely across studies but common themes include consideration of factors of a subjects current environment, developmental environment, genetic background and parental background or SES status. Therefore, in this study groups have been partitioned into a more and less deprived group based upon the SIMD criteria which defines the subjects current SES. Thus, while there are clearly developmental and parental factors which may affect SES these are not explicitly considered in this study [150]. Instead, a cross-sectional assessment of two groups of differing SES was performed with the aim establishing the brain imaging analysis methodology and to acquire pilot data which may in future be used to reliably identify differences in hippocampal volumes between the more deprived and least deprived SES groups. Defining groups more formally in terms of their SES is more comprehensive measure than simply considering someones financial wealth or or single variable reflecting affluence [149]. One of the benefits of multi-faceted measures such as SES is the encompassing nature of these measures. As SES is undoubtedly a multi-variable problem this approach seems appropriate, though the challenge with tackling the problem in this manner comes to trying to hone in and untangle the relative contributions of the component parts of SES.

As discussed in the introduction, lower SES may be considered as a chronic stressor, perhaps through an inflammatory pathway [149, 226]. In adult humans small hippocampal volumes

have been observed in cases of post traumatic stress disorder [148, 57]. Similarly, smaller hippocampal volumes have also been recorded in adults exposed to childhood maltreatment [336, 58]. It has been known since the late 1960's that the hippocampus was a region of the brain specifically targeted by stress hormones [228]. It is also known that stressful experiences and the resultant release of stress hormones produce adaptive and maladaptive effects on the hippocampus as well as the hypothalamus and other brain regions [223]. There is growing evidence that a spectrum of socioeconomic factors contribute an appreciable amount of the variance in disease specific morbidity and mortality rates as well as for risk factors for chronic medical conditions [4, 3, 5]. Several models of SES-related health disparities propose that life experiences related to SES at various levels, either at the individual, family or community level, may influence disease risk by means of a stress related pathway [3, 107, 177]. While any perceived stress pathway from SES to increased morbidity or mortality is clearly complex with many mediating factors, here we present a novel opportunity to assess distinct community based samples of differing socioeconomic status, with the aim of comparing hippocampal volumes in these groups.

In addition to solely comparing hippocampal volumes between differing SES groups, in order to explain any factors mediating volumetric differences, selected markers of inflammation will be investigated. Inflammatory markers recorded for the PSOBID study included measures of Interleukin 6 (IL6), a protein that acts as a pro-inflammatory and anti-inflammatory cytokine. A cytokine being a signalling protein excreted by the glial cells of the nervous system and by the immune system. IL-6 has been proposed as part of the pathway by which hippocampal volumes may be reduced as a result of stress [216, 241]. Similarly, cortisol has also been implicated in the stress- inflammation- hippocampal volume pathway and these measures were also recorded [291, 241]. Cortisol is a steroid hormone released by the adrenal gland. It is excreted in response to stress. Therefore to investigate the proposed link between deprivation as a stressor leading to hippocampal volume reduction as a result of inflammatory processes, the relationship between hippocampal volumes and IL-6 and cortisol will be explored in the two differing SES groups.

To provide an initial estimation of the relationship between hippocampal volumes and measures implicated in the inflammatory pathway, IL-6 and cortisol were measured in both the LD and MD groups. Subjects were asked to fast from 10pm the night prior to having their blood sample taken the following morning. Following acquisition of the blood sample, the sample separated within an hour and then frozen at -80 degrees. IL-6 was measured by sandwich ELISA (R&D Systems Europe Ltd., Abingdon, United Kingdom). Measuring cortisol from a blood sample in this way is described as taking a single, morning measure.

2.5.2.2 Hippocampal volume and alcohol

It has been known for some time that the hippocampus is affected by chronic dependence on alcohol [38]. Moreover, alcohol has been shown to have a detrimental effect on the whole brain volume and on the hippocampus, where the volume reduction between the whole brain volume and hippocampal volume was found to be proportional [6]. While there appears to be a consensus in the peer reviewed literature regarding the detrimental effects of chronic alcohol abuse, there are less emphatic views around the damage done by low to moderate alcohol consumption. This topic was the subject for reflection in an article in the Lancet in 2004 [284]. A recent review focussing on low to moderate consumption has suggested however that there is a preferential decrease in brain volume of subjects [357]. While there appears to be some room for further discussion in this area, there was certainly enough evidence to suggest that alcohol consumption may at some level play a role in contributing to decreases in brain volume and potentially hippocampal volume. Therefore, self reported measures of alcohol consumption of the each of the subjects was recorded as part of the PSOBID dataset. Similarly to the scores of general psychological well being, subjects were not screened for excessive alcohol use, i.e. any excessive users of alcohol were not rejected from being imaged. Instead the self report measure of alcohol consumption was as a covariate in the later analysis such that differences in SES may be explored, irrespective of differences in alcohol consumption between each group. To ensure the suitability of alcohol as a covariate in later analysis alcohol levels between the two groups were compared and there was found to be no difference in the level of alcohol consumption.

2.5.3 Imaging hardware

All MRI scans were acquired using a 3T GE Medical systems, Signa Excite HD system (Milwaukee, USA) with software version 12m5. Furthermore, an eight channel phased array (receive only) head coil was used to acquire the imaging data.

2.5.4 Imaging Protocol

Following localiser scans a T1- weighted IR-FSPGR was acquired with TR = 6.8ms; TE = 1.5ms; TI = 500ms; flip angle=12°; FOV = 26cm; phase FOV= 70%; matrix: 320 x 320; bandwidth 31.25kHz and slab thickness = 1mm. The acquisition time for this scan was 8min 54s. The T1-weighted IR-FSPGR sequence has high resolution in all three imaging planes and good white matter to grey matter contrast which facilitates both the extraction of the whole

hippocampus and tissue type segmentation necessary for voxel based morphometry (VBM) (further described in Chapter 3). An example of a T1-weighted IR-FSPGR acquisition is shown in figure 2.1 on the following page.

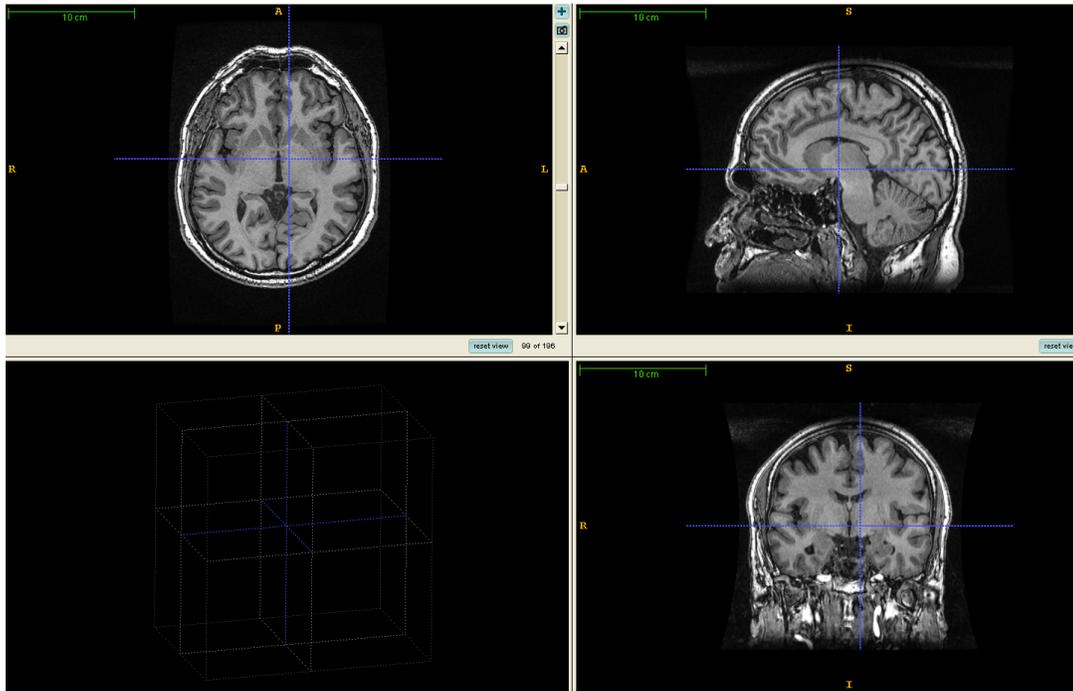


Figure 2.1: PSOBID study: 3D T1-weighted image displayed in 3-planes using ITKsnap (v1.8), Whole brain and hippocampus, T1-weighted 3D Whole Brain image

2.5.5 Manual outlining of the hippocampus

As discussed above, many studies have assessed whole hippocampal volumes based on varying protocols and these have been summarised by Konrad et al [185]. Based on the observations made by Konrad et al, a protocol for manual outlining of the hippocampus was devised and is detailed below.

Manual outlining was performed using the image processing software suite ITKsnap (v1.8) (<http://www.itksnap.org>) [377]. While there are many programs which allow manual outlining and editing of brain, the combined four pane display allowing visualisation of the three imaging planes and the 3D rendering window made this a suitable choice. Furthermore, related itk software (convert3d) offers additional functionality for assessing images, such as the ability to calculate the volume overlap measure known as the Dice metric.

The raw DICOM T1-weighted 3D data were converted to NIFTI format using Chris Rorden's dcm2nii tool (<http://www.cabiatl.com/micro/>). DICOM stands for: Digital Imaging and Communications in Medicine and is the standard medical imaging file format. NIFTI stands for Neuroimaging Informatics Technology Initiative and this file format was the result of a project to support and develop the use of informatics tools in neuroimaging. The NIFTI image format is widely used throughout neuroimaging.

2.5.5.1 Manual segmentation strategy

The strategy to manually segment the hippocampus began on the sagittal plane. Following this initial outlining further corrections and refinements were made on the axial and coronal planes. The decision was made to include the hippocampal WM as part of the hippocampus, therefore the resultant hippocampal volumes reflected both its GM and WM components. An example of the whole hippocampus, displayed in 3-planes, is shown in figure 2.2 below.

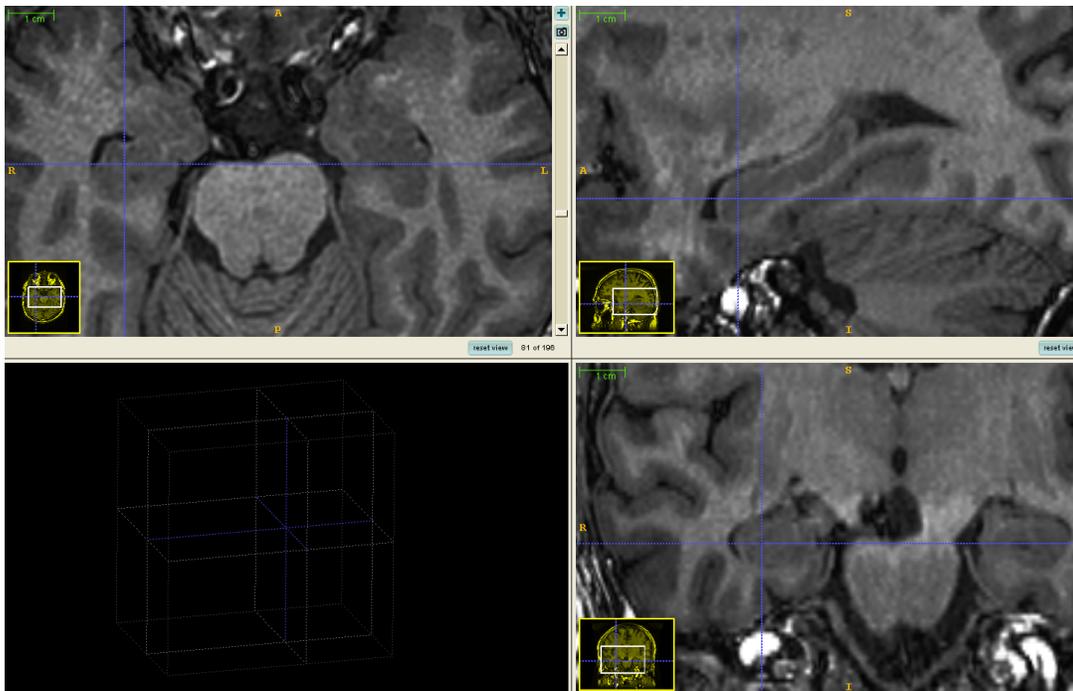


Figure 2.2: PSOBID study: 3D T1-weighted image displayed in 3-planes using ITKsnap (v1.8), Whole brain and hippocampus, 3D T1-weighted view of the hippocampus

2.5.5.2 Anterior border

Using 3T MRI the amygdala may be clearly differentiated from the hippocampus, this is shown in figure 2.3 on the following page. This border is most easily visualised on the imaging

sagittal plane and this was partly the reason that segmentation began on this imaging plane, see point A at the intersection of the two white planes shown in figure 2.3. The alveus is a WM layer that defines the border between these structures and was included as part of the gross hippocampal volume.

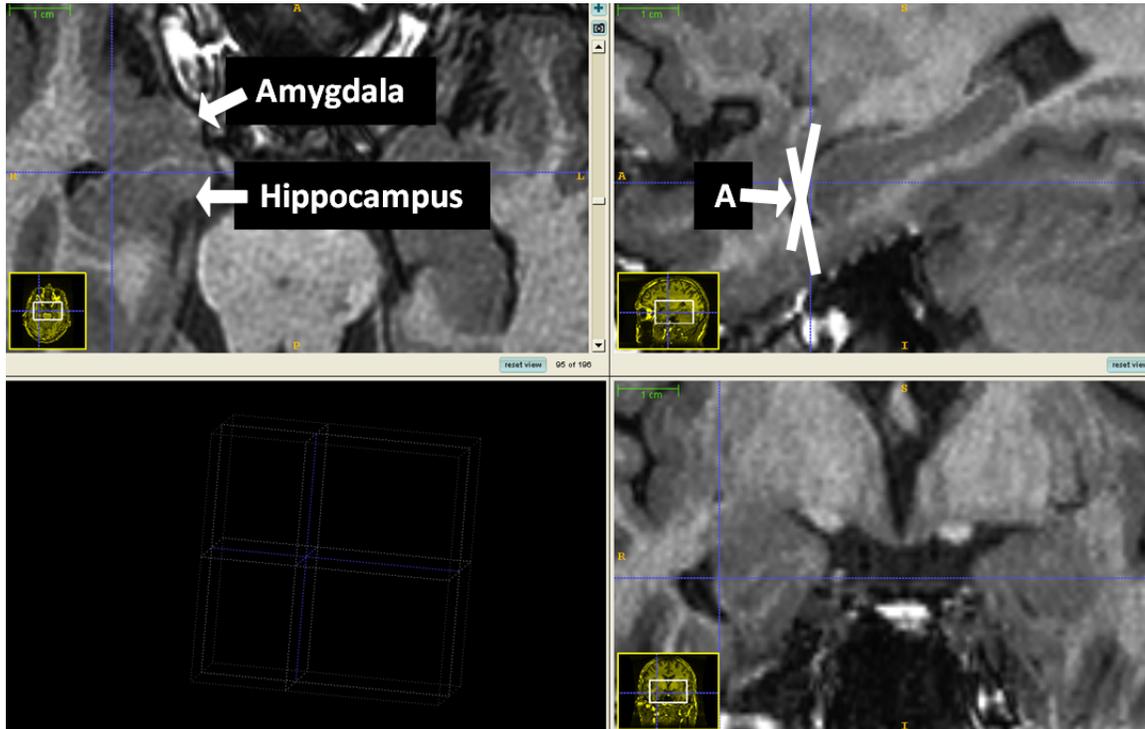


Figure 2.3: Anterior border of the hippocampus

2.5.5.3 Superior borders

As the alveus was included as part of our hippocampal volume, the alveus acts as the border between the hippocampus and amygdala at the anterior-superior aspect of the hippocampus. The superior border of the body and tail of the hippocampus was demarcated by the CSF at the medial (Point B at the intersection of the two white planes on figure 2.4) and lateral posterior aspects of the hippocampus. The plane anterior-superior border of the hippocampus and the amygdala is illustrated by plane A and at the intersection of the two white planes at point C on figure 2.4 on the following page.

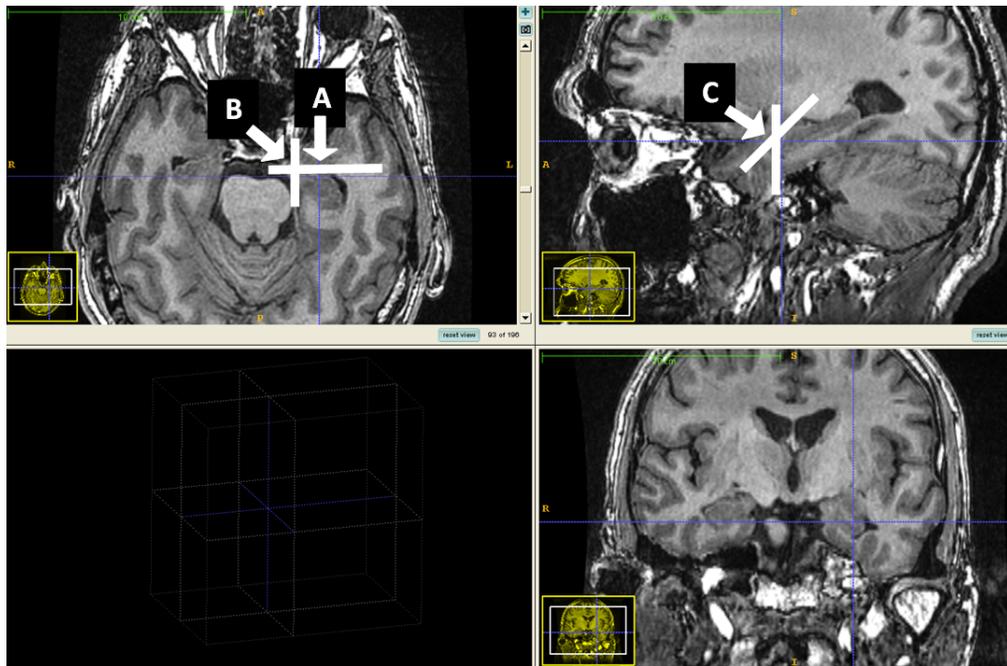


Figure 2.4: Anterior-superior border of the hippocampus

The medial-superior borders of the hippocampus are shown by planes A and B in figure 2.5 below.

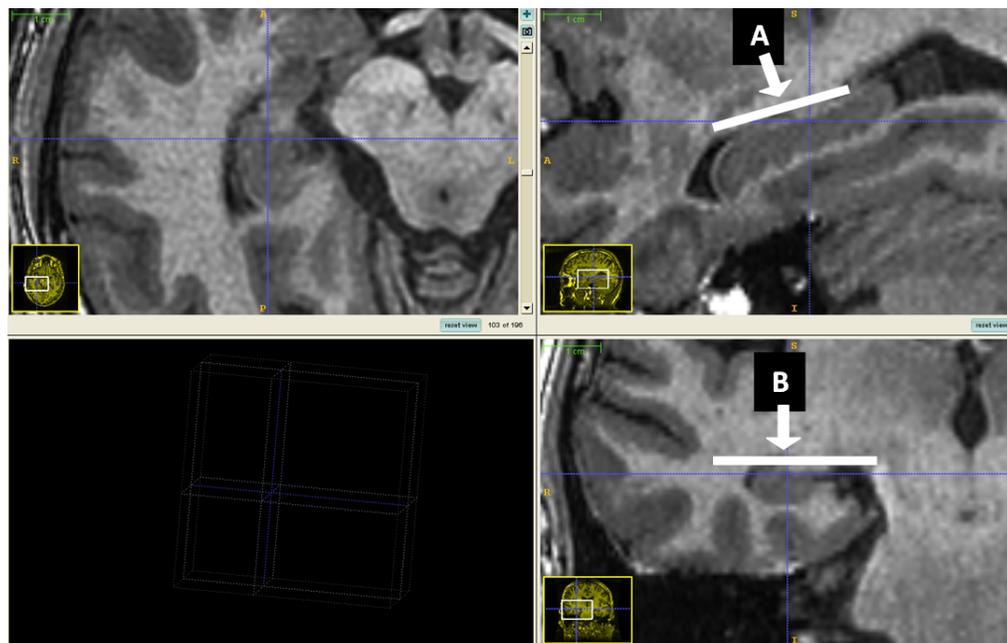


Figure 2.5: Medial-superior border of the hippocampus

2.5.5.4 Posterior border

The CSF of the lateral ventricle acts as a natural border at posterior aspects of the hippocampus (see planes A and B on figure 2.6 below), however, this is only satisfactory to describe the lateral aspects of the posterior border. Medially the posterior aspect of the hippocampus is more difficult to define due to the projection of the fornix. While it was decided to include hippocampal WM regions as part of the whole segmented volume, the fornix WM tract clearly extends well beyond the hippocampus. Therefore, it was necessary to define an anatomical landmark to act as the boundary beyond which the fornix would not be segmented.

The last axial slice on which the fornix was outlined, was the slice prior to the entire convergence of the white matter (medially) around the tail of the hippocampus i.e. the convergence of the WM into a single WM bundle. This may otherwise be described as the slice beyond which the grey matter ovoid of the hippocampal tail is no longer visible. The posterior boundary described here has been described in several protocols for hippocampal segmentation and was most easily visualised and segmented on the coronal and axial imaging planes.

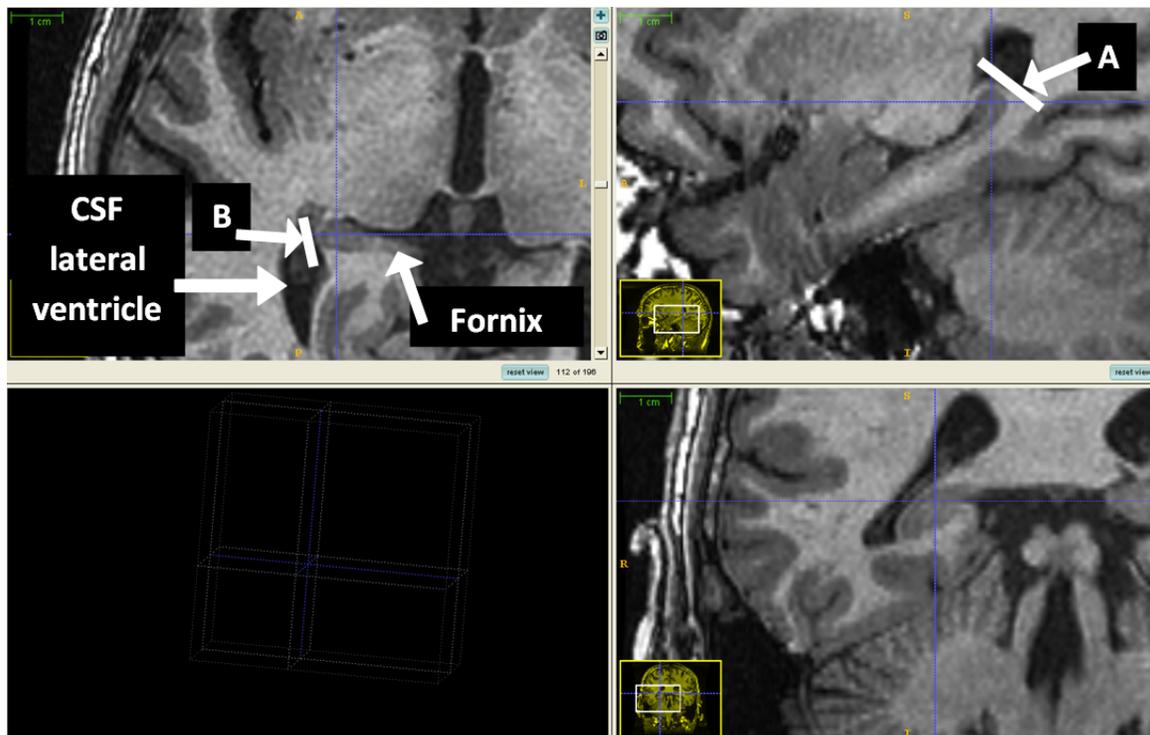


Figure 2.6: Posterior-medial border of the hippocampus

2.5.5.5 Inferior border

The WM of the parahippocampal gyrus below the subiculum was clearly discernible from the hippocampal GM and was used as the inferior border of the hippocampus. This border was most easily visualised by using a combination of all three imaging planes. Konrad et al describe the inferior medial border of the hippocampus, as defined in neuroimaging articles to delimit the hippocampus along the inferior part of the cornu ammonis and the subiculum [185]. The medial border is shown on the axial image as plane A on figure 2.7 below, while the inferior-medial border is shown on the coronal image at the intersection of the two white planes at point B on figure 2.7. This border is the most infrequently and poorly defined in the literature, similarly there is no consensus over where the border should be and as a result there are large variations between methods.

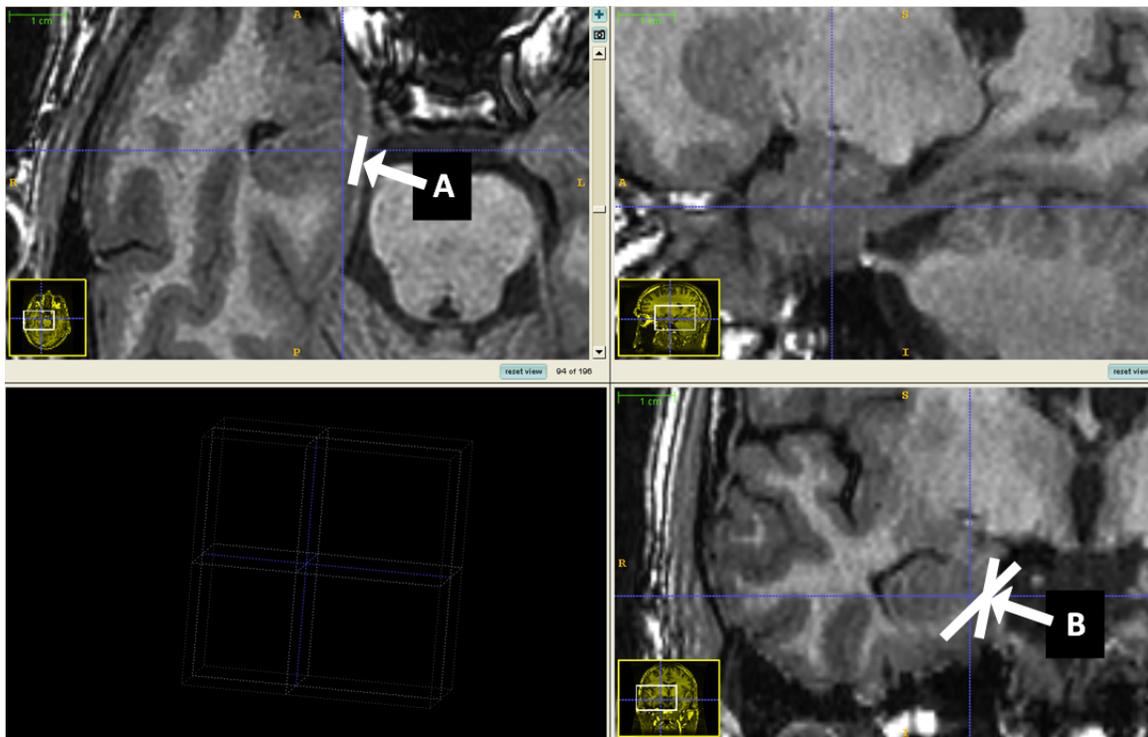


Figure 2.7: Inferior border of the hippocampus

2.5.5.6 Superior-medial border

The consensus across most of the imaging protocols defines this boundary as the CSF of the

cisterna ambiens. This is most easily visualised on the coronal and axial planes.

2.5.5.7 Lateral border

The CSF of the lateral ventricle was used as the lateral border of the hippocampus. This border was most easily visualised on the axial and coronal planes.

2.5.5.8 The complete manually segmented hippocampus

An example of the output from the complete manual segmentation process is shown in figure 2.8 below. This figure shows the right and left hippocampal segmentations overlaid on the T1-weighted 3D image. The image also shows the three 3D mesh rendering of both hippocampi.

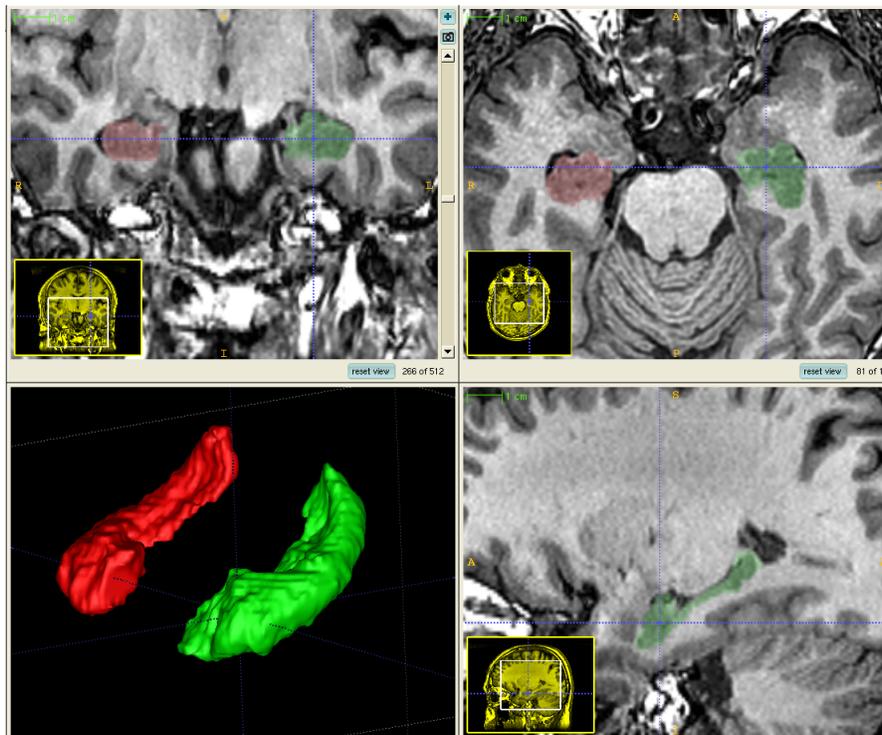


Figure 2.8: Manual segmentation of the right and left hippocampi as delineated using ITK-snap software

2.5.6 Automated segmentation of the hippocampus

2.5.6.1 Background to automated subcortical brain segmentation

At the time of starting this thesis, manual segmentation of the hippocampus was deemed the gold standard method. However, manual segmentation is time consuming not only in the time it takes to perform the segmentation itself, but also in the time it takes to define a protocol, and subsequently train an operator to a high standard. Whole hippocampal manual segmentation performed in this study by the author took approximately 30-45 minutes per hippocampus.

Given the time penalty for performing manual segmentation of the hippocampus, considerable efforts has been put into the development and refinement of automated procedures. Perhaps the most notable work in this area is the work of Fischl et al [123], implemented in the Freesurfer (FS) project (<http://surfer.nmr.mgh.harvard.edu/>). Freesurfer is a widely used software suite and offers many quantitative measures of structural MRI data, including subcortical structure volumes, cortical parcellation volumes and cortical thickness measures. Despite its many features initially FS was found not to be the easiest software to install nor the most intuitive interface to use. Computer processing demands for the FS analysis pipeline are also considerable.

An alternative to FS was to use a newly developed subcortical segmentation algorithm known as FIRST, from the Functional MRI in the brain (FMRIB) group in Oxford, released in 2007 [278]. This group has an extensive software library (FSL) of high calibre MRI analysis tools, the background to which is published in the peer reviewed literature [328].

2.5.6.2 Choosing an automated segmentation method

At the time of starting to perform the image analysis for this study there was little information in the literature comparing the results of FS and FIRST. As FS had been around longer, more data was available on this program and good correspondence between automated and manual hippocampal volumes has been demonstrated [123]. However, as the FIRST algorithm was new, this presented a novel opportunity to test the capabilities of this software. Moreover, given that the FSL software was familiar to the author and deemed easier to use, a decision was made to use FIRST to obtain automated hippocampal volumes which could then be compared to the manual delineated values.

2.5.6.3 How FIRST works

The Bayesian model uses principles of active shape and appearance models, allowing probabilistic relationships between shape and intensity to be utilised. In practice this means that the Bayesian model incorporates within and between structure variability while taking account of the limited size of the training dataset. A further major benefit of this type of model is that it allows shape differences between groups to be assessed, highlighting regions of local volume difference rather than simply the total volume [278].

2.5.6.4 Training data

FIRST uses 336 pairs of images. The original T1-weighted 3D images and their manually labelled counterparts. The training dataset encompasses both normal and pathological brains including data from schizophrenia and Alzheimer's patients. From the set of manually labelled regions, 15 structures were chosen to be modelled, these were: the brainstem, right and left amygdala, hippocampus, caudate, nucleus accumbens, putamen, pallidum and thalamus.

2.5.6.5 Linear image registration

Prior to creating the shape and intensity model FIRST requires that the data is registered to common space. Image registration methods can come in many forms. The options typically offered by brain imaging software are translation only registration (3DOF), rigid-body registration (6DOF), affine registration (12 DOF) and non-linear registration methods (12 DOF or more). For each registration method there are also several cost functions available for use. In FSL options include: correlation ratio, mutual information, normalised mutual information, normalised correlation (intra-modal) and least-squares (intra-modal). Depending on the type of images involved and the circumstances under which registration is to be applied, different options may be chosen. Historically cost functions have generally fallen into one of two categories to solve the image registration problem. Some have used geometrically defined features to quantify the degree of similarity or disparity between the images. Other methods have used techniques based on the intensity of the images. A comparative study between the two types of cost function methodologies found that intensity based methods were generally more accurate and reliable than geometric based methods [368].

For the circumstances here, T1-weighted MRI images were registered to an T1-weighted MRI template. For this purpose an affine, 12DOF model with an intensity based, correlation ratio cost function was used [173, 172]. The non-linear T1-weighted MNI152 template with

voxel size 1mm x 1mm x 1mm was used for this purpose. For image registration FIRST utilises another one of the FSL's tools known as FLIRT, that is, FMRIB's Linear Image Registration Tool [173, 172]. Registration is performed in two stages to achieve the optimal alignment of subcortical structures. Firstly a 12 degrees of freedom (DOF) affine registration is performed for the whole head to the non-linear MNI152 template. The second stage is initialised by the first and then employs a subcortical mask image in MNI space to improve the affine registration. The purpose of this mask is to exclude areas outside the subcortical regions from the registration procedure such that the registration procedure can then focus on the subcortical alignment. Following the two-stage registration procedure to the MNI152 template the inverse transformation is applied to the modelled data to bring it back into native space. Subsequent segmentation steps may now be performed in native space. This is advantageous as the original, non-interpolated voxel intensities can be used.

Following registration any new dataset may be compared to the trained data where the aim is for the similarity function (the posterior probability) to be maximised. This is a measure of the goodness of fit for a proposed segmentation. Further to the posterior probability being maximised a boundary correction is applied to the segmented subcortical volumes. To perform the boundary correction the mesh is converted to a volume which defines a mask for which voxels should be included excluded. A three-class tissue classification method based on the intensities of GM, WM and CSF was determined by Patenaude et al to work best [278].

An example of the right and left hippocampal segmentations generated by FSL's FIRST are shown in figure 2.9 below.

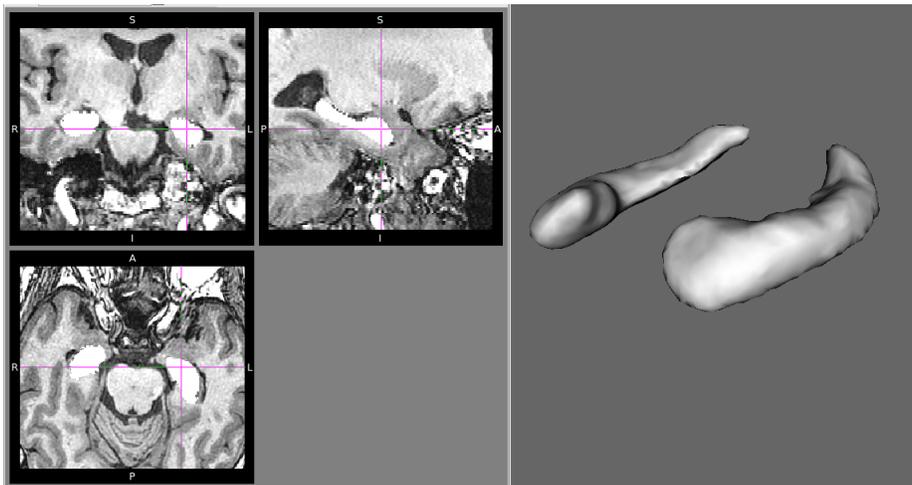


Figure 2.9: FSL's FIRST, example of hippocampal segmentation, hippocampal mask volume overlaid on the T1-W 3D image and 3D rendering of the hippocampal meshes

Following the completion of FIRST, volumes for the segmented hippocampal meshes were obtained for both the poorer and more affluent study groups. Volumes generated from the automated FIRST process were then compared to the manually defined volumes.

2.6 Results

Prior to comparing the manual and automated segmented hippocampal volumes it was necessary to first assess the validity and quality of the manually segmented volumes. Such measures will reflect the comprehensiveness of the manual segmentation protocol and how well the protocol was adopted by the manual raters. Quality was assessed by establishing intra and inter rater Intraclass correlation coefficients (ICC) and Dice metrics, two measurement concepts which are introduced below.

2.6.1 Introducing the intraclass correlation coefficient (ICC)

The most commonly used correlation measure is typically Pearson's correlation coefficient (typically known in the literature as either the r or R value). This is almost universally used to assess the strength of linear relationships between two variables. At first glance the ICC, a descriptive statistic to assess how well measures in groups resemble each other, looks equally straightforward. However, there are several variations of the ICC measurement, calculated in slightly different ways, and having different meanings. Thus one must understand which ICC measure is the most appropriate to apply depending on the specific situation. Furthermore, one must be sure of firstly, which ICC value is being calculated if statistical software is being used, how the ICC value is being calculated, what this means in practice, and finally that when ICC values from the literature are being compared, care must be taken to ensure only similar ICC values are being used. The ICC may be described by equation 2.1.

$$ICC = \text{Variance}(S) / (\text{Variance}(S) + \text{Variance}(R) + \text{Variance}(r)) \quad (2.1)$$

Where variance(S) is the variance due to subjects, variance(R) is the variance due to raters and variance(r) is the residual variance.

The seminal work on ICC was performed by Shrout and Fleiss, 1979 [323]. In this paper the authors detail six different variations on the ICC. The ICC in its most general sense is a reliability or consistency measurement and is the ratio of two variances, these being the variance of the measurement and the sum of the variances of measurement and all other factors contributing to the variance. The reliability measures defined by Shrout and Fleiss are labelled

ICC(1,1), ICC(2,1), ICC(3,1), ICC(1,n), ICC(2,n) and ICC(3,n). The first three measures being the expected reliability of single raters measure and the latter three being the expected reliability of the mean of a set of 'n' raters. Typically it is the ICC (2,1) measure which is desired for most applications, this measure is more simply known as the ICC(Agreement) measure. For the purposes here of comparing within rater and later between rater reliability, the most appropriate ICC measure to use was the ICC(2,1) or ICC (Agreement) measure. This measure is based upon a two-way random effects ANOVA design.

The SPSS statistical software package v18 (SPSS Inc, Chicago, IL. <http://www.spss.com>) can measure various ICC measures depending on the design option chosen. SPSS uses terminology more closely associated with the work of McGraw and Wong, 1996 [230] but their definitions are equivalent to the Shrout and Fleiss descriptions. Thus to obtain the ICC (2,1) measurement one must define a two-way random effects ANOVA design to assess absolute agreement. Thus SPSS not only provides a value for the ICC but also determines if the ICC is different from a given value and if this difference is statistically significant.

A further ICC measure which may be of interest in some instances, is the ICC(3,1) or ICC (consistency) measurement. This is distinct from the ICC measure of agreement in that it treats variability due to raters as irrelevant and thus this variability is not considered in the denominator of the ICC calculation and this leads to an assessment of consistency. The need to include an assessment of consistency, other than for a purely mathematical abstraction, depends on how one wishes to use the data, what one is hoping to prove from the ICC assessment, and how one wishes to proceed with further analysis based upon the initial ICC assessment. For example, in some situations it may be the case that many multiple raters are assessing the subjects, we may not be concerned that raters score the subjects differently as long as their practices are consistent, i.e. systematic differences between raters are not a concern, thus it seems in some instances it is sensible to assess ICC (consistency).

In other circumstances however, the purpose of ICC analysis is to prove that different raters are scoring / measuring in a similar manner (i.e. agreement). Ideally agreement between different raters should be such that their data are interchangeable without issue. In this chapter the aim was to prove that results from within and between different raters agree i.e. that results are not significantly different. Assessing rater quality in this way will in part validate the manual segmentation hippocampal analysis protocol that was developed. Furthermore, ensuring that measures from either of two raters can be treated as being from the same population means that either rater could be used to analyse hippocampal volumes in subsequent studies.

2.6.2 Introducing the Dice metric

In comparing intra or inter rater segmentation results or different segmentation methods it was important to compare more than just the raw volumetric data. For example, two segmentations may be identical in volume but may be considerably different shapes and may also cover different spatial locations. Therefore, comparing the agreement between the volumes alone, as would be given from the ICC measurement, would be an incomplete description of the conformance of the structures. For these reasons the author decided to perform a further measure to assess intra and inter rater volume and displacement differences. This was the Dice metric which determines the overlapping fraction of segmented volumes.

The Dice metric was calculated using an extension to the ITKsnap [377] software known as The Convert3d toolbox (c3d). This tool has a number of useful processing commands which can perform relatively simple mathematical operations on medical imaging data.

The Dice metric itself takes the following form, see equation 2.2.

$$D = V/[(n_{z1} + n_{z2})/2] \quad (2.2)$$

Where D is the Dice similarity metric, V is the number of voxels equal to z in both images, i.e. the overlapping volume/number of voxels and n_{z1} and n_{z2} are the number of voxels in volumes 1 and 2 of intensity value z . Note in our case as binary images were being assessed, then the intensity value of both regions was 1.

A schematic, 2D description of the principle surrounding the Dice metric is shown in figure 2.10 on the following page.

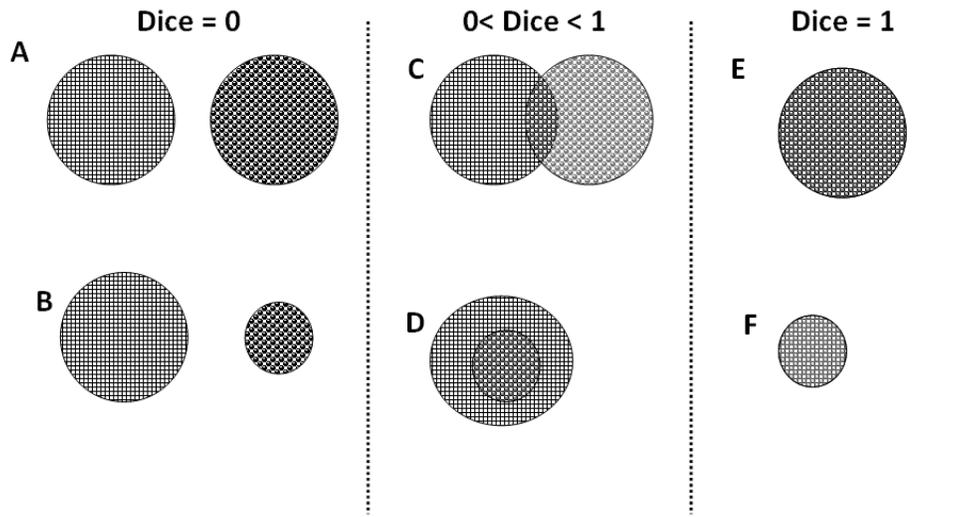


Figure 2.10: Illustration of the Dice metric

In figure 2.10 A and B show that if two volumes do not intersect, regardless of their relative sizes, the dice metric will be zero. C and D show that for volumes that partially intersect and depending on the degree of this intersection, the Dice metric will lie between 0 and 1. For E and F the volumes are equal and the shapes intersect completely, in this circumstance the Dice metric is 1.

2.6.3 Intra operator quality measures: ICC and Dice metric results

To validate the robustness of the manual hippocampal segmentation protocol it was important to assess how similar repeat measurements by a single operator of whole hippocampal volumes were. As discussed above, the Dice metric is a good measure to assess the similarity of volume measures as it considers the degree to which two volumes overlap. Repeat whole hippocampal volume measures were made from 5 randomly selected subjects from the 42 subjects recruited from the imaging study of the PSOBID project.

In addition to the Dice metric the ICC for agreement was also determined. The results of the intra operator assessment for ICC (agreement) and the Dice metric are shown in table 2.2 on the following page.

Quality measures of manually segmented whole hippocampal volumes: Intra-rater agreement and Dice metric		
	Left Hippocampus	Right Hippocampus
ICC (agreement)	0.86	0.85
Dice metric	0.86	0.86

Table 2.2: Whole hippocampal volume measures of quality: Intra-rater agreement and the Dice metric

As is shown in table 2.2 both the ICC and Dice values for intra rater agreement were determined to be greater than 0.85. The Dice intra rater overlap metrics acquired here using the 3T system were comparable to that which has been observed by others using 3T MRI and which have deemed to be of a good standard i.e. greater than 0.75 [174]. This lends confidence to the quality and detail of the manual hippocampus segmentation protocol and in the skill of the main rater. This provided a base on which comparisons between manually segmented volumes may be compared to automatically segmented volumes, justifying the choice of using the manually segmented hippocampal volumes as the gold standard measurement of hippocampal volumes.

2.6.4 Inter operator quality measures: ICC and Dice metric results

To validate the robustness of the manual hippocampal segmentation protocol it was important to assess how similar measurements of whole hippocampal volumes made by two independent operators were. As discussed above, the Dice metric is a good measure to assess the similarity of volume measures as it considers the degree to which two volumes overlap. Whole hippocampal volume measures were made by two raters on 5 randomly selected subjects from the 42 subjects recruited from the imaging study of the PSOBID project.

In addition to the Dice metric the ICC for agreement was also determined. The results of the inter-rater ICC for agreement and Dice metric are shown in table 2.3 below.

Quality measures of manually segmented whole hippocampal volumes: Inter-rater agreement and Dice metric		
	Left Hippocampus	Right Hippocampus
ICC (agreement)	0.90	0.84
Dice metric	0.85	0.85

Table 2.3: Whole hippocampal volume measures of quality: Inter-rater agreement and the Dice metric

With the exception of the ICC (agreement) value for the right hippocampus being 0.84 all other ICC and Dice metric values were determined to be greater than 0.85. Though for the purposes of this study the inter rater measures for the whole hippocampus are less important than obtaining confidence in the intra rater measurement, these measures do serve to show that the manual segmentation protocol was adopted well by a 2nd rater. This implies that the protocol was sufficiently well described and well illustrated to be implemented by someone other than the principle rater and author of this thesis.

2.6.5 Comparing manual and automated hippocampal volume measurements

Having reviewed the quality of the manual segmentation procedure it was then possible to compare automated and manual hippocampal volumes in the more and least deprived groups where the gold standard method for measuring hippocampal volumes was assumed to be manual outlining process. The results of the manual and automated hippocampal volumes measures from 3T MRI were then compared. The mean volumes with 95% confidence intervals as error bars are shown in figure 2.11. Corresponding mean and standard deviation values are shown in table 2.4.

There are various ways in which error bars can be calculated and represented. Therefore it is important to briefly discuss how the error bars were calculated throughout this thesis. Error bars typically represent one of three things, the standard deviation, the standard error, or the 95% confidence interval.

The standard deviation describes a measure of variance around the mean value. Based on the z-distribution an error bar of +/- 1 standard deviation encompasses 66% of the data around the mean. This relationship is described mathematically in equation 2.3.

$$StdDev = \sqrt{Variance} \tag{2.3}$$

Whereas the standard deviation describes a measure of variance around the mean value based on the z-distribution i.e. 95% lie within +/- 1.96 standard deviations , the standard error describes an estimate of certainty in the mean value. As is intuitive, the larger the sample size (n), then the less variability in this measurement and the more certain one can estimate the true mean. Therefore the smaller the standard error value, the smaller the subsequent error bars will be. The relationship between the standard error and standard deviation is described mathematically in equation 2.4.

$$StdError = StdDev/\sqrt{n} \tag{2.4}$$

Finally we consider confidence intervals (CI). A 95% CI tells us that there is a 95% probability that the interval contains the true mean. The relationship between confidence intervals and the standard error is described mathematically in equation 2.5.

$$95\%CI = StdError * t_{df} \tag{2.5}$$

The error bars shown in figure 2.11 are the 95% confidence interval for the mean. Where t_{df} is the degrees of freedom the t-distribution and tends to 1.96 for large sample numbers (>100). 95% confidence intervals will be used consistently throughout this these as error bars.

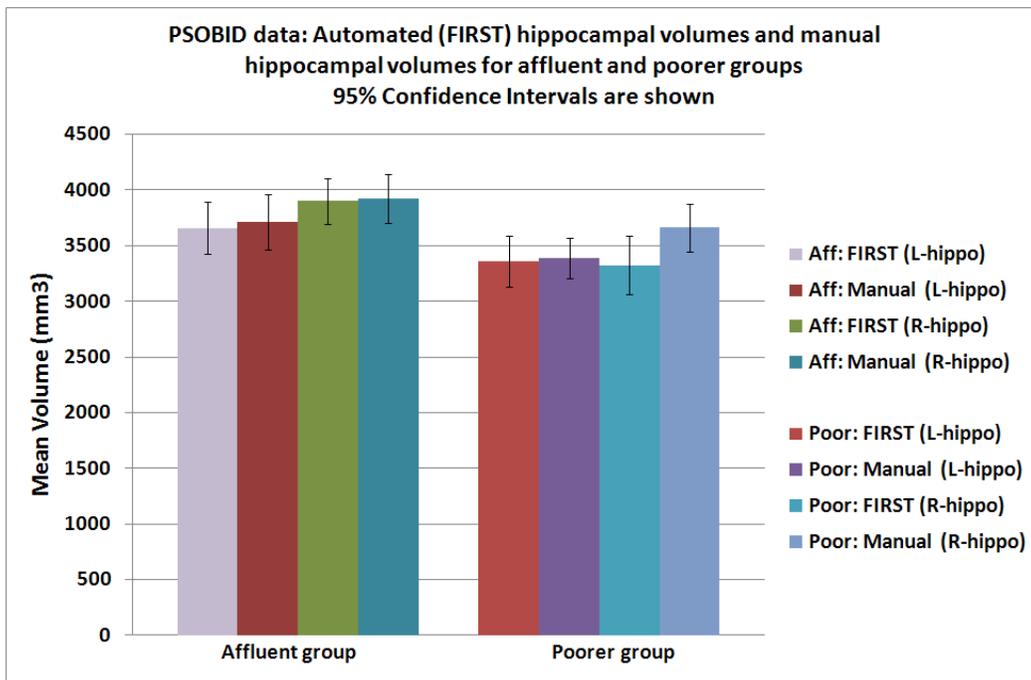


Figure 2.11: Automated and manual volumetry measures of hippocampal volumes for affluent and poorer groups

There are several pieces of information that can be derived from figure 2.11. Firstly, the automatic and manual volumes were comparable and were not significantly different though there was a trend for the FIRST automatically segmented volumes to be smaller than the equivalent manual measurements. This was more evident for the right hippocampal volume measurements than the left. Secondly, comparing right and left hippocampal volume

measurements, there was a trend for left hippocampal volumes to be smaller than right hippocampal volumes, but this was not statistically significant. The mean and standard deviation values are shown in table 2.4 below.

	PSOBID data: mean hippocampal volumes with standard deviations							
	Left		Left		Right		Right	
PSOBID results	FIRST	Std dev	Manual	Std dev	FIRST	Std dev	Manual	Std dev
Affluent	3659	540	3714	572	3900	463	3925	506
Poorer	3360	523	3388	421	3323	604	3661	494

Table 2.4: Automated and manual, left and right hippocampal volumes for affluent and poorer groups

To further compare the two segmentation methods Bland-Altman (BA) difference plots were calculated, see figures 2.12 below and 2.13 on the following page [45]. BA plots are a recognised method for comparing the agreement of two measurement techniques. BA plots are widely used to measure agreement in medical statistics and are distinct from correlation plots which may be misleading as any method developed to measure the same parameter are likely to correlate. That is, the Pearson correlation 'r' value measures the strength of a relationship between two variables, it does not assess the agreement between them.

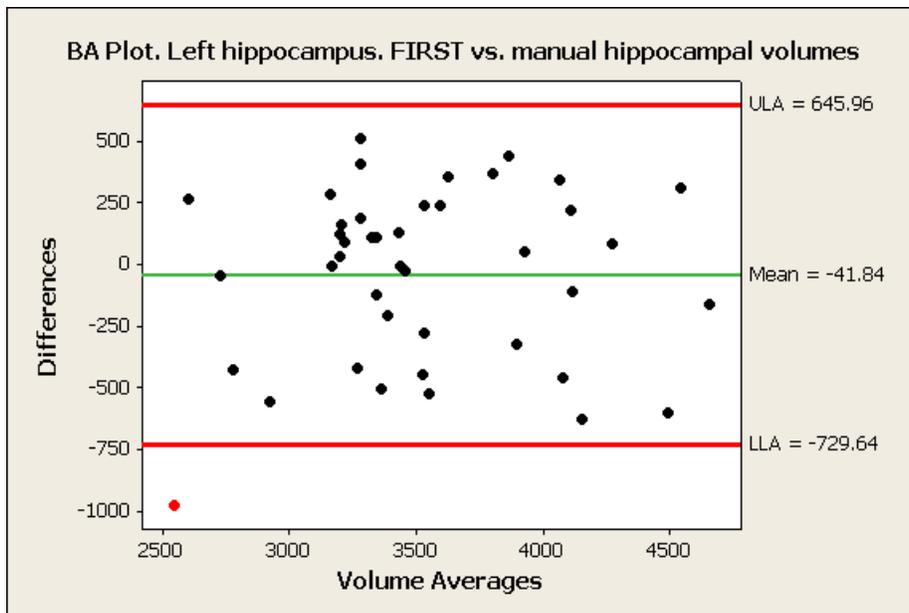


Figure 2.12: Bland-Altman difference plot to compare agreement between FIRST output and manually segmented hippocampal volumes

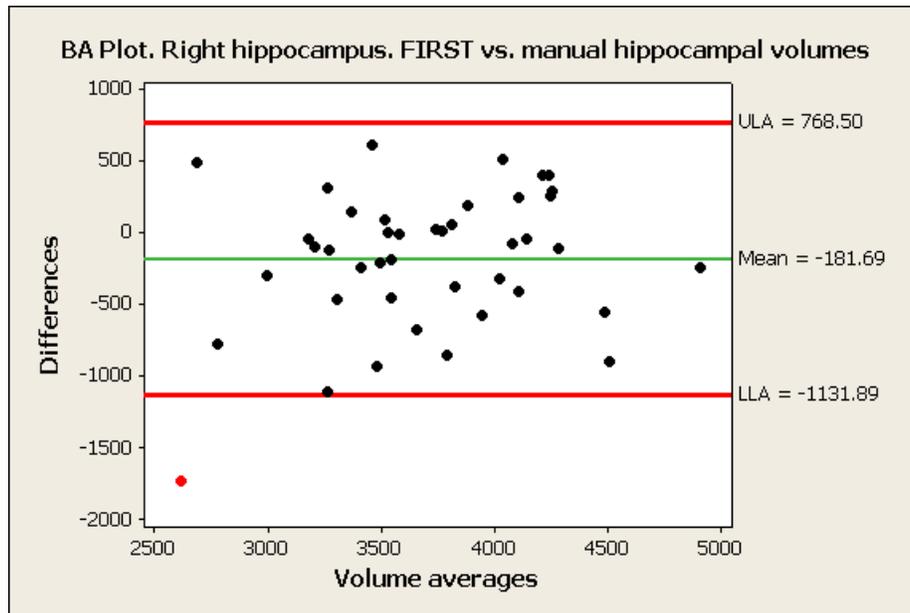


Figure 2.13: Bland-Altman difference plot to compare agreement between FIRST output and manually segmented hippocampal volumes

The BA plots show the difference between the automated and manual volume measurements and the estimated mean difference. On average, automated left hippocampal volumes were -42mm^3 smaller than the equivalent manual segmented volumes and automated right hippocampal volume measurements were -182mm^3 smaller than the equivalent manual segmented volume. The upper and lower levels of agreement (ULA and LLA) are also shown in the figures 2.12 and 2.13. These represent the average difference \pm the standard deviation of the difference.

The BA plots confirm what was suspected from the initial inspection of the mean and standard deviation of the hippocampal volume results from the two methods. That is, the results for the left hippocampus by the manual and automated methods were in much better agreement than for the right hippocampus though both left and right measurements showed a tendency for the automated method to underestimate the hippocampal volume. The BA plots shown in figures 2.12 and 2.13 include data for both the affluent and poorer groups. By separating the groups for analysis, to focus on the right hippocampal data for the poorer group, the mean difference between the automated and manual method was even worse being -338mm^3 i.e. on average the automated measures for the right hippocampus in the poorer group were 338mm^3 smaller than their manually outlined counterparts, see figure 2.14 on the following page. It is worth noting that the equivalent difference between the left hippocampal volumes for the poorer group by the automated and manual methods was 29mm^3 . This is

important as it would imply that one cannot blame issues of image quality for obtaining poor results between different methods. From the BA plot of the right hippocampal volumes of the poorer group (figure 2.14), seven subjects were observed to show a mean difference between the automated and manual methods of greater than 500mm^3 (i.e. in the range: -500mm^3 or lower). This strategy was useful for identifying volumes which may have been poorly segmented by the automated method and thus would inform the user of which volumes are in greatest need of manual correction.

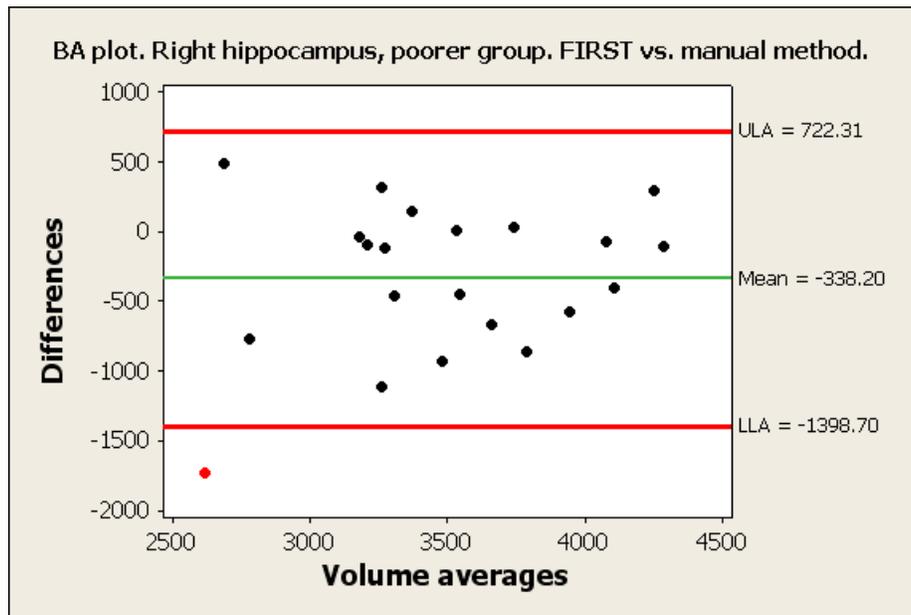


Figure 2.14: Bland-Altman difference plot to compare agreement between FIRST output and manually segmented hippocampal volumes

In practice of course, the purpose of using the automated method was to negate the need for performing manual measurements on the entire dataset and thus the correction strategy discussed here could not be implemented. Similar strategies could be developed to compare the automated volumes with a database of normal data, where one could investigate the outer lying values of the range but this on its own would not necessarily imply that a poor quality automated segmentation had been performed. Ultimately, unless a superior method is developed and found to compare better to the gold standard manual measurements, it would be essential to manually inspect all automated segmentation volumes for errors. Implementing schemes to create and visually inspect large numbers of volumes in a quick and simple manner would greatly benefit this work and all works of this type.

2.7 Comparing the manual and automated volumes using the Dice metric

Hippocampal volumes have been determined by automated and manual segmentation techniques. Before being able to perform the Dice metric calculations, it was first necessary to prepare the automatically segmented hippocampal data. Thus the automatically segmented hippocampal data from FSL's FIRST algorithm were saved as meshes and then corporealised. A consequence of this process was that the new image from the mesh had to be binarised such as to remove any intensity identifiers used by FSL's FIRST program leaving only an image of one's and zero's. It was then possible to calculate the Dice metric to compare the two segmentation methods.

For the 82 hippocampi, i.e. 42 subjects, which were manually outlined and automatically segmented the Dice similarity index was found to be 0.81 +/- 0.05 (standard deviation), results are shown in table 2.5 below.

	DICE SI
Mean	0.81
Std dev	0.05
Std Error	0.01
95% CI	0.01

Table 2.5: PSOBID study: Dice similarity index for manual vs. automated hippocampal volume measurements

2.7.1 Displaying differences between the manual and automated segmentation processes

In the measures applied above automated and manual hippocampal volumes have been compared with respect to the raw volumes and also by the overlapping Dice metric. Including the Dice metric in this assessment is more than many have previously performed yet this still does not give a complete comparison between the methods. While the volume comparison and Dice are good summary measures and reflect the global nature of the measures, they do not provide any illustration as to the location of volumetric differences that exist. In essence, the Dice metric tells us that the manual and automated voxels overlap, on average, by approximately 80% but does not however highlight where in the hippocampus this 80% overlap or more importantly the 20% discrepancy occurs.

The areas of discrepancy could be the result of various factors, namely errors in the manual outlining, or errors in the automated methodology or both. To improve either of these methodological approaches, be it through an improved protocol for manual outlining or through an improved automated algorithm, then non-overlapping regions must first be identified. Therefore, the hippocampal image masks were used to illustrate areas of discrepancy between the automated and manual images.

Comparing the segmented images in native space for a particular subject only generates a subtraction image for a single individual. While this is informative to a degree, it does not necessarily reflect the overall picture of both the automated and manual procedures. Thus it was better to compare the segmented images in template space (MNI152) such that the group subtractions could be performed to illustrate the discrepancy between the automated and manual methods for the entire group.

It has been assumed that the manual segmentation of the hippocampus is more accurate than automated volume measurement and was therefore the gold standard measurement. This is an assumption which was supported by the evidence from the literature. Thus, the graphical representation showing the voxels of volume differences will be considered to focus on regions where the automatic segmentation algorithm was deficient and therefore should ideally be improved.

Having created segmentation images in the same space for both the automatic and manually segmented hippocampal volumes, binary images were then created for each individual segmentation. The rationale to create these so called 'difference images' i.e. the difference between the manual and automatic methodologies was to take the difference of the sum of the masks for both the manual and automatic methodologies. By doing this, regions which consistently overlapped would show no difference, or put another way, images where differences were consistently detected would be highlighted.

As we have determined from the Dice metric that on average 80% of voxels overlap then much of what was left on the difference images will be zero in intensity i.e. no difference. Moreover, the voxels where differences have occurred regularly had a small range. Therefore, in order to visualise the regions of difference in a clear way, a colour scale look-up table (.lut) specific to the limited range of intensity values, was created using ImageJ software (<http://rsbweb.nih.gov/ij/>). The range of values observed included both negative and positive values. Thus the difference image mask was created and overlaid on the T1 152 MNI template brain to localise the regions of difference to an anatomical frame of reference. The differences images highlight regions of difference in hippocampal volume measurements cre-

ated by manual and automatic segmentation methods, examples of the differences images are shown in figures 2.15 below and 2.16 on the following page where these figures focus on regions where the greatest discrepancies occurred.

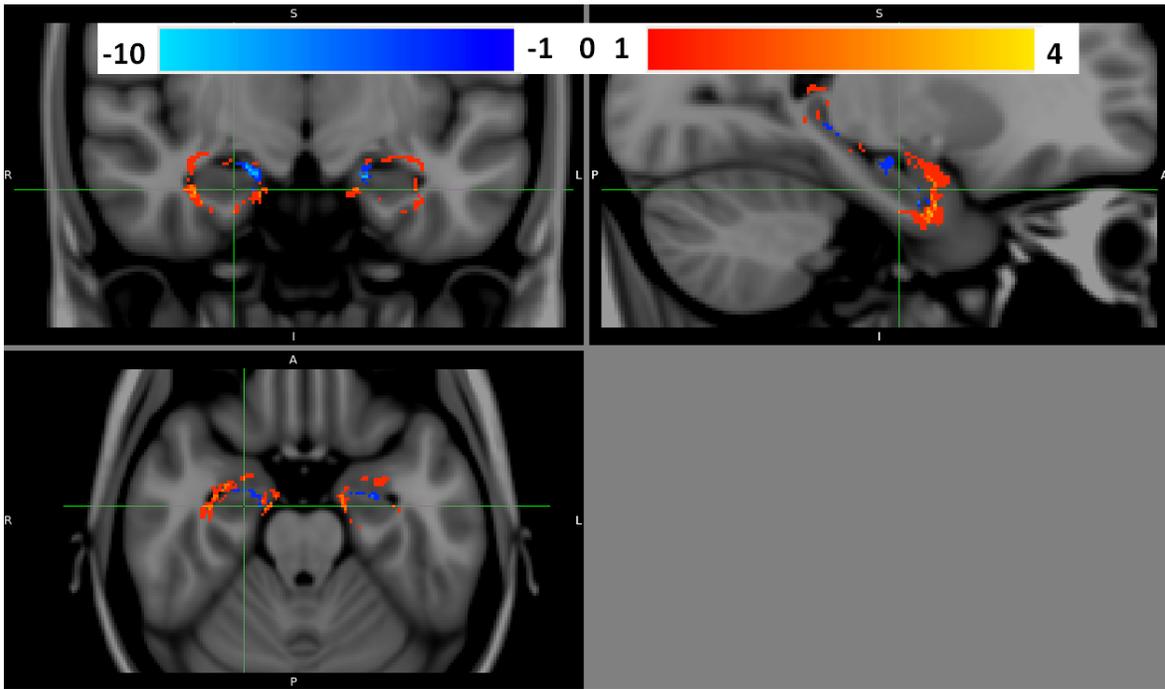


Figure 2.15: Anterior boundary discrepancies (1)

The 'difference images' were derived by subtracting the automatically segmented images from the manual segmented images. The colour scale shown on these figures ranges from -10 to -1 and from 1 to 4, voxel intensity values between -1 and +1 are transparent i.e. no colour is associated with voxels between these intensities. Thus, the 'hotter' or more intense the orange colour on the image, the more often a discrepancy between the automated and manual measurements occurred at that location. These represent the voxels that are greater than zero. The range of positive values was from 1 to 4. That is, the hotter the region the more often voxels at that location were identified as being part of the hippocampus by the manual method than the automatic method i.e. regions that were underestimated by automatic segmentation procedure. From figures 2.15 and 2.16 shown on the following pages such discrepancies tended to occur around the hippocampal head at the anterior and anterior-superior borders between the hippocampus and the amygdala and also at the posterior and posterior-lateral boundary between the hippocampal tail and lateral ventricles.

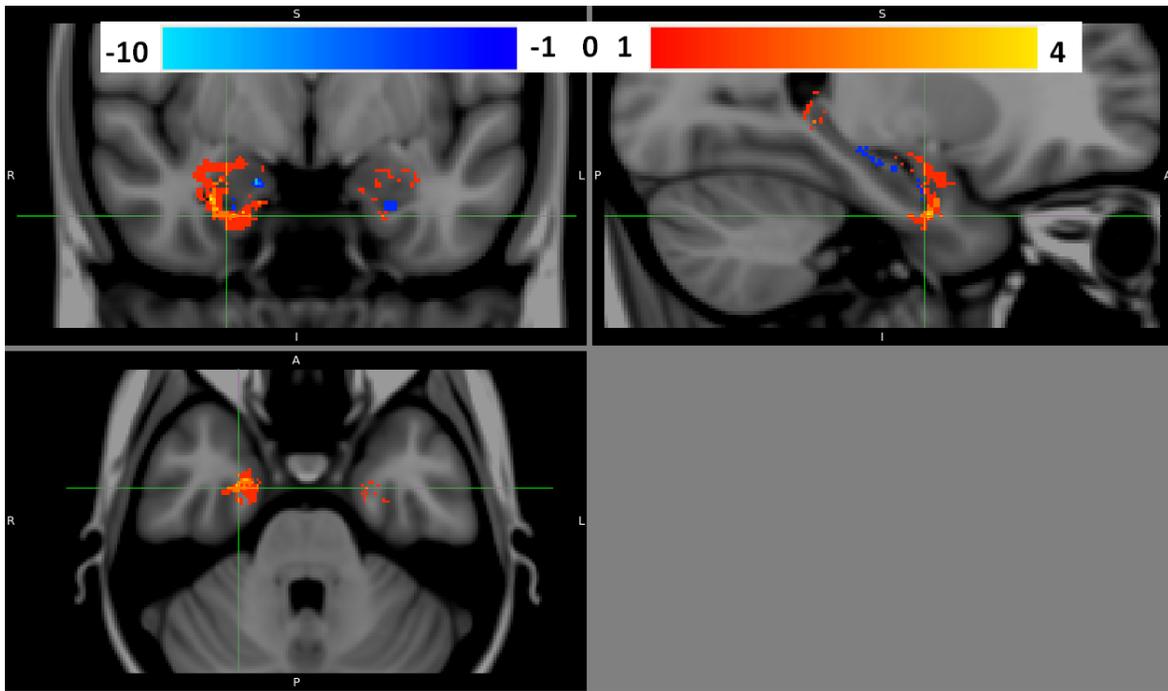


Figure 2.16: Anterior boundary discrepancies (2)

In the difference images shown in figures 2.17 and 2.18 on the following pages, the focus of the figures is on the intensity values in the range from -1 to -10. As can be seen from the colour scale the 'cooler' or more intensely blue the colour the more voxels which have consistently been underestimated by the automatic segmentation method. These represent the voxels that are less than or equal to -1. From figures 2.17 and 2.18, discrepancies of this type tended to occur at the medial boundary between the hippocampus and the brainstem and CSF and at the medial-posterior boundary at the fornix of the hippocampus.

On closer inspection of the difference images one might notice that some of the differences lie outside of the hippocampus. As the difference images were normalised to the T1 152 MNI template one might be surprised to see that not all the discrepancies lie within the hippocampus. The observed discrepancies exist because in some cases there were errors in the initial automatic segmentation procedure, this would include, errors as a result of the fact that the normalisation of the subjects to the MNI template wasn't perfect. Whether the errors were a result of poor normalisation or poor segmentation or both is not clear, but the error as a result of automatic procedure as a whole is precisely what we aimed to illustrate. Regardless of the source of the error in creating these maps, they do serve to illustrate where

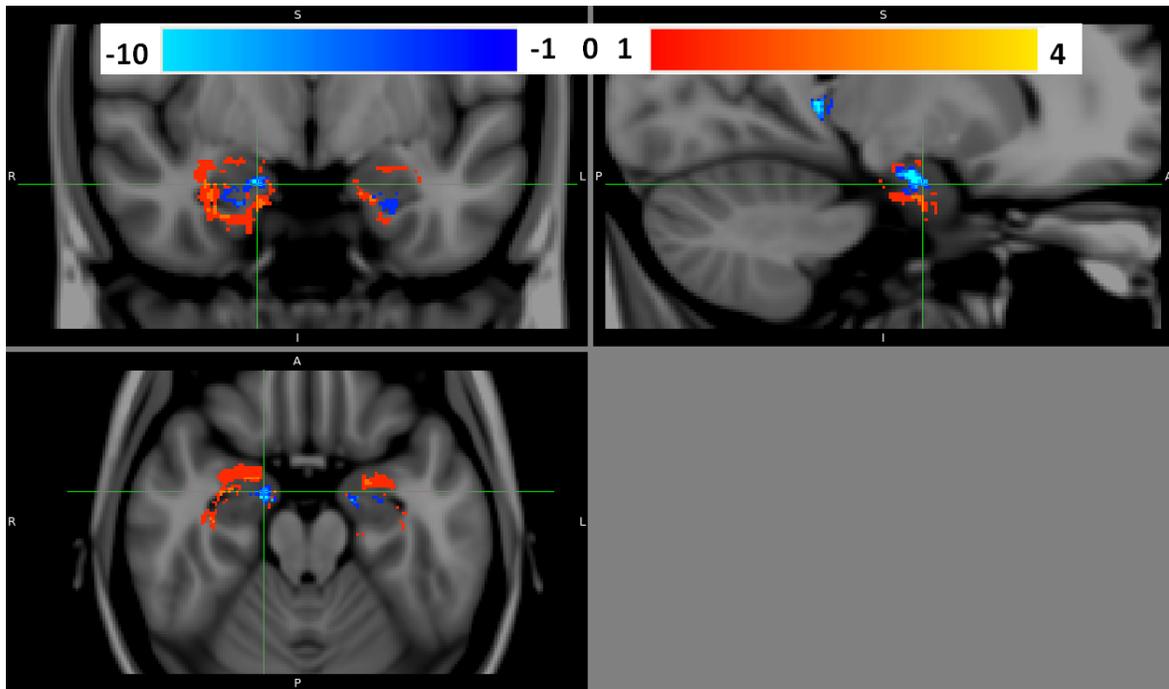


Figure 2.17: Hippocampal head: medial discrepancies

The range of negative values was from -1 to -10.

the differences in the full automatic segmentation analysis pipeline relative to the manual tracing occurs. There may be a component of spatial error in the difference maps as a result of normalisation errors but the difference images are still useful as an indicator of the broad trends as to whereabouts, in and around the hippocampus, errors are likely to occur between manual and automatic segmentation methods.

2.7.2 Establishing the mean and standard deviation of hippocampal volume measures in the poorer and more affluent groups

The mean and standard deviation of the affluent and poorer groups was established. Given that the gold standard method for hippocampal volumes was the manual method, these values were considered initially. The results of the manual segmentations for the left and right hippocampal volumes are shown in table 2.6 below.

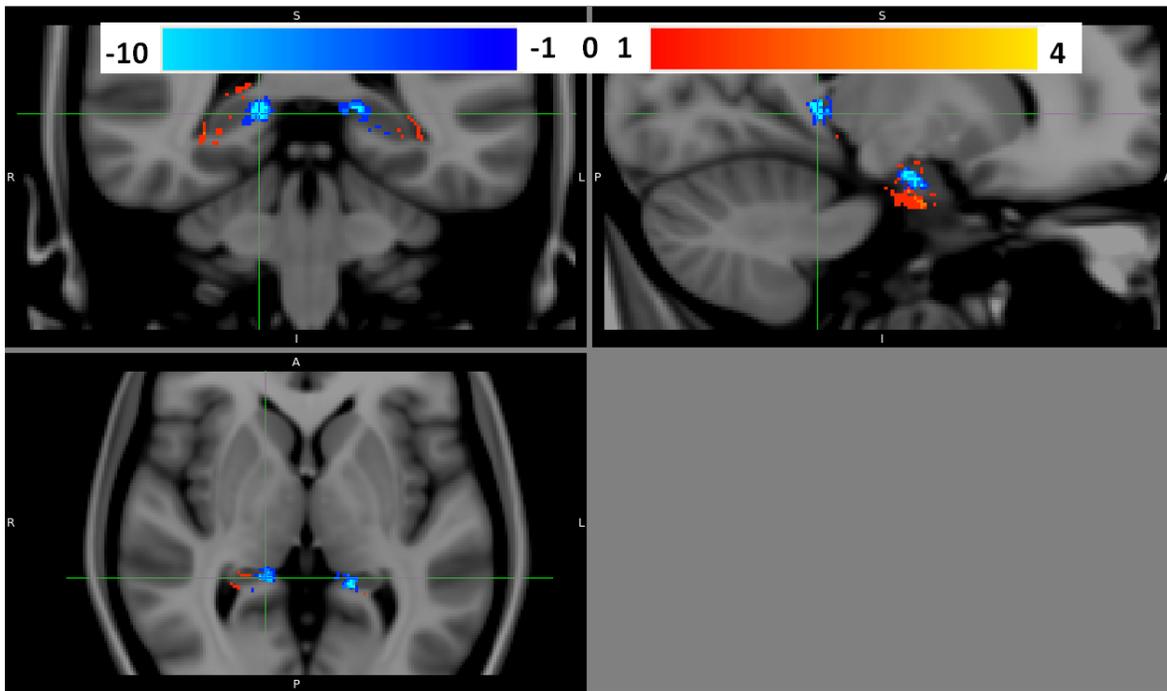


Figure 2.18: Fornix discrepancies

PSOBID data: mean hippocampal volumes & standard deviations				
	Left		Right	
PSOBID results	Manual	Std dev	Manual	Std dev
Affluent	3714	572	3925	506
Poorer	3388	421	3661	494

Table 2.6: Results of manually segmented hippocampal volumes for poorer and more affluent groups

The results of the manual volumetry suggest that both the left and right hippocampal volumes are smaller in the poorer group. At this stage it was important to consider the intracranial volume (ICV) of the subjects when comparing subcortical structure volumes. ICV values were obtained using SPM5's segmentation routine, this will be discussed in more detail in chapter 3 where the ICV values were determined as part of the VBM processing pipeline.

The raw manually segmented hippocampal volumes were corrected for ICV, the corrected results are shown in table 2.7 on the following page.

PSOBID data: ICV corrected mean hippocampal volumes				
	Left		Right	
PSOBID results	Manual	Std dev	Manual	Std dev
Affluent	0.231	0.039	0.244	0.036
Poorer	0.216	0.039	0.233	0.040

Table 2.7: PSOBID data: ICV corrected, manually segmented hippocampal volumes presented as a % of ICV

The ICV corrected volumes reflected the trend observed for the uncorrected volumes in that both the left and right hippocampal volumes were smaller in the poorer group.

2.7.2.1 Power calculations for hippocampal volume data acquired using 3T MRI

An important aspect in defining any experimental design is to estimate the necessary number of subjects to obtain a statistically significant result between two group measurements. This is known as determining the power of the study and depends on the variability in the measurement, and the estimated minimum difference between the measurements. Power defines the likelihood or probability of recording a perceived difference between two groups. Therefore, for a desired degree of power, one can calculate the required sample size necessary to obtain statistical significance given a known (or estimated) standard deviation and an known (or estimated) minimum difference. To put it another way, given two of either the sample size, standard deviation and minimum difference the third parameter can be calculated for a desired degree of power.

Power has come to prominence in recent years and is typically desired in grant applications, presumably (amongst other things) to avoid studies being performed with such low power that they are unlikely to record statistical significance. It is common for pilot studies to be used as a vehicle to obtain estimates of standard deviation and the minimum difference between groups which can be used to obtain estimates of sample size for a given degree of power, alternatively one can use the peer reviewed literature from similar studies to obtain these measurements to estimate power. Thus an important aspect in using any new or improved technology or methodology is to consider the variance in normality. For the purposes of the work in this thesis the evolving technology is that of 3T MRI and we may speculate that the improved field strength may lead to an improved accuracy in measuring hippocampal volumes. Where an automated segmentation algorithm is being applied then we are considering the variance in the combination of these technologies. Power calculations were performed using the statistical software package Minitab v.15.1, Minitab Inc.

For the two groups discussed in this chapter power and sample size calculations were performed by Minitab using a 2 sample t-test method as shown in figure 2.19. The input values

for estimations of sample sizes must be integer values. In performing power or sample sizes we aim to compare the means of two sample populations where the samples forming the group are assumed to be representative of the underlying population being investigated. Minitab will only perform power and sample size calculations when the groups are populated with equal numbers of subjects. By default, the null hypothesis for 2 sample t-test is that there is no difference between the means of the sample populations. Thus there are three remaining options for the alternative hypothesis, these are, that the groups are not equal, that the first group has a greater mean than the second or that the first group has a lower mean than the second. The alternative hypothesis that is chosen will depend on what is required to be assessed. As sample sizes must be integers, target power values may not be found exactly however Minitab will display the actual power for a requested sample size.

For comparing the hippocampal volumes between the affluent and poorer groups a two-sample t-test was performed. The minimum difference was defined as the difference between the sample means of the affluent and poorer hippocampal volumes as determined by the automated segmentation algorithm and this was recorded as 299mm^3 . The standard deviation of the measurement (averaged between the two groups) was 538mm^3 . The affluent volumes were larger than the poorer hippocampal volumes and thus this allows a directionality to be applied to our power calculation. That is, the power and sample size estimation here was based on the alternative hypothesis that the mean of the affluent group was larger than the mean of the lower SES group. This calculation informs further hypotheses aiming to detect significant hippocampal volume differences in studies with subjects of differing socioeconomic status. Given the observed group difference and variance, for power values of: 0.5, 0.6, 0.7, 0.8, 0.9 and 0.95, necessary sample sizes were calculated as: 19, 25, 32, 41, 57 and 71 respectively. This data is summarised in figure 2.19 on the following page.

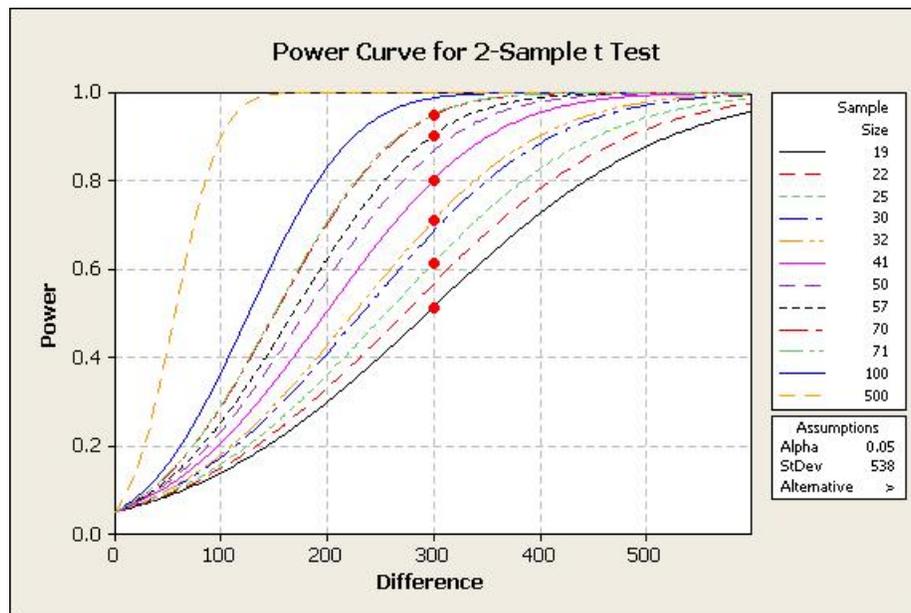


Figure 2.19: Power calculation for automatically segmented hippocampal volume difference at $3T$

Based on the data from the PSOBID study, to determine a statistically significant difference in hippocampal volume between the poor and affluent groups with a power of 90%, the necessary sample size was estimated as requiring 57 subjects in each study group.

2.7.3 Hippocampal volumes in relation to IL-6 and cortisol

To investigate the relationship between the IL-6 inflammatory marker and ICV corrected hippocampal volumes between the two SES groups scatterplots were created. For IL6 an outlying value of 25 was removed to avoid skewing the correlational analysis. The results are plotted in figure 2.20 on the following page.

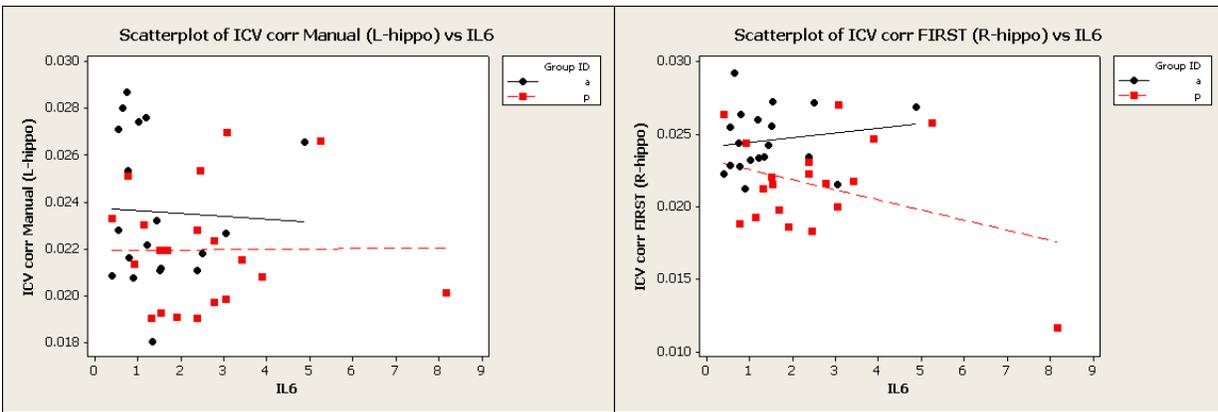


Figure 2.20: IL6 vs. ICV corrected left and right hippocampal volumes for affluent and poorer groups

Correlations were also performed to assess the strength of the linear relationship between the variables. No significant correlations existed between IL-6 and ICV corrected hippocampal volume. For both affluent and poorer groups none of the correlations with IL-6 were statistically significant.

Similarly the relationship between cortisol and hippocampal volumes were assessed. The results are shown in figure 2.21 below.

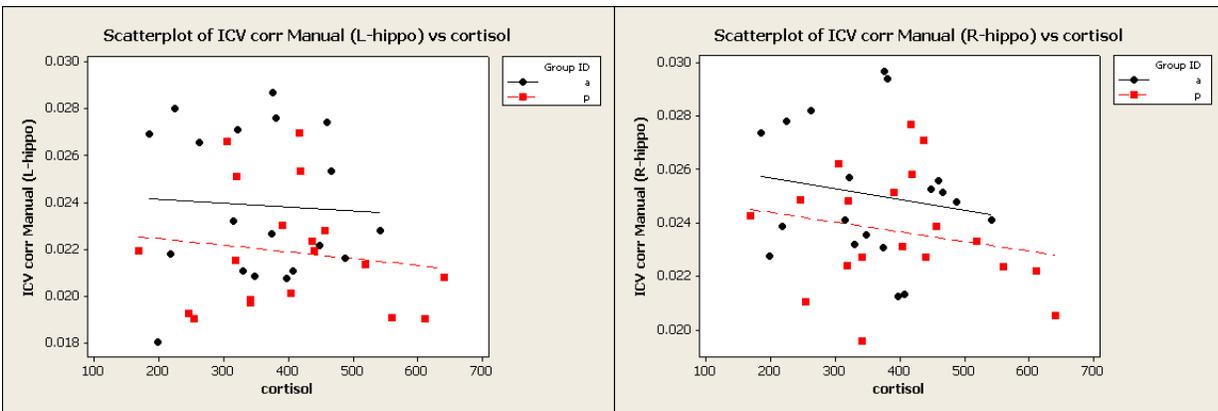


Figure 2.21: Cortisol vs. ICV corrected left and right hippocampal volumes for affluent and poorer groups

Correlations between the cortisol and hippocampal volume were calculated. For both the affluent and poorer group no significant correlations were observed for either the ICV corrected right or left hippocampal volumes.

2.8 Discussion

2.8.1 Assessing the quality of the manually segmented hippocampal volumes

While the primary aim of this chapter was to compare the manually delineated hippocampal volumes with the automatically segmented volumes from the FIRST algorithm it was important to establish the reliability of the manual segmentation procedure. To address this, Intraclass correlation coefficient (ICC) values for agreement and Dice metric values were calculated on the manually segmented hippocampal volumes for both intra and inter rater scenarios.

For both the intra and inter rater assessments two measurements were made on 5 randomly selected subjects from the larger PSOBID dataset. Both the left and right hippocampi were manually segmented. For the intra rater scenario the ICC (agreement) for the left and right hippocampi were 0.86 and 0.85 respectively. The Dice metric for the left and right hippocampi were 0.86 and 0.86 respectively. For the inter rater scenario the ICC (agreement) for the left and right hippocampi were 0.90 and 0.84 respectively. The Dice metric for the left and right hippocampi were 0.85 and 0.85 respectively.

From these results we can say that the agreement between repeat measurements is similar for both the intra rater and inter rater cases. That is, repeat measurements made by the first rater are similar to the volume measurements between the two raters. This suggests that the second rater adopted the manual segmentation protocol well, producing segmented volumes which were similar in both size and position to the first rater. A good degree of agreement, as has been achieved here, for the manually segmented hippocampal volumes across different raters implies that the two rater results may be used exchangeably. This may have practical implications in that if future work requires manually segmented volumes on larger numbers of subjects then the dataset could be split between the two raters having negated concerns of introducing rater bias.

The inter rater agreement and Dice metric assessment is of some interest in that it informs us to a degree about how well the manual segmentation protocol has been presented in terms of its clarity such that similar volumes may be reproduced by different raters. However, in the broader context of this chapter, it is less important as the focus of the work was to assess the automated volumes relative to the manual segmented volumes as delineated by a single

rater (the author of the thesis). In this respect, the intra rater agreement measures are of greater importance.

The reliability of hippocampal volume measurements at 3T have previously been discussed by Jeukens et al [174]. In this paper they set the acceptable criteria for ICC agreement measures as being greater than or equal to 0.85 while for the Dice overlap metric they set as being greater than or equal to 0.75. ICC for hippocampal volumes have also been reported elsewhere in the literature and these values range from 0.75 to 0.95 [15, 202, 330, 274, 366]. In this thesis we have demonstrated that both intra rater and inter rater ICC measures of agreement greater than 0.85 are achievable, where we only failed to meet this criteria for the inter rater agreement measurements of the right hippocampus. For the purposes of this thesis where the intra rater ICC results are of greater importance the values of 0.86 and 0.85 for the left and right hippocampus respectively are acceptable and comparable to the best of the intra rater values cited by Jeukens et al of 0.88 [174].

As was discussed when the Dice metric was introduced the ICC measure only considers the size of the structure and not the relative space the volume inhabits it can potentially be misleading to present ICC values on their own. Jeukens et al, also state this fact where for one of their intra-rater measurements they observed an ICC agreement value of 0.59 but deemed that reproducibility was still reasonable as the Dice metric value for the same intra rater assessment was determined to be 0.87 [174]. Thus they recommend that ICC values of manually segmented hippocampal volumes should always be presented in combination with measures of the Dice overlap metric. The intra rater Dice metric values observed in this chapter of 0.86 for both the left and right hippocampi were comparable to values observed by Jeukens et al, and better than the acceptable criteria of 0.75 as set by Jeukens et al [174]. Dice metrics are not often cited, where even those studies which do include it often only consider the inter rater assessment [274, 73], though some studies have assessed intra rater Dice measures [154]. The first of these studies were performed at 1.5T where inter rater Dice measures were observed to be poorer than that which was recorded in this thesis [274, 73] though the improvement may not necessarily be solely attributable to the increased field strength where intra rater measures acquired using 1.5T MRI systems have been obtained which are slightly better (between 0.87 and 0.90) than the values acquired in this thesis (0.86) [154]. Differences in the Dice metrics may be attributed to a number of factors which could include field strength, rater skill, protocol clarity and choice of what parts of the hippocampus constitute the volume to be segmented.

To summarise, the Dice intra rater overlap metrics acquired here using the 3T system were comparable to that which has been observed by others using 3T MRI and which have deemed

to be of a good standard i.e. greater than 0.75 [174]. This lends confidence to the quality and detail of the manual hippocampus segmentation protocol and in the skill of the main rater. This provided a base on which comparisons between manually segmented volumes may be compared to automatically segmented volumes, justifying the choice of using the manually segmented hippocampal volumes as the gold standard measurement of hippocampal volumes.

2.8.2 Comparing automated and manual measurements made using 3T MRI

Given that performing manual outlining of hippocampal volumes has a considerable time burden, it was worth considering whether automated methods of hippocampal volumes could be used in their place. However, in order to have confidence in using the automated data, a comprehensive comparison between the automated and manual data was performed.

With perhaps the exception of the right hippocampal volumes in the poorer group, FIRST's automated measurements returned similar mean volumes to the manual method for both the left and right hippocampal volumes. The levels of variance in the measurements were also comparable between methods. This initial inspection of the mean volumes and the spread of the data are encouraging. A BA analysis was also applied to the data. The mean difference between the two methods for obtaining left and right hippocampal volumes was -42mm^3 and -182mm^3 . Therefore in both instances the automated volumes were smaller than the manually segmented regions with the methods having considerably better agreement for the left hippocampal then the right. Given the mean volume of the hippocampus was approximately 3700mm^3 having a mean difference of 42mm^3 and 182mm^3 corresponds to a percentage error of between approximately 1% and 5%. These are modest values.

To further assess the quality of the conformance between the two methods Dice overlapping metrics were calculated. The mean Dice was 0.81, where values of around 0.8 or better are typically considered reasonable. A comparison between automated measures from Freesurfer (FS) and FIRST to manual tracing was performed by Morey et al [246]. They calculated a Dice metric of ~ 0.82 and ~ 0.82 for the left and right hippocampal volume, comparing FS to manual tracing, while comparing FIRST to manual tracing was slightly poorer with Dice metric of ~ 0.78 and 0.79 . Therefore, the Dice metric values calculated in this thesis are comparable but slightly better than that observed for FIRST in Morey et al, and only slightly poorer than the Dice metric achieved by comparing FS to manual segmentation. The publication detailing the FIRST methodology itself contains Dice metric which were ~ 0.81 for the hippocampus [278]. The original FS subcortical segmentation publication by Fischl

et al reported overlap measures of ~ 0.80 for the hippocampus. In this regard the values calculated here are in keeping with what is presented in the literature both for FIRST and for other automated methods such as FS [123].

In another work assessing automated hippocampal segmentation in temporal lobe epilepsy (TLE) patients both FS and FIRST were compared to manual segmentation [275]. They calculated a Dice metric of 0.73 between FS and manual tracings and 0.71 between FIRST and manual tracings in control subjects and 0.66 between FS and manual tracings and 0.62 between FIRST and manual tracings in TLE patients. These values are lower than was observed in this thesis and also by Morey et al and could be a result of differences in the manual tracing protocol that was used [246].

To further augment the volume, variance and Dice summary measures discussed above it was deemed useful to create 'difference images' to highlight regions where discrepancies between the methodologies consistently occurred. This proved an insightful comparison as these images are not something which are commonly documented in the literature, with one exception to this being the work performed by Morey et al [245]. In this thesis it was observed that the aspects of the hippocampus where discrepancies between the automated and manual methods occurred were at the anterior and anterior-superior portions of the hippocampus. Similarly, deficits were also frequently observed around the tail of the hippocampus where it meets the fornix. In a similar way Morey et al exploited a vertexwise analysis to display regions of volume difference on a 3D surface mesh with vectors highlighting the direction and significance of any discrepancies. The results of this work match what was observed in this thesis, with the largest errors being at the anterior boundary, medial surface of the hippocampal head and the medial aspects of the tail where good agreement was observed around the lateral and superior aspects of the hippocampal body and tail. In some ways these observations are not a great surprise as typically the experience of segmenting the volume will inform the operator of the challenges faced and difficult aspects to segment. Nonetheless, the data is useful to confirm these ideas and to again highlight regions of common errors in hippocampal segmentation. It would be advisable for future work to focus on these challenging areas and perhaps to refine the automated methods to improve the segmentation in these regions, this will likely improve the conformance of the automated methods to the manual operators.

Despite the relatively good agreement between the automated and manual methods there was still situations where the automated segmentation failed to accurately segment the hippocampal volume. Therefore it is recommended that in future work the automated methodology should be accompanied by a manual inspection and correction of the segmented volume. This would take advantage of the speed of the automated method while negating the need

for a manual operator to spend time segmenting the whole hippocampus. By also including a manual refinement to the automatically segmented hippocampus the overall accuracy of the final hippocampal segmentation will be improved.

The power calculation performed here was informative for a number of reasons. The variance in both the automated and manual methodologies has been obtained, this would be useful when performing sample size calculation using these methodologies. Mean differences between hippocampal volumes of distinct groups of differing SES have also been obtained. While this is not as useful as the variance in the measurement in that it is specific to the groups assessed here, it would inform future studies assessing the hippocampus in the area of deprivation and SES. At the time of writing this study, to the best knowledge of the author, no other study assessing the statistical power of adult human hippocampal volumes in deprivation has been performed. To recap, for the group differences and observed variance of the automated measurement of the hippocampus, to achieve a study power of 0.9, approximately 57 subjects would be required in each distinct group with differing SES.

Considering the methodology alone, Morey et al considered the power of FS, FIRST and manual segmentation to detect a range of volumetric differences for varying sample sizes based on a desired study power of 0.8 [246]. By their measurements they estimated slightly lower subject numbers were necessary to achieve the same power when manual and FS methods were used compared to FIRST. Their estimates for power were comparable to what was observed in this thesis where ~ 40 subjects would be deemed necessary to achieve a power of 0.8 if a group difference of 300mm^3 was observed.

2.8.3 Hippocampal volumes in the more and least deprived SES groups: discussion of results and limitations of this study

To address the main hypothesis of this chapter a detailed protocol was defined to manually segment the hippocampus in 21 subjects in a more deprived group (low SES) and 21 subjects in less deprived group (higher SES). These resultant volumes were then compared to automatically segmented hippocampal volumes from the same two groups. In future work it will be necessary to compare in a robust manner hippocampal volumes between lower and higher SES groups. The results and limitations of the current pilot study are discussed below with a view to highlighting improvements which would be beneficial in further work in this area.

Where low socioeconomic status (SES) was considered as a chronic stressor, it was observed that there was a trend for the more deprived (lower SES) group to have smaller hippocampal

volumes than the higher SES, more affluent group. At first glance the results here would add weight to the evidence already observed that environmental stressors have led to reduced hippocampal volumes [336, 57]. Of novelty in this thesis was the fact that these differences were observed in a community based sample, as opposed to institutionalised individuals or those subjects who have experienced specific traumatic events. While these results are encouraging it is noted that given the pilot nature of this study and the subsequent suboptimal power in the design these results must be treated tentatively. Having established the methodology it was then possible to estimate the sample size that would enable greater confidence in comparing hippocampal volumes between two groups of different SES. It was determined that approximately 57 subjects would be required in each separate group if the study were constructed in a 2 sample t-test design to achieve a power of 90%.

In perhaps the largest study of its kind relating to the effects of SES on hippocampal volumes 317 subjects were recruited via the National Institute for healths (NIH) normal brain development database [156]. In this study the parents income was used as the principle factor determining the child's SES. Hanson et al found that children from lower income backgrounds had lower hippocampal grey matter density [156]. While assessing income is a more simplistic measure of SES than the multi-faceted measures that have been applied in some studies it does make the interpretation of the results more straightforward. Moreover, the large sample size examined lends additional weight to the findings. A further benefit of the Hanson et al study is that the imaging data has been acquired from children and therefore the impact of adult stress factor need not be considered. This is a particularly good study and highlights the benefit that can be drawn from establishing large, central archives of data. It is likely that in future more research by such methods will be necessary to address the ever increasing subtlety and refinements of disease processes.

The majority of other studies examining the effects of stress on the hippocampus typically have fewer numbers of subjects. Much of this work has focused on the effects of PTSD either from physical or sexual abuse [58] or from military combat [57, 148] where the subject numbers in these studies were 17 and 26 respectively with both studies suggesting hippocampal volumes are smaller in the group exposed to stressful life events. In another study, perhaps the only of its kind to assess chronic life stress in subjects without a clinical syndrome, perceived stress scores, acquired over a 20 years period, predicted decreased grey matter volume in the right orbitofrontal cortex (OFC) and in the right hippocampus [134]. This study was performed on 48, healthy, post menopausal women and implies that even in healthy individuals, chronic life stress may be responsible for morphological changes in the brain, and in particular in this case the right OFC and right hippocampus. Thus there are few studies

of large numbers assessing differences in brain volumes and the hippocampus with differing SES. This supports the need for the pilot work performed in this thesis to be extended into a more refined and larger study, where this thesis has laid the foundation for any subsequent neuroimaging analysis that would need to be performed.

While it was important to discuss the results of this chapter within the context of the peer-reviewed literature it is also necessary to highlight many study design factors, recruitment and screening which may have limited the results of the PSOBID pilot imaging study. From the experience of performing this pilot study there are several confounding factors in respect of the study design which if a further larger study were to be performed should be amended. As stated in the methods section, subjects completed a GHQ questionnaire [165, 270]. The GHQ-12 results for both the least and most deprived groups were assessed following imaging. This situation is less than ideal if one wished to enforce robust control for the degree of mental illness between the two groups. Firstly, if one had truly wished to screen for mental illness and in particular, depression, this should have been performed prior to imaging such as to avoid acquiring potentially confounding data and wasting valuable imaging time and staff resources in the process. Moreover, the GHQ-12 for depression is not the most commonly used or likely to be the optimum tool for assessing depression. More explicit methods include the Hamilton depression scale [153] and the Structured Clinical Interview for Diagnostic and Statistical Manual for mental disorders (SCID: DSM-IV) [122]. Ultimately, if one wished to hone in on brain differences between a more deprived and least deprived group then ideally subjects should be screened out prior to imaging for factors such as depression and alcohol. Alternatively, the two groups could be matched in terms of these factors such as to control for these factors in a more prospective manner.

Similarly to the discussion above surrounding the recording and control of the likelihood of psychosis in the sampled subjects, levels of alcohol were assessed following acquisition of the imaging data. Once again, if further work is to be performed in a larger study with similar aims to build upon the pilot data acquired here, it would be preferable to control the alcohol of subjects under investigation in a more prospective manner. One method would be to match subjects for imaging between the more and least deprived groups in terms of the weekly number of units of alcohol they consume. This could include screening out subjects with excessive alcohol consumption. Matching the groups in this way would negate the need to include alcohol measures as a covariate in later statistical analysis. Of course on a more practical note, imposing an increasingly strict inclusion criteria may have implications for the numbers of subjects that can be recruited so in some circumstances, in order to achieve the desired number of study subjects one may not wish to be too prescriptive. However,

given the large number of subjects from the PSOBID study who opted-in for imaging, it would probably have been possible to be more strict regarding the inclusion criteria. This is something that should be addressed in further work.

Finally, in partitioning the groups using SIMD index, the educational level or intelligence of the subjects was not explicitly accounted for. It has been shown in other work that intelligence correlates with hippocampal volume [318]. In the PSOBID study the NART was used a proxy measure for subject IQ following a retrospective analysis of the data there was a trend for subjects in the lower SES group to have poorer NART scores. Thus in future work the role which intelligence plays as part of SES and its impact on the hippocampus will require careful consideration by a person suitably experienced with this aspect of the literature, most likely a psychologist or psychiatrist. This is non-trivial given the different types and definitions of intelligence and the various tests which exist to assess the different aspects of intelligence. In the SES literature, a parameter often assessed to account for educational level is the given number of years a subject was in full time education i.e. the total number of school/college years [149]. The issues of dealing with NART or indeed any measure of intelligence or education as a component of SES is challenging not only with respect to selection and testing of the subjects but also in determining the appropriate analysis method. That is, one might simply wish to partition two differing SES groups by some method which includes educational level and assess the outcome in broad terms, however, the challenge comes when aiming to apportion the relative effects of the various components of SES. One might wish to investigate factors via a multiple regression analysis to assess the relative amount of variance in the design that might be explained by the different SES factors. Alternatively, one might wish to assess the degree to which a particular factor modulates or mediates a given outcome measure. Whatever the factors for consideration or outcome measures are it is strongly advisable that in future studies such discussions and decisions regarding analysis are made prior to performing the study and preferably with considerable input from a statistician.

2.8.4 Hippocampal volumes and inflammatory markers: discussion of results and limitations of this study

Having tentatively observed a trend of smaller hippocampal volumes in the lower SES group we also aimed to explore which factors might mediate this relationship. Is it possible that the impact of SES as a stressor in the poorer group has indeed led to smaller hippocampal volumes or are there other factors at work? While a full statistical treatment of the many factors involved in teasing apart the proposed relationships was beyond the scope of this

thesis, an overview of some of the key considerations combined with some post-hoc analysis are presented to direct further, more comprehensive analysis.

Given the proposed mechanism for a reduction in hippocampal volume as a result of an inflammatory pathway, correlation analysis was performed between hippocampal volume and two key substrates of inflammation: IL-6 and cortisol. Neither IL-6 or cortisol showed a significant correlation with ICV corrected left and right hippocampal volumes. Thus it was not possible to replicate the results of Marsland et al, who had shown a correlation between IL6 and left hippocampal volume in normal subjects [216]. Similarly, it was not possible to replicate the results of Pruessner et al, who had, in a study assessing hippocampal volumes, memory and cortisol in young healthy volunteers, shown that cortisol levels correlated with right-sided hippocampal volumes [291].

There may be a number of reasons why the results in this chapter failed to replicate that which was observed by Pruessner et al [291]. It could simply be that the data are subject to a type 2 error and that there were not enough subjects to detect the trend in the data. Moreover, cortisol, as it can be measured in a number of ways (assayed from saliva, blood plasma or urine) and can be broken down into further subcomponents of cortisol has already resulted in conflicting results throughout the literature [109]. The relative scientific merits of the different methods of cortisol measurement is still under debate.

Dowd et al conducted a review of twenty-one peer reviewed publications surrounding cortisol, SES and allostatic load [109]. Of the twenty-one studies reporting associations between SES and cortisol, two analysed serum cortisol, five analysed urinary cortisol and fourteen analysed salivary cortisol. Of the twenty-one papers, seven reported a significant association between lower levels of SES and higher cortisol [18, 119, 120, 198, 87, 88, 195], four found mixed results [199, 339, 372, 190], eight studies found no relationship between SES and cortisol [140, 108, 132, 302, 105, 303, 295, 338] and two studies found a relationship between lower SES and lower cortisol [55, 82]. Thus there is a large variation in reported results of the relationship between cortisol and SES. To further increase the disparity in these measures, there does not appear to be a trend associated between the method used to measure cortisol and the resultant relationship with SES. Of the two serum cortisol studies, one found a relationship between lower SES and higher cortisol and one did not. Of the five urinary studies, three found no relationship between lower SES and higher cortisol and two did. Finally, of the fourteen studies measuring salivary cortisol measures, six found no relationship between lower SES and higher cortisol, four found a mixed relationship and 4 found at relationship between low SES and higher cortisol. It is worth highlighting that the majority of these studies had large subject numbers with only five studies having less than 200 subjects ($n = 31, 81, 144, 181,$

193), nine studies had between 200 and 1000 subjects and two even larger studies had 2256 and 6335 subjects. Of greatest concern is the fact the two very large studies, one from a urine cortisol measure and one from a salivary measure provide conflicting results when assessing the relationship between low SES and high cortisol.

In addition to issues surrounding the uncertainty in the relationship between cortisol and SES as detailed above, there also remains issues over what measurement of cortisol is the most appropriate and also the most biologically relevant [193]. From a similar standpoint to Dowd et al i.e. from a perspective of cortisol measures within a psychobiological context, the various types of measurement of cortisol were discussed by Levine et al [193]. Cortisol is known to vary throughout the day, and therefore, even within a single type of measurement such as measures from plasma, there are different schemes which may be applied to record a single outcome measure of cortisol. In this thesis cortisol was measured from a plasma sample at a single time point in the morning. Cortisol typically peaks in the morning and gradually decreases during the day [193], therefore a morning measure will be close to the maximum value observed in any individual on a given day. Unbound blood plasma cortisol has been reported to correlate well with salivary measures, where salivary measures are often preferred as they are less invasive and can be done at home by the subjects under study [181]. Salivary measures are also more conducive to enabling repeat recordings throughout the day [181]. A measurement of the area under the cortisol - time curve may then be obtained, this method is preferred by some in the literature [87, 195].

Taking all of the above into account there are clearly issues within the variability in the measurement of cortisol. When this uncertainty is superimposed within a psychobiological framework such as in assessing cortisol in subjects with differing SES, given the variation and differences in methods created to assess SES, these uncertainties are further compounded. Given that differences in outcomes that have been observed for the levels of cortisol in low SES subjects in studies with subjects number of over 1000, it is highly unlikely one could establish anything meaningful in a pilot study of 42 subjects such as was performed here. However, the issue of the validity of cortisol runs deeper than any error in this thesis and is still an issue across the peer-reviewed literature where much work is still necessary to improve the understanding and biological significance of the different methods of cortisol and to reach a consensus over how it might best be measured in studies assessing the impact of SES and beyond [109].

The relationships which may mediate and moderate the impact of biological factors on the hippocampus are clearly complex. It is perhaps too simplistic a view to try and look for relationships between singular inflammatory factors and volumetric changes in the hippocampus.

When adding the additional complexity of socioeconomic status, this further complicates these issues. Larger scale studies, preferably including genetic assessments as well as a range of biological, inflammatory and metabolic markers on subjects from differing social classes may prove more insightful.

2.9 Conclusion

In a pilot imaging study to assess hippocampal volumes using 3T MRI in two distinct SES groups, manual and automated protocols of assessing the hippocampus were developed and implemented. The author developed a protocol for segmenting the hippocampus based mainly on observations recorded in Konrad et al around a review of differing hippocampal volumetry protocols [185]. Intra and Inter rater quality measures including the Intraclass correlation coefficient for agreement and the Dice metric were determined for repeat measures of the manually delineated hippocampal volumes. Both Intra and Inter rater measures were typically of a good standard (>0.85), providing confidence in the clarity and comprehensiveness of the manual segmentation protocol and rater skill. Performing manual segmentation of the hippocampus is time consuming and challenging where even experienced operators may take between 30 and 45 minutes to complete a single segmentation. Thus the author applied an automated method of hippocampal segmentation where the aim was to determine whether or not it was possible to replace the manual method. The mean volumes and also the variance in these measurements were comparable between the two methods. The Dice overlapping metric comparing the two methods was 0.81, which is comparable to what has already been observed in the literature for automated segmentation methods of the hippocampus [123, 278]. Despite the reasonable results that were obtained by running the automated segmentation algorithm, in some circumstances the automated method had clear, gross errors. Therefore it is recommended that if automated segmentation algorithms are employed in future work that they are manually inspected and volumes are corrected wherever necessary. It is hoped this approach will provide the best balance between operator analytical time and volumetric accuracy.

Having established the methodology for measuring hippocampal volumes from 3T MRI data the mean hippocampal volume and the standard deviation of the two differing SES groups were determined. It was observed that there was a general trend for the more deprived group to have smaller hippocampal volume than the more affluent group. However, given the exploratory nature of this study a sample size calculation was performed to determine how many subjects would be necessary in further work to achieve a study power of 0.90

and thus confirm the trend observed in the pilot data. It was found that approximately 57 subjects would be required in each group to achieve this level of study power when assessing hippocampal volume differences as a result of deprivation. It is hoped that this calculation will inform future work which may be performed to assess if significant volume differences exist between differing SES groups.

Post-hoc correlations between IL-6, cortisol and hippocampal volumes found no significant linear relationship between these measures. It was suggested that a larger scale, more complex analysis of genetic, inflammatory and metabolic makers may be necessary to illicit the mediating and moderating factors between SES and its relationship to hippocampal volume.

The time taken to manually segment hippocampal volumes was significant and could be restrictive to future work, particularly if large scale studies are being performed. Therefore, a comprehensive comparison between an automated method of hippocampal volumetry and a manual segmentation method was done. The agreement between the manual and automated method was reasonable with a Dice metric of 0.81. These values were comparable to what has been presented in the literature for other automated methods as well as for FSL's FIRST algorithm used here. Despite the modest success of the automated method it was observed that it can occasionally fail to give accurate results. Therefore, it was recommended that in future to maximise volumetric accuracy while minimising the time penalty for analysis, automated segmentation should be followed by a manual inspection and correction of the segmented volumes. This approach was adopted in chapters 5 and 6 of this thesis to assess hippocampal volumes and malrotation in a normal group of healthy volunteers.

Chapter 3

Voxel Based Morphometry (VBM) in groups with different socioeconomic status

3.1 Abstract

From a large, community based sample of data on subjects from different socioeconomic backgrounds, a subset were selected for imaging using a 3T MRI system. This presented a novel opportunity to assess the volumetric differences in the brains between the two differing socioeconomic groups. It was proposed that by considering those individuals in the lower SES group to have been exposed to a chronic stress, perhaps by means of inflammatory pathways, this would result in volume reductions in this group when compared to a more affluent, higher SES group. To test this hypothesis, voxel based morphometry (VBM) was used to compare the volume of grey matter between the two groups. In applying VBM it is critical to have good inter-subject normalisation. To satisfy this requirement, a recently developed normalisation scheme known as DARTEL was used in the VBM processing pipeline [20]. DARTEL-VBM results were found to be superior to the earlier 'optimised' VBM method. Following a small volume correction to the DARTEL-VBM results, the data were suggestive of a GM volume reductions in both the right and left hippocampi of the lower SES group. This agreed with what the tentative results for the manual and automated segmented hippocampal volumes reported in chapter 2 of this thesis. As VBM is a voxel based method, it was possible to localise the regions suggestive of volumetric difference between the two groups to the head and anterior portion of the hippocampus.

In order to try and explain why the hippocampal volumes were smaller in the lower SES group, a VBM multiple regression model was created and included the inflammatory markers Interleukin-6 and cortisol. These factors had previously been shown to correlate with hippocampal volumes in normal, healthy subjects [217, 291]. However, we found no correlation between hippocampal grey matter and IL-6 or cortisol.

Finally, further post-hoc analysis was performed to assess other regions in the brain which may show a GM volume reduction in the lower SES group. The most significant region of volume reduction in the lower SES group was observed in the cerebellum. When assessing the whole brain, this was more statistically significant than the differences observed in the hippocampus. Global GM and WM volumes were also compared between the two groups and both were on average smaller in the lower SES with the WM volume reduction being statistically significant.

Further work will be required to elicit the mechanisms by which deprivation affects the brain. The inflammatory markers assessed here did not explain an appreciable portion of the variance in the observed volumetric measurements and given the small number of subjects studies here there are considerable concerns over the reliability of these measures both in this chapter and in the peer reviewed literature pertaining to SES. A larger study with utilising an improved biological model of SES and a more detailed assessment of inflammatory and metabolic factors in deprivation and the links to brain volumetry are required. It will also be important to consider brain networks in further work as it is perhaps too simplistic a view to consider only volumetric changes in a single region of brain to be affected by biological agents for processes such as inflammation.

3.2 Introduction

As was described in chapter 2, a subset of forty-two subjects from two differing SES groups were selected for imaging. This subset was part of a larger study of the psychological, social and biological factors in deprivation, known as the PSOBID study [356]. In chapter 2 the aim was to define a manual protocol for segmenting the hippocampus and then to assess hippocampal volume measures between the affluent and poorer groups. Furthermore, the manual method was compared to an automated segmentation algorithm for segmenting hippocampal volumes between the two groups. Here an alternative method of analysis to investigate group differences in hippocampal volume is presented. The method applied here is known as voxel based morphometry (VBM) [138, 21].

Thus, while the analysis method differs in this chapter, the motivation and rationale for applying the method is the same as it was in chapter 2. That is, to establish the optimal methodology to investigate GM volume differences between differing SES groups, where it was proposed that the more deprived group may have smaller hippocampal volumes.

3.3 Aims and hypotheses

The aim of the imaging aspect of the PSOBID study was to establish the methodology necessary to assess GM volume differences between more deprived (low SES) and more affluent subjects. Several methods of analysis were applied to investigate this aim. As discussed above hippocampal structure volumes were considered in chapter 2, and hippocampal metabolites are considered in chapter 4. This chapter will explain how VBM was used to compare grey matter (GM) volume between poor and affluent groups. Further analysis will assess the role of inflammatory markers on GM volumes using a VBM regression model.

As was the case in chapter 2, the purpose of this thesis was not only to consider volume differences between SES groups but to investigate the quality of the methods employed to answer this question. To this end different aspects of the VBM methodology itself were assessed. This focused on different image normalisation methods used to warp images into a common template space and different methods of correcting the statistical parametric maps for multiple comparisons.

3.3.1 Primary null VBM hypothesis

There are no focal deficits in grey matter volume in the lower SES group when compared to the more affluent group

3.3.2 Secondary null VBM hypothesis

Measures of subject inflammatory markers are not correlated with focal deficits in grey matter volume

3.4 Study methodology

The study was approved by the local NHS research ethics committee.

3.4.1 Volunteers, materials, methods and study design

Study design, subject recruitment, inclusion and exclusion criteria have been described previously in chapter 2.

3.4.2 Data acquisition

A T1-weighted IR-FSPGR was acquired, TR = 6.8ms; TE = 1.5ms, Inversion preparation time = 500ms; flip angle=12°; FOV = 26cm; phase FOV= 70%; matrix: 320 x 320; bandwidth 31.25kHz; slab thickness = 1mm. The data in this acquisition was subsequently used for the whole hippocampal volume segmentation and VBM analysis. This acquisition sequence with high resolution in all three imaging planes and good white matter to grey matter contrast facilitates well both the extraction of the whole hippocampus and tissue type segmentation necessary for VBM.

3.5 Data analysis

The T1-weighted 3D FSPGR MRI data was pre-processed using the N3 inhomogeneity algorithm and then both optimised VBM and the DARTEL-VBM approach were applied to assess volumetric differences between the affluent and poorer groups. These steps are discussed in more detail below.

3.5.1 Data pre-processing

All MRI data were converted from Dicom to Nifti image format using Chris Rorden's dcm2nii GUI program (<http://www.cabiatl.com/mricro/mricro/index.html>). This enabled the data to be processed using the SPM5 software.

3.5.1.1 Manually defining the image origin

As the position of different subjects may vary within the head coil in both origin and orientation it was necessary to correct for these discrepancies. Thus, the first stage of data pre-processing was to manually define the origin of the image in SPM. A visual assessment of the data was performed to locate the position of the anterior commissure (AC), this was then defined as the origin of the image. Furthermore, the images were manually realigned to remove any unwanted tilts or rolls. While these manual corrections to the data are not always

necessary, these steps come highly recommended, and by defining a consistent starting point for the data, this reduces the likelihood of errors from subsequent normalizations steps.

3.5.1.2 N3 inhomogeneity correction

Image intensity inhomogeneities are inherent in all MR images and are a consequence of the fact that coil sensitivity varies with position within the coil. On modern MR scanners there are typically on-board algorithms to correct for image inhomogeneities, for example, GE applies a surface coil intensity correction (SCIC). This is typically sufficient to correct for intensity variations making the image suitable for clinical reporting by a Radiologist. However, for the purposes of many image processing techniques there is a requirement to further reduce the image inhomogeneities across the field of view [326]. For example, in producing segmented tissue type maps, the intensity of a voxel partly defines the probability of which tissue type each voxel belongs to. Therefore, there was a need to remove as much of the inhomogeneity in the T1-weighted 3D images as possible before further processing of the data.

N3 stands for nonparametric, non-uniform, intensity normalisation and benefits from being able to be applied to any pulse sequence and is insensitive to pathological data which might confound other correction methods [326]. N3 implements an iterative approach to estimate the bias field and the distribution of the true image intensities.

While early methods of bias corrections involved homomorphic filtering these methods have subsequently been superseded. Current methods can generally be classed within two types, those that use parametric representations of image intensity distributions (for example as in a mixture of Gaussians) and those that use nonparametric representations (such as histograms).

N3 has been employed over a number of years within the early pre-processing stage of the Freesurfer data processing pipeline, a well known software package for analysis of structural and functional MRI data. In recent years there has been some work detailing the optimal parameters to use within N3 [378] and in particular optimising the use of this algorithm on 3T MRI data [51].

The intensity correction used in the unified segmentation approach by Ashburner and Friston, 2005 involves a parametric bias mixture of Gaussians (MOG) correction model where the noise is assumed to be due to variations in the tissue properties as opposed to being from the scanner itself. As image bias tends to be spatially smooth, a linear combination of a low number of basis functions are used to parameterise the bias field [23]. Models that

incorporate scanner and tissue noise do exist and may be better, for example, the single noise source model is mainly used because of its simplicity [124]. A further commonly used inhomogeneity correction is N3.

The issue of image inhomogeneity can begin to be tackled if it is modelled by a smooth bias field of multiple, multiplicative components. This models the nonuniformity arising from variations in the sensitivity of the receiver coil and to a lesser degree induced currents and nonuniform excitation.

Consider the following model of image formation in MRI, see equation 3.1.

$$v(x) = u(x)f(x) + n(x) \quad (3.1)$$

Where at location x , v is the measured signal, u is the true signal emitted by the tissue, f is an unknown smoothly varying bias field and n is the white gaussian noise assumed to be independent of u . Therefore, f must be estimated if the data are to be corrected, though the problem of additive and multiplicative inference makes this task difficult.

Consider a noise free case where the true intensities u at each voxel location x are independent identically distributed random variables. Using the notation $\hat{u}(x) = \text{Log}[u(x)]$, the image formation model becomes additive, see equation 3.2.

$$\hat{v}(x) = \hat{u}(x) + \hat{f}(x) \quad (3.2)$$

Consider the distribution of values that \hat{f} takes over the region of interest to the probability distribution of a random variable. For example if \hat{f} is a linearly increasing field aligned on a rectangular region then \hat{f} will have a uniform distribution.

Let U , V and F be the probability densities of \hat{u} , \hat{v} and \hat{f} respectively. Making the approximation that \hat{u} and \hat{f} are independent and uncorrelated random variables, the distribution of their sum is found by convolution as follows, see equation 3.3.

$$\hat{V}(\hat{v}) = \hat{U}(\hat{v}) * \hat{F}(\hat{v}) \quad (3.3)$$

That is, the nonuniformity distribution F can be interpreted as a blurring of the intensity distribution U .

The blurring due to the field reduces the high frequency components of U and therefore the task of correcting for the field lies in restoring these high frequency components of U .

However, as the shape of the blurring kernel of F is not known, it is not clear which frequency components of U need to be restored to get from the observed distribution V to the true distribution U . The solution lies partly in the fact that the nonuniformity field \hat{f} is restricted to being smooth and slowly varying, thus the space of the possible distributions U corresponding to a given distribution V is small enough that the problem becomes workable. That is, the approach employed by N3 works by determining the smooth, slowly varying, multiplicative field that maximises the frequency content of U .

It would be possible to search through all possible \hat{f} fields to find the value which maximises the frequency content of U , however, there are two problems with this approach. The search space of all 3D fields \hat{f} are very large, and spectral estimates of related measures such as entropy are difficult to compute with sufficient accuracy to detect subtle changes in U . The N3 approach to this problem is to propose a distribution for U by sharpening the distribution V and then to estimate a corresponding smooth field which produces a distribution U close to the one proposed. A key aspect in the implementation of this process is that N3 capitalises on the simple form of the F field, if we suppose that the F field is gaussian, then we need only search the space of all distributions U corresponding to gaussian distributed F having zero mean and a given variance. In this way the space of all distributions U collapses into 1D space, the width of the F distribution.

In practice, the F distribution is only approximately gaussian and some of the assumptions, such as zero noise, are violated. To tackle these difficulties, an incremental approach is taken to estimating the true distribution of intensities U and corresponding \hat{f} field. Since any gaussian distribution can be decomposed into a convolution of narrower gaussian distributions, the space of all U distributions corresponding to gaussian distributed F can be searched incrementally by deconvolving narrow gaussian distributions from subsequent estimates of U . A key benefit to this approach is that after each increment and estimate of U a new estimation of the smooth field \hat{f} is determined. The constraint that the field must be smooth changes the shape of the proposed distribution U to one that is consistent to the field. These perturbations of U perturb F from its gaussian shape and compensate for the distortion of V caused by noise and other factors. This iterative process can be viewed as travelling in the space of all U distributions along a path corresponding to smooth fields \hat{f} with increasingly wider distributions. The iterations continue until no further changes in \hat{f} or U result from deconvolving narrow gaussian distributions from V . Within the unified segmentation model, registration, tissue segmentation and bias correction are all performed within a probabilistic framework [23]. To avoid any dependence on the underlying anatomy N3 implements an iterative approach to estimate both the multiplicative bias field and the distribution of the

true tissue intensities.

The N3 inhomogeneity correction to the T1-weighted 3D dataset was applied using Freesurfer software (<http://surfer.nmr.mgh.harvard.edu/>). There have been two papers discussing the optimal N3 parameters to be used for 3T MRI data [51], [378]. They state that using a smoothing correction distance of between 30mm and 50mm to estimate the bias field was optimal. This was in contrast to the default smoothing correction value for N3 of 200mm. Thus a correction distance of 50mm was chosen. It is also suggested that improved results from the N3 algorithm may be achieved by running it multiple times, in practice this is equivalent to increasing the number of iterations allows the algorithm more time to converge on a solution. Thus the number of iterations was increased from the default value of 50 to 200, this was equivalent to running the default number of iterations four times on the dataset.

In practice the application of the N3 inhomogeneity correction means that some of the inherent, artificial variation in intensity across regions of the same tissue type are reduced. This had the effect of homogenising the tissue intensities which will have a positive impact on subsequent tissue segmentation processes. The improvement in image quality as a result of applying N3 can be seen in figure 3.1.

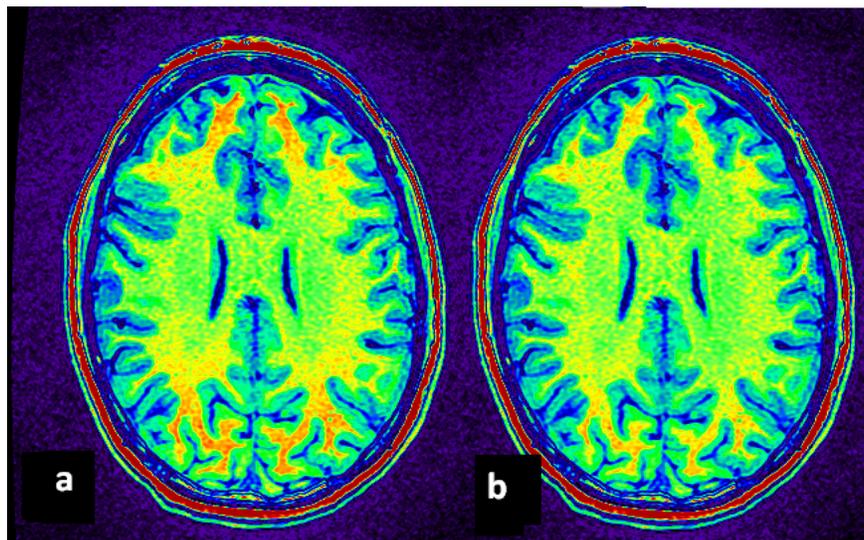


Figure 3.1: T1-weighted 3D FSPGR image pre (image a) and post (image b) N3 inhomogeneity correction

In comparing the pre and post N3 corrected images shown in figure 3.1, the variation in the intensity of the WM, in particular around the frontal lobe and occipital/parietal lobes

was reduced on the post N3 image. Such image intensity variations are subtle, and would not always be obvious from a visual assessment of the grey colour scale data. However, the intensity variations are real and do have an appreciable effect on how well subsequent segmentation processes perform.

3.5.2 Data processing: 'Optimised' VBM

Voxel based morphometry (VBM) has been used in imaging studies since the mid 1990's though it wasn't until slightly later that the image processing steps were formalised into what is now known as the VBM methodology [21]. VBM enables voxel based statistical analysis to be applied to groups of structural MRI data. This method enables group comparisons of grey matter and white matter tissue to be done on the whole brain such that differences in tissue class volumes may be localised to specific regions of the brain. VBM may also be used to perform regression analysis to assess correlations of other variables with respect to GM or WM volume differences. Until recently (approximately <2009) the most commonly applied implementation of VBM was the optimised VBM approach implemented in the SPM5 software. This unified segmentation approach included normalisation, tissue segmentation and bias correction all performed within a probabilistic framework [23]. While the methods within the VBM pipeline continue to evolve and improve as developments are made within the respective image analysis fields the core components of the VBM pipeline are relatively constant. The methodology described here is known as the optimised VBM procedure. The seminal publication on optimised VBM is the work by Good et al [139].

The original T1-weighted 3D structural images are segmented in native space. The resulting grey and white matter images are then normalised to grey matter and white matter templates in template space to derive the optimised normalisation parameters. These parameters are then applied to the original, whole brain image in native space prior to a new segmentation procedure. This iterative process reduces the risk of misrepresentation of significant differences with respect to the original VBM procedure [139].

The optimised VBM workflow is shown in figure3.2 on the following page.

3.5.2.1 Segmentation in native space and removal of non-brain structures

The T1-weighted 3D images in native space were segmented into their GM, WM and CSF

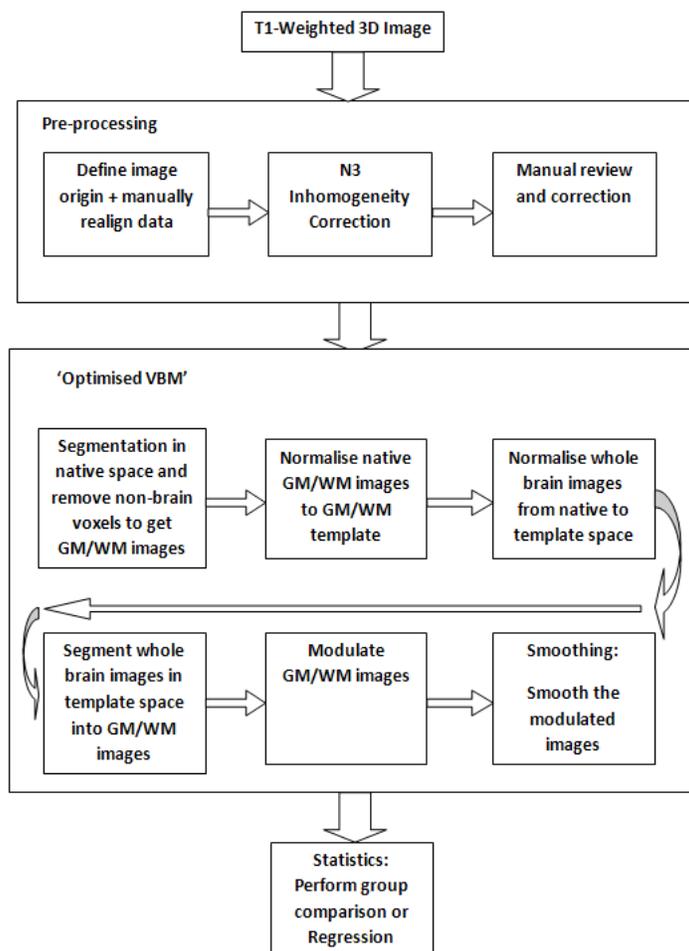


Figure 3.2: 'Optimised' VBM image processing pipeline

components. This procedure also ran a series of morphological operations to remove non-brain voxels. Normalised MRI data are then segmented into GM, WM, CSF and other non brain tissue images. SPM segmentation applies a mixture model cluster analysis to identify voxel intensities matching a particular tissue type. This is combined with apriori knowledge of the spatial distribution of the main tissue types in normal subjects, derived from probability maps. The segmentation also applies an inhomogeneity correction to address image intensity variations as a result of different cranial surface positions within the MRI head coil [21].

Following segmentation of the T1-weighted 3D image, the GM, WM, CSF and total intracranial volumes (ICV) were computed for the poorer and more affluent populations, the results are shown in figure 3.3 on the following page.

In order to compare the volume of the tissue classes between the affluent and poorer groups

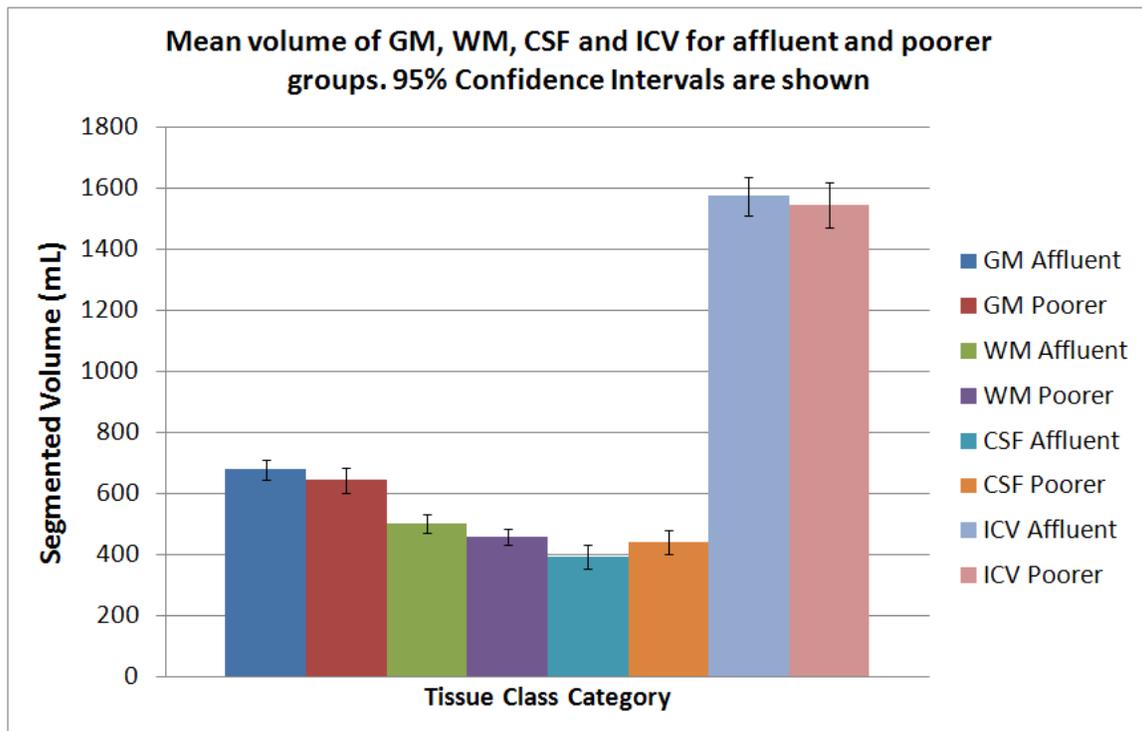


Figure 3.3: Global GM and WM volume measures for affluent and poor populations

a general linear model (GLM) was set up to perform an analysis of variance (ANOVA) on the GM, WM and ICV volumes. Despite the fact that the affluent and poorer groups were approximately age matched, subject age was included as a covariate in the GLM design to ensure the most precise statistical comparison of the tissue type volumes between the two groups. The ANOVA results showed that there was no statistically significant difference in the ICV volume between the two groups (p-value = 0.517).

3.5.2.2 Normalisation of GM/WM images to template space

The GM and WM images were affine normalised to the apriori GM and WM template maps. By removing the non-brain structures during the initial segmentation procedure the risk of non-brain voxels corrupting the GM and WM maps is minimised, thus optimising the normalisation procedure.

3.5.2.3 Normalisation of whole brain images to template space

As the GM and WM probability maps were in template space, to achieve the optimum segmentation, the whole brain images was affine normalised and then segmented. Thus,

using the normalisation parameters determined for the GM and WM images the whole brain images in native space were normalised to template space.

3.5.2.4 Segmentation of normalised whole brain image and removal of non-brain structures

The normalised whole brain images in template space were segmented. Non-brain structures were also removed as part of this procedure.

3.5.2.5 Modulation

As a result of the affine normalisation procedures some regions of GM and WM may grow or shrink. Therefore, in order to maintain the correct volume of a particular tissue type within a voxel a further processing step is required, this known as the modulation step. This involves multiplying the image intensities of the segmented images by the Jacobian determinants of the affine normalisation step. Analysis performed on modulated data effectively compares regional differences in absolute tissue class volume. If data were not modulated (unmodulated) then a voxel by voxel comparison would provide regions of difference in tissue class concentrations. For the purposes of this thesis modulated data was always used.

3.5.2.6 Smoothing

Following modulation of the data, the data were smoothed. This conditions the data to conform more closely to the gaussian field model underlying the statistical procedures used for making inferences about regional specific effects. Smoothing renders the data more normally distributed which is advantageous in applying statistics. The intensity in each voxel of the smoothed data is the weighted average of grey matter density from a region of surrounding voxels, the size of the region being the size of the gaussian smoothing kernel [21]. Following smoothing of the data, a study design was established and the model was estimated in preparation for the final statistical analysis being performed.

3.5.3 The DARTEL-VBM processing pipeline

The optimised VBM processing pipeline which was widely used before the introduction of DARTEL was discussed above and the processing pipeline has been shown previously in figure 3.2. In SPM5 a new normalisation algorithm, known as DARTEL, was introduced in approximately 2007 and is a fast Diffeomorphic Anatomical image Registration algorithm that uses Exponential Lie algebra (DARTEL).

One of the key steps in the VBM processing is how individual T1-weighted 3D brains are normalized to the underlying template. The aim of normalisation with respect to VBM is to ensure that a brain approximately matches the template image while only leaving the differences between the subject and template brain that are a result of true volumetric differences. That is, it is possible that one brain could be morphed entirely to match a template but this would not leave any measurable volumetric differences between the images such that a proper statistical analysis could be applied. Therefore, within the normalisation process there is a constraint known as regularisation which inhibits the degree to which images are normalised. Regularisation is necessary to stop overfitting of the nonlinear registration process thus avoiding unnecessary deformations being introduced.

Image registration typically has the objective of trying to obtain the optimum set of parameters that allows a smooth and continuous mapping between the points of one image to another. There are many ways to model this situation but they broadly fall into two categories of parameterisation. There are small-deformation frameworks, which does not necessarily preserve topology, although if its deformations are relatively small then they may be preserved and there are large deformation frameworks or diffeomorphisms that have a number of elegant mathematical properties, such as enforcing the preservation of topology.

In practical terms there are some additional refinements which can be made to a VBM workflow such that the quality of the data are improved. This includes prior application of an inhomogeneity correction to the dataset and a six degree of freedom transformation with subsequent manual adjustments.

A diffeomorphism is an isomorphism in the category of smooth manifolds, where a manifold is simply a topological space of small enough scale to resemble Euclidean space i.e. it can be described by Euclidean geometry. An isomorphism is a mapping between two objects. It is a bijective map which implies that the mapping function and its inverse preserve structure and that for every element of the mapping there is a single equivalent mapped location. Thus a diffeomorphic registration method uses an invertible function to map one differentiable manifold, which in our case is a brain image, to another.

Small deformation frameworks typically parameterise a displacement field (u) that is then added to an identity transform (x), see equation 3.4.

$$I(x) = x + u(x) \tag{3.4}$$

In these parameterisations, the inverse transformation is sometimes approximated by subtracting the displacement. It is worth noting that this is only an approximate inverse which

fails badly on larger deformations. Small deformation models therefore do not enforce a one-to-one mapping, particularly if the model assumes the displacements are drawn from a multivariate gaussian probability.

The diffeomorphic framework is a much more elegant approach. A diffeomorphism is a global, one-to-one, smooth and continuous mapping with derivatives that are invertible (i.e. a non-zero Jacobian determinant) . If the mapping is not diffeomorphic, then topology is not necessarily preserved. A key element of the diffeomorphic framework is that it enforces consistency under compositions of the deformations. A composition of two functions is essentially taking one function of the other in order to produce new function. For two functions I_1 and I_2 this would be denoted as follows, see equation 3.5.

$$I_2 \circ I_1 \circ x = I_2(I_1(x)) \quad (3.5)$$

For deformations, the composition operation is achieved by resampling one deformation field by another. If the deformations are diffeomorphic then the resulting composition will also be diffeomorphic. In practice, deformations are typically represented discretely by a finite number of parameters, so some small violations may still be possible, particularly if the composition is done using low degree interpolation methods. Perfect (i.e. infinitely dimensional) diffeomorphisms for a Lie group, under the composition operation, as they satisfy the requirements of closure, associativity, inverse and identity. Where a Lie group is a group which is a differentiable manifold, with the property that the group operations are compatible with smooth structures.

Recently developed large deformation registration frameworks aim to find the smoothest possible solution. DARTEL works by parameterising using velocity fields where the velocity fields describe the different time periods across the course of the evolution of the diffeomorphic mapping. If $u^{(t)}$ is the velocity at time t , then the diffeomorphism evolves as described by equation 3.6.

$$d\Phi/dt = u^{(t)}(\Phi^{(t)}) \quad (3.6)$$

Diffeomorphisms are generated by initialising with an identity transform ($\Phi^{(0)} = x$), and integrating over unit time to obtain $\Phi^{(1)}$. DARTEL employs a single flow velocity field that remains constant over unit time. It is similar to the log-euclidean framework used by Arsigny et al, [19].

DARTEL has the advantage over small deformation methods in that is diffeomorphic, easily invertible and can be rapidly computed.

The DARTEL model assumes a flow field (u) that remains constant over time. With this model, the differential equation that describes the evolution of the deformation is shown in equation 3.7.

$$d\Phi/dt = u(\Phi^{(t)}) \quad (3.7)$$

A deformation is generated by using an identity matrix ($\Phi^{(0)} = x$) and integrating over time to obtain $\Phi^{(1)}$. The Euler method is a simple integration approach, which involves computing new solutions after many successive small steps time steps (h), see equation 3.8.

$$\Phi^{(t+h)} = \Phi^{(t)} + hu(\Phi^{(t)}) \quad (3.8)$$

Each of these Euler steps is equivalent equation 3.9.

$$\Phi^{(t+h)} = (x + hu) \circ \Phi^{(t)} \quad (3.9)$$

It has been shown that DARTEL achieves an improved non-linear normalisation over previous affine normalisation methods applied within SPM5 [23]. DARTEL uses a more sophisticated higher dimensional mathematical framework to produce deformations. This ensures that the Jacobian is always positive as the exponent of a real number is always positive. This condition therefore ensures that the mapping is diffeomorphic, and that the forward and inverse transformations can be generated from the same flow field, allowing the transformation to be inverse consistent. This is an important advance for image registration algorithms since this class of registration is expected to best represent the normalization of the brain to a template image. DARTEL works by creating a customised, study specific template of the input images based upon the registration results as well as all the relevant flow fields which represent the transformations of each individual image to the template. The DARTEL-VBM image processing pipeline, including pre-processing steps, is summarised in figure 3.4 on the following page.

In comparing the optimised and DARTEL VBM procedures one can see that the core elements of the processing pipeline are quite similar. Both pipelines involve spatial normalisation, segmentation, modulation and smoothing. The significant difference between the two methodologies is the application of DARTEL as an improved normalisation procedure. This leads to study specific templates being created for both the GM and WM tissue classes for

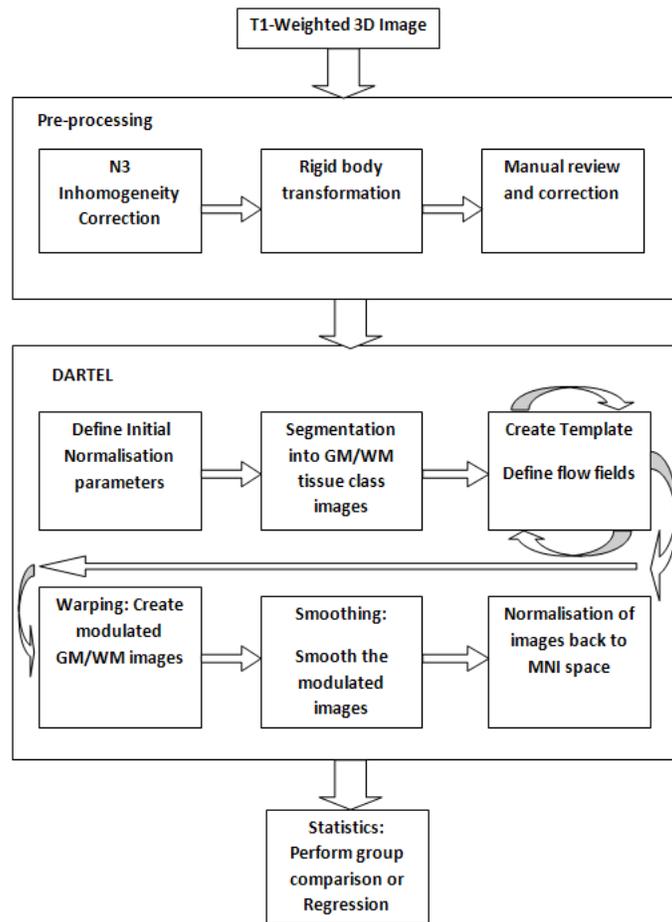


Figure 3.4: DARTEL-VBM image processing pipeline

the PSOBID data. It has been shown that DARTEL performs better than previous methods implemented within SPM [182] where the practical implications of this include an improved sensitivity to volumetric differences, for example in regions such as the hippocampus [40].

DARTEL has been shown to perform better when it is combined with pre-processing steps such as an inhomogeneity correction [1, 282]. DARTEL and a range of nonlinear registration methods were tested recently [182]. The methods were subjected to the same tests based on an agreed criteria. The results of the comparison recognised DARTEL as one of the best nonlinear registration methods, with only ART [16], IRTK [306] and Syn [24] performing better. Fifteen algorithms were tested in total. Thus, given the ease by which the use of DARTEL could be incorporated into the wider VBM processing pipeline, it was a suitable choice for image normalisation.

DARTEL, of course, is not without its limitations. DARTEL is far more computationally

intensive than previous normalisation methods implemented within SPM and as such takes a considerably longer time to run to completion.

3.5.4 Criticisms of VBM

Despite its widespread use there are a number of issues and limitations with VBM. The use of the VBM process when images are imperfectly registered has been criticised by Bookstein et al in a paper entitled: “Voxel-based morphometry should not be used with imperfectly registered images” [50]. As the title suggests, the authors cite issues with the VBM methodology where uninformative and incorrect voxel inferences are made at the locations of imperfect registration. The criticism was largely addressed in the reply to this paper by Ashburner and Friston in a paper entitled “Why VBM should be used” [22]. The reply agrees that VBM should not be used on imperfectly registered images and acknowledges that the normalisation method adopted within SPM at that time was ‘far from perfect’. Since those early criticisms of VBM, the method has continued to evolve and to be refined. The main criticism directed at VBM surrounded the normalisation method implemented within VBM. However, the introduction of the DARTEL normalisation scheme to the VBM pipeline represents perhaps the largest step improvement in the VBM methodology within SPM for some time, though it is still advisable to check the image registrations for errors.

3.5.5 Statistical design, analysis and the correction for multiple comparisons

3.5.5.1 Statistical analysis: The General Linear model (GLM)

The voxel-based statistics of MRI images later in the chapter will utilise the General Linear Model (GLM). The GLM is a versatile framework where many more familiar statistical tests such as the analysis of variance (ANOVA) and multiple regression are special cases of the more generic GLM. When conducting an experiment a dependent variable will be measured (Y_j) where $j=1, \dots, J$ indexes the observation. For the dependent variable there are likely to be one or more explanatory or independent variables, we will consider that we have a set of K independent variables ($K < J$) denoted by x_{jk} where $k=1, \dots, K$ indexes the independent variables. The independent variables may be continuous or discrete and may represent covariates, functions of covariates or dummy variables to indicate the levels of an experiment. The GLM explains the dependent variable Y_j in terms of a linear combination of the independent variables plus an additional error term (ε). Therefore, the generic form of the GLM is as follows, see equation 3.10.

$$Y_j = x_{j1}\beta_1 + x_{jk}\beta_k + x_{jK}\beta_K + \varepsilon_j \quad (3.10)$$

Where β_k are unknown parameters corresponding to each of the K independent variables x_{jk} . The errors ε_j are independent and identically distributed random variables with zero mean and variance σ^2 .

The parametric statistical framework assumes that data are normally distributed around a mean that can be parameterised by a general linear model (GLM) [371, 130]. One of the great benefits to the GLM is its versatility as it can encapsulate t-test, F-tests, paired t-tests, ANOVA, correlation, multiple regression and ANCOVA statistical designs. The estimated parameters of the GLM are contrasted to produce a test statistic at each voxel which have a Student's t-distribution under the null hypotheses. The resulting t-statistic image is then assessed for statistical significance using continuous random fields to identify voxels or clusters where there is significant evidence against the null hypothesis.

3.5.5.2 The GLM in imaging studies

The GLM has been used extensively to analyse functional MRI data [361] and structural imaging data using VBM [21]. One of the reasons the GLM works so well for dealing with imaging data is because the GLM may be expressed as a matrix formulation. If we take the generic form of the GLM introduced in chapter 2, see equation 3.11.

$$Y_j = x_{j1}\beta_1 + x_{jk}\beta_k + x_{jK}\beta_K + \varepsilon_j \quad (3.11)$$

This may be re-written for the case of multiple observations as:

$$Y_j = x_{j1}\beta_1 + x_{jk}\beta_k + x_{jK}\beta_K + \varepsilon_j \quad (3.12)$$

$$Y_1 = x_{11}\beta_1 + x_{1k}\beta_k + x_{1K}\beta_K + \varepsilon_1 \quad (3.13)$$

$$Y_J = x_{J1}\beta_1 + x_{Jk}\beta_k + x_{JK}\beta_K + \varepsilon_J \quad (3.14)$$

This can then be written in matrix form:

$$\begin{pmatrix} Y_1 \\ \cdot \\ \cdot \\ Y_j \\ \cdot \\ \cdot \\ Y_J \end{pmatrix} = \begin{pmatrix} x_{11} & \cdot & \cdot & x_{1k} & \cdot & \cdot & x_{1K} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{j1} & \cdot & \cdot & x_{jk} & \cdot & \cdot & x_{jK} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{J1} & \cdot & \cdot & x_{Jk} & \cdot & \cdot & x_{JK} \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \cdot \\ \cdot \\ \beta_k \\ \cdot \\ \cdot \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \varepsilon_j \\ \cdot \\ \cdot \\ \varepsilon_J \end{pmatrix} \quad (3.15)$$

or in the preferred matrix notation as:

$$Y = X\beta + \varepsilon \quad (3.16)$$

Where Y is the column vector of observations, ε is the column vector of the error term and β the column vector of parameters $[\beta_1, \dots, \beta_k, \dots, \beta_K]^T$. The $J \times K$ matrix X with j_k^{ith} element x_{jk} is the design matrix. The design has one row per observation and one column per independent variable or model parameter. The important point about the design matrix is that it is a near complete description of the model with the remainder between the model and the observation data being the error term.

Therefore, when we utilise the GLM in imaging studies such as in VBM we seek to determine the least square estimates of the parameter estimates to minimise the residual error term. Thus for a VBM study such as the one we are investigating here by contrasting a lower SES group with a more affluent SES group, the GLM design is as follows:

$$Y(GM) = \beta_1(lowSES) + \beta_2(higherSES) + \beta_3(age) + \beta_4(ICV) + \beta_5(Alc) + \varepsilon \quad (3.17)$$

Where *lowSES* and *higherSES* are the design matrix terms and (*age*), (*ICV*) and (*Alc*) are three covariates relating to the age, intracranial volume and weekly alcohol units added to the model to help explain the variance in the observation for subject Y 's grey matter (GM).

3.5.5.3 Multiple comparison corrections

In many fields of science, largely as a result of the development of computing processors, the ability to deal with large datasets and to perform mass statistical comparisons is now achievable on a modest desktop computer. With this development, specific problems arise which had not needed to be dealt with previously. For neuroimaging datasets such as those from MRI, SPECT and PET data on which voxel-wise statistics are performed, the major issue which must be addressed is that of the multiple comparison problem. A full treatment

of the multiple comparisons problem and methods to correct it is beyond the scope of this thesis, however, the major problems, solutions and any assumptions made are summarised briefly below.

In many quantitative neuroimaging methods, statistics are done at the voxel level. Given that each image may consist of tens to hundreds of thousands of voxels then clearly the multiple comparison correction (MCC) problem is a genuine concern. For each statistical test an arbitrary threshold is generally set (typically 0.05) to control the false discovery rate (FDR) otherwise known as the type 1 error. This is the error which governs false positives, that is, the null hypothesis is rejected when it is actually true. The problem with applying a basic correction such as this is that at this level of thresholding 1 in 20 voxels will be false positives. If we consider 100,000 voxels, then there would be 5000 false positive voxels. Such error rates are wholly unsatisfactory and thus there is a need for a more appropriate solution to correcting the false positive rate, not just at the voxel level, but at the level of the entire collection (or family) of tests that make up an experiment.

3.5.5.4 The Bonferroni correction and familywise error (FWE)

The Bonferroni correction and other related methods were among the first to be proposed as a means of dealing with the MCC problem in imaging. The correction is based on the idea that if an experimenter is testing n dependent or independent hypotheses on a set of data, then one way of maintaining the familywise error rate is to test each individual hypothesis at a statistical significance level of $1/n$ times what it would be if only one hypothesis were tested. Thus, if for example, we have 100 voxels to be tested, and we desire a FWE threshold of 0.05 for each voxel, the Bonferroni correction would make the FWE threshold: $0.05/100 = 0.0005$. It is worth noting that for values as low as this the power for any single test is greatly reduced. It is for this reason that for neuroimaging purposes at least, the Bonferroni correction has been deemed overly conservative. Moreover, the Bonferroni correction does not incorporate a spatial dependence to statistical significance which would be preferable given that neighbouring voxels are in many cases likely to have related properties.

3.5.5.5 Random field theory, FWE and the false discovery rate (FDR)

Random field theory (RFT) proved a significant advance in considering imaging statistics. RFT is the discretisation of a continuous underlying random field. RFT allows a threshold to be found in a dataset where it may be difficult to find the number of independent variables. There are several key concepts in using RFT to describe the statistical image. The Euler

characteristic (EC), this is equivalent to the number of clusters that exist above a given threshold. The concept of resolution elements (resels) is also important as it describes the size of a given block of an image following smoothing, where the resel is typically the size of the FWHM of the image and therefore, the number of resels depends only on the smoothness of the image and total number of voxels. At a high threshold value, the EC is 0 to 1 and is equivalent to the probability of the FWE. That is, for a given resel count and threshold, the EC defines the probability of getting one or more clusters above the given threshold. In RFT values usually have some degree of spatial correlation, i.e. neighbouring voxels are more likely to have similar values than values which are further apart. Bonferroni correction does not account for this. In practice SPM calculates the statistically significant thresholded Z-scores corresponding to the EC value.

Another concept that can be applied as a means of a MCC is the false discovery rate principle (FDR). The FDR is the fraction of false positives among all tests declared significant. The motivation behind this method is to ensure all possible true results are not removed from the analysis while one accepts that 5% of these values will be false positives (for $\alpha = 0.05$). The FDR correction is widely applied in the neuroimaging literature and is a viable option. It is considerably more liberal than bonferroni, but its weakness lies in the number of false positives which survive thresholding.

3.5.5.6 Non-stationary (NS) statistics

There are various statistical problems which occur when interpreting imaging data. The most common problem of dealing with multiple comparisons is well documented and the tools to deal with this problem within SPM are fully integrated into the default installation of the program. However, perhaps a less well known issue, is how to treat images where there are significant variations in smoothness across the image. One of the assumptions when applying cluster-size statistics within the RFT is that the image has stationarity i.e. it has uniform smoothness. If this assumption is violated and there is variation in smoothness over the image, then cluster sizes tend to be large in smooth regions, resulting in increased false positives, while in rough regions, clusters tend to be small, resulting in decreased sensitivity. This is known as the non-stationary (NS) problem [370].

While for the majority of applications in neuroimaging analysis NS is not a problem, VBM is particularly susceptible to errors of this kind as even after overall smoothing of the image, significant intensity variation in the structural image may still remain, more than, for example, would be present in an fMRI image [159]. This can lead to invalid cluster correction

statistics being performed with cluster sizes being either over or under estimated. It has therefore been recommended that cluster p-values should not be used in VBM [21].

A solution to this problem was first proposed by Worsley et al [370]. They proposed a random field theory (RFT) method that adjusted cluster sizes according to local roughness of the images. A measure of local roughness may be obtained in SPM from the resels per voxel (RPV) image. This image can then be used to correct the data into an isotropic space. Thus the sizes of clusters are adjusted according to the local smoothness and this is recorded in the SPM results output table.

Accounting for NS is important as it has been shown to provide erroneous results in VBM studies [244]. The non-stationary correction toolbox (<http://fmri.wfubmc.edu/cms/software#NS>) is available as an extension for SPM to allow an improved approach to dealing with VBM data (or any image with significant variation in smoothness). The toolbox implements the RFT version of the cluster size correction under non-stationarity [159, 158]. To ensure that the final statistical analysis of the PSOBID data were not affected by non-stationary issues, the NS toolbox was used to compute the final SPM maps and statistics. Moreover, a comparison between the NS corrected data and the FDR corrected data was performed to estimate the difference between the two statistical correction schemes.

3.6 Results

3.6.1 Assessing grey matter volume differences between deprived and more affluent groups

The general study design that was used was a two sample t-test design. In fact, more correctly, the model established was an ANOVA GLM analogous to what was used in chapter 2. The covariates used were age, ICV and self reported alcohol levels.

In order to assess the hippocampus it was possible to focus the final results stage of the VBM pipeline by applying a small volume correction (SVC) to improve the statistical interpretation of the results. Statistically if we consider the whole brain ($\sim 400,000$ voxels) for a voxel by voxel based analysis then a vast number of t-tests would be performed. This can lead to an increase in type 1 errors and is the reason that multiple comparison corrections are necessary during this type of analysis. However, by enforcing the SVC on the data the volume was reduced to ~ 2500 voxels, decreasing the likelihood of type 1 errors.

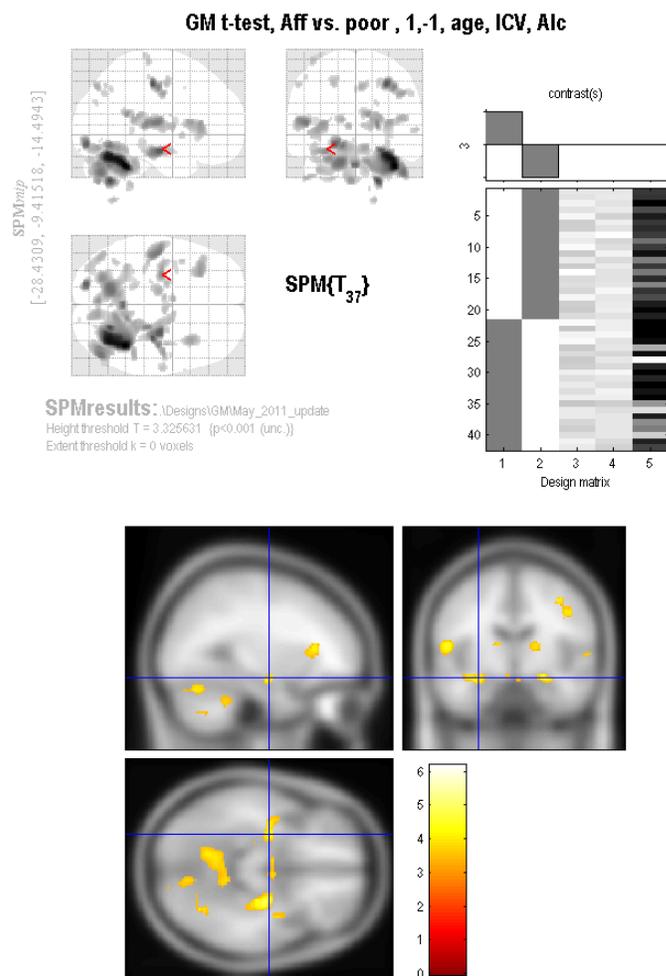


Figure 3.5: SPM map and section of the result of contrasting the affluent and poorer groups and inspecting the left hippocampus

The SVC that was applied to the hippocampal region was a spherical ROI of radius 15mm. Figures 3.5 and 3.6 on the following pages show the initial localisation of the statistically significant volume differences in the left and right hippocampi respectively.

The statistics and co-ordinates of the results associated with figures 3.5 and 3.6 are shown in figures 3.7 and 3.8 on the following pages.

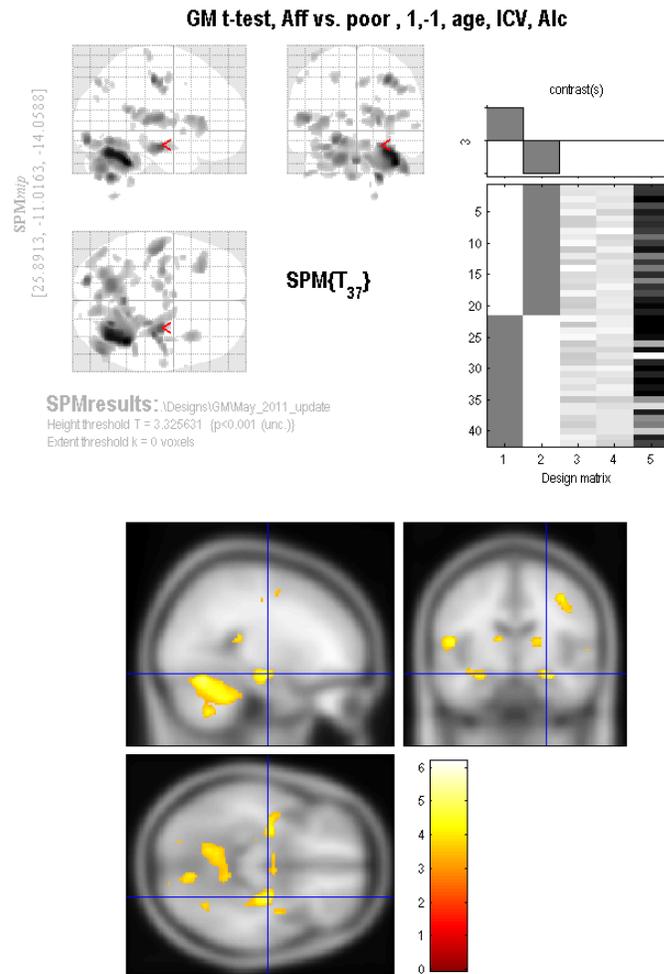


Figure 3.6: SPM map and section of the result of contrasting the affluent and poorer groups and inspecting the right hippocampus

Figures 3.5, 3.6, 3.7 and 3.8 show statistically significant GM hippocampal volume reductions in both the left and right hippocampal volumes. The regions of greatest statistical significance were in and around the head of the hippocampus, though the body of the right hippocampus also appears to have reduced GM volume.

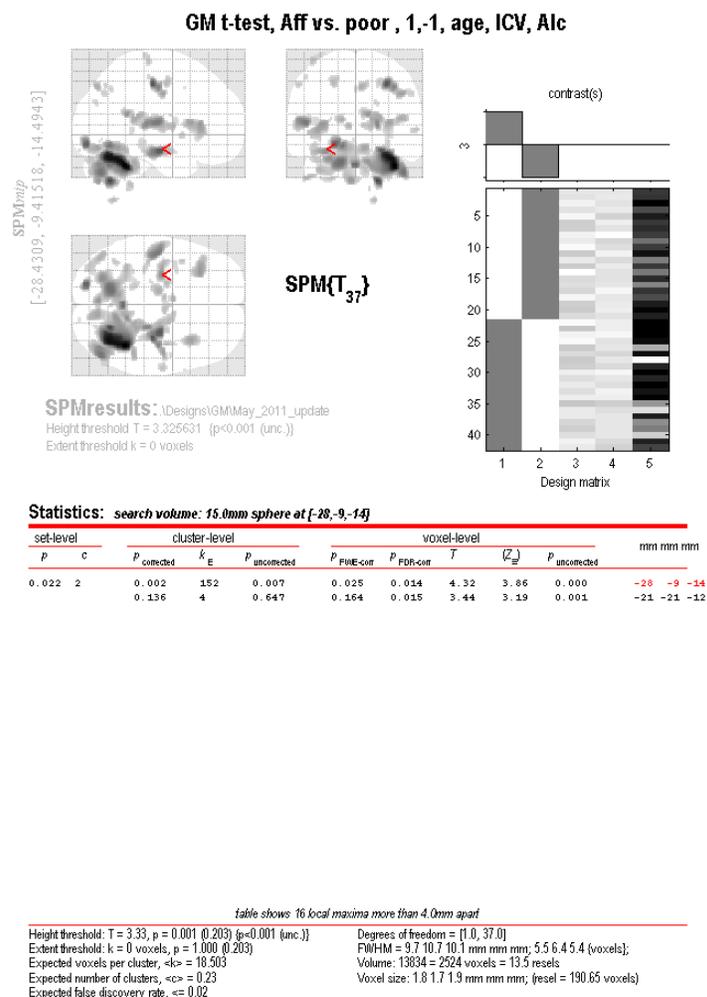


Figure 3.7: SPM with statistics and co-ordinates from the results of contrasting the GM volume of the left hippocampus of the affluent and poorer groups

3.6.2 Regression analysis of inflammatory markers and grey matter volume of the affluent and poorer groups

No statistically significant correlations between GM volume and IL-6 or cortisol were observed in the affluent and poorer groups.

3.6.3 Comparing the results of the 'optimised' VBM approach to DARTEL-VBM

As discussed in the methods section, it has recently been proposed that the traditional optimised VBM approach can be improved by the introduction of a new normalisation scheme

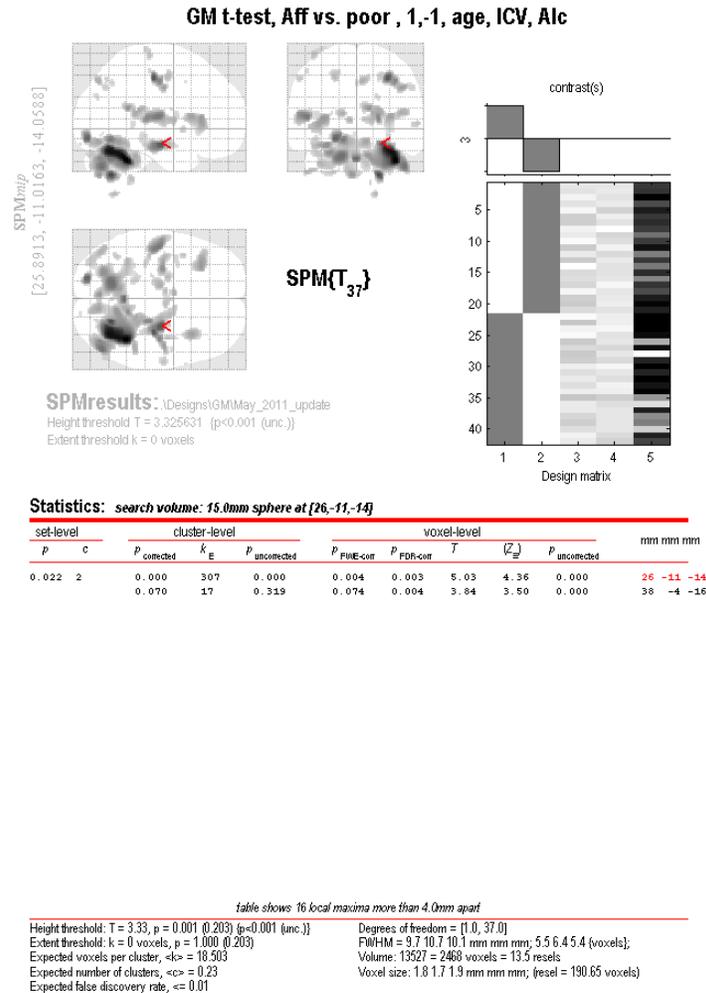


Figure 3.8: SPM with statistics and co-ordinates from the results of contrasting the GM volume of the right hippocampus of the affluent and poorer groups

known as DARTEL [20]. To test whether this is true or not the data from the PSOBID study were processed using both analysis pipelines and the results compared. Figures 3.9 and 3.10 show the results from both the optimised VBM approach and the DARTEL-VBM approach respectively.

For both methods the same smoothing filter (FWHM: 8mm x 8mm x 8mm) and statistical model were used, i.e. the same study design and covariates were used: these being age, ICV and self reported alcohol levels.

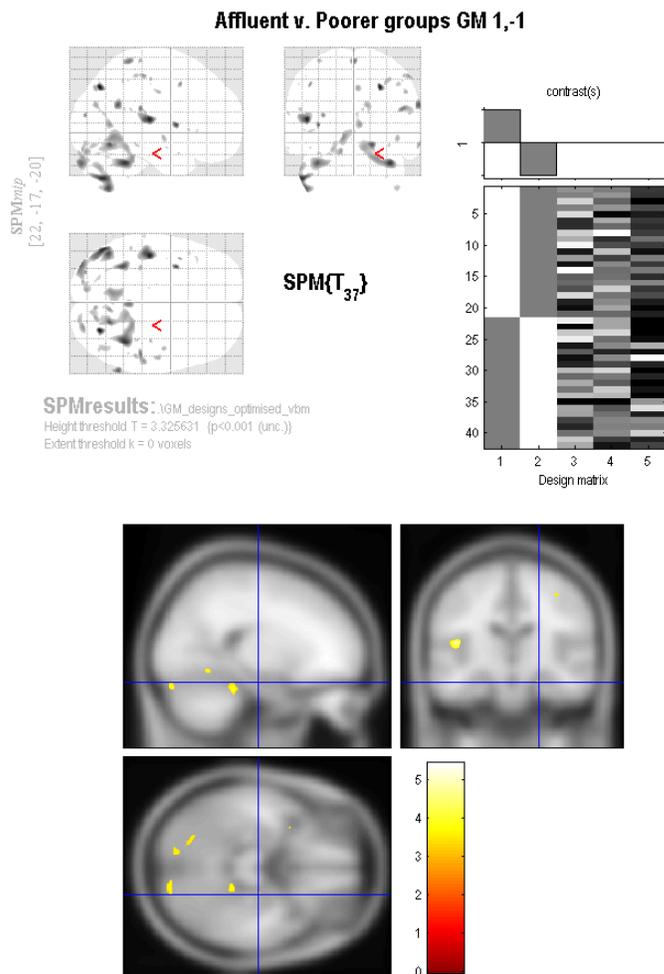


Figure 3.9: Optimised VBM results for affluent vs. poorer groups, GM, t-test

Results of the optimised VBM method and DARTEL-VBM can be compared by inspection of figures 3.9 and 3.10. Notice that the global maxima were deemed more statistically significant i.e. the maximum z-score is greater, when DARTEL-VBM was used. Similarly, using the older optimised VBM approach, there was no statistically significant voxels remaining in the left or right hippocampus following the correction for multiple comparisons whereas both the left and right hippocampus had significant regions following the correction. In general, throughout the brain, more significance and larger clusters were observed when the DARTEL-VBM approach was used.

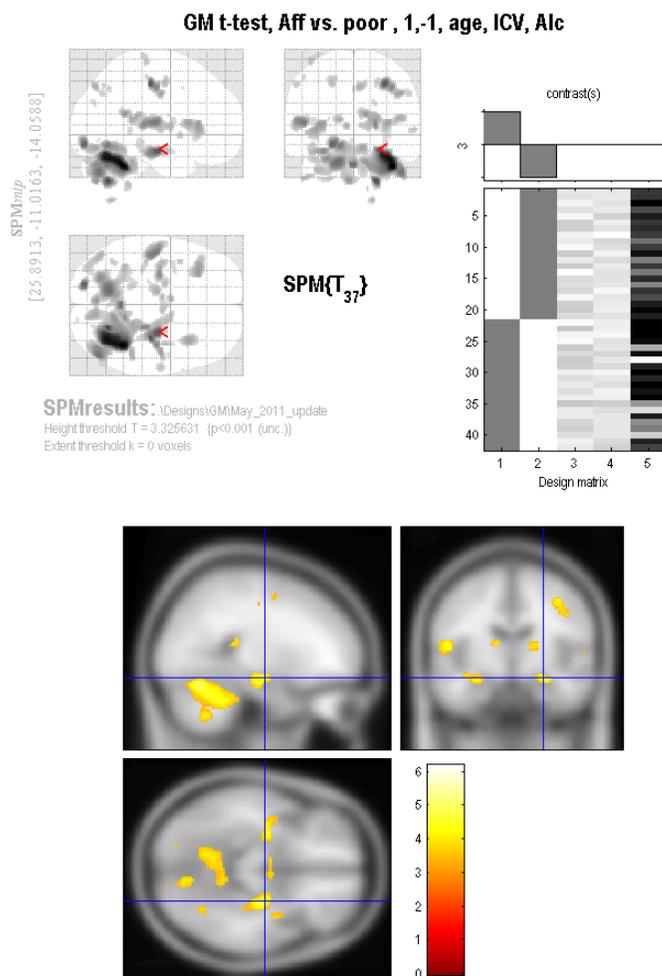


Figure 3.10: DARTTEL-VBM results for affluent vs. poorer groups, GM, t-test

3.6.4 Comparing the results of VBM with and without the NS cluster correction

To illustrate the utility of the NS cluster correcting statistics toolbox. SPM's were created for circumstances where the default FDR parameters ($p = 0.05$) and NS statistics default values were used ($p=0.001$). This was done for data from both the unified segmentation approach to VBM and from the DARTTEL-VBM approach.

3.6.4.1 The utility of NS cluster correction statistics on the DARTTEL-VBM

It has been stated that it is not ideal for the standard SPM multiple comparison corrections statistics to be applied to VBM data. This is because it has been shown that some of

the assumptions necessary to apply the default statistics within SPM are not always valid. Specifically, the assumption to satisfy the RFT that smoothing is consistent throughout the brain has been shown to be violated. Therefore, within the VBM toolbox there is an option to apply a non-stationarity cluster correction.

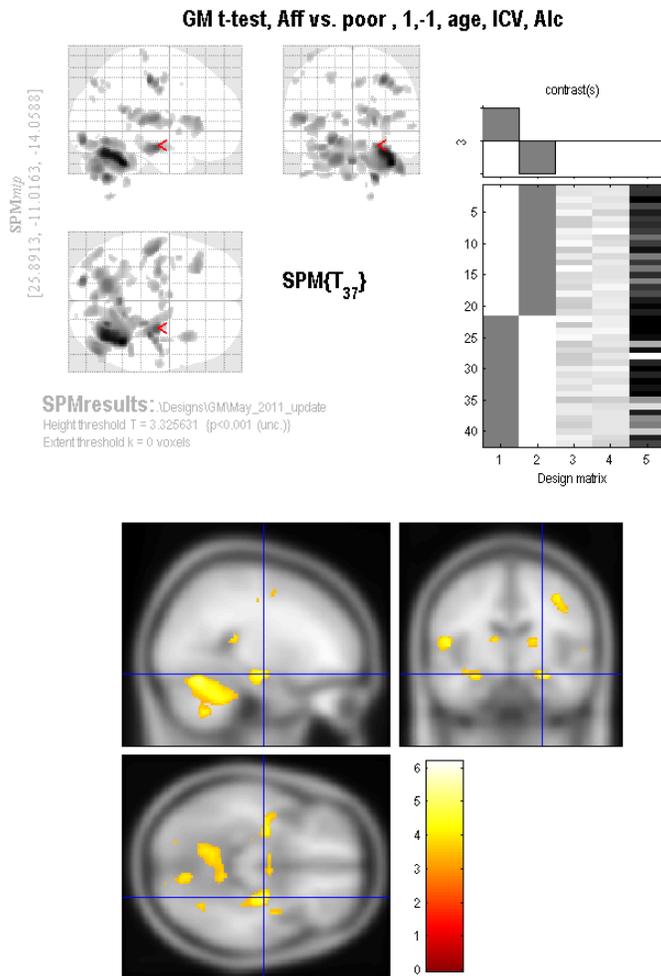


Figure 3.11: NS corrected statistics for DARTEL-VBM

From figures 3.11 and 3.12 we can see that the SPM's from the two statistical approaches are very similar. The NS SPM maps are slightly more refined than the FDR counterparts but essentially the same regions of statistically significant volume differences were identified.

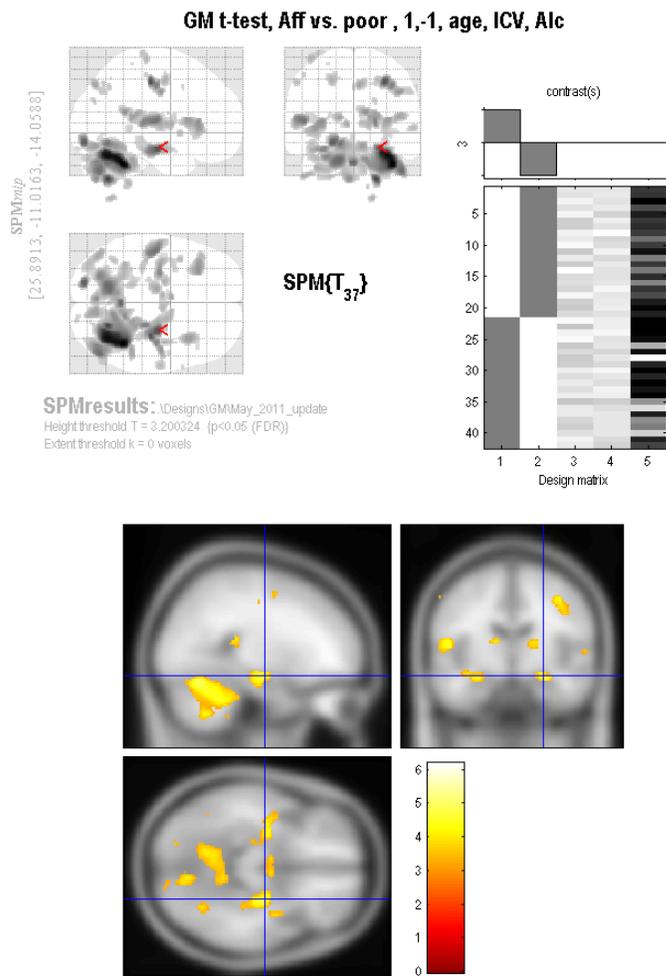


Figure 3.12: FDR corrected statistics for DARTEL-VBM

Thus it would appear that although there is still some improvement in interpreting the data from using NS statistics the effect of this is perhaps less dramatic when it is applied to DARTEL-VBM than previous VBM approaches. This is likely to be because the improved registration offered by DARTEL has reduced the variance in the smoothness across the image.

3.6.5 Post-hoc analysis of other brain regions of statistically significant GM volume differences

Whilst the VBM analysis focused on the hippocampus, as VBM is a technique which can analyse the whole brain, it was also possible to assess other brain regions to see if differences

existed between the affluent and more deprived groups. By applying a whole brain statistical correction to the data, the region volume difference between the groups of greatest statistical significance was the GM in the right cerebellum, see figure 3.13. The model applied here was the same as was applied to the hippocampus discussed above where the model includes covariates for: age, ICV and alcohol intake.

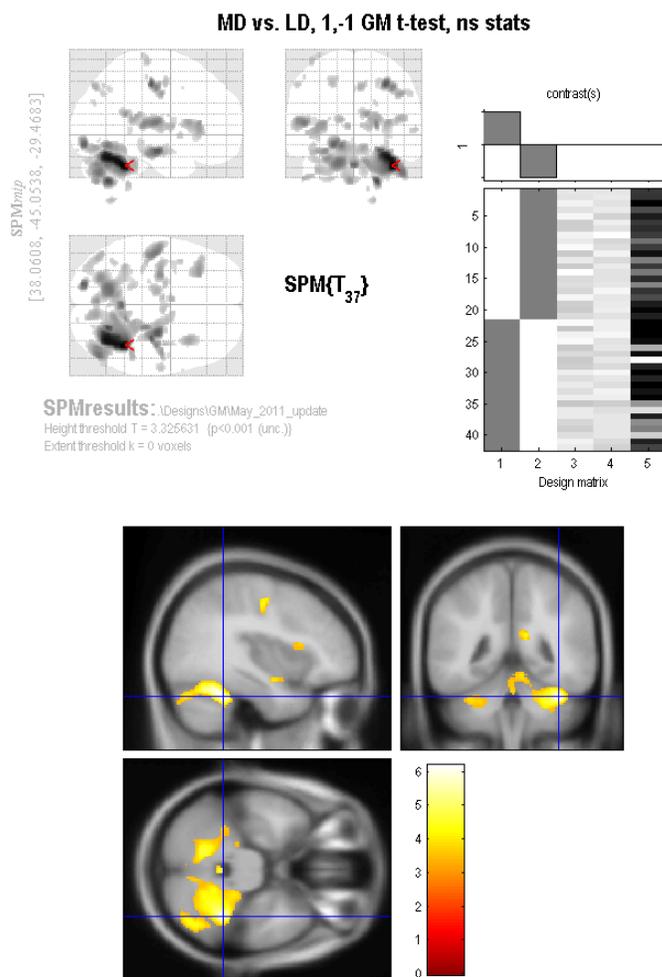


Figure 3.13: VBM GM result. Affluent vs. Poorer group. NS corrected SPM. Right cerebellum.

As can be seen from figures 3.13 and 3.14 there is a large statistically significant cluster of GM volume difference in the right, anterior portion of the cerebellum (FDR $p < 0.005$).

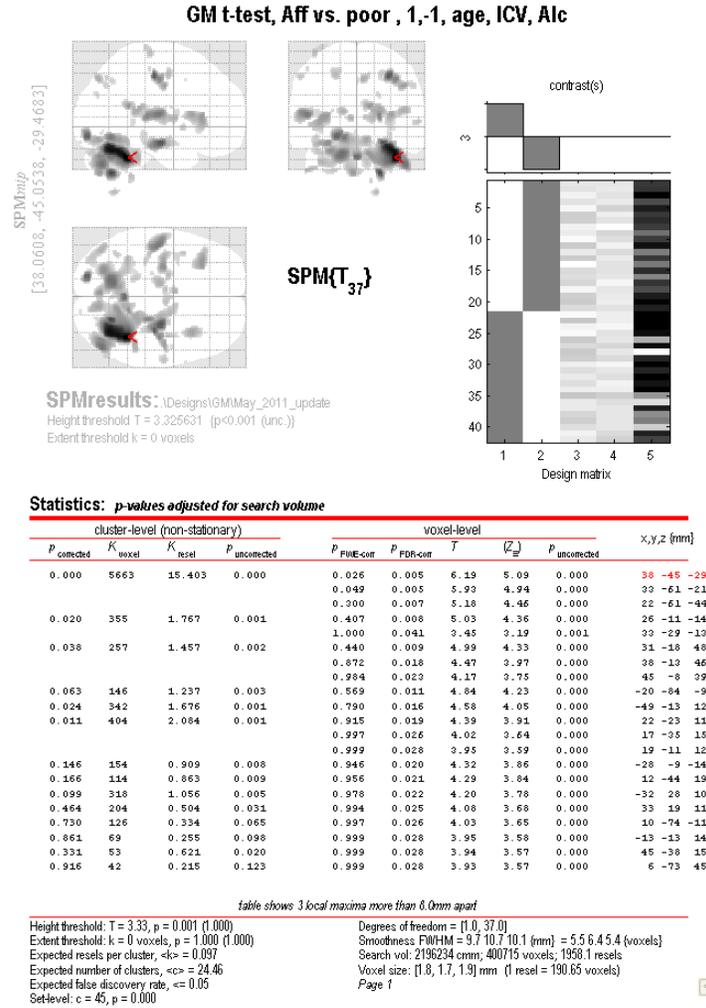


Figure 3.14: VBM GM result. Affluent vs. Poorer group. NS corrected statistics. Right cerebellum.

3.6.6 Post-hoc analysis of whole brain tissue class volumes

As discussed above, during the segmentation procedure as part of the VBM processing pipeline, volumes for the GM, WM and CSF tissue types were determined. While the hippocampal volumes (chapter 2) and GM volume in the hippocampus (chapter 3, above) have been established, it was also of interest to compare the whole brain tissue class volumes for the affluent and poorer groups. From inspection of figure 3.3 it was observed that while the total ICV was not significantly different between the groups, there was, on average, less GM and WM in the poorer group than the affluent group. Statistical tests were done to determine whether this difference was statistically significant. The GM and WM volumes between the two were compared using an ANOVA test. There was no statistically significant global

GM volume difference between the two groups ($p = 0.174$). However, there was less WM globally in the poorer group than in the more affluent group and this was deemed statistically significant ($p=0.029$).

3.7 Discussion

A limitation of this study was that the early life origins of the subjects were not recorded, that is, it is not known whether or not subjects from the least deprived group were born into a more affluent environment. Subjects who would have started life in a poorer environment and have ended up being considered affluent in society are known as positive deviants. In many ways these subjects are themselves a singularly interesting group, in that if one is postulating that developmental effects contribute to eventual social status, then what is it about this group that galvanises them to the stresses to which their birth peers are more susceptible. This area in itself warrants further, more explicit examination in future work.

3.7.1 Discussion of DARTEL-VBM GM volume differences between poorer and more affluent groups

The primary hypothesis of this chapter was to assess GM volume differences between community based samples of affluent and more deprived subjects. To address the primary hypothesis voxel based morphometry, including a novel normalisation method known as DARTEL was used. Statistically significant clusters suggestive of GM volume reduction were observed in both the right and left hippocampi. More specifically the volume reduction was localised to the head of the hippocampus.

It has recently been shown that children from poorer income backgrounds have lower hippocampal GM density, the data was taken from 317 children across the socioeconomic spectrum [156].

DARTEL-VBM was used to determine whether or not inflammatory markers: IL-6 and cortisol might explain the observed reduction in hippocampal volume in the affluent and poorer groups. A multiple regression model was created using VBM with age, ICV, alcohol, IL-6 and cortisol all included in the model. No correlations were found between hippocampal GM and IL-6 or cortisol.

These results are in contrast to what has been observed by Marsland et al [216]. In their work, using the optimised VBM approach to assessing GM, higher IL-6 values were correlated

with smaller hippocampal volumes. The subjects studied were middle aged, normal healthy males taken from the community. 76 subjects were studied which was more than three times as many as were used in the analysis for this thesis, adding significant power to the data. However, the study by Marsland et al did use the optimised VBM method and not DARTEL-VBM which as we have shown above, is less sensitive to volume differences in the hippocampus. Overall, it is difficult to explain why different results were observed between the Marsland et al study and in this thesis.

The relationship between GM volume and cortisol was also considered in a similar way. It has been shown that increasing cortisol levels in thirteen normal, healthy volunteers have been correlated with hippocampal volumes, where increasing manually outlined hippocampal volumes were found to correlate with increasing cortisol [291]. Once again, the work in this thesis failed to replicate this finding. However, as discussed extensively in chapter 2 (section 2.8.4), cortisol measures have been shown to be notoriously variable and therefore much greater confidence would be needed in this measurement before one could reliably assess correlations between brain volume measures and cortisol levels [109].

In the case of both IL-6 and cortisol and their purported relationship to hippocampal volume, much further work on a larger sample size is required to confirm the presence and nature of these relationships. There may well be differences across differing socioeconomic groups with these factors, but the work done in this thesis was not able to determine these robust relationships. Many other inflammatory and metabolic markers may also play a role in affecting the hippocampus, and these should also be assessed in any further work [226].

3.7.2 Comparing results of optimised VBM and DARTEL-VBM

Recalling the results from chapter 2, there was a trend for hippocampal volumes to be smaller on the left side and to a lesser degree on the right side of the brain. Therefore the DARTEL-VBM results presented here were suggestive of agreement with the results observed in chapter 2. Moreover, it could be argued that the DARTEL-VBM was more sensitive to the volumetric change between the groups than comparing the gross hippocampal volumes. This may be a result of the fact that VBM compares volumes down to the level of a voxel (or a few voxels after smoothing) whereas the segmented volume facilitates only a comparison of the gross volume which may be less precise. Thus, given the time penalty involved with manually segmenting the hippocampus, VBM would be a preferred option to performing volume based analysis. While automated methods for hippocampal segmentation are also quicker, these still typically require some manual intervention at some level to correct the volume, though

it could be argued that manual checking and intervention is also necessary with the VBM pipeline. Thus, from the evidence from chapters 2 and 3, there are definitely benefits to having used both approaches where the data compliment each other.

In a recent paper by Bergouignan et al, different VBM methods as well as manual and automated segmentation methods were compared for detecting volume differences within the hippocampus of patients with acute depression compared to normal controls. They found that DARTEL-VBM was an improved method as it was able to detect volumetric differences in the hippocampus, while the standard VBM methodology could not. The automated segmentation, manual segmentation of the hippocampus and DARTEL-VBM were found to have equal sensitivity to volumetric changes [40]. The hippocampal volumetric differences observed between the patient and normal control groups were approximately 10%, this was greater than the volume differences observed in this thesis for the affluent and more deprived groups which were between 6% and 9% which might explain why the trend observed in the volumetric differences between the manually segmented right hippocampus (i.e. smaller difference) was not as strong, but DARTEL-VBM was able to detect areas suggestive of grey matter volumetric differences in this area.

3.7.3 Discussion of post-hoc analysis on DARTEL VBM results across the whole brain

The most statistically significant region suggestive of GM volume reduction between the affluent and more deprived groups was observed in the cerebellum, specifically the right anterior lobe of the cerebellum. There is evidence in the literature indicating that the cerebellum is a part of the brain which is particularly vulnerable to environmental factors [135]. A study has recently shown that neglected children have a smaller cerebellar volume and that larger cerebellar volumes mediated improved performance on memory and planning tasks [33]. Thus, the cerebellum GM volume deficits observed in the poorer group may be the result of a cocktail of environmental factors inherent to their lower social class, however, a larger, more refined study much is necessary to confirm and understand the differences observed in the pilot study performed in this chapter and the underlying mechanisms explaining any perceived differences. Moreover, to further separate the nature and nurture components, or perhaps more correctly, the genetic and environmental factors that moderate and mediate these relationships, neonatal or paediatric development could be investigated to assess cerebellar volumes in early life. No cerebellar volume differences were reported in the study by Hanson et al, 2011, but it is unclear as to whether the cerebellum was considered or whether no differences were found [156]. In a study explicitly assessing cerebellar volumes

in fifty-eight adolescents with PTSD, volume reductions in the cerebellum were found when compared to ninety-eight normal controls [104]. Interestingly, in the study by DeBellis et al, SES was recorded for both the PTSD and normal groups. The SES was significantly lower in the PTSD group, which they state may be partly because of the result of lower IQ related to maltreatment and a poorer level of education. The cerebellar volume differences between the groups in this study were observed over and above the effect of SES which was used as a covariate in the analysis. Whether the environmental factors themselves mediate the cerebellar volume difference via inflammatory or metabolic processes remain to be proven and this area too, merits further investigation .

3.7.4 Discussion of results of global tissue type volume differences

It was determined that, on average, there was less GM and WM globally in the poorer group than in the more affluent counterparts, the WM difference being statistically significant while the GM difference was not. These brain tissue volume deficits were compensated to some degree by an increased volume of CSF in the poorer group and this was why there was no significant difference in the total intracranial volume between the poorer and more affluent groups. The global GM and WM parameters are less often cited in publications, perhaps because most studies focus on particular cortical regions or subcortical structures within the brain. Nonetheless, this parameter is of interest, particularly when assessing socioeconomic factor which might affect a number of disparate regions of the brain. In a study of institutionalised adolescents, global GM and WM were reduced when compared to normal controls. Amygdala volume reductions were also observed. Hippocampi and corpus collusum volumes were also considered but no volume differences were observed [235]. In contrast, in a study using 3T MRI to assess hippocampal volume differences between twenty-two PTSD patients and twenty-two matches controls no significant difference in total brain volume was observed, though smaller volumes were observed at the posterior aspects of the hippocampus in the PTSD group [49].

3.8 Conclusion

VBM methodology was developed to assess whether or not grey matter volume differences existed in the hippocampus between groups of more deprived and more affluent subjects. This was a novel opportunity to assess two differing SES groups from a community based sample.

Both optimised VBM and DARTEL VBM, a method which implements an improved normalisation scheme were used. The DARTEL-VBM method was found to be an improvement over the older, optimised VBM method. Regions suggestive of grey matter volume reductions were localised to both the left and right hippocampi of the more deprived group when compared to the more affluent group. The GLM model to compare the groups included age, ICV and alcohol consumption as covariates.

It has been shown previously that increases in inflammatory markers such as IL-6 and cortisol correlate with decreasing hippocampal volumes in healthy volunteers [217, 291]. This had been posited as supporting evidence for hippocampal volume reductions being the result of chronic environmental stress factors. Both IL-6 and cortisol measures were included in a VBM model of multiple regression but these factors were found not to correlate with grey matter volume in the hippocampus in this case.

During a post-hoc VBM analysis of the rest of the brain, the most significant region of volume reduction in the poorer group was found to be in the grey matter of the cerebellum. This is an interesting finding as cerebellar volume reduction has been observed in subjects suffering from PTSD which it has been proposed may result in brain volume reduction by chronic stress, a similar mechanism which has been proposed in subjects from deprived backgrounds. However, further, larger studies are necessary to lend greater confidence to the initial findings of the pilot work performed in this chapter.

While volume reductions in a group with lower socioeconomic status are suggestive of grey matter volume deficits in important brain regions such as the hippocampus and cerebellum, the mechanisms by which these reductions occur are still unclear. It is hoped the work here will add to the body of knowledge in this area and to guide further, larger studies, with more specific goals around eliciting the mediating and moderating factors between SES and volumetric changes in the brain.

Chapter 4

Magnetic Resonance Spectroscopy (MRS) measures of the hippocampus in groups of differing socioeconomic status

4.1 Abstract

Using the same two study groups from differing socioeconomic backgrounds discussed in chapter 2 and 3, MR spectroscopy was utilised to assess the concentration of metabolites within the hippocampus. The poorer socioeconomic group were considered to have been exposed to chronic stress and therefore via inflammatory processes it was anticipated that the NAA/Cr ratio would be reduced in this group when compared to the more affluent group. To answer this question the LCmodel was used to quantify NAA, creatine and choline containing compounds by the water scaling method. Metabolite concentrations in Institutional Units (IU) were established. As the estimate of creatine did not differ between the two socioeconomic groups, ratios of NAA/Cr and Cho/Cr were subsequently assessed. Both NAA/Cr and Cho/Cr hippocampal metabolite ratios were not significantly different between the two groups. Correlation analysis between NAA/Cr and Cho/Cr hippocampal metabolites and cortisol and interleukin-6 inflammatory markers was also performed. In all but one of the correlations tested there were no statistically significant results. However, having removed a severe outlying point from the IL-6 dataset, a statistically significant correlation was observed between the right hippocampal NAA/Cr ratio and IL-6 ($r = 0.53$, $p = 0.005$).

Further post-hoc analysis was performed to assess changes in hippocampal metabolites with

age and with hippocampal volumes. No significant correlations between the hippocampal metabolites and age or hippocampal volume were observed. In summary, it was not possible to reject the null hypothesis that the NAA/Cr and Cho/Cr ratio would be the same across the two different socioeconomic groups. Furthermore, no linear relationship between hippocampal metabolites and inflammatory markers was observed. Further work, perhaps best performed in animal studies, to better understand the biological processes underlying the effects of stress, inflammation and their effects on hippocampal metabolites may prove useful. This would go some way to confirming the biological model underpinning any subsequent changes. It may be the case here with the two different socioeconomic groups that simply the NAA and Cho hippocampal metabolites tested are not sensitive markers to changes as a result of stress. Alternatively it could be that there are simply no differences between the hippocampal metabolism between the two groups or that the initial hypothesis that the lower socioeconomic group may be treated as being exposed to chronic stress is incorrect or incomplete. New developments in in-vivo imaging technology such as the developments in spectroscopy or in molecular imaging may prove useful in the future in establishing improved biological models and teasing the relationships between stress, inflammation and hippocampal metabolism apart.

This was the first time hippocampal metabolites have been assessed using 3T MRI between two groups of differing SES. While a group difference between hippocampal metabolites was not observed, the metabolite ratios and the variance in those measurements may well inform other, similar projects as to what is likely to be achievable using 3T MRI. On removal of a severe outlying value a statistically significant correlation was observed between the right hippocampal NAA/Cr ratio and IL-6, this initial result should be treated tentatively and it is recommended that the result be replicated in a larger sample size.

4.2 Introduction

The same volunteers discussed in chapters 2 and 3 were also subject to Magnetic Resonance Spectroscopy (MRS) of their hippocampi. The aim was to assess differences between the more deprived and affluent groups to key brain metabolites implicated in the effects of chronic stress.

There are various brain metabolites which it is possible to quantify using MRS. The three main metabolites which can be measured using hydrogen MRS are choline (Cho), creatine (Cr) and N-acetylaspartate (NAA) containing compounds. Other metabolites which are less abundant in the brain have been more tentatively measured, these include: myo-inositol,

glutamate and glutamine. Furthermore, lactate is a compound which can often be measured in the brain in diseases such as stroke, MS and higher grade neoplastic lesions, similarly fatty lipid contributions may be observed in higher grade tumours and in brain abscesses. However, as this study dealt with generally normal, healthy adult brains, both lactate and lipid were not considered further. For a comprehensive review of the metabolites in the brain and their roles observed using NMR, see Govindaraju et al (2000) [143].

4.2.1 Overview of hippocampal metabolites and the relationship to SES and stress

Given that there are few ways of investigating the in-vivo chemistry of the brain, MRS has been an obvious choice to investigate conditions such as depression, schizophrenia and PTSD [47, 347, 264]. There has been little work done to assess metabolites from subjects from different socioeconomic backgrounds and thus the work in this thesis had the opportunity to investigate metabolite levels in the hippocampus of such subjects.

The hippocampus plays a critical role in forming new memories and in memory retention itself. The hippocampus contains the highest concentration of corticosteroid binding sites in the entire brain [225, 311]. Animal studies indicate that chronic stress and the associated increase in corticosteroids are toxic to a subsection of hippocampal cells [310].

In an early work to assess cerebral metabolites in a psychosocial model, male, adult tree shrews were subjected to stress and changes in cerebral metabolites were recorded [101]. Decreases in the forebrain of 13, 15 and 13% were observed in NAA, Cr and Cho concentrations respectively when compared to control animals. In an in-vivo human study in PTSD and normal controls hippocampal NAA was measured. Cortisol levels were recorded in these subjects. NAA values were reduced in PTSD subjects when compared to normal controls. NAA values were also found to correlate positively with cortisol in both PTSD subjects and normal controls. There was no hippocampal volume difference between the two groups [264]. In a further in-vivo human study, hippocampal volumes and NAA values on patients on corticosterone therapy were compared to normal control subjects. Patients on corticosteroid (asthma and rheumatoid disease) were found to have smaller hippocampal volumes and lower levels of NAA when compared to control subjects. Poorer cognitive scores and higher ratings on the Hamilton rating scale for depression and brief psychiatric rating scales were also recorded [65]. Brown et al suggest that these results are evidence that exposure to corticosteroid affects hippocampal volume and functioning. In a further study assessing MRS metabolites in patients with PTSD, NAA/Cr was reduced in both the right and left

hippocampus in the absence of volume deficits. PTSD was also associated with reduced NAA in the right ACC [316]. This study replicated earlier work done by the same group [317].

It does therefore appear that across a range of studies there is a reduction in NAA in different parts of the brain, including the frontal brain, hippocampus and ACC as a result of stress. This supports the idea that inflammatory responses may in some way play a role in affecting the underlying metabolism on the brain. However, the nature of the relationship between MRS metabolite differences and brain volumetric differences remains unclear.

4.3 Aims and hypotheses

The aim of this MRS component of the PSOBID study was to assess NAA, Cr and Cho, and in particular, the NAA/Cr and Cho/Cr ratios between the affluent and poorer socioeconomic study groups. To investigate this the author implemented an MR imaging and spectroscopy acquisition protocol, trained radiographers to acquire the data or in some cases acquired the data himself. The author also performed all of the subsequent analysis on the imaging and spectroscopy data.

4.3.1 Primary null hypotheses

NAA/Cr in the hippocampus will not be lower in the poorer group than the affluent group

4.3.2 Secondary null hypotheses

Cho/Cr in the hippocampus will not be lower in the poorer group than the affluent group

4.3.3 Additional research questions

Hippocampal concentrations of NAA and Cho will be correlated with inflammatory markers implicated in mechanisms of stress, these are: cortisol and IL-6. Furthermore, correlations between hippocampal volume and NAA/Cr and Cho/Cr metabolite ratios will also be assessed.

4.4 Literature review and theory

4.4.1 Overview of MRS

MRS evolved from its parent technique known as Nuclear Magnetic Resonance (NMR). NMR

was first developed in the 1950's and is still a widely used technique in modern chemistry departments and industry. It is mainly used to identify the chemical constituents and purity of substances as well as the molecular structure of samples. NMR exploits the fundamental magnetic spin properties of atomic nuclei. According to quantum mechanics subatomic particles such as protons, electrons and neutrons have associated spin values. In some atoms such as ^{12}C and ^{16}O the spins are paired and therefore cancel each other out so that the overall atom has a net spin of 0. However, an atom such as such a hydrogen ^1H with a single proton nuclei has a net spin value of $1/2$. If these atoms are placed within a large external magnetic field quantum mechanics states that the nuclear magnetic moment of the nucleus will align in one of two ways i.e. atoms will align parallel or anti-parallel to the external magnetic field. For a single nucleus of spin equal to $1/2$ only a single transition is possible between the two energy states. The energetically preferred option is for the magnetic moment to be aligned parallel to the external magnetic field (spin $+1/2$) whereas the higher energy state is orientated anti-parallel to the external magnetic field. It is worth noting that the rotational axis of the spins does not align exactly parallel (or anti-parallel) to the external field but precesses about the field at the Larmor frequency. By introducing an external radiofrequency (RF) source at the Larmor frequency the proton will absorb the energy and be promoted to a less favourable energy state. The process of matching the applied RF frequency to the precessional Larmor frequency is known as inducing resonance state. For ^1H nuclei placed in an external magnetic field strength of 3T (as was used throughout this thesis) the Larmor frequency is 128MHz.

Spin states which are orientated parallel to the external field are lower in energy than in the absence of the external field. Spins states orientated anti-parallel to the external field are in a higher energy state than in the absence of the external field. Where the energy separation exists, as it does when the nuclei is placed within an external magnetic field, it becomes possible to induce transitions between the two spins states. By irradiating the nuclei with RF energy of the correct frequency a nucleus with a low energy orientation may be induced into a transition to a higher energy state.

It is important to note that in a nuclei such as ^1H , the nuclei will be distributed throughout both the lower and higher energy states. As the energy states between the states is comparatively small, energy from thermal collisions is sufficient to place many nuclei into higher energy spin states. This relationship is governed by the Boltzmann distribution, see equation 4.1.

$$N_{upper}/N_{lower} = e^{-\gamma B_0/kT} \quad (4.1)$$

Where N_{upper}/N_{lower} is the population ratio, γ is the gyromagnetic ratio, B_0 is the external magnetic field strength, k is the Boltzmann distribution and T is the temperature.

Therefore, for an 1H nuclei at room temperature and an external field strength of 3T this equates to a population ratio of 0.999868 i.e. the upper and lower spin states are almost equally populated with only a very small excess in the lower energy state that represent spins aligned parallel to the applied magnetic field.

The fact that the population distribution across the two states are very similar has an important consequence. The signal intensity involved that can be observed is proportional to the population difference between the two energy levels. In NMR where the energy difference between the two states is very small then at room temperature (or body temperature), in quantum mechanical terms, the technique is considered to be insensitive. Thus in both NMR and in MRI the relatively weak signals observed need to be signal averaged to obtain suitable signal to noise.

Despite the insensitivity at the quantum mechanical level NMR is still a useful tool for assessing the constituents of chemical samples though this is limited in medical research by the need for samples to be assessed ex-vivo. Throughout the 1970's the principles of NMR were exploited and then developed to include schemes for spatial localisation which enabled NMR to become a technique used to perform in-vivo MR spectroscopy and ultimately to the birth of MRI itself.

As we know from MRI, we are concerned with imaging the 1H protons which are in large abundance in the body due to the high water concentration of human tissue. However, the number of protons which actually contribute to the MRI signal is comparatively smaller, as only a small number of protons at any one time actually achieve a higher energy state when placed in the static magnetic field. A measurable signal can be obtained nonetheless. The basis of this signal is that when placed in a large static magnetic field a greater number of 1H protons align with the field than are aligned anti-parallel. These 1H protons have a characteristic spin known as the Larmor frequency, where the frequency is proportional to the static magnetic field strength. If we finish this description by adding a radio frequency pulse at this characteristic frequency this completes the description of how the 1H protons are stimulated such that an MR signal may subsequently be obtained.

We can now modify and extend the description above to explain the phenomenon of chemical shift. Returning to the 1H protons placed within a static magnetic field, for 1H protons belonging to the abundant H_2O molecules, these will be resonating at the characteristic

frequency of $\sim 64\text{MHz}$ at 1.5T and 128MHz at 3T . If we consider the structure of a basic atom, with the positive charge at the centre and the circulating electron cloud possessing a negative charge surrounding it, then the circulating negative charge is like a small current that exerts its own magnetic field. This alters the magnetic field which the central proton experiences and in turn modifies, albeit to a small degree (about 1 part per million, i.e. 1Hz for every 1MHz), the characteristic frequency of the ^1H proton. Thus the characteristic frequency of the ^1H protons is altered as a result of its fundamental chemistry (i.e. its electron orbits). Extending this concept still further, as different molecules possess different chemical bonds, these molecules will exhibit different chemical shifts, or to put it another way, these chemical shifts will be specific to the molecule that created it. Within the human brain there are many chemical processes occurring all the time, such chemicals involved in metabolism are known as metabolites. Many of these metabolites possess ^1H -protons though they are typically present within more complicated molecular structures. Furthermore, signal contributions from ^1H protons metabolites are in much lower abundance than that of water and so the concentration of brain metabolites are of the order of several mM, with water having a concentration of 55M . Thus, in order to visualise and assess the relative contributions of brain metabolites the contributions from water protons must be suppressed. The individual spectral contributions from the key metabolites of interest are discussed in greater detail below.

An example spectrum taken from the hippocampus is shown in figure 4.1 on the following page.

4.4.2 Metabolites of Interest

4.4.2.1 NAA

NAA is an amino acid and is present in relatively high concentrations in the normal brain. NAA's role in the brain is not completely understood though it is believed to act as an osmolite (an organic compound affecting osmosis), a storage form of aspartate and a precursor to NAAG [43],[348]. Due to its relatively high natural abundance in the brain NAA has been much studied and is believed to be a marker for neuronal density, though this is not considered a complete description as NAA concentrations differ among neuron types [325]. It has also been observed in other cell types [350] and dynamic changes of neuronal concentrations have also been observed indicating that NAA levels may reflect neuronal function and dysfunction as well as simply cell loss.

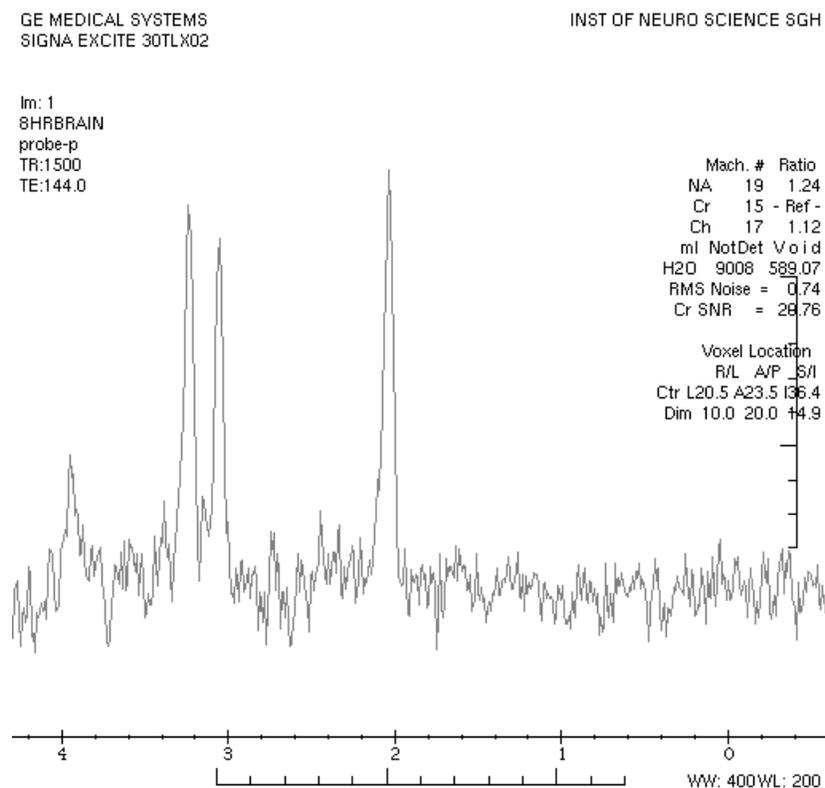


Figure 4.1: An example of an MRS spectrum from the hippocampus, PRESS, TR 1500ms, TE 144ms

Examples of NAA changes observed in disease include: the restoration of NAA levels following an ischemia insult [66] and brain injury [334]. Decreased NAA levels have also been observed in multiple sclerosis though neuronal loss itself was not observed [348]. The chemical structure of NAA is shown in figure 4.2 on the following page.

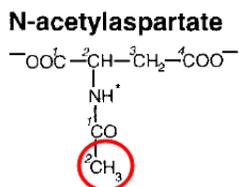


Figure 4.2: Chemical Structure of NAA

NAA has seven ¹H protons which contribute to the NMR signal, the most dominant of which is the signal from the 3 protons of the 2CH₃ group that produces a singlet resonant peak at 2.01ppm. This is the signal which is most commonly observed in ¹H MR spectroscopy.

and is a marker for the energetic status of cells. The synthesis of creatine takes place in the kidneys and the liver. In-vivo MRS measurements of the brain detect a combined creatine and phosphocreatine peak as a singlet resonance at 3.03ppm as a result of their methyl (CH₃) protons. Contributions in the normal human brain range from 4.0 - 5.5mmol/Kgww for phosphocreatine and between 4.8 and 5.6mmol/Kgww for creatine [115], where the total creatine signal has been reported to be higher in grey matter at 6.4-9.7mM than in white matter at 5.2-5.7mM [187, 288]. The creatine peak is relatively stable and no change in concentration has been observed with age [312]. Therefore in many in-vivo measurements assessing NAA and choline levels, these metabolite concentrations are often presented as a ratio with respect to the Cr+PCr peak. However, it should be noted that one must take care when dealing with metabolite ratios as it is not always the case that Cr+PCr is unchanged in disease. In particular it has been observed that Cr+PCr is decreased in tumours and stroke and may be increased in myotonic dystrophy (muscle wastage) [81].

As well as the methyl proton singlet peak at 3.0ppm, Cr + PCr also has a prominent methylene (CH₂) peak at 3.9ppm. In both instances it is not possible to resolve, in-vivo, Cr and PCr. The chemical structures of Cr and PCr are shown in figure 4.4 below.

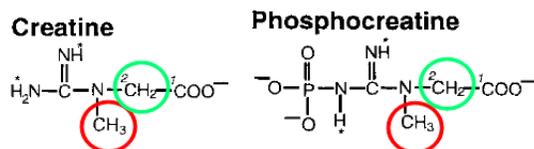


Figure 4.4: Chemical structure of Cr and PCr: methyl (CH₃) group is circled in red, methylene (CH₂) group circled in green. Note the Cr + PCr methyl group produces a singlet at 3.0ppm and the methylene group produces a singlet at 3.9ppm.

4.4.2.4 Choline (Cho), Phosphocholine (PCho) and Glycerophosphorylcholine (GPCho)

The singlet choline peak observed in MRS examinations at 3.2ppm is a result of contributions from free choline, glycerophosphorylcholine and phosphorylcholine, so we will consider it as the 'total' choline peak [342, 239, 238]. The total concentration of choline in the brain ranges from 1-2mmol/Kgww, though contributions across the brain may vary [288, 364, 342]. The

main contributors to the total choline signal are PCho and GPCCho, where the concentrations of these metabolites have been measured using ^{31}P MRS as 0.6mM and 1.0mM respectively [342, 46]. Free choline contributes little to the signal in the normal brain with a concentration of <0.03 mmol/Kgww, though this may be significantly increased in tumours [239].

Choline is an essential nutrient which is obtained from phospholipids through the diet. It is necessary for the synthesis of the neurotransmitter acetylcholine, and of phosphatidylcholine, a major constituent of cell membranes. There are 13 non-exchangeable protons in choline, 9 from a trimethylamine group ($\text{N}(\text{CH}_3)_3$) and 4 from two methylene (CH_2) groups. The 9 protons of the trimethylamine group are magnetically equivalent and thus all contribute to the singlet peak at 3.19ppm. GPCCho has a total of 18 protons from its glycerol and choline moieties. The trimethyl protons resonate at 3.21ppm as a singlet. Similarly to Cho, PCho also has 13 protons, these again contribute to the singlet peak observed at 3.21ppm as a result of the trimethyl protons. The chemical structures of Cho, PCho and GPCCho are shown in figure 4.5 below.

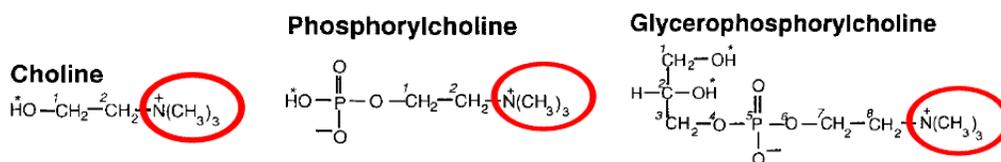


Figure 4.5: Chemical structure of choline, PCho, GPCCho. Circled in red are the trimethylamine groups with nine protons which contribute to the 'total choline' singlet resonance at 3.20ppm.

Given that the primary singlet peak observed in MRS spectra around 3.20ppm are a result of three different metabolites, this can make the biochemical interpretation of changes in this peak difficult. Changes are generally attributed to alterations in membrane composition, be it from increased signals observed in neoplastic lesions, ischemia, head trauma and MS and decreased signals associated with liver disease and stroke [305].

Throughout this chapter, from here onwards, the term NAA will be used to describe both NAA and NAAG, similarly the term Cho will be used to describe Cho, PCho and GPCCho, and the term Cr will be used to describe Cr and PCr.

4.5 Methodology

4.5.1 Volunteers, materials, methods and study design

Study design, subject recruitment, inclusion and exclusion criteria have been described previously in chapter 2.

4.5.2 Acquiring anatomical MR images on which to plan the MRS voxel

One limitation to implementing MRS on the GE Signa HD system was that the MRS voxel had to be planned parallel to the orientation of the planning image acquisition. That is, it was not possible to plan oblique voxels by rotating the voxel on the MR console. This was particularly restrictive when attempting to acquire spectra from the hippocampus as it typically lies on a steep angle within the temporal lobe. The solution to this problem was to acquire MRS structural planning acquisitions on angles relative to the hippocampus itself. Axial acquisitions were tilted onto an oblique plane such that they were parallel to the hippocampus with coronal acquisitions also being tilted onto an oblique plane such that they were perpendicular to the hippocampus. This not only improved visualisation of the hippocampal structure but also enabled MRS voxels to be planned such that the large proportion of the hippocampus was within the voxel, minimising the contributions to spectra from the CSF.

4.5.3 MRS acquisition protocol

There were several decisions made on imaging parameters in order to optimise the MRS acquisition to meet the aims of the study, these are discussed below.

4.5.3.1 Pulse Sequence: PRESS or STEAM

Historically, two pulse sequences have been used for localising MRS data, the Point Resolved (PRESS) and the Stimulated echo (STEAM). Both sequences have their relative merits, for example, PRESS has an inherently greater SNR as a result of the fact it employs three 90° localisation pulses whereas STEAM was deemed to have better spatial localisation of the prescribed voxel. STEAM had also been preferred by many for acquiring short TE data. In recent years however, PRESS sequences have improved at shorter TE values, with good

spatial localisation and superior SNR so it is now broadly accepted as the preferred pulse sequence in most circumstances and thus a PRESS sequence was chosen to be used here.

4.5.3.2 TE

Due to the varying T2 relaxation time of the brain metabolites, different metabolites are observed when different echo-times are used. For example, for a short TE sequence (typically 35ms) acquired in the brain will show NAA, Cr and Cho as well as contributions from Glu, Gln and myo-inositol whereas a longer TE sequence (typically 144ms) will tend to only show contributions from NAA, Cr and Cho. While one might instinctively assume that as more metabolites are present on shorter TE acquisitions that this would be the better choice of echo time, the presence of Glu and Gln can make the baseline of the MRS acquisition more difficult to identify and thus make it more difficult to measure the NAA peak as these peaks lie in close proximity. Moreover, there is still a much greater body of evidence of MRS data in the literature acquired at longer TE's such as 144ms, and historically at our institution, 144ms acquisitions have been most commonly used both for clinical and research purposes. Therefore, for these reasons, a TE of 144ms was chosen such that NAA may be measured in the optimal way.

4.5.3.3 TR

A repetition time of 1500ms was chosen for the PRESS pulse sequence. This was a compromise between obtaining a good estimate of the metabolites of interest, while ensuring the acquisition time was as short as practically possible. The 1500ms repetition time has been used successfully for a number of years at our institution for clinical cases and was therefore deemed appropriate for this study.

4.5.3.4 Voxel size and placement

One of the key problems in trying to acquire MRS data from the hippocampus is that the hippocampus is irregularly shaped. As MRS voxels can only be cuboidal in shape it is difficult to achieve good conformance between the voxel and the hippocampus. To ensure the majority of the signal in the resultant spectrum was the result of stimulating protons from the hippocampus it was crucial to correctly place the MRS voxel. To achieve this a clear description of the hippocampus and its boundaries was created.

The hippocampus is an elongated structure with a distinct head, body and tail, with the longest axis being in the anterior-to-posterior direction. Therefore, the size for the hippocampal voxel was chosen to be 15mm (S/I) x 10mm (R/L) x 20mm (A/P). While the

hippocampus itself can be longer than 20mm, the voxel was chosen such that it would fit within the head and body of the hippocampus, avoiding as far as possible regions of CSF. The 15mm, superior-inferior size of the voxel was chosen to include the hippocampus, in some circumstances the voxel was too large and as such included parts of the parahippocampal gyrus and entorhinal cortex. This compromise was acceptable and retaining the total volume of the hippocampal voxel was important to achieve a good spectral signal-to-noise ratio.

4.5.3.5 Number of signal averages

To improve spectral quality by increasing SNR many signal averages are typically acquired to create the resultant MRS spectra. Increasing the number of signal averages comes at the cost of increasing scan time. For most brain examinations 128 signal averages is appropriate to obtain a good quality spectrum, however, from some initial trials it was decided that 256 signal averages was necessary to obtain good quality data from the hippocampus. The time for the MRS acquisition given the TR of 1500ms and 128 averages was: 7mins 45s. An example of the MRS voxel placement within the hippocampus is shown in figure 4.6 below.

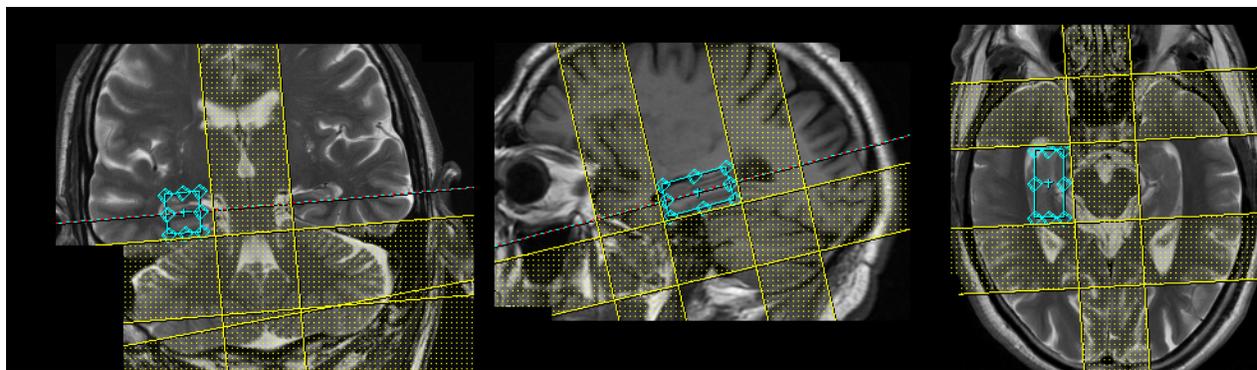


Figure 4.6: PSOBID MRS planning images: coronal, sagittal and axial planes

4.5.4 Linear Combination (LC) model

The LCmodel (version 6.1) is a well established and widely used MRS analysis package written by Stephen Provencher [289]. It works by comparing the acquired spectroscopy data to a modelled linear combination of basis sets of metabolite spectra, where the basis sets have been established for all common MRI manufacturer scanner types, field strength and

typical TE values. Subsequently, an output of concentration values, confidence/error (in the form of Cramer-Rao bounds) are presented along with a graphical representation of the raw, modelled and residual data, where the residual data are the least squares differences between the raw and modelled data. An example of the one-page output from the LC-model is shown in figure 4.7.

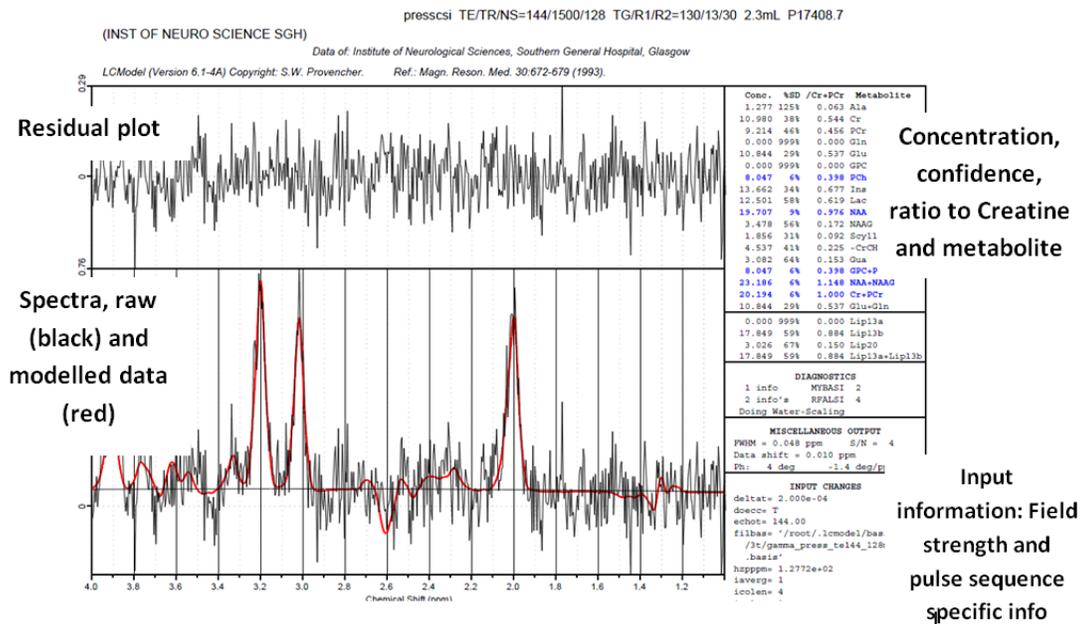


Figure 4.7: LCmodel output from the PSOBID study

The LCmodel was quick and easy to use, and while this may be perceived as a benefit, the ease of use can be deceptive with much of the data processing being performed hidden from the operator, though the LCmodel does come with a comprehensive manual. Robust, absolute quantification of MRS metabolite values are notoriously difficult to perform within practically acceptable scanning times, where often far reaching assumptions need to be made. Thus, it is common for MRS metabolite concentrations to be presented in 'Institutional Units (IU)', this makes it difficult for multi-centre studies to be performed using MRS. Similarly, care must be taken when comparing results to IU or absolute concentrations of metabolites from the peer review literature. However, within centre studies and metabolite ratios do not suffer to the same degree from the aforementioned limitations of absolute quantification and thus robust study designs can still be performed.

There are two options available within the LCmodel which can be applied to obtain absolute quantification of metabolite concentrations. One uses the internal water reference as a means

of scaling the metabolite concentrations, another uses an external calibration standard which may then be used to scale the metabolite values appropriately. Both of these methods have their own advantages and disadvantages.

Use of an external standard or calibration phantom has been widely applied throughout the MRS community [238]. This is where a known concentration of a particular metabolite or metabolites is scanned at a separate time from the actual MRS acquisition. A scaling factor would then be determined to convert the observed metabolites from the subject under study to absolute concentrations in milli-mol (mM). One of the disadvantages of this method is that the calibration factor needs to include factors for voxel size (if it is different between the calibration scan and subjects scan) and changes in total gain due to coil loading.

An alternative to the external phantom calibration method is to use the water scaling method [30, 329]. In practical terms, the largest benefit from using the water scaling method is that no additional acquisition or calibration phantom is necessary. The MRS acquisition data itself is used to perform the scaling. A further benefit of this scaling method is that voxel size corrections, gain correction from coil loading and other hardware factors are inherently corrected. The major disadvantage to performing calibration by this method revolves around the assumptions made about the unsuppressed water peak. The underlying assumption is that water concentration is approximately the same between the hippocampi of different subjects. Moreover, the water concentration in the hippocampus is assumed to be approximately the same as the water concentration which was determined for the basis set. Thus, it is difficult to know with a great deal of certainty the water concentration and the attenuation of the water peak due to relaxation effects. Despite this, these factors will present themselves as systematic errors and while it would be wrong to present the results in mM, data can be presented in IU, which are reliable provided the acquisition scheme is applied consistently throughout the study.

While the implementation of both of these methods are described within the LCmodel manual, it became apparent that the only option which was feasible with the GE hardware available for the study was to use the internal water referencing method. That is, the external calibration standard method was not recommended when scaling data acquired from a 3T GE phased array coil as it does not accurately represent the behaviour of the gain correction at 3T. Therefore, the internal water referencing method was used.

Even if the choice had not been made due to the hardware limitations, it is likely that the internal water referencing method would have been the preferred option. As this method of scaling correction uses the water value from the standard MRS acquisition, there is no need

for additional scans and therefore additional imaging time, in addition, gain corrections are implemented in the LCmodel. More commonly concerns might be raised with the internal water referencing method when it is being used to assess gross pathology such as tumours where the water value may vary significantly making any subsequent scaling equally erroneous. However, for the purposes of this study we are imaging normal healthy brains with no gross pathology and therefore utilising the internal water referencing method with the LCmodel was considered appropriate.

When implementing the water scaling method with the LCmodel, the program adjusts the data by a scaling factor f_{scale} to scale the data consistently with the basis set. This ensures that the signal strength per proton resonance is the same for both datasets. This is the ratio of the normalised signal strength in the basis set to the normalised unsuppressed water signal of the data.

$$f_{scale} = Basis_{norm}/Water_{norm} \quad (4.2)$$

A normalised signal is the measured water resonance area divided by the concentration of protons and by the the factor by which the measured signal is attenuated as a result of relaxation.

$$Basis_{norm} = Area_{met}/[N1HMET * Conc_{met} * ATTMET] \quad (4.3)$$

Where $Area_{met}$ is the observed resonance area of the selected singlet in the basis set, by default this is Cr CH₃. $N1HMET$ is the number of equivalent protons contributing to the singlet, $Conc_{met}$ is the model metabolite concentration in the basis set and $ATTMET$ is the attenuation factor of the singlet used in the in vitro basis set spectra.

Similarly,

$$Water_{norm} = Area_{water}[2 * WCONC * ATTH20] \quad (4.4)$$

$N1HMET$ and $Conc_{met}$ are known, the two area terms ($Area_{met}$ and $Area_{water}$) are computed by the LCmodel from the spectra. However, $ATTH20$, the attenuation of the NMR visible water spectra due to relaxation effects and $WCONC$, the NMR visible water concentration in the voxel, have to be estimated.

As $ATTH20$ and $WCONC$ are difficult to guess correctly this means when trying to establish absolute concentrations that all of the results will be incorrect by an unknown factor. To

achieve concentration values in absolute units of mM, ideally corrections for T1 and T2 relaxation must be implemented, or the acquisition scheme must be appropriate such as to minimise the contribution of these factors, typically this would involve using a short TE < 35ms and TR > 3000ms. However, for reasons such as achieving a flatter baseline at longer echo time (TE =144ms) and for the practicality of the total time subjects can be in the scanner, a TR of 1500ms was used. Thus one might correct for T1 and T2 retrospectively, but this in itself is a non-trivial task. The compromise that was reached from considering all of the above was to ensure the imaging parameters, the TE and TR are consistent throughout the study, thus while our measurements of concentration may differ in absolute terms from other values obtained in the literature, the comparison between the institutional units for the two differing socioeconomic subject groups was viable.

An important advantage worth noting about how the LCmodel works is the fact that it uses basis sets, scaled with respect to the water value to obtain concentrations of metabolites. Basis sets provide a comprehensive model for the data acquired and as such, incorporate the effects of secondary, contributing metabolite peaks. This is a more comprehensive and powerful approach to describing the spectra and improves the validity of the data taken from the LCmodel over simply a description of only the primary metabolites such as: NAA, Cr and Cho. This method is superior to, for example taking only the ratios of the concentrations as determined by the on-board GE software which only considers the primary peaks of interest.

4.6 Results

Metabolite values as determined by the LCmodel were only used in further statistical analysis when the model itself had a significant degree of confidence in its measure of the metabolites, i.e. where the model itself was unsure of the contribution for a particular metabolite, the value was not used. This measure of confidence was obtained from the LCmodel output. Standard deviation of metabolite values with Craemer-Rao bounds of less than 20% were determined to be of good quality. This ensured that subsequent statistical tests were performed on data of sufficient quality. Therefore, having removed the poorer quality datasets, there remained 17 MRS datasets from the affluent group and 14 from the poorer group for comparison. The results of the affluent vs poorer group comparison in IU and metabolite ratios are shown in figures 4.8 below and 4.9 on the following page respectively.

4.6.1 Comparing NAA/Cr and Cho/Cr between the affluent and poorer groups

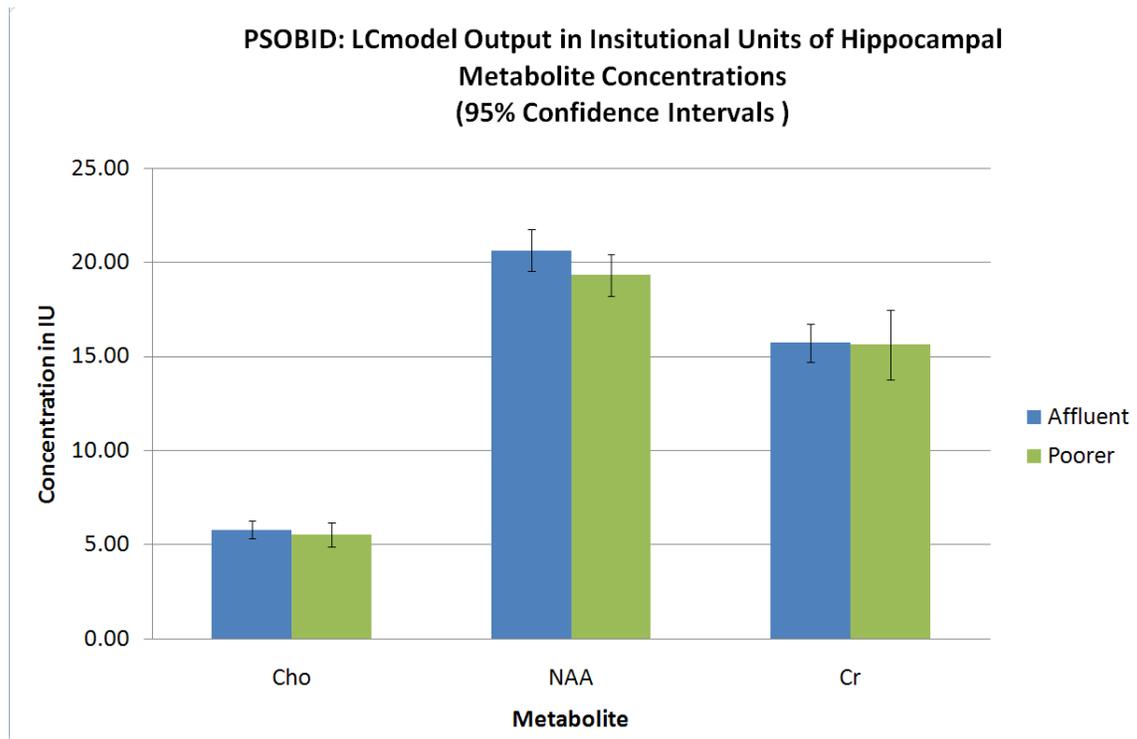


Figure 4.8: PSOBID study: Metabolite concentrations as determined by LCmodel in Institutional Units

The NAA/Cr and Cho/Cr ratios were tested for normality and to a good approximation, both datasets were found to follow the normal distribution. Using a one-way ANOVA the NAA/Cr and Cho/Cr for the right and left hippocampi were compared between the affluent and poorer groups. There were no statistically significant differences between the two groups.

4.6.2 Assessing correlations between hippocampal metabolites and inflammatory markers

Scatterplots of the NAA/Cr and Cho/Cr against IL-6 and cortisol for both the left and right

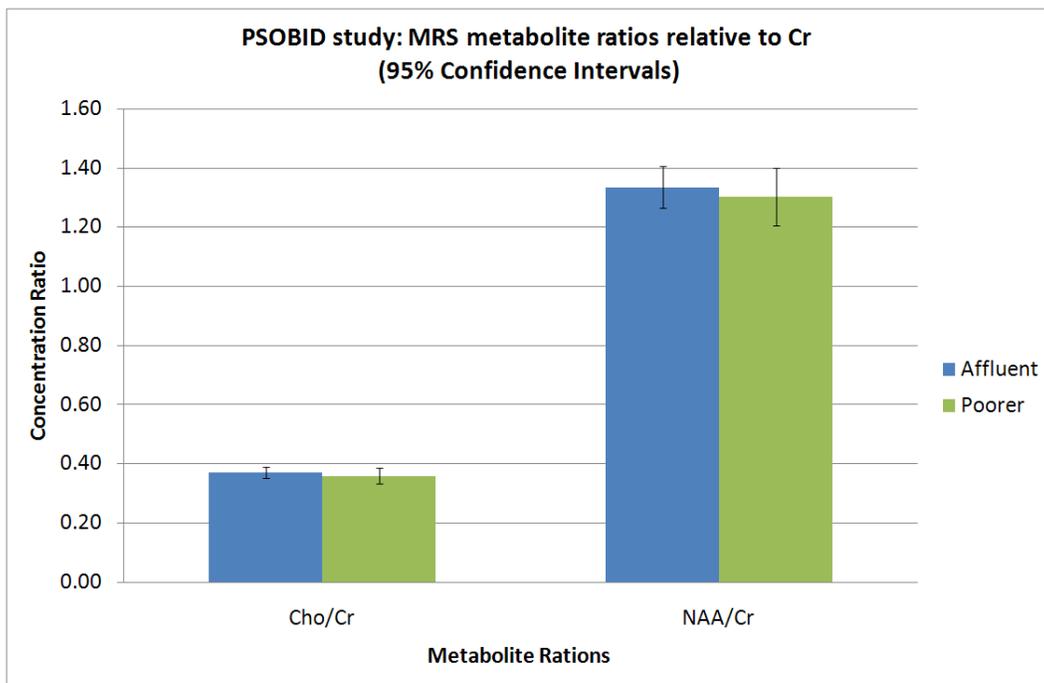


Figure 4.9: PSOBID study: Metabolite concentrations as determined by the LCmodel presented relative to Creatine

hippocampus of the affluent and poorer groups are shown in figures 4.10, 4.11, 4.12 and 4.13 on the following pages.

4.6.2.1 NAA/Cr vs IL-6

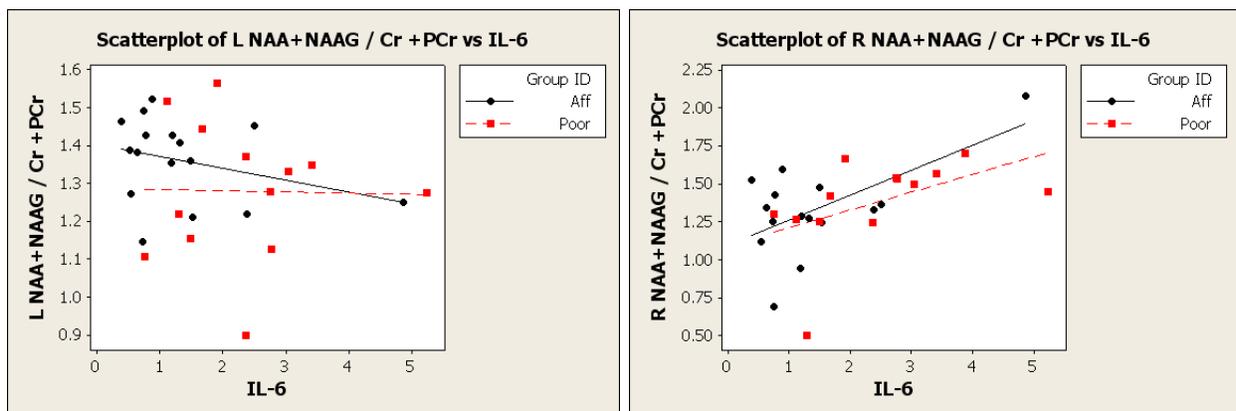


Figure 4.10: Left and right hippocampal NAA/Cr vs. IL-6

Comparing the NAA/Cr ratio with the IL-6 data, there was one significant outlier in the IL-6

dataset, with a value of 25.2, compared to the next highest value of 5.2. This outlier was having a disproportionate affect on the correlation analysis and was therefore removed. Having removed the outlier, no significant correlation was observed between the left hippocampal NAA/Cr ratio and IL-6, however, there was a significant correlation between the right hippocampal NAA/Cr ratio and IL-6 ($p = 0.005$, $r = 0.53$). The strength of the correlation was similar between the affluent and poorer groups.

4.6.2.2 Cho/Cr vs. IL-6

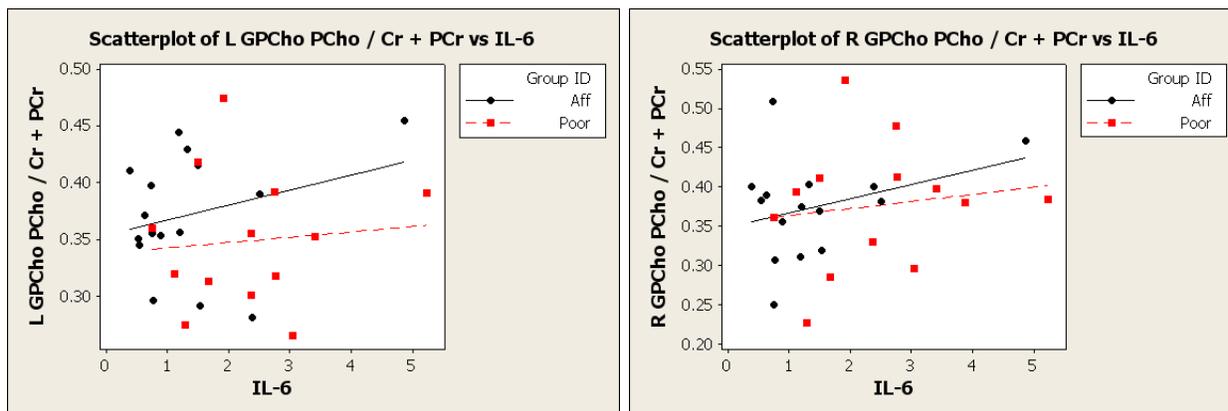


Figure 4.11: Left and right hippocampal Cho/Cr vs. IL-6

Comparing the Cho/Cr ratio with IL-6, no significant correlations were observed for the right or left hippocampi of the affluent or poorer groups.

4.6.2.3 NAA/Cr vs. cortisol

Comparing the NAA/Cr ratio with cortisol, no significant correlations were observed for the right or left hippocampi of the affluent or poorer groups.

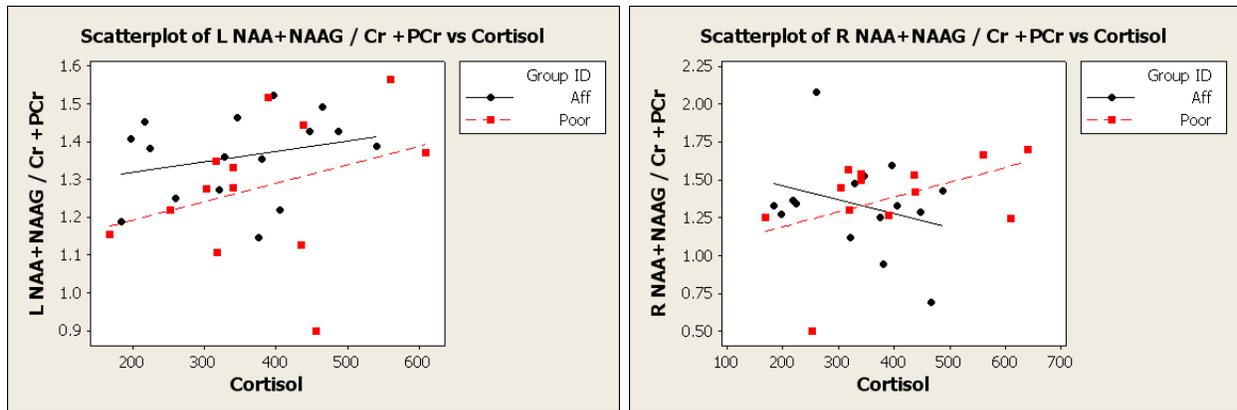


Figure 4.12: Left and right hippocampal NAA/Cr vs. cortisol

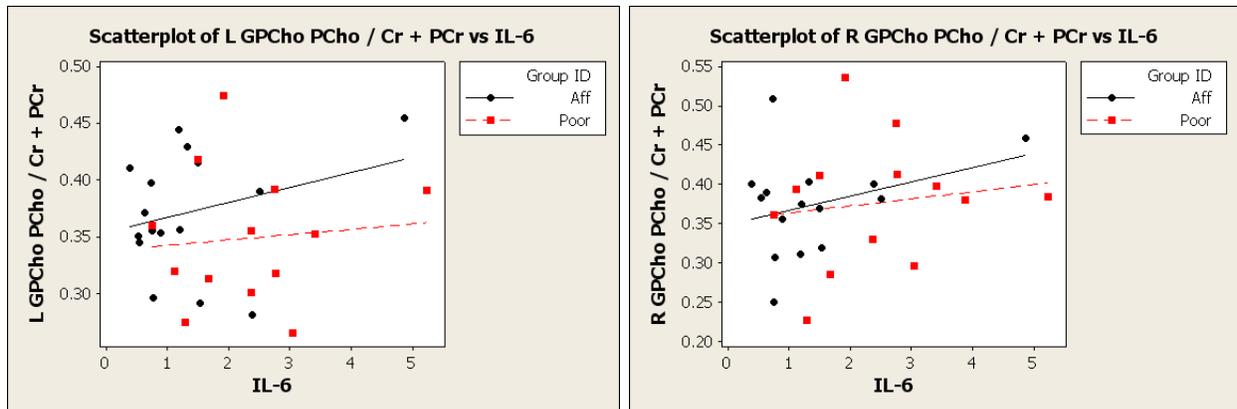


Figure 4.13: Left and right hippocampal Cho/Cr vs. cortisol

4.6.2.4 Cho/Cr vs. cortisol

Comparing the Cho/Cr ratio with cortisol, no significant correlations were observed for the right or left hippocampi of the affluent or poorer groups.

4.7 Post-hoc analysis of MRS data

4.7.1 Metabolites with age

The correlation plots between NAA/Cr and Cho/Cr are shown in figures 4.14 and 4.15 below.

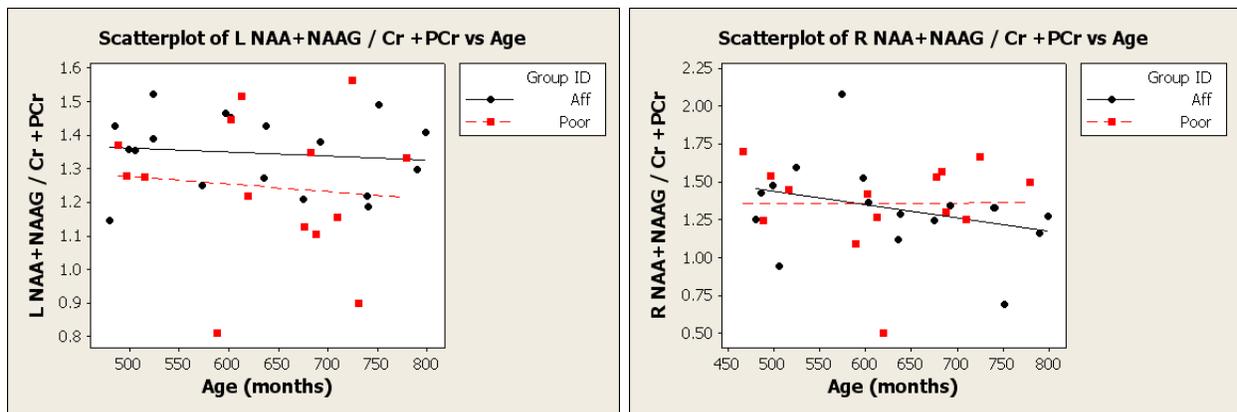


Figure 4.14: Left and right hippocampal NAA/Cr vs. age (months)

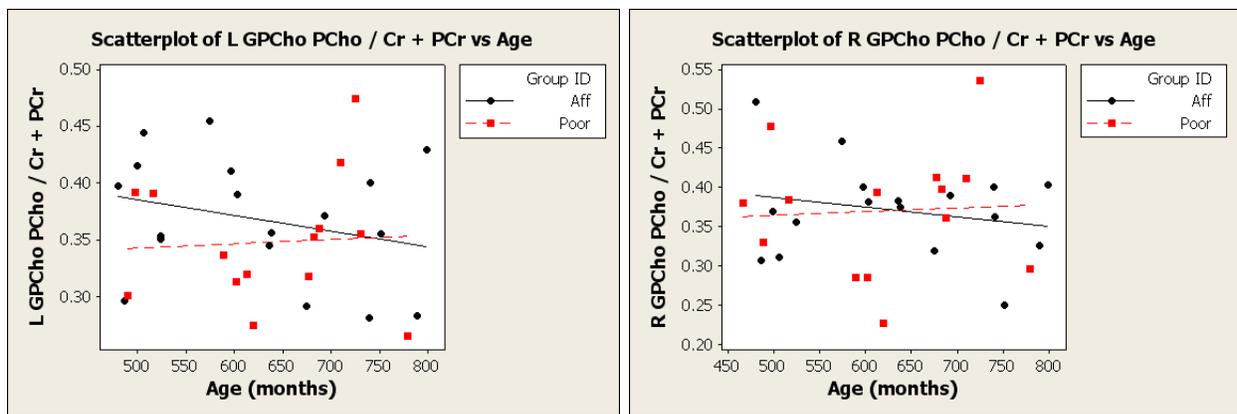


Figure 4.15: Left and right hippocampal Cho/Cr vs. age

Neither right or left hippocampal NAA/Cr or Cho/Cr ratios correlated significantly with age for either the poorer or more affluent groups.

4.7.2 Hippocampal metabolites and hippocampal volumes

4.7.2.1 NAA/Cr vs. hippocampal volume

The scatter plots for comparing hippocampal metabolites and hippocampal volumes are shown in figures 4.16 below and 4.17 on the following page.

After correcting the hippocampal volumes for total ICV, there were no significant correlations

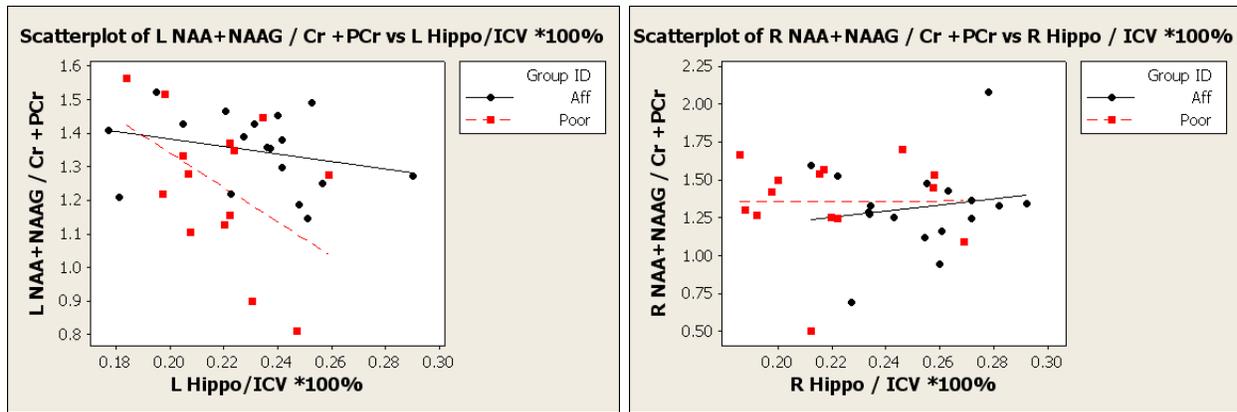


Figure 4.16: Left hippocampal NAA/Cr vs. left hippocampal volume and Right hippocampal NAA/Cr vs. right hippocampal volumes

between the left and right hippocampal volumes and the levels of NAA/Cr.

4.7.2.2 Cho/Cr vs. hippocampal volume

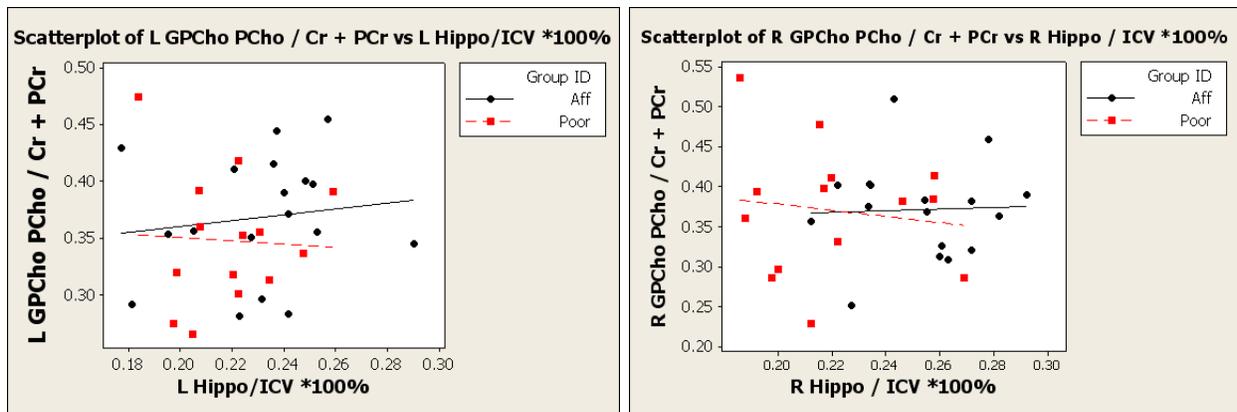


Figure 4.17: Left hippocampal Cho/Cr vs. left hippocampal volume and Right hippocampal Cho/Cr vs. right hippocampal volumes

After correcting the hippocampal volumes for total ICV, there were no significant correlations between the left and right hippocampal volumes and the levels of Cho/Cr.

4.8 Discussion

4.8.1 Comparing hippocampal metabolites between affluent and poorer subjects

What has been presented here was the first study to assess hippocampal metabolites in two differing socioeconomic groups from a community based sample. It was posited that subjects from a lower socioeconomic background may be considered to have been exposed to chronic stress and as such, by means of inflammatory pathways, and therefore that the hippocampus may be preferentially affected. MRS offered the opportunity to assess in-vivo the metabolism of the hippocampus. The primary and secondary null hypotheses were that the NAA/Cr ratio and Cho/Cr ratio would not be lower in the poorer group than the more affluent group.

The NAA/Cr and Cho/Cr ratios were compared between the poorer and affluent groups. There were no significant differences between the groups and therefore it was not possible to reject the null hypothesis. Much of the literature which has assessed hippocampal metabolites in subjects exposed to stress relates to studies of PTSD patients. A meta-analysis of the methods and findings of MRS studies in PTSD patients was recently performed by Karl et al [175]. The results of the meta-analysis are by no means conclusive but they suggest that NAA/Cr may be reduced in the hippocampus of PTSD patients. However, the exact nature of the underlying biological mechanism which might facilitate these changes still remains unclear.

4.8.2 Hippocampal metabolites and inflammatory markers

There have been a number of studies which have assessed the affects of corticosteroids on MRS. As corticosteroids are often used as a treatment (asthma, allergies, arthritis and skin disorders), this offers the opportunity to assess patients hippocampal metabolites in response to therapy [65]. Some studies have exploited the fact that patients with Cushing's syndrome have chronic hypercortisolemia and these patients exhibit various functional deficits. In a group of these patients cortisol levels were found to correlate with hippocampal volumes, smaller hippocampal size was also associated with poorer results on language and memory tests. [332]. Interestingly, the same researchers showed that the effect of hypercortisolemia on the hippocampi were reversible following the reduction in cortisol levels, though no improvements in the associated functional tests from the first study were reported [333]. A study by another group also choose to investigate patients with Cushing's disease and to

compare the patient group to normal controls. MRS was used to assess brain metabolites and voxels were placed in frontal, thalamic and temporal regions. None of the metabolites were significantly different between the patient and normal control groups [179]. Therefore, there are conflicting reports as to whether changes in inflammatory measures such as cortisol are related to the underlying hippocampal metabolites such as NAA/Cr.

It has been known for some time that glucocorticoids can damage the hippocampus [309]. Building on their earlier work in rats Sapolsky et al [308] went on to demonstrate the damage to the hippocampus from glucocorticoids in non-human primates [310]. Later, Czeh et al assessed the effects of chronic stress on the hippocampal metabolites in the adult tree shrew and found decreased levels of NAA (-13%), Creatine (-15%) and Choline (-13%) [101]. In an another study, this time in rats, hippocampal NAA reductions were observed in animals reared in isolation [157]. More recently, Coplan et al have assessed the effects of early life stress on non-human primates on hippocampal metabolites [93]. A validated variable foraging demand (VFD) paradigm to induce a state of stress in the animals. MRS was used to assess metabolites in the temporal lobes of the animals and it was found that NAA/Cho was reduced in the animals subjected to the VFD paradigm. In another recent study MRI and MRS were used to assess the hippocampus in a well documented chronic, mild stress (CMS) rat model [271]. Three separate subgroups of animal were defined within the model, an anhedonic group, a resilient group and a normal group. Each group contained 8 animals. Increases in the Glu/Cr ratio and NAA/Cr ratio were observed in the anhedonic rats. While previous studies had not assessed Glu/Cr levels, the NAA/Cr increase was contradictory to previous reports, where one might have anticipated that the anhedonic rats i.e those animals responding to the stressor might have had a reduced NAA/Cr ratio. Therefore it would appear that there is as yet no consensus from the animal literature on the effects of stress on hippocampal metabolites metabolites.

Here we report no significant correlation between NAA/Cr in the left and right hippocampus and cortisol levels. Similarly, no significant correlations between NAA/Cr and IL-6 in the hippocampus were observed expect for the case where an IL-6 outlying value was removed. Following this amendment to the data, a significant positive correlation between NAA/Cr and IL-6 was observed in the right hippocampus i.e. higher NAA/Cr values were observed at higher IL-6 values ($p=0.005$, $r = 0.53$). In this section of the thesis, of course, we were not assessing diseased patients, and therefore any impact on chronic stress might be too subtle to be detected in hippocampal metabolites. We speculated that hippocampal NAA might be lower in the lower socioeconomic group but this was not found to be the case. As this is the first study to assess hippocampal metabolites in two separate groups differing only

by their socioeconomic status, there is little data in the literature to compare these results with. Despite the lack of correlations observed between hippocampal metabolites and cortisol and IL-6, given the animal evidence discussed above, further investigation is warranted to the effects of stress, inflammation and deprivation on hippocampal metabolites. However, as discussed extensively in chapter 2 (section 2.8.4), cortisol measures have been shown to be notoriously variable and therefore much greater confidence would be needed in this measurement before one could reliably assess correlations between brain volume measures and cortisol levels [109].

As mentioned above, further work on animals to greater understand the biological mechanisms at work may be more fruitful. For example, if one wishes to test if neurogenesis is related to NAA metabolic signature, then this would need to be confirmed in the animal before attempting to measure such changes in the humans.

It often appears that in studies assessing hippocampal metabolites, and in MRS in general, NAA, Cho and Cr are given undue prominence largely because they are relatively easy to measure rather than for their biological significance. In this respect much greater work needs to be done to develop biological models and the measurement techniques to interrogate, *in-vivo*, the metabolism of the brain. This may come through further development of MRS techniques or perhaps through greater development of PET and SPECT radioligands.

4.8.3 Hippocampal metabolite ratios vs. age and hippocampal volume

Decreased NAA/Cr with age has been presented in several papers [110, 11, 315]. Schuff et al, acquired MRS data using a 1.5T MRI scanner from the hippocampus on a range of subjects from age 36 to 85 in normal, healthy volunteers. The NAA/Cr ratio was significantly correlated with age ($r = -0.61$ and $p < 0.005$). In this chapter we observed no significant correlation between NAA/Cr and age over the age range of the subjects imaged in the PSOBID study between 40 and 67 [315]. From inspection of the Schuff et al data it could be the more elderly nature of their sample group might be responsible for the correlation between age and NAA/Cr. They also observed that only hippocampal reductions in NAA/Cho correlated with decreasing hippocampal volume. From our data in this chapter, no correlations between hippocampal volume and metabolites concentrations were observed.

In the work by Angelie et al, significant correlations between NAA/Cr and aging were observed in the cortical, semioval and temporal regions assessed, however, while the NAA/Cr

did decrease with age in the hippocampus, this relationship was not statistically significant ($r = -0.23$). In this study a 1.5T MRI system was used to acquire MRS metabolite data from 32 healthy subjects aged between 21 and 61, however, metabolites ratios were not compared to hippocampal volumes [11].

In a further study assessing hippocampal metabolites with age, Driscoll et al recruited young (20-39) and elderly groups of subjects (60-85) [110]. The subjects were scanned with a 1.5T MRI system and elderly subjects were found to have significantly lower NAA/Cr than the younger group of subjects. Driscoll et al also considered the contribution of hippocampal volume, as well as metabolism to try and explain task human subject performance on a virtual Morris water task and a transverse patterning discrimination tasks. They did not explicitly compare hippocampal NAA/Cr to hippocampal volume but found that both of these measures predicted significant amounts of the variance in the subsequent task performance where both lower NAA/Cr and smaller volumes contributed to poorer task performance. Relating to task performance, on both tasks, improved performance was correlated with higher NAA/Cr levels. Interestingly, when a multiple regression analysis was performed, the level of the performance of the water task was correlated with age and NAA/Cr but not with hippocampal volume. For the transverse patterning discrimination task, age, NAA/Cr and hippocampal volume all accounted for sizable portions of the variance in the task performance.

More recently from the work of Zimmerman et al, poor task performance of a free and cued selective reminding test with immediate recall (FCSRT-IR) correlated with smaller hippocampal volumes [381]. Individuals with lower NAA/Cr also demonstrated poorer verbal memory performance. Furthermore, individuals with smaller hippocampal volumes demonstrated lower NAA/Cr. After controlling for age, FCSRT-IR performance was correlated with NAA/Cr. Indeed, a model for prediction of FCSRT-IR task performance revealed significant effects for both hippocampal volume and hippocampal NAA/Cr levels while prediction of the logical memory task was not significant with any of these variables of interest.

The results from these studies and from the work performed in this thesis are mixed, with NAA/Cr being reduced with age in some studies and not in others. However, where significant relationships have been reported, more elderly subjects have been imaged and it may be the case that much of the NAA/Cr reduction with age occurs in later life, for instance, in subjects older than 60 rather than as a gradual decline. The results comparing hippocampal volume and NAA/Cr and Cho/Cr are also mixed though from the work done in this chapter there appears to be little or no correlation between metabolite ratios and ICV corrected hippocampal volumes.

4.8.4 Future technological developments to improve MRS

As introduced in chapter 2, one of the primary motivations behind this thesis was to exploit the use of the 3T MRI to assess hippocampal size, shape and metabolism. In this chapter 3T MRI was used to obtain estimates of hippocampal metabolite ratios in two groups differing in their socioeconomic status. While no explicit comparison between 3T and any other field strength was performed here, comparisons between fields strengths have been performed elsewhere [29]. As discussed in the work of Barker et al, some improvements in sensitivity were observed when 3T MRS was performed when compared to 1.5T though these improvements fell short of what was predicted by theory. Any additional benefit 3T MRI had to offer was largely offset by increased linewidths and reduced T2 metabolite times. Therefore, only marginal improvements in spectral quality are typical when using 3T MRI when compared to 1.5T.

It was not possible in assessing the 42 subjects of the PSOBID study to acquire useful spectra in all circumstances. The final data for comparison of the 42 subjects was 17 affluent subjects and 13 poorer subjects, where the rest of the datasets were not deemed to have sufficient confidence to assess the hippocampal metabolites. There is no doubt the hippocampus was a particularly challenging part of the brain to acquire MRS data from, given its contorted form, its proximity to CSF and the ear canal which can produce susceptibility related effects. While we have demonstrated that it is possible to achieve good quality spectra from the hippocampus, further developments may focus on acquiring data from more specific regions of the hippocampus. For example, it would be advantageous to acquire multi-voxel MRS data from the head, body and tail of the hippocampus to improve the spatial localisation of the spectra, though this would greatly reduce the signal to noise in each spectra, as well as increasing the time for each acquisition.

Other developments which will allow MRS Imaging (MRSI) to be performed in acceptable time frames include greater exploitation of parallel imaging techniques as greater numbers of receiver coil elements are used, indeed, 32-channel devices are already being used [380]. Moreover, the introduction of compressed sensing techniques to greatly reduce the acquisition time (or improve quality if scanning time is maintained) are also likely to offer benefits to MRSI. Compressed sensing MRSI acquisitions have already been acquired in animals, where in dynamic nuclear polarisation (DNP) experiments, rapid acquisition schemes are even more beneficial [166, 167]. Therefore it is likely improvements in MRS will come on several fronts, with developments in receiver hardware and pulse sequences combining to enable faster acquisitions times or improved SNR.

4.9 Conclusion

Hippocampal metabolites were measured using 3T MRI in two groups of differing socioeconomic status from a community based sample, we believe this is the first time this has been done. No significant difference was observed between the NAA/Cr and Cho/Cr ratios between the poorer and more affluent groups. The motivation behind assessing hippocampal metabolites originated from the fact that the hippocampus may be preferentially affected in subjects exposed to chronic stress, where in this case we considered social deprivation as the stressor. To further investigate this proposed mechanism, the hippocampal metabolites were correlated against cortisol and interleukin-6, two markers for inflammatory processes.

There were no significant correlations between NAA/Cr or Cho/Cr and cortisol. On removal of an outlying IL-6 value, a statistically significant correlation was observed between the right hippocampal NAA/Cr ratio and IL-6 ($p = 0.005$, $r = 0.53$). Therefore, in general these results did not support the theory that inflammatory processes might affect hippocampal metabolism, although the correlation between right hippocampal NAA/Cr and IL-6 warrants replication in a larger sample size. Future work to develop a clearer biological framework and a more comprehensive investigation of metabolic and inflammatory markers may be more informative. However, it could also have been the case that as we were dealing with an otherwise healthy population, the anticipated metabolic changes in the hippocampus simply do not exist as a result of the posited inflammatory pathway.

Further to the primary aims of this chapter, hippocampal metabolites were correlated with age and hippocampal volume. No significant correlations were observed between hippocampal metabolite ratios and age or between hippocampal metabolite ratios and ICV corrected hippocampal volumes. Perhaps the reason hippocampal metabolites were not observed to correlate with age, was due to the limited age range of the population studied where other studies have assessed more elderly subjects [315]. Nonetheless, little trend between NAA/Cr, Cho/Cr and age was observed here.

Finally, there are various technological advances which may occur over the next several years. This includes the potential for faster imaging by further exploitation of multi-element coils, and improved pulse sequences. However, technological advances may only be useful if they are framed within plausible biological models and to this end, further studies in animals may be useful for interrogating the underlying metabolism as a result of stress and inflammatory processes.

Chapter 5

Normal adult hippocampal volumes as measured using 3T MRI

5.1 Abstract

Following the installation of a 3T MRI system, and the anticipated programme of research into Psychobiological research, it was deemed necessary to acquire a normative dataset of brain images. A region of interest within the brain on which much of the research into mood disorders focuses on is the hippocampus. Thus the aim of this work was to determine a normal range for hippocampal volumes at 3T and to investigate the relationship to factors which may affect hippocampal volumes such as age, total intracranial volume, laterality and sex. 3D T1-weighted IR-FSPGR images were acquired in 39 healthy, normal volunteers in the age range from 19 to 64 (mean age = 34). There were 23 females and 16 male subjects. The 3D T1-weighted image data were then pre-processed, including application of the N3 inhomogeneity algorithm before normalisation and segmentation of the data into subcortical volumes, including the hippocampus using FSL's FIRST algorithm. Following the automated procedure hippocampal volumes were manually inspected and edited. This approach was described in greater detail in chapter 2 of this thesis. The mean and standard deviation of the left and right hippocampal volumes were determined to be: $3421\text{mm}^3 \pm 399\text{mm}^3$ and $3487\text{mm}^3 \pm 431\text{mm}^3$ respectively. After correcting for total ICV the volumes were: $0.22\% \pm 0.03\%$ and $0.23\% \pm 0.03\%$ for the left and right hippocampi respectively. Multiple regression analysis was performed and models were created to predict hippocampal volumes, given the subject age and ICV. The model for the left hippocampus explained 18% of

the variance while the model for the right hippocampus explained 23% of the variance. After correcting for ICV there was no significant difference in hippocampal volume between the male and female subjects. As was in keeping with what has previously been observed in the literature the left hippocampus tended to be slightly smaller than the right but this difference was not statistically significant. Thus, a normative database of hippocampal volumes was established. The normative data here will in future act as a good baseline on which other methods of determining hippocampal volumes may be compared to. To date only a few studies have acquired hippocampal volumes in normal healthy subjects using 3T MRI. The values obtained here compare well with what has already been observed in the literature. At higher field strengths there is some evidence to suggest that larger volumes are observed when compared to results from 1.5T MRI, perhaps due to the hippocampus being more accurately defined when higher field strength magnets are being used.

5.2 Introduction

Chapters 2 and 3 established methodologies using 3T MRI to assess hippocampal volume differences between two different socioeconomic groups. While the focus of these chapters was then to compare hippocampal volumes between the two differing SES groups it would also be informative to have a separate, normal range of hippocampal volumes which could be used as a baseline for future work. As has been discussed earlier in this thesis, hippocampal volumes have been much investigated in diseases such as Epilepsy [117], Alzheimer's [248] and Depression [203]. A review of normal hippocampal volumes was performed in chapter 2 where it was observed that at higher field strengths there was a tendency for hippocampal volume estimates to be larger. This may be a result of better resolution and contrast at 3T, enabling a more accurate delineation of hippocampal volumes at higher field strengths. It could also be the case that as time has moved on, and 3T scanners have come online, hippocampal protocols have improved and become more reliable, which has no doubt been helped by the improved image quality.

Much of the work in this thesis and the focus of the research of the Sackler Psychobiological research group at Glasgow University is concerned with applying neuroimaging techniques to assess depression in the medically ill and the biological etiology and pathophysiology of various mood disorders. Therefore it is important to establish the normal variance in the measurements of concern for the local equipment and methodology being applied. Throughout much of this thesis, the measurement of concern has been the volume of the hippocampus and so acquiring a normal database of subjects using 3T MRI with a combined automated and manually refined volume approach will be the subject of this chapter.

A protocol for manually delineating the hippocampus was defined in chapter 2. The protocol included clear definitions of which tissue regions should be included as well as defining the anatomical boundaries around the various aspects of the hippocampal formation. FSL's automated subcortical segmentation algorithm FIRST was also compared to manual outlining and while the results were generally satisfactory it was observed that occasionally the automated algorithm produced gross errors and therefore it was necessary in some circumstances to manually correct the automated volume.

Beyond assessing only the hippocampal volume, it was important to assess the relationship between hippocampal volume and other factors such as age, intracranial volume and memory. A clear appreciation of the relationships between these factors will determine which factors ought to be controlled for in subsequent analysis and that those parameters are controlled for in the correct manner.

5.3 Aims

The aim of this chapter was to determine the normal range for the hippocampal volumes in healthy volunteers, where we sought to exploit the higher static magnetic field strength to improve delineation of the hippocampus. Furthermore the relationship between hippocampal volume, age, ICV, sex and laterality will also be assessed.

5.4 Methodology

The study was approved by the local NHS research ethics committee.

5.4.1 Volunteer recruitment and inclusion/ exclusion criteria

We aimed to recruit subjects who were healthy volunteers, with no history of mental illness, head injury or neurological disorder. Subjects were chosen such that they were suitable for MRI scanning and had no medical implants that were contraindicated within the MRI environment.

45 volunteers were recruited for imaging using posters placed around the campus at the Southern General Hospital and by word of mouth through staff associated with the project. The poster used for recruitment formed part of the ethics application which was subsequently approved (the poster is shown in appendix C). Volunteers who expressed an interest in the

study were given a patient information sheet to read. The patient information sheet explained all that would be required of the subjects if they decided to proceed with participation in the project. The patient information sheet also stresses that volunteers who are also members of staff are under no obligation to participate. The author of this thesis also discussed aims of the project and the inclusion and exclusion criteria with potential volunteers either over the phone or in person prior to them agreeing to participate in the study.

The procedure for dealing with incidental findings on MRI scans was for a consultant radiologist to provide a report on all scans. If an incidental abnormality was found the neuroradiologist would discuss the finding with an appropriate clinician (typically a surgeon or neurologist) to discuss how best to proceed. Where necessary the consultant neuroradiologist would contact the volunteer directly to discuss their ongoing medical care. Two subjects were excluded from the study as a result of incidental findings (1 acoustic neuroma, 1 Chiari malformation). All other subject scans were reported as being normal from a neuroradiological perspective.

Volunteers were also taken through the Structured Clinical Interview for Diagnostic and Statistical Manual for mental disorders (SCID: DSM-IV). This was performed by a psychiatrist. This assessment was initially developed by the American Psychiatric Association and has been widely applied in both clinical and research environments [122, 106]. Though subject information documentation and in discussions with the author of this thesis it was expressed that volunteers should not have a history of psychiatric illness or neurological disorders, nevertheless, two subjects who were scanned were subsequently found to have a history of depression and therefore their data was removed from further analysis within the study. In retrospect it would have been preferable to perform the SCID examination prior to MRI scanning to avoid this scenario. Such procedures will be implemented in future work.

Following two initial scans on volunteers in which whole head T1-weighted 3D acquisition were not acquired a reliable protocol was then confirmed, imaging parameters are detailed in section 5.4.3.1 of this chapter.

Therefore, following the exclusion of 6 subjects for the reasons detailed above, 39 whole head T1-weighted 3D scans were acquired in total and were useable for further analysis. The age range of the volunteers was 19 - 64 years old, with a mean age of 34. 23 subjects were female and 16 were male.

The study was funded by the Sackler Institute of Psychobiological Research.

5.4.2 Imaging hardware

All MRI scans were acquired using a 3T GE Medical systems, Signa Excite HD system (Milwaukee, USA) with software version 12m5. Furthermore, an eight channel phased array (receive only) head coil was used to acquire the imaging data.

5.4.3 Imaging Protocol

5.4.3.1 Data acquisition

A T1-weighted IR-FSPGR was acquired, TR = 6.8ms; TE = 1.5ms, Inversion Preparation time = 500ms; Flip angle=12°; FOV = 26cm; Phase FOV= 70%; matrix: 320 x 320; Bandwidth 31.25kHz; Slab thickness = 1mm. The acquisition time for this scan was 8min 54s. This acquisition sequence with high resolution (0.81mm x 0.81mm x 1mm) in all three imaging planes and good white matter to grey matter contrast facilitates well both the segmentation of GM and WM tissue types and whole hippocampal segmentation.

5.4.4 Segmenting the hippocampus

Following on from the work performed in Chapter 2, while it was shown that there was a good correspondence between hippocampal volumes generated by the automated FIRST algorithm, when compared to manually defined volumes, it was also determined that there were instances when the automatic method produced appreciable errors, clearly visible on manual inspection. Similarly, there were aspects of the hippocampus which were either over or under estimated by the automated method. Thus a scheme by which automated segmentation methods were augmented with visual inspection and manual editing was implemented. This combination of automatically generating volumes supplemented by manual refinement aims achieve a protocol where the practical ease and speed of the automatic method was initially exploited and was then combined with the accuracy of manual refinements.

Throughout this chapter hippocampal volumes will be presented as both raw whole hippocampal volumes and as corrected volumes with respect to each individuals corresponding Intracranial volume (ICV). Thus the ICV corrected hippocampal volumes will be expressed as a percentage of ICV (%ICV).

5.5 Results

5.5.1 Data normality

Histograms of the left and right hippocampal volumes, intracranial volumes and ICV corrected hippocampal volumes were plotted to gain an impression of the normality of the data, see figure 5.1 below and figures 5.2 and 5.3 on the following page. Furthermore, summary statistics were also acquired which included the Anderson-Darling (AD) normality test. The null hypothesis for the AD was that the data are from a normal population distribution. For these variables here, the null hypothesis could not be rejected (i.e. for all parameters the A-D normality test p-value was much greater than 0.05) and therefore we can assume that all the data was part of a normal distribution. Thus it was acceptable to continue to assess these data using parametric statistics. Statistics were performed using Minitab v16 (Minitab Inc.) software.

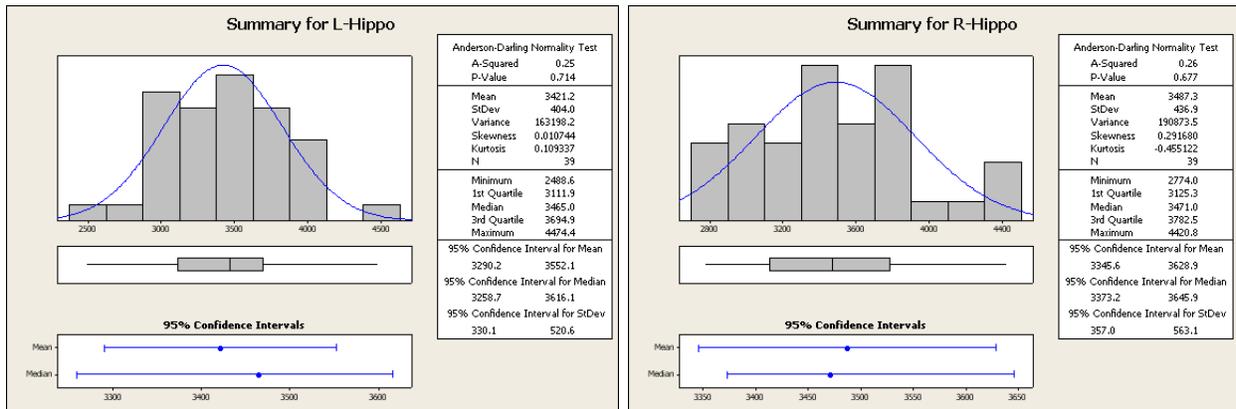


Figure 5.1: Histogram and summary statistics of uncorrected left and right hippocampal volumes

5.5.2 Hippocampal volumes and their relationship to ICV

It has been well known for some time that there is a need to correct for subject head size

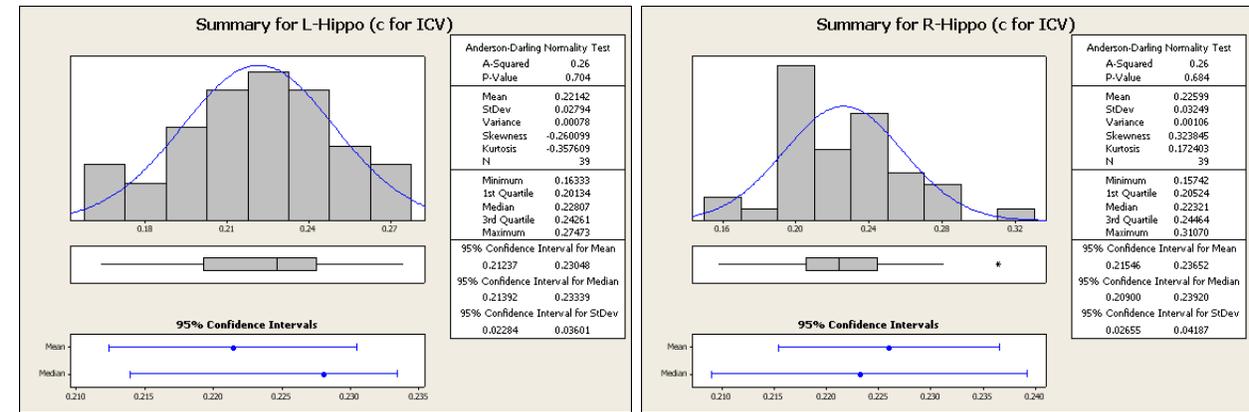


Figure 5.2: Histogram and summary statistics of ICV corrected left and right hippocampal volumes

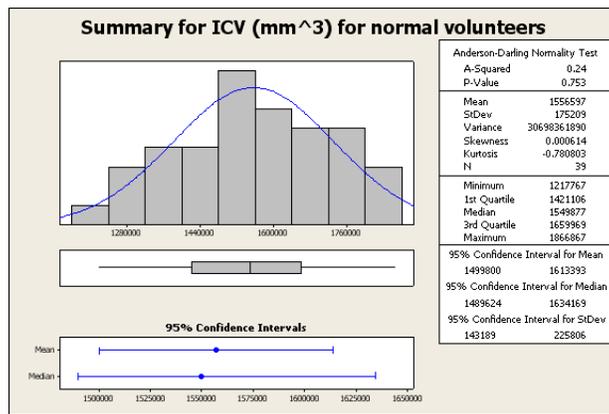


Figure 5.3: Histogram and summary statistics of intracranial volume measurements

when measuring brain structure volumes [90]. Thus it is common for hippocampal volumes to be expressed both independently and with a correction for ICV. The relationship between hippocampal volumes and ICV was determined for the normal volunteers in this study and are shown in figure 5.4 on the following page.

The corresponding Pearson correlation coefficients between hippocampal volumes and ICV were 0.40 for the left hippocampus and 0.31 for the right hippocampus, with significance values of 0.01 and 0.06 respectively. Thus, the left hippocampal volume was significantly correlated with ICV while the right hippocampal volume is just below the level of statistical significance. Therefore, the results here are in keeping with what has previously been observed in the literature and therefore support the decision to present ICV corrected hippocampal volumes.

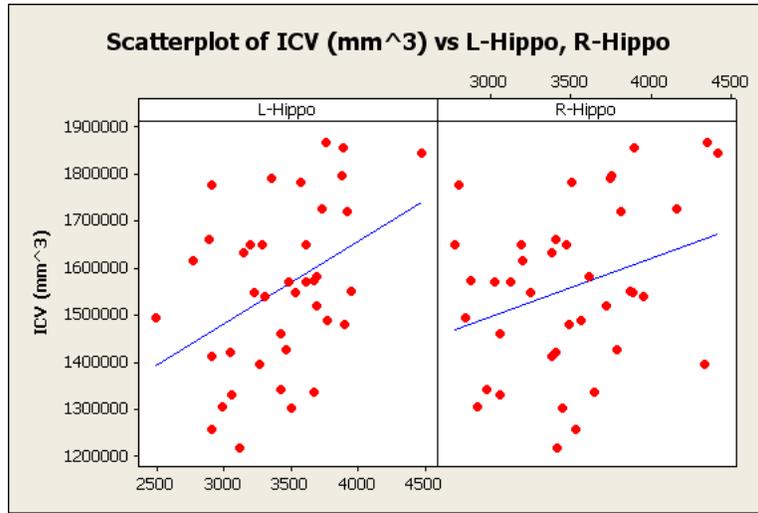


Figure 5.4: Scatterplot illustrating the relationship between ICV and hippocampal volume

As an aside, one would expect ICV measures to be independent of age, while relationships between the component parts of ICV, namely the GM, WM and CSF may differ. In particular, one would expect a decline in GM with age. Figures 5.5 below and 5.6 on the following page illustrate that indeed, for the age range assessed here, age and ICV are independent and that while GM volumes decline with age, CSF volume increases presumably filling the intracranial spaces left behind by the GM volume reduction.

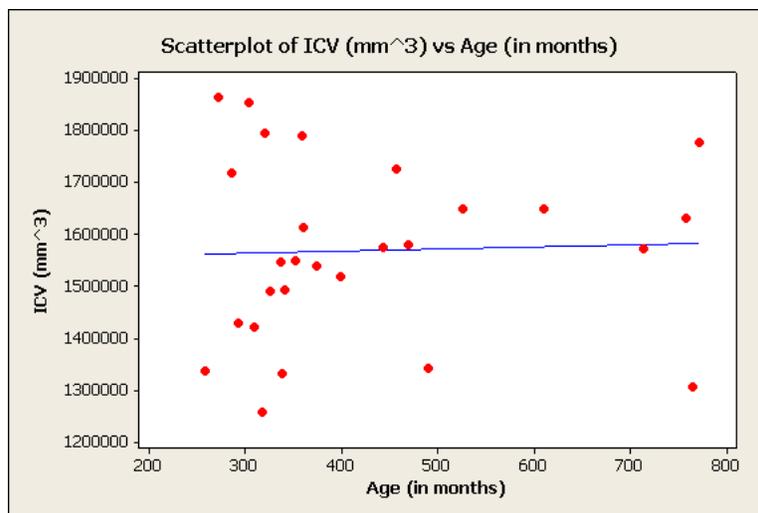


Figure 5.5: Scatterplot of ICV (mm³) data plotted against age (in months)

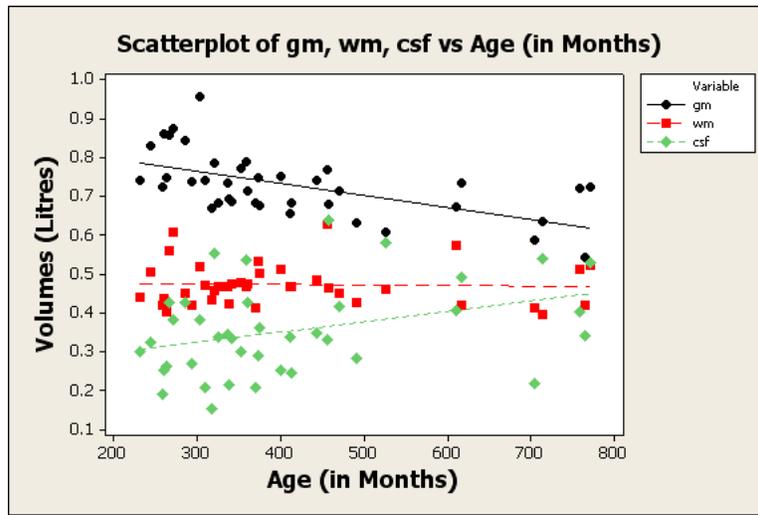


Figure 5.6: Scatterplot of GM, WM and CSF volumes (in Litres) plotted against age (in months)

5.5.3 Hippocampal volume with aging

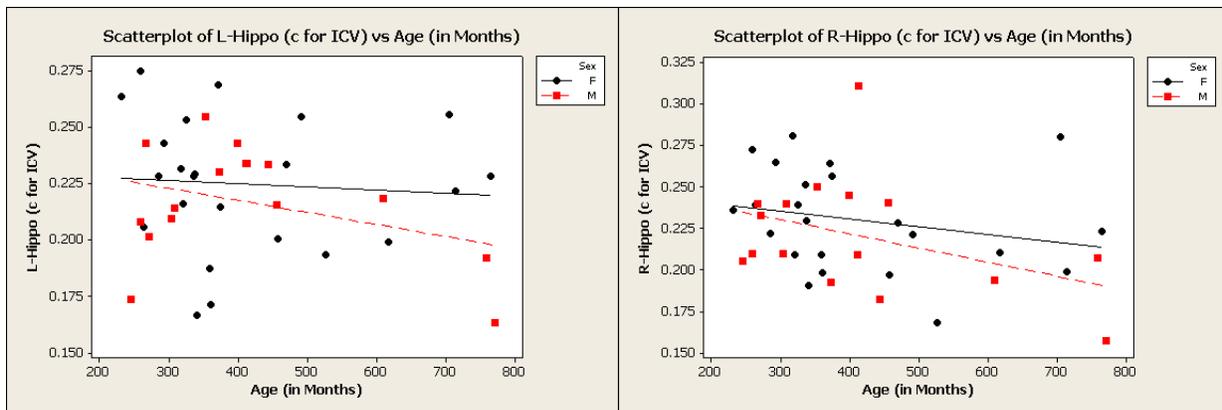


Figure 5.7: Left and right hippocampal volumes with age

Figure 5.7 above shows the ICV corrected hippocampal volumes against age. From figure 5.7 there appears to be a gradual decrease in hippocampal volume with age, and this may be slightly more pronounced for men than women though it is worth noting that there are fewer data points at the upper end of the age range. One can be more confident however that the hippocampal volume does not appear to decrease markedly between the ages of 20 and 40 years old. Over the entire age range assessed here the Pearson correlation between age and hippocampal volumes were, for the left hippocampus $r = -0.28$ ($p = 0.16$) and for

the right hippocampus $r = -0.54$ ($p \ll 0.05$). Thus, from figure 5.7, a decrease in left hippocampal volume correlates with increasing age, though not to statistical significance, whereas a decrease in right hippocampal volume correlated with increasing age and was deemed statistically significant. Thus, the right hippocampus seems to be more affected by normal aging than the left hippocampus.

The male and female data were combined and the ICV corrected left and right hippocampal volume data were plotted against age. The Pearson correlation for the left and right hippocampal volumes were -0.28 and -0.54 with associated p -values of 0.156 (not statistically significant) and 0.003 respectively. Therefore, after correcting for ICV, the right hippocampal volume appears to be negatively correlated with age.

5.5.4 Hippocampal volume measures in normal volunteers

Having established that the hippocampal volumes and ICV are normally distributed and having confirmed that ICV does indeed correlate with hippocampal volumes it was then possible to establish a normal range for hippocampal volumes as measured using our 3T MRI system and the automated segmentation with manual refined methodology. Of the 39 normal volunteers studied in the age range 19 to 64, the mean uncorrected volume was $3421\text{mm}^3 \pm 399\text{mm}^3$ and $3487\text{mm}^3 \pm 431\text{mm}^3$ for the left and right hippocampi respectively. When corrected for intracranial volume (ICV) the hippocampal volumes were $0.22\% \pm 0.03\%$ and $0.23\% \pm 0.03\%$ for the left and right hippocampi respectively.

Thus the normal ranges for the ICV corrected left and right hippocampal volumes with age are shown in figure 5.8 on the following page. Figure 5.8 shows the mean ICV corrected hippocampal volume for a subject of a given age (black line) and the associated confidence interval (CI) for the mean (dashed red line). Furthermore, the prediction interval (PI) which represents the range within which a single, new observation is likely to fall is also shown (dashed green line).

A linear regression analysis was also performed to estimate the relative contributions of age and ICV to the variance in estimating the left and right hippocampal volumes. In order to ensure this approach is valid the residual plots were assessed and were found to be both randomly scattered with approximately constant variance around the best-fit line.

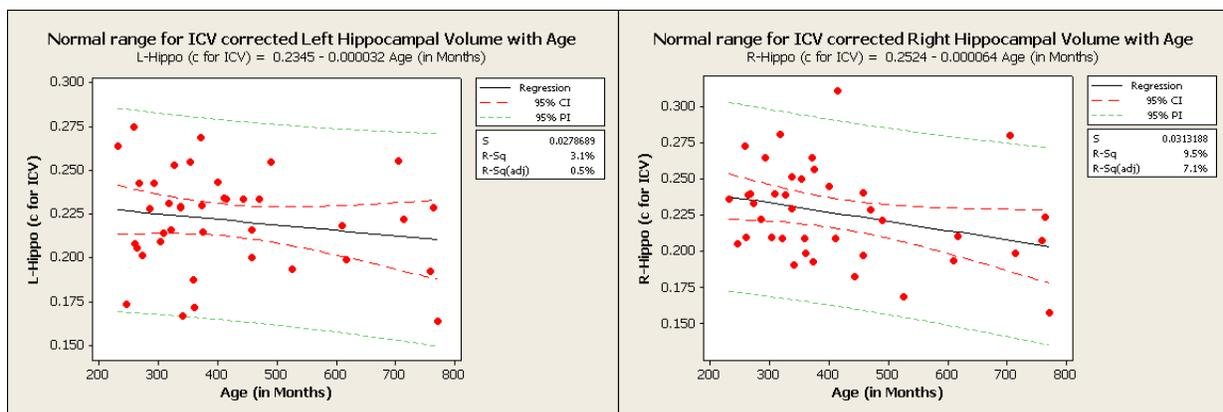


Figure 5.8: Normal range ICV corrected left and right hippocampal volumes with age

For the left hippocampus the linear regression equation was:

$$LeftHippocampusVolume = 2279 - 0.626xAge(months) + 0.000900xICV(mm^3) \quad (5.1)$$

For the left hippocampus, age was not deemed a statistically significant regression factor with a p value of 0.107, though ICV was determined to be statistically significant ($p=0.01$). Together, age and ICV combine to explain 17.8% (R^2adj) of the variance in left hippocampal volume measurements. Where R^2adj is a measure of how much variance in the response variable (hippocampal volume) is accounted for by the model which includes the two predictor variables, which in this case are age and ICV.

For the right hippocampus the linear regression equation was:

$$RightHippocampusVolume = 2849 - 1.15xAge(months) + 0.000717xICV(mm^3) \quad (5.2)$$

For the right hippocampus, age was a statistically significant regression factor with a p value of 0.006, and ICV was determined not to be statistically significant ($p=0.052$). Together, age and ICV combine to explain 22.6% (R^2adj) of the variance in right hippocampal volume measurements.

5.5.5 Hippocampal volumes and sex

Of the 39 normal volunteers, 23 were female and 16 were male. Uncorrected hippocampal volumes for males had a mean value of $3516mm^3 \pm 426mm^3$ for the left hippocampus and $3571mm^3 \pm 544mm^3$ for the right, while females had uncorrected volumes of $3355mm^3 \pm 383mm^3$ for the left hippocampus and $3429mm^3 \pm 344mm^3$ for the right. The ICV corrected

hippocampal volumes for males had a mean value of $0.22\% \pm 0.03\%$ for the left hippocampus and $0.22\% \pm 0.04\%$ for the right, while females had a mean value of $0.22\% \pm 0.03\%$ for the left hippocampus and $0.23\% \pm 0.03\%$ for the right.

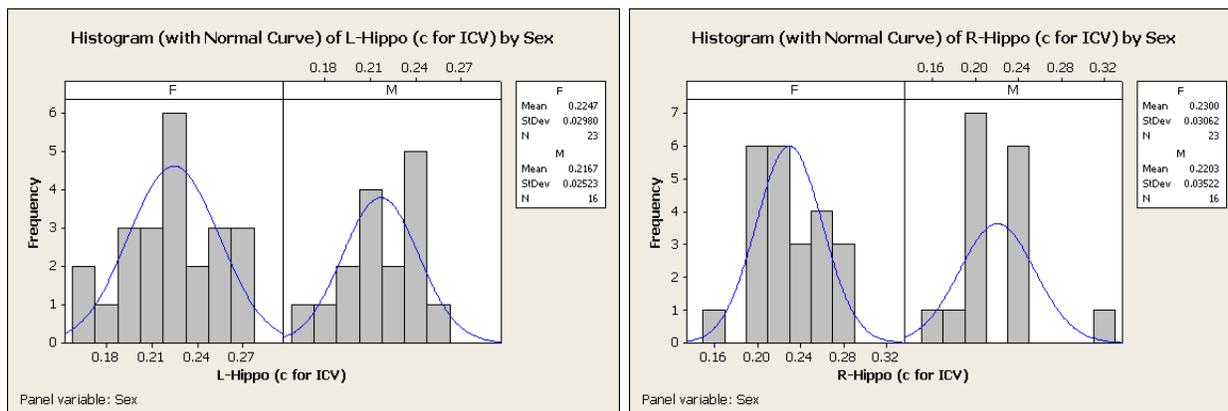


Figure 5.9: Left and right, ICV corrected hippocampal volumes for females and males: Histogram and summary statistics

Figure 5.9 above illustrates the distribution of ICV corrected hippocampal volumes for both female and male subjects. A two sample t-test was conducted to compare hippocampal volumes between males and females. The null hypothesis was that male and females hippocampal volumes were from the same underlying populations. Where corrected volumes were compared between males and females, the left hippocampus p-value was found to be 0.377 and the right hippocampus p-value was 0.381, thus in both cases the null hypothesis cannot be rejected.

5.5.6 Hippocampal volume: laterality

As was shown from the earlier literature review 2.1, in general, in healthy normal populations, left and right hippocampal volumes are similar, with a tendency for the right hippocampus to be slightly larger. The correlation between left and right hippocampal volumes from the data acquired in this study was plotted and can be seen in figure 5.10 on the following page.

These volumes were strongly correlated as one would expect with a Pearson correlation coefficient of 0.62 with a p-value of less than 0.01. Performing a two-sample t-test, where

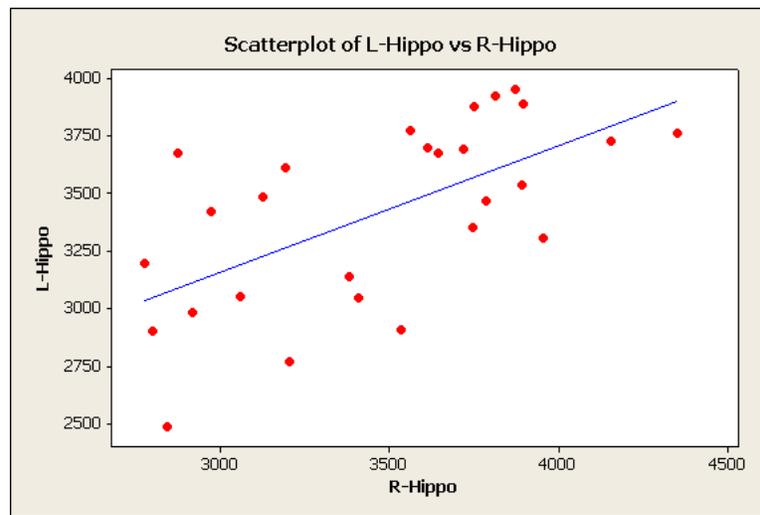


Figure 5.10: Left hippocampal volume vs. right hippocampal volume

the Null hypothesis was that the left and right hippocampal volumes were from the same underlying population, the t-test p-value was found to be 0.49, i.e. the null hypotheses cannot be rejected and therefore it is highly likely that the volumes do indeed form part of the same normal distribution.

5.6 Discussion and limitations

5.6.1 The relationship between hippocampal volume and ICV

The Pearson correlation coefficients between hippocampal volumes and ICV were 0.40 ($p = 0.01$) for the left hippocampus and 0.31 ($p=0.06$) for the right hippocampus. This relationship is well founded and is the reason that in the overwhelming majority of studies hippocampal volumes are presented in a way which accounts for differences in head size. Here hippocampal volumes were presented both in raw volume form and as a fraction of the total intracranial volume. This allows the most complete comparison to other studies in the literature.

5.6.2 The relationship between hippocampal volumes and age within the age range of 19 to 64

Over the entire age range assessed here (19-64) the Pearson correlation between age and hippocampal volumes were, for the left hippocampus $r = -0.28$ ($p= 0.16$) and for the right hippocampus $r = -0.54$ ($p \ll 0.05$). Recent work has shown a similar decline in hippocampal

volume with age [299]. Where right and left hippocampal volumes were considered together the correlation value was -0.28, with a p-value of < 0.05 . This was over an age range of approximately 20 to 84 years old, where 113 subjects were studied.

In another study a longitudinal analysis was performed (N=40) and showed a decline in hippocampal volume in subjects of age range 50 to 85 years old [297]. Subjects in this study were normal, healthy volunteers, where an initial scan was followed by two follow-up scans fifteen months apart. Significant shrinkage of the hippocampus was observed over these time frames and in general the correlation of volume with age was steeper in this older age group when compared with studies which have assessed hippocampal volumes in younger populations. In other words, Raz et al observed that in this older age group hippocampal volumes appeared to decline more rapidly with increasing age [297]. In one of the largest studies of its kind, and perhaps one of the most evolved in terms of the form of the automated analysis, hippocampal volumes were assessed with a view to discriminating patients with Alzheimer's dementia from patients with mild cognitive impairment and normative, age matched controls [85]. In this study, 166 cognitively normal subjects with a mean age of 76 ± 5 were assessed. The mean volume for the hippocampal volumes was 2430mm^3 . In the study by Chupin et al, no explicit correlation of hippocampal volume and age was performed, this was perhaps because of the relatively small range of ages being assessed. While these volumes are smaller than the ones recorded in other studies, and indeed smaller than volumes acquired in this work, given the work of Raz et al [297], it could be that the smaller volumes in the more elderly group are the result of the effect of aging, this adds to the evidence suggesting that age is an important factor in comparing hippocampal volumes. Moreover, given the accelerated nature in which hippocampal volumes decrease within an older age group [297], it would appear particularly necessary to ensure appropriate control data is being used when assessing diseases in an elderly population.

Thus, for the age range assessed in this study, decreases in both right and left hippocampal volumes were observed with increasing age. The correlation value of $r = -0.28$ for the left hippocampus is in keeping with what has already been observed in the literature though we observed a slightly stronger correlation between decreasing right hippocampal volume and age, $r = -0.54$. Therefore, in both instances it is important to take into account a subject's age when assessing hippocampal volume. While not assessed in this study, but is worthy of note, it would appear that in normal, healthy, older populations (>50 years old) hippocampal volume may decrease more rapidly with time and that for studies assessing hippocampal volumes in this aged group, one might expect an even stronger correlation between decreasing volume with time than is observed in the 20 to 50 year old age range.

5.6.3 The normal range for hippocampal volumes in normal subjects between the ages of 19 and 64 using 3T MRI

Of the 39 normal volunteers studied in the age range 19 to 64, the mean uncorrected volume and standard deviation were $3421\text{mm}^3 \pm 399\text{mm}^3$ and $3487\text{mm}^3 \pm 431\text{mm}^3$ for the left and right hippocampi respectively. When corrected for intracranial volume (ICV) the hippocampal volumes were $0.22\% \pm 0.03\%$ and $0.23\% \pm 0.03\%$ for the left and right hippocampi respectively.

Recalling the work performed in chapter 2 (section 2.4.1), the meta-analysis of hippocampal volumes in MDD patients performed by Mackinnon et al was updated to include more recent results [232]. Based on assessing the hippocampal volumes of the control subjects from this meta-analysis the mean volume was found to be $3079\text{mm}^3 \pm 377\text{mm}^3$ and $3159\text{mm}^3 \pm 381\text{mm}^3$. Therefore, both the left and right hippocampal volumes of the normal subjects acquired here were approximately 10% bigger than the mean values observed from the meta-analysis. There may be several reasons that contribute to differences observed between the normal values determined in this thesis and the results of the meta-analysis.

Included in the meta-analysis were a number of studies assessing hippocampal volumes in older age groups [335, 196, 343], one could consider that this might bring the average value observed in the meta-analysis down, however, the mean hippocampal values for these studies were similar to the mean observed in the meta-analysis, indeed the mean hippocampal volumes observed by Steffens et al, (mean age 67) were 3170mm^3 (left) and 3300mm^3 (right). Thus, one cannot attribute the lower volumes observed in the meta-analysis to age differences in the subjects. Moreover, a recent study assessing manually segmented hippocampal volumes in 422 individuals using 1T MRI found no significant correlation between hippocampal volumes and age [250]. This study covered a wide range of ages from 20 to 84 years of age. The results of this study would imply that the hippocampus is more resistant to aging than grey matter cortical areas and therefore that there may be no need to adjust or correct hippocampal volumes for age. The implications of this are likely to be of particular importance to studies of pathological aging such as Alzheimer's [14].

A meta-analysis assessing hippocampal volume in normal volunteers across the life span has been done previously [283]. Estimates of hippocampal volume loss with age vary in the peer reviewed literature. In a review of fifteen such studies by Raz et al in the Handbook of Aging and Cognition, correlations between adult hippocampal volume and age ranged from -0.03 to -0.67 with a median effect size of $r = -0.3$ [95]. Thus while the correlation strength varied,

all correlations were negative correlations. However, there are also many studies which cite no significant correlation between hippocampal volume and age as reported in Van Petten et al's meta-analysis [283].

Therefore across the peer-reviewed literature there are conflicting reports on the effects of aging on hippocampal volume. The data acquired in this thesis is more in agreement with the median value ($r = -0.3$) reported from the meta-analysis by Raz et al where we observed correlation for the left hippocampus and right hippocampus of $r = -0.28$ ($p = 0.16$) and $r = -0.54$ ($p << 0.05$) respectively. However, the recent work by Mouiha et al, in one of the largest studies of its kind ($n = 422$) where no correlation between hippocampal volume and age cannot be ignored [250].

In view of the conflicting evidence in the literature there continues to be a need to assess the effect of aging on the hippocampus on a study by study basis ensuring that appropriate controls groups, matched for age, are acquired. This will be of particular relevance in older age groups when comparisons are being made between normal controls subjects and diseases such as Alzheimer's and mild cognitive impairment. As more advanced analysis methods evolve, such as to create surface models of hippocampal volume difference, the development of age range specific templates may be helpful to minimise the degree of normalisation to the template of the individual subjects brains.

As was discussed when introducing the need for a clear hippocampal protocol in chapter 2, when defining the hippocampal segmentation protocol it is essential to decide whether hippocampal white matter is going to be included in the total hippocampal volume or not. In studies such as Sheline et al, only grey matter hippocampal volumes are presented and therefore the volumes were significantly lower than the average value for the meta-analysis [321]. Thus, differences in protocols across studies may account for differences between the mean values of studies .

As time has moved on, imaging quality has improved and a consensus for a protocol to segment the hippocampus has been proposed in the literature [185]. There is some evidence to suggest that moving to higher field strengths has improved the segmentation accuracy of hippocampal volumes, where volumes from these studies tend to be larger than the mean observed from the Mckinnon et al meta-analysis and thus more in keeping with the data that was observed in this thesis. Perhaps the most similar set-up to the study performed in this thesis i.e. a 3T scanner and a similar number of subjects, the mean and standard deviation for the left and right hippocampi were: left hippocampus $3576 \text{ mm}^3 \pm 342$ and right hippocampus $3679 \text{ mm}^3 \pm 351$ [262]. Thus both the volumes and variances are comparable

to what was observed in this thesis. Further studies at 3T also reported mean hippocampal volumes larger than the mean observed in Mckinnon et al's meta-analysis. This includes, Haller et al's study where left and right hippocampal volumes were 3780mm^3 and 3890mm^3 and the study by Qui et al, with left and right volumes of 3520mm^3 and 3607mm^3 . In a further study, one of the few studies which explicitly compares hippocampal volumes of the same subjects at different field strengths (1.5T and 7T) [83], segmented volumes were $2884 \pm 297\text{mm}^3$ and $3112 \pm 332\text{mm}^3$ at 1.5T and $3232 \pm 232\text{mm}^3$ and $3340 \pm 323\text{mm}^3$ at 7T, where again, the values at the higher field strength were more comparable to what we observed in this thesis at 3T.

Thus there are a number of factors which may contribute to differences in hippocampal volumes when comparing results across studies. This includes subject age, study protocol and field strength. Given that differences may result from these factors the difficulty is then in trying to attribute volume differences to a particular process or disease. All of this reiterates the need for control groups to be acquired for each study. Given the growing need for imaging studies with large numbers of subjects being acquired over multiple centres, there is a strong argument for analysis protocols be agreed and implemented centrally if manual segmentation is necessary, though this in itself is probably the biggest argument in favour of the need for automated methodologies to be developed and applied.

5.6.4 The relationship between hippocampal volume and sex

Of the 39 normal volunteers, 23 were female and 16 were male. Uncorrected hippocampal volumes for males had a mean value of $3516\text{mm}^3 \pm 426\text{mm}^3$ for the left hippocampus and $3571\text{mm}^3 \pm 544\text{mm}^3$ for the right, while females had uncorrected volumes of $3355\text{mm}^3 \pm 383\text{mm}^3$ for the left hippocampus and $3429\text{mm}^3 \pm 344\text{mm}^3$ for the right. The ICV corrected hippocampal volumes for males had a mean value of $0.22\% \pm 0.03\%$ for the left hippocampus and $0.22\% \pm 0.04\%$ for the right, while females had a mean value of $0.22\% \pm 0.03\%$ for the left hippocampus and $0.23\% \pm 0.03\%$ for the right. While these was a tendency for women to have smaller hippocampal volumes, this was not statistically significant given the inherent variance in the population. Moreover, after correcting the volumes for ICV there was no difference in the hippocampal volumes between men and women. This matches the results which were obtained by Maller et al [211] where hippocampal volumes were measured on $N = 150$ subjects (74M, 76F), with raw volumes for males being $3442\text{mm}^3 \pm 429\text{mm}^3$ and $3388^3 \pm 377\text{mm}^3$ for the right and left hippocampi respectively, and for females being $3155\text{mm}^3 \pm 372\text{mm}^3$ and $3121^3 \pm 374\text{mm}^3$ for the right and left hippocampi respectively. These raw volumes and standard deviations compare very well with the data acquired here. Maller

et al also observed that correction for ICV resolved the majority of any volume differences between the sexes, in fact, after correction for ICV, male volumes were slightly smaller.

Therefore, while there is a slight preponderance for females to have smaller hippocampal volumes than males, this is almost entirely resolved with correcting for ICV. Thus, after correcting for ICV, it is reasonable to assume male and female hippocampal volumes are from the same underlying population distribution.

5.6.5 Hippocampal volumes and laterality

The left and right hippocampal volumes were strongly correlated as one would expect with a Pearson correlation coefficient of 0.62 with a p-value of less than 0.01. Performing a two-sample t-test, where the null hypothesis was that the left and right hippocampal volumes were from the same underlying population, the t-test p-value was found to be 0.49, i.e. the null hypotheses cannot be rejected and therefore it is highly likely that the volumes do indeed form part of the same normal distribution. When corrected for intracranial volume (ICV) the mean and standard deviation of the hippocampal volumes were $0.22\% \pm 0.03\%$ and $0.23\% \pm 0.03\%$ for the left and right hippocampi respectively. Thus there is little difference between the right and left hippocampal volumes, with only a slight tendency for the right hippocampus to be bigger, but this was not statistically significant. These results mirror that which was observed in the meta-analysis of normal hippocampal volumes [232]. Similarly, in the work by Maller et al [211] it was recorded that there was a slight preponderance for right hippocampal volumes to be larger than left hippocampal volumes, though here as in the majority of other studies of normal hippocampal volumes, this difference was not statistically significant. Indeed in many studies, left and right hippocampal volumes are treated without distinction, though for the purpose of the work throughout this thesis, right and left hippocampal volumes were generally treated separately given the potential cognitive associations with laterality.

5.6.6 Further limitations of the normative 3T hippocampal volume data

5.6.6.1 Educational and intellectual effects on the volume of the hippocampus

Individual general intelligence quotient (IQ) has been attributed largely to differences in whole brain, GM and WM volumes [221, 286, 280]. A large study ($n = 90$) of normal healthy adults found that there was a small but significant positive correlation between volumes in

the frontal ($r = 0.25$, $p < 0.05$) and temporal ($r = 0.28$, $p < 0.01$) lobes and IQ [126]. More specifically hippocampal volumes in healthy adults have been shown to correlate significantly and positively with IQ [10] though a more recent study using VBM failed to identify any correlation between IQ and volume of GM and WM in mesolimbic structures [151]. In a further more recent study ($n = 34$), hippocampal volumes were assessed by a surface analysis method and local volume changes in the hippocampus were assessed for correlations with IQ [9]. In contrast to earlier reported studies total brain volume and total intracranial volume did not correlate with any of the IQ measures. Moreover, a statistically significant negative correlation was observed between hippocampal volume and IQ ($r = -0.49$, $p < 0.007$). The inverse correlation between hippocampal volume and IQ appeared to be localised to the anterior portion of the hippocampus. This implies that higher IQ was accompanied by an inward deformation of the hippocampal head or a localised reduction in volume of the surface of the hippocampus.

The results of the work by Amat et al are in contrast to earlier works and in some ways are counter intuitive to what one would suspect [9]. Amat et al point out that their results do not reflect any causal link between a smaller hippocampal volume and intelligence, the inverse correlation could be the result of increased hippocampal volumes in subjects with lower IQ. Moreover, the inverse correlation could be the result of correlation between hippocampal volume and cognitive factors with a third unknown variable. While Amat et al do highlight the limitations and also other possible interpretations of their results this study does serve to show that the relationship between intelligence and hippocampal volume and shape has yet been fully established. While many studies assessing hippocampal volumes consider the effect of IQ and in some cases a measure of educational level such as the number of years in education, many do not [250]. Until such time as a consensus arises around these relationships it would be advisable to record and investigate these factors for any individual study. If one was to believe the earlier work assessing brain volume and IQ and that these factors do indeed correlate then by controlling for ICV one would also be partially controlling for IQ. Therefore it would be inappropriate to include both IQ and head size as covariates in an assessment of hippocampal volume if this was the case.

For the study described in this chapter of the thesis we would ideally have obtained a normal range for hippocampal volumes in a cross section of the population of Glasgow. However, as limited resources for recruiting subjects were available a targeted approach to recruitment was not possible i.e. subjects who responded to the request in the poster regarding the study were all considered for inclusion regardless of their education or intellectual qualities. It was the case, perhaps as a result of where the posters were placed or perhaps as a result of the

motivation of the subjects that the majority of subjects that ultimately participated in the study were either around the mean or higher than average level of educational years i.e. most were university graduates. This was not formally controlled for and is a source of potential bias in this study.

The lack of consideration of the educational background and intelligence of the subjects limits the utility of the normative data acquired in this thesis. For example, any comparison between the data acquired here and other study groups which included a range of educational backgrounds would be confounded by the fact the educational level and intelligence had not been determined in the normal group discussed in this chapter of the thesis. Any future work assessing hippocampal volume will ensure that the educational level and intelligence of the subjects are recorded such that they may at least be controlled for in subsequent analysis. Alternatively, it may be preferable to match the groups in a way that controls for educational level although though this may be challenging to implement in practice.

5.7 Conclusion

Using data acquired from 3T MRI and a combined automatic and manual editing approach, left and right hippocampal volume were segmented for a group of 39 normal, healthy volunteers. The subjects ranged in ages from 19 to 64 years old. The relationships between hippocampal volumes, ICV and age were investigated where it was found that hippocampal volumes were larger with increasing ICV and smaller with increasing age. The ICV results are in keeping with what has already been observed in the peer reviewed literature though the conflicting reports on the effect of aging on the hippocampus make results difficult to compare with any degree of confidence. Given the relationships between hippocampal volumes, age and ICV a multiple regression model was created to determine an equation to describe the size of the hippocampus for a given ICV of a subject of known age. Thus the normal range for hippocampal volumes, for subjects aged 19 to 64 was determined. That is, the mean and standard deviation of the left and right hippocampal volumes were $3421\text{mm}^3 \pm 399\text{mm}^3$ and $3487\text{mm}^3 \pm 431\text{mm}^3$ respectively.

In further comparisons between males and females, after correcting for ICV, no significant difference was observed between the hippocampal volumes across the sexes. In considering laterality differences the results obtained here reflected that which have been reported elsewhere in the literature in that there was a trend for the left hippocampal volume to be slightly smaller than the right.

Thus, a robust methodology has been produced and a normal database of hippocampal volumes for 3T MRI has been established for subjects in the age range between 19 and 64. This normative data will prove useful in future for further comparisons with psychosocial study groups and diseased populations, though this does not negate the need for appropriate control groups to be acquired.

Chapter 6

A quantitative assessment of hippocampal malrotation in normal subjects using 3T MRI

6.1 Abstract

Hippocampal malrotation or as it is also known, Incomplete hippocampal inversion (IHI) is an observation that has been made regarding the shape of the hippocampus. It is believed to owe its origins to the development of the hippocampus. There has been some controversy within the neuroradiology community over whether IHI may play a role in disease processes such as in temporal lobe epilepsy (TLE) or that it is simply part of the spectrum of normality in hippocampal development. Here results are presented for an assessment of IHI in a healthy, normal, adult population. A criteria for defining IHI was determined by three consultant neuroradiologists and they assessed the images based on this criteria. IHI was only attributed to a particular hippocampus when all three neuroradiologists agreed. Few studies have assessed the normal incidence of IHI and of those studies the analysis of IHI extended only to a radiological assessment of the hippocampi. Here we present a more comprehensive and quantitative assessment of IHI. This builds on the techniques discussed in chapters 2 and 5 with the aim of comparing hippocampal volume and morphological differences between normal and IHI hippocampi. IHI was found on 31 of the 84 subjects assessed (37%). This had a strong left-sided preponderance with 25 of the 31 cases having only left-sided IHI (i.e. 81% of IHI cases). 5 subjects (16% of IHI cases) had bi-lateral IHI with only a single subject having right sided IHI (3% of IHI cases).

The quantitative assessment compared the ICV corrected IHI left-sided hippocampal volumes against the ICV corrected normal left-sided hippocampal volumes (25 vs. 52 hippocampi). The IHI hippocampal volumes were determined to be smaller than the normal hippocampal volumes ($p < 0.05$). However, on further inspection it was observed that the ICV of the IHI brains were significantly larger than the ICV of the normal group, confounding the previous result. Further analysis compared the raw hippocampal volume of the IHI hippocampi against their normal, contralateral hippocampi (i.e. 25 vs. 25 hippocampi). The left-sided IHI hippocampi were not significantly smaller than the contralateral, right-sided hippocampi. Given the morphological changes that occur in IHI it was posited that in IHI cases, the volume and shape of the amygdala may be affected. ICV corrected left-sided amygdala volumes adjacent to IHI hippocampi were compared to normal, left-sided amygdala volumes. Raw left-sided amygdala volumes, adjacent to IHI hippocampi were also compared to the right hemisphere amygdala volumes. In both cases there was no significant difference between amygdala volumes adjacent to IHI hippocampi and amygdala volumes adjacent to normal hippocampi.

The quantitative approach was extended by performing a morphological assessment of the IHI hippocampi. A model was created to compare the left-sided IHI hippocampal volume shape to the normal left-sided hippocampal volume shape. The model included ICV and age as covariates. The same model was created for amygdala volume shape. While the uncorrected statistics implied that there may be hippocampal volume differences at the lateral-inferior aspects of the hippocampal body and inferior aspects of the hippocampal head between the normal and IHI volumes, none of the statistical differences survived the false discovery rate (FDR) correction for multiple comparisons that was applied. Therefore, the morphological analysis was unable to detect the average shape differences between IHI hippocampi and normal hippocampi. No statistically significant morphological differences were observed between amygdala volumes adjacent to IHI hippocampi and amygdala volumes adjacent to normal hippocampal volumes.

To the best knowledge of the author of this thesis, this is the first time such a comprehensive, quantitative assessment of hippocampal IHI in normal volunteers has been performed. This includes for the first time an assessment of amygdala volumes adjacent to IHI hippocampi and a morphological treatment comparing both IHI hippocampi and adjacent amygdala shape to normal hippocampal and amygdala volumes. The work performed in this chapter is currently under discussion to decide if a single manuscript covering the radiological and quantitative methodological aspects of this work should be produced, or whether two separate manuscripts would better cover both of these topics. Once this decision has been made the manuscript(s)

will be submitted to a peer reviewed journal.

6.2 Introduction

As greater magnetic field strengths have become available for structural MRI it has been possible to assess complex structures such as the hippocampus in ever increasing detail. Chapters 2, 3 and 4 have assessed developed methodologies to assess two differing socioeconomic groups, where chapter 2 focused on hippocampal volumetry in these groups. Chapter 5 also assessed hippocampal volumes in a group of normal subjects which were used to establish a normative database. Despite the fact the data from the PSOBID study considered data from two different socioeconomic groups, the subjects themselves were otherwise healthy volunteers, combining this dataset to the normal database of hippocampal volumes, a relatively large sample ($N = 84$) of normal images was collected. In this chapter the combined dataset of normal volunteers was used to assess hippocampal volumes for incomplete hippocampal inversion.

A somewhat contentious issue within neuroradiology [293] has been defining what is known as hippocampal malrotation (HIMAL) [32], although some prefer the term incomplete hippocampal inversion (IHI) [27]. Part of the major interest in better understanding the hippocampus has come through a desire to better understand the etiology of temporal lobe epilepsy (TLE) [131]. TLE is the most common form of partial epilepsy and has been commonly associated with hippocampal sclerosis [236]. Further to assessing only hippocampal sclerosis, it has been considered that hippocampal malrotation may be associated with the incidence of TLE. However, few studies have sought to establish the incidence of malrotation in the normal population, this being a necessary step before one is able to have confidence in asserting any direct or indirect link between IHI and epilepsy or other disease.

In describing an abnormal feature in the brain it is critical that the description is clear and without ambiguity. Thus, the two most frequently used terminologies to describe a malformed hippocampi, that is HIMAL and IHI are discussed below.

6.3 Literature review

6.3.1 What is IHI / hippocampal malrotation?

Gamss et al described HIMAL as “In normal fetal development, the hippocampus inverts within the medial temporal lobe. In cases of HIMAL hippocampal inversion fails to occur”

[131]. This is a brief description and doesn't fully explain what is meant by 'inverts'. Preceding the work of Gamss et al, Barsi et al demonstrated 32 cases of HIMAL from a population of 527 patients with a suspicion of epilepsy [32]. Therefore around 6% of the people suspected of epilepsy had a demonstratable abnormality. It could be argued that as each subject has two hippocampi and Barsi et al describe HIMAL as a unilateral effect, that only 3% of the hippocampi are abnormal. Furthermore, even within Barsi et al's criteria for HIMAL, not all of the criteria were met in all cases. In fact, the only criteria which was satisfied in all of the 32 so-called HIMAL subjects was the observation of a "blurry" internal structure. This is a fairly non-specific description to consider and it was unclear how such a factor could be reliably attributable to the underlying anatomy and not simply a consequence of poorer image quality as the result of patient motion.

Barsi et al's criteria for describing HIMAL was as follows:

1. Incomplete inversion of the hippocampus with an abnormally round shape
2. Unilateral involvement of the whole hippocampus
3. Normal signal intensity and size
4. Blurred internal structure
5. Abnormal angle of the collateral sulcus
6. Abnormal position and size of the fornix
7. Normal size of the temporal lobe
8. Enlargement and particular configuration of the temporal horn typical of corpus callosum agenesis
9. Normal corpus callosum

This was a more extensive list than was described by Gamss et al, with nine separate criteria being used to define HIMAL. However, despite the obvious desire for a comprehensive and, at least, semi-qualitative description of HIMAL, some of the terminology still lacks precision and is inherently flawed. For example, consider "an abnormally round shape", this was not measured in any quantitative manner, moreover, if one is describing an abnormally round

shape, then what is the definition of normality? Similarly, for example, another criteria defining HIMAL was “the angle of the collateral sulcus”, which in 50% of positive subjects in Barsi et al’s study (16/32) the collateral sulcal angle was described as being 45°. Strangely, one subject had a collateral sulcal angle of 30 degrees, with lower angles being described as ‘deformed’, so presumably then this was not the criteria for HIMAL. If the normal angle for the collateral sulcus was, as described by Barsi et al, as being horizontal to the hippocampus at the level of its body and tail, then shouldn’t all deviations from this be considered abnormal? “The inversion of the hippocampus during normal development was lacking, hence the term malrotation: the perpendicular axis of the subiculum and the hippocampus was turned medially. In cross section (presumably meaning in the coronal plane), the hippocampus appeared round, and its upper surface looks into the widened choroid fissure”. This was Barsi et al’s most complete description of what is meant by HIMAL, and hints that at the heart of the observation of HIMAL is incomplete inversion of the hippocampus. Focusing more on this aspect of the hippocampus formation than on the many other supplementary measures may be a better approach.

6.3.2 What is incomplete hippocampal inversion?

Raininko and Bajic, commenting on the work of Gamss et al [131], suggested that the term HIMAL was questionable and that it would be preferable to use the term incomplete hippocampal inversion (IHI) [293]. Earlier work by Bronen and Cheung, had considered the hippocampal formation and described that in 12/58 hippocampi that the head of the hippocampus was not oval shaped, with the shape being either described as rectangular (10 hippocampi), circular (1 hippocampus), or square (1 hippocampus) [61]. Similarly Bajic et al observed IHI, defined as a round or pyramidal hippocampus with a vertical collateral sulcus, in 19% of healthy subjects and in a population of patients without epilepsy or obvious developmental brain abnormality [27]. Furthermore, no correlation with round hippocampi and age was observed, indeed the incidence of IHI appears to be similar at adulthood as at the gestational age of 25 weeks and onwards [25]. Despite the incidence of IHI in healthy and non-epileptic populations being non-negligible, a recent study by Bajic et al found a statistically significant difference between the populations without and with seizures (18% vs. 30%, $p < 0.05$) [26]. The incidence of IHI in some epileptic syndromes such as cryptogenic generalised epilepsy was very high, but there was no significant difference between TLE patients and the control group when IHI was the only differing factor in the temporal lobe.

Thus Bajic et al, reached the conclusion that there is no causal link between IHI and TLE [26].

6.4 Aims and hypotheses

6.4.1 Hippocampal volumes and shape analysis: can a more quantitative analysis add information to the debate on incomplete hippocampal inversion?

As has been discussed elsewhere in this thesis (chapter 2), interest and analysis of the hippocampus goes well beyond a consideration of the morphology of the hippocampus in epilepsy alone. While the hippocampus has been implicated to play a role in depression and other diseases [72, 358], what seems to be unique about epilepsy is the proposed causal role of malrotation of the hippocampus to the disease, as opposed to gross anatomical differences being a subsequence of the disease. Assessments of malrotation of the hippocampus in epilepsy and normal populations to date have largely been qualitative, radiological assessments with few rigorous measurements [32, 27, 131]. In contrast, assessments of the hippocampus in depression have evolved from basic manual outlining techniques into sophisticated, quantitative segmentation methods and surface modelling techniques [35, 245].

We propose here to take the novel approach of providing a comprehensive, quantitative assessment of the volume and morphology of IHI hippocampi and their local environment in healthy volunteers. This was done with the aim of addressing the following questions.

6.4.2 Volume analysis of the hippocampus and amygdala

From the radiological definition of IHI, one of the key criteria by which malrotation has been defined is for the hippocampus to be circular rather than elliptical. What is unclear is whether or not these morphological differences correspond to volume differences between normal and malrotated hippocampi? Thus, the initial question to ask for the first time is: are IHI hippocampi volumes smaller than normal hippocampal volumes?

As there is a perceived laterality to malrotation as observed in the FEBSTAT study [194] and also from the radiological assessment done here, a supplementary question asked was: are left hemisphere malrotated hippocampi smaller than the normal right sided hippocampi in those individuals with unilateral left sided hippocampal malrotation?

A further question was posed that relates more to the anatomical environment around the hippocampus than the hippocampus itself. Given that morphological differences of the malrotated hippocampus are often observed in the hippocampal head, it was reasonable to posit that such changes in morphology might also have a 'knock-on' effect on the development of the amygdala. The amygdala has received little detailed consideration in the literature surrounding hippocampal malrotation. Whether this is because it has been deemed unremarkable and therefore of little interest or simply that the hippocampus has always been the focus of the studies around this topic remains unclear. The means now exists to apply a comprehensive assessment of the amygdala structure in the context of hippocampal malrotation. Thus in a similar manner to the question posed above we asked for the first time: are amygdala volumes ipsilateral to malrotated hippocampi smaller than normal amygdala volumes?

Similarly to the case of the hippocampus, this was followed up by asking: are left hemisphere amygdala volumes smaller than the normal right-sided amygdala volumes in those individuals with unilateral left-sided incomplete hippocampal inversion?

The hippocampus and amygdala are both shown on a T1-weighted 3D image acquisition in figure 6.1 on the following page. On this figure the red arrow points to the amygdala, the green arrow points to the head of the hippocampus and the blue arrow points to the tail of the hippocampus.

6.4.3 Morphological analysis of the hippocampus and amygdala

Rather than stop at only a volumetric analysis of the whole hippocampus and amygdala, the analysis was further refined to assess morphological differences between malrotated and normal hippocampi. When assessing hippocampal shape, differences between IHI and normal cases may not necessarily be best elicited through a simple comparison of whole hippocampal volumes between the groups as two objects of the same volume may be a different shape. Thus, what may be important is not the gross volumetric differences of the hippocampus as a whole, but the specific location within the structure of any volumetric differences. Identifying where hippocampal volumetric differences occur may give a more detailed insight into the underlying subfields that are affected which may help to improve the understanding of the origins of IHI. Therefore, the next question asked was: can the morphological volume differences of the hippocampus used to define IHI as observed radiologically be elicited from a quantitative surface analysis that compares IHI hippocampi and normal hippocampi? To

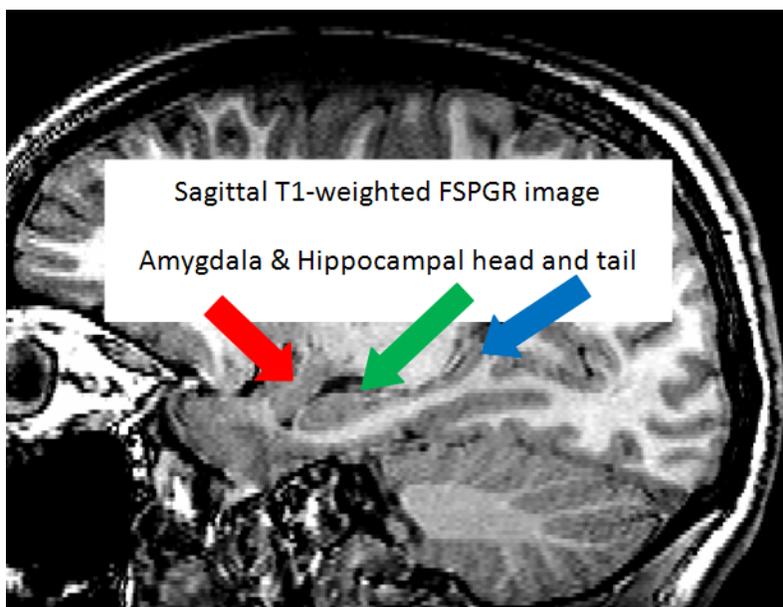


Figure 6.1: T1-weighted 3D acquisition with the amygdala and hippocampal head and tail shown

put it another way, can the changes observed in the hippocampus by radiologists, such as narrowing of the width of the body from the a normal hippocampus to a more rounded IHI be detected when a morphology of a group of IHI hippocampi are compared to a group on non IHI, normal subjects.

Following the same argument as above with respect to the potential impact of IHI on the amygdala, again we asked for the first time: are morphological volume differences present in the amygdala structure when a group of amygdala structures ipsilateral to malrotated hippocampi are compared to normal group of amygdala shapes?

6.5 Methodology: defining hippocampal malrotation radiologically

6.5.1 Subjects

To maximise the number of subjects which may be assessed for IHI the subject datasets from the studies described in chapters 2 and 5 were combined to form a larger study group of 84 subjects. All subjects used in this chapter had been reported as being radiologically normal prior to the assessment of IHI. As the subjects from chapter 2 were recruited from two distinct socioeconomic groups it was a concern that this unorthodox, but pragmatic

approach of combining the data one might introduce a bias towards either the more deprived or more affluent socioeconomic group. To negate these concerns the incidence of IHI was compared across the two separate SES groups. No single group displayed a significantly increased incidence of IHI where 38% (8/21) of affluent SES subjects displayed IHI and 33% (7/21) of poorer SES subjects displayed IHI. The incidence of IHI in the normal group was 28% (11/39).

In total from the combined PSOBID and normal datasets there were 61 males and 23 female healthy volunteers with a mean age of 44 years old (age range 19-67 years).

6.5.2 Image acquisition

All the images were acquired on a GE healthcare 3T Signa excite HD MR scanner (Milwaukee, USA). The MR imaging protocol included a 3D T1-weighted FSPGR and high resolution T2-weighted FSE imaging with coronal and sagittal sections acquired perpendicular to and parallel to the plane of hippocampi respectively. A matrix size of 512 x 512 and a 20cm x 20cm FOV giving an in-plane voxel size of 0.39mm x 0.39mm was used to acquire the images. Coronal images were acquired with 4mm thick slices and a 1mm slice gap. Sagittal images were acquired with 3mm thick slices and a 1mm slice gap.

6.5.3 Radiological assessment of malrotation

Each subject's data was reviewed by three consultant neuroradiologists. The neuroradiologists utilised the coronal and sagittal T2-weighted FSE images as well as the 3D volume T1-weighted FSPGR acquisition. These images offered a comprehensive visualisation of the hippocampus. Examples of the coronal and sagittal T2-weighted FSE images are shown in figures 6.2 below and 6.3 on the following page respectively.

The aim was not to investigate inter-observer variability and thus a diagnosis of IHI was only made when all three neuroradiologists agreed. The neuroradiologists agreed to broadly use the predetermined nine-point morphological criteria suggested by Barsi et al to determine IHI [32]. Barsi et al's criteria included: abnormal round shape of the hippocampus, unilateral or bilateral involvement, the part of hippocampus involved, blurring of the internal structure, vertical angulation of collateral sulcus, abnormal position of the fornix, prominence of

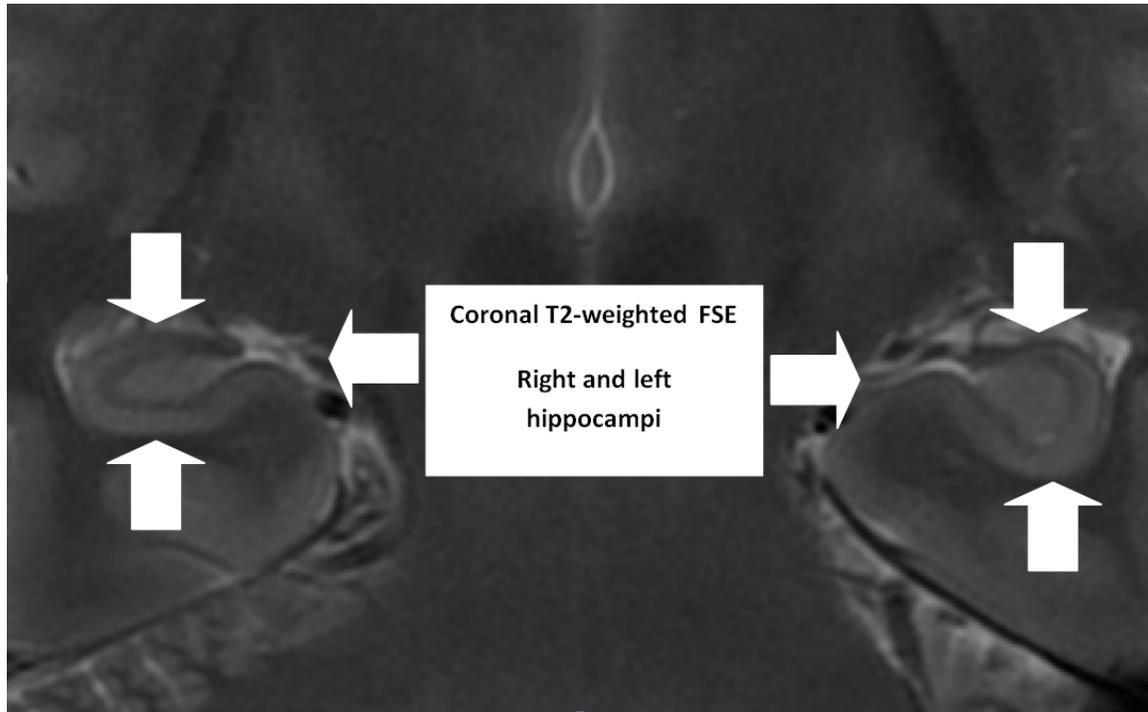


Figure 6.2: Right and left hippocampi and visualised using 3T MRI and a coronal T2-weighted FSE acquisition

temporal horn and collateral sulcus. Thus the factors recorded in this study when defining hippocampal malrotation were: 1. hippocampus shape 2. parts of hippocampus affected 3. signal intensity 4. blurring of internal structures 5. orientation of collateral sulcus 6. collateral eminence 7. position of fornix 8. temporal horn 9. corpus callosum 10. other structural abnormality

6.6 Quantitative image analysis methodology

The automated quantitative image analysis utilised only the T1-weighted FSPGR 3D volume data.

6.6.1 Pre-processing of MRI data

Image intensity inhomogeneities are inherent in all MR images and are a consequence of the fact that coil sensitivity varies with distance from the coil. On modern MR scanners there are typically manufacturer supplied algorithms to correct for image inhomogeneities, for example, GE applies a surface coil intensity correction (SCIC), and this sufficiently corrects the image intensities making them suitable for clinical reporting by a radiologist. However, for the

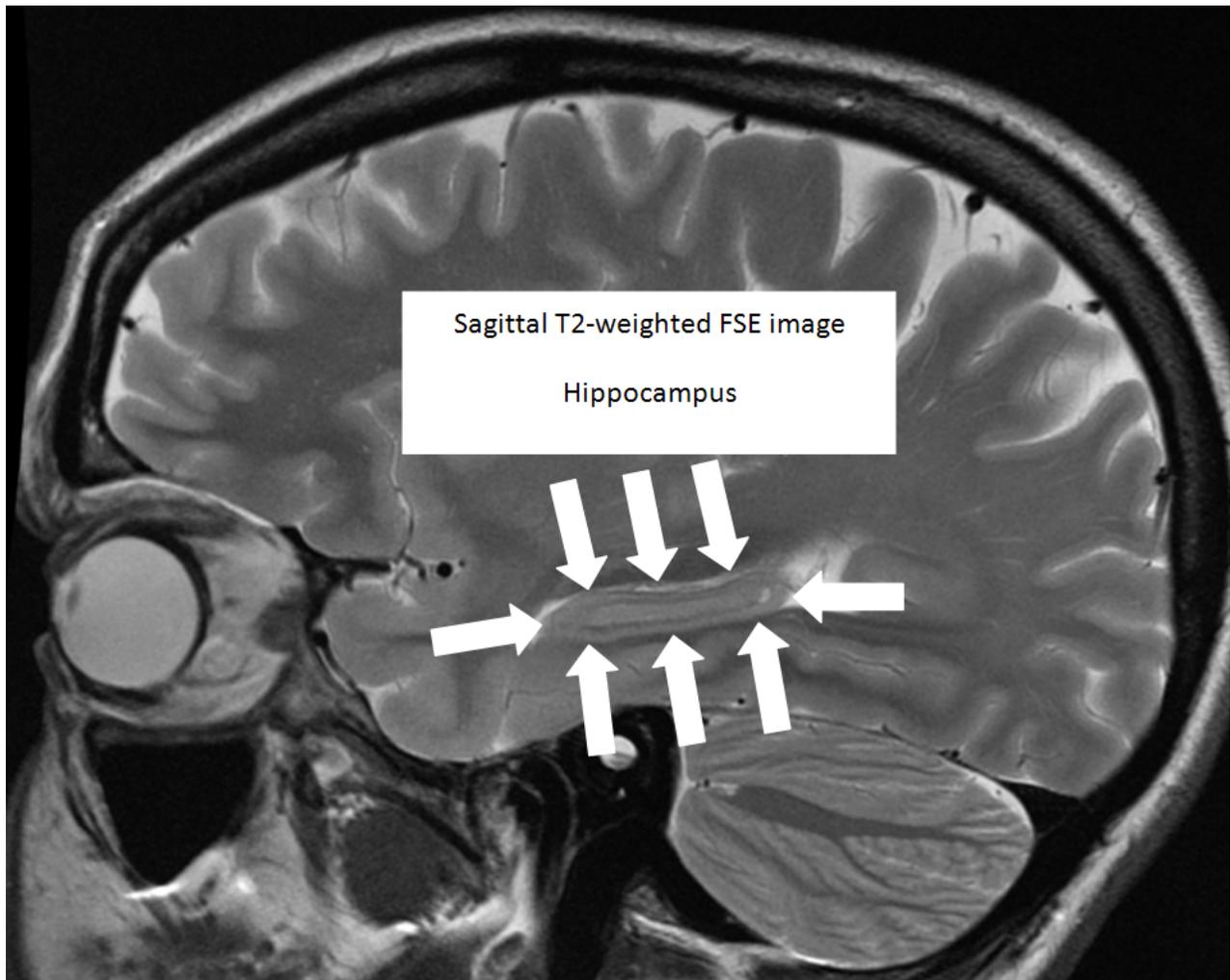


Figure 6.3: The hippocampus as visualised using 3T MRI and a sagittal T2-weighted FSE acquisition

purposes of many image processing techniques it is advantageous that these inhomogeneities are further corrected. For example, in producing segmented tissue class maps, the intensity of a voxel defines the probability of which tissue class it belongs to. Therefore, poor image inhomogeneity may lead to poorly segmented images which can directly affect the calculated tissue type volumes.

A commonly used inhomogeneity correction is N3 and this was the inhomogeneity correction algorithm applied on the data discussed in this chapter [326]. N3 stands for nonparametric, non-uniform intensity normalization and benefits from being able to be applied to any pulse sequence and is insensitive to pathology. To avoid any dependence on the underlying anatomy N3 implements an approach to estimate both the multiplicative bias field and the distribution of the true tissue intensities.

N3 has been employed over a number of years within the early pre-processing stage of the Freesurfer data processing pipeline, a well known software package for analysis of structural and functional MRI data [123]. In recent years there has been some work detailing the optimal parameters to use within N3 [378] and in particular optimising the use of this algorithm on 3T MRI data [51].

6.6.2 Data processing: automated segmentation of the hippocampus with manual editing

Following the inhomogeneity correction the data was entered into FSL's FIRST analysis pipeline. The first stage of this pipeline performs affine normalisation of all the data to the Montreal Neurological Institute (MNI) 152 template using FSL's FLIRT algorithm [220, 172]. FIRST is a component of the FMRIB's software library. FIRST was introduced in chapter 2 of this thesis but is briefly summarised here.

FIRST employs a Bayesian shape and appearance model to allow subcortical brain structures to be segmented, and for subsequent analysis to be performed at a more localised level around a structure. The Bayesian model is informed and limited by a large training dataset. Conditional distributions are calculated by the model. The appearance model is framed as the conditional distribution for a given shape, which is useful as the posterior solution can then be used to fit the model to new data. The conditional distribution may also be used in terms of predicting one shape distribution given the location of another shape. This framework generalises to other types of data beyond shape and intensity. The analytic form for the conditional distribution scales appropriate covariance in such a way that one does not need an empirical or arbitrary weighting for relating intensity variance to shape variance as is typically required [278].

The training data for the FIRST algorithm uses 335 brain images. This program has been used to assess hippocampal volume in Alzheimer's [85] and also amygdala volumes during the acute phase of depression [354]. It is worth noting that the training data includes both normal and pathological brains. 19 structures in total are modelled, these are: the brainstem and then for each hemisphere: the hippocampus, amygdala, caudate, putamen, lateral ventricles, nucleus accumbens, thalamus, pallidum and cerebellum.

Volume labels are parameterised by deforming a 3D mesh representation of the most typical structure to each subject. The cross-subject vertex correspondence is conserved within the surface motion constraints and minimal smoothing forces within the deformable model. By

sampling normalised intensities along the surface normal at each vertex, appearance models can then be generated. Thirteen samples per vertex were used at 0.5mm intervals. Intensities were normalised by subtracting the median intensity value across a given structure.

From the data it was now possible to extract both the left and right hippocampal volumes for all the normal volunteers. The automatically segmented data was then manually edited to remove any gross errors in the segmentation procedure. These volumes were then separated into their malrotated and normal (non-malrotated) groups based upon the neuroradiological criteria defined previously. The volumes were then subjected to further statistical analysis.

6.6.2.1 Vertex analysis

The work presented in chapters 2 and 5 of this thesis displayed FIRST's ability to generate automatically segmented hippocampal volumes. However, Patenaude et al developed further analysis which can be performed on the segmented volumes using FIRST [278]. The further analysis takes the form of vertex analysis which enables volume differences to be assessed at a more local level within the segmented volume at the level of the voxel. Vertex analysis assesses the difference in the mean vertex position between two or more groups. Vertex analysis is performed by multivariate testing of the three dimensional coordinates of corresponding vertices. Each vertex is analysed independently and following initial comparison between the groups FDR multiple comparison correction may be applied in the same manner as for standard brain image analysis.

Vertex analysis may be performed in standard space or in group space. Once the coordinates are transformed into the desired space, a multivariate F-test is performed for each vertex using the Multivariate General Linear Model (MVGLM) with Pillai's trace as the test statistic. A summary of the equations describing the MVGLM with Pillai's trace can be found in the appendix B.

The flexibility of the GLM not only facilitates comparison between groups but also enables the effects of multiple factors to be considered such as age, ICV or disease state. F-statistics which are calculated are sensitive to both volumetric increases and decreases. While other methods such as VBM have also been used to localise region of volumetric difference, as it is based on segmented tissue class types it is sensitive to errors in the segmentation procedure and the extent of smoothing that is applied. The vertex analysis uses a combined shape and appearance model to determine the structural boundary. As the vertex method is based on the local geometry and shape and has no additional smoothing it has the potential to identify focal regions of volumetric change more precisely.

6.7 Results of radiological assessment of hippocampal malrotation in normal volunteers

Shape of hippocampus: The shape of the hippocampus was assessed on the coronal T2-weighted images. A round shaped hippocampus was found in 37% (31/84) of healthy subjects.

Sex: IHI was observed in 23 males and 8 females. There was no significant gender preponderance with approximately a third of both males and females having IHI.

Lateralization: The IHI was seen on left side in 81 % (25/31), bilateral in 16 % (5/31) and isolated right side in 3% (1/31). In the majority of subjects exclusively the body of the hippocampus was affected 72% (26/36), head and body in 14% (5/36), head only in 8% (3/36) and all parts of hippocampus were affected in 6% (2/36).

Blurring of internal structures: Blurring of the internal structure was observed in 11% (4/36) of subjects and was normal in 86% (31/36). There were no abnormal signal intensities observed in any of our healthy volunteers.

Collateral Sulcus and prominence of collateral eminence: The orientation of collateral sulcus was assessed on coronal T2-weighted images. The normal collateral sulcus is usually horizontal at the level of body and tail. In hippocampal malrotation it is vertically orientated on the coronal section. Barsi et al used various images to demonstrate its vertical orientation [32]. However, a single observation was used here i.e if the angle of the collateral sulcus was greater than 70° in any of the coronal slices from head to tail of the hippocampus this was deemed to meet the criteria for malrotation. This singular observation was used to avoid confusion introduced by variations in the entorhinal sulcus, collateral and occipito-temporal sulcal configurations. A vertical collateral sulcus of greater than 70° was found in 72% (26/36) of cases and a similarly in the occipito-temporal sulcus in 25% (9/36) of cases. The results of two subjects were uncertain because of variation in the sulcal pattern. A prominent collateral eminence was recorded, that is, where the grey matter of the collateral sulcus rises above the level of hippocampal fissure, in 72% (26/36) of cases.

Position of Fornix: The position of the fornix was assessed on coronal T2-weighted images at the level of the third ventricle. The fornix is normally bilaterally symmetrical. Barsi et al's series showed an abnormal position of fornix on the same side as the IHI in 90% of cases of which 88% were low lying [32]. 56% (20/36) were found to have an abnormally placed fornix which was ipsilateral low lying in 41% (15/36) of cases, contralateral low lying in 11% (4/36)

of cases and ipsilateral side and high lying in 3% (1/36) of cases. A normal placed fornix with IHI was found in 44% (16/36) of cases. Gamss et al found only 0.35% of cases with low fornix had one or more criteria of IHI and hypothesized that a low fornix might be a normal variant [131]. In our series all the subjects with an abnormally placed fornix showed one or more features of IHI.

Temporal Horn: The temporal horn was larger in 47% (17/36) of cases and normal in 44% (16/36) of cases. We found a large, bilateral temporal horn in two cases and a large temporal horn on the contralateral hemisphere in a single subject.

None of the subjects had an associated corpus callosum or other structural abnormality.

6.8 Results of quantitative assessment of hippocampal malrotation

6.8.1 Results of hippocampal volume analysis

6.8.1.1 Compare malrotated hippocampal volumes against normal hippocampal volumes

The aim here was to compare left malrotated hippocampal volumes against normal left hippocampal volumes. As we are comparing volumes from across different subjects it was necessary to correct volumes for total intracranial volume (ICV). The segmented hippocampal volumes were corrected for ICV in the same manner as discussed in chapters 2 and 5, and a two-sample t-test was done. The null hypothesis here was that there was no difference between the malrotated ($n = 25$) and normal ($n = 52$) left hippocampal volumes. The mean volume for the malrotated hippocampi was 0.212 (%ICV) while the mean volume for the normal hippocampi was 0.232 (%ICV), this difference was statistically significant with a p-value of 0.006. As this was less than 0.05 we can reject the null hypothesis and infer for the first time that after correcting for differences in head size malrotated hippocampal volumes are smaller than normal hippocampal volumes. Figure 6.4 below shows graphically the difference between the mean ICV corrected malrotated volumes and the normal hippocampal volumes.

It is worth noting that, following this result the ICV values between the two groups were compared, to rule out any group difference of ICV which might explain the group difference

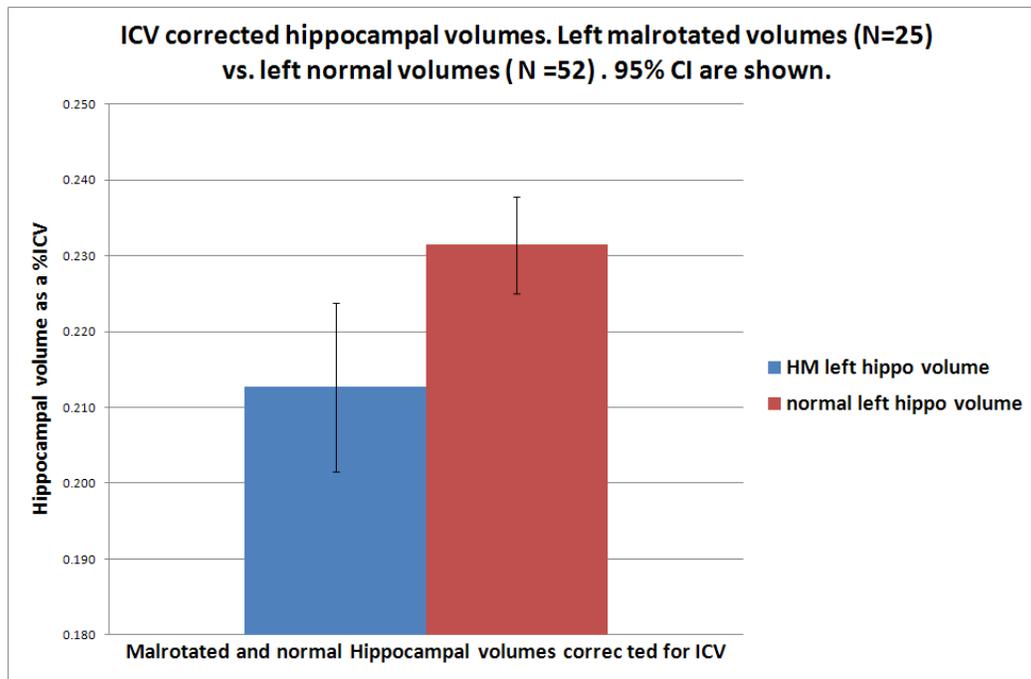


Figure 6.4: ICV corrected, left malrotated volumes vs. normal, left hippocampal volumes

observed. It was found that the ICV values between the malrotated and normal groups were larger in the malrotated group on average and that this group difference was statistically significant at a p-value of 0.03. The ICV volumes in the malrotated group were $\sim 5\%$ larger than the normal group.

6.8.1.2 Comparing left malrotated hippocampal volume with right hippocampal volumes in the same subjects

Performing a simple comparison between the left and right hippocampal volumes in subjects with a left malrotated hippocampus had its benefits. As volumes within the same subject were compared the raw hippocampal volumes were used as there was no need to correct the volumes for age or ICV.

Of the 31 subjects in which malrotation was identified, 25 were left-sided, 5 cases were bilateral and 1 was right-sided. Therefore, in order to compare malrotated volumes against normal hippocampal volumes, the 25 left-sided malrotated volumes were compared to the 25 right-sided volumes. The results are shown in figure 6.5 below.

A two sample t-test was used to assess whether the observed differences between the malrotated left and right hippocampal volumes were statistically significant. The null hypothesis

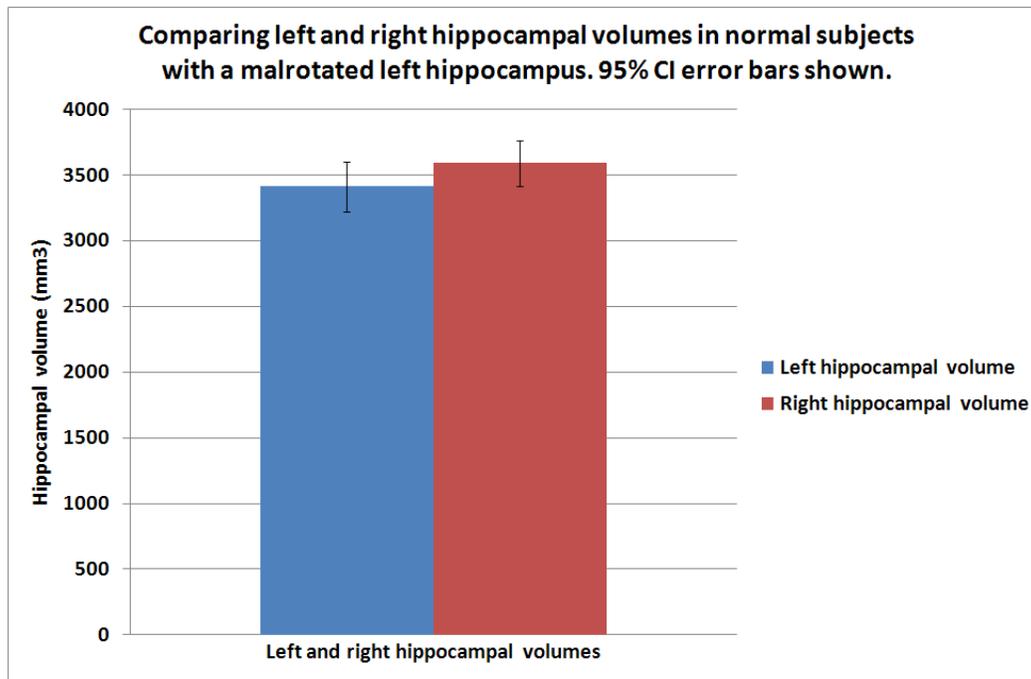


Figure 6.5: Comparing left and right hippocampal volumes in subjects with a malrotated left hippocampus

was that there was no difference between left and right hippocampal volumes. The p-value was 0.19 and as this was greater than 0.05, the null hypothesis was not rejected.

6.8.2 Results of amygdala volume analysis

6.8.2.1 Comparing amygdala volumes, adjacent to malrotated hippocampi to amygdala volumes adjacent to normal hippocampi

The aim here was to compare left amygdala volumes, i.e. the amygdala that was adjacent to the malrotated hippocampi versus normal amygdala volumes. The results are shown in figure 6.6 below.

The amygdala volumes, corrected for ICV were not significantly different between the malrotated and normal groups.

6.8.2.2 Compare left-sided amygdala volumes adjacent to malrotated hippocampi against right-sided amygdala volumes adjacent to normal hippocampi

To compare the left and right amygdala volumes only the left sided hippocampal malrotated

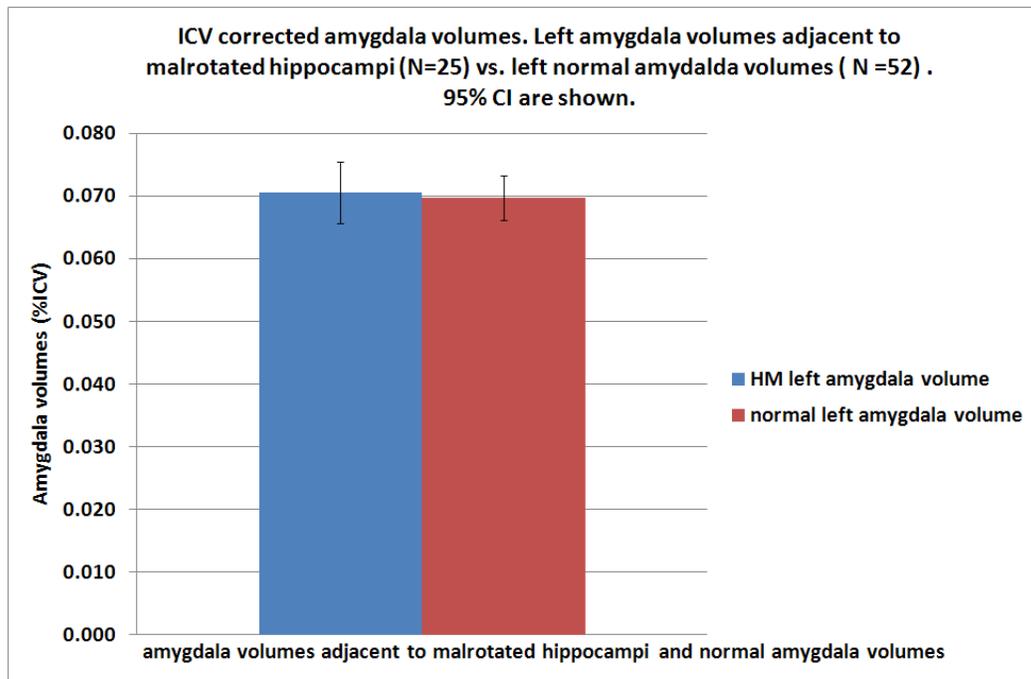


Figure 6.6: ICV corrected amygdala volumes, comparing left amygdala volumes adjacent to malrotated hippocampi against normal left sided amygdala volumes

subjects were assessed i.e. $n = 25$. The results are shown in figure 6.7 below.

As was the case for assessing raw hippocampal volumes, performing a simple comparison between the left and right amygdala volumes in subjects with a left malrotated hippocampus has its benefits. As volumes were compared within the same subject there was no need to correct the raw hippocampal volumes for age or ICV.

A two sample t-test was used to assess whether the observed differences between the left and right hippocampal volumes were statistically significant. The null hypothesis was that there was no difference between the left and right amygdala volumes. The p-value was 0.08 and thus as it was greater than 0.05 the null hypothesis was not rejected.

6.8.3 Results of hippocampal vertex analysis

6.8.3.1 Surface analysis of hippocampal volumes: malrotated vs. normal volumes

Following the automated segmentation of the individual hippocampal volumes, where necessary the volumes were manually corrected and raw volumes were then obtained. A comparison

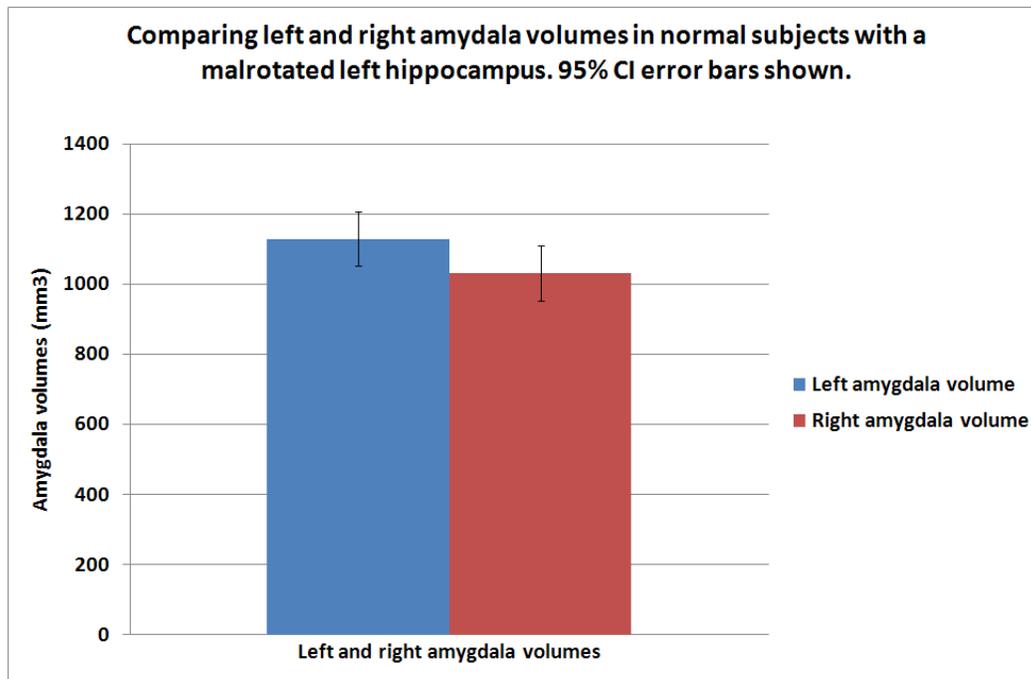


Figure 6.7: Comparing left and right amygdala volumes in subjects with a malrotated left hippocampus

of the raw volume data between the IHI and normal hippocampal cases was shown above. To further refine the analysis a vertex-wise surface analysis was performed using FSL's FIRST program. A MVGLM model was created to compare the malrotated left and normal left hippocampal volumes. Age and ICV were added to the model as covariates such that volume differences may be identified regardless of the impact of these factors. The results of the surface analysis are shown in figure 6.8 on the following page. The analysis consisted of 78 hippocampi in total, 26 with left sided malrotation and 52 which were deemed normal hippocampal volumes.

Figure 6.8 shows 4 views of regions of volume difference between the malrotated and normal groups. The size of the vectors shown is arbitrary and is for illustration purposes only. The arrow vectors simply highlight the direction of volume change between the two groups. The colour scale here was a measure of statistically significant differences in the form of Z-scores and ranges from 2 to 5. Values less than approximately 2 imply little or no volume change while values around 5 are those areas where statistically significant volume differences are present. Here the vectors point from the regions of difference between the normal volumes to the malrotated volume.

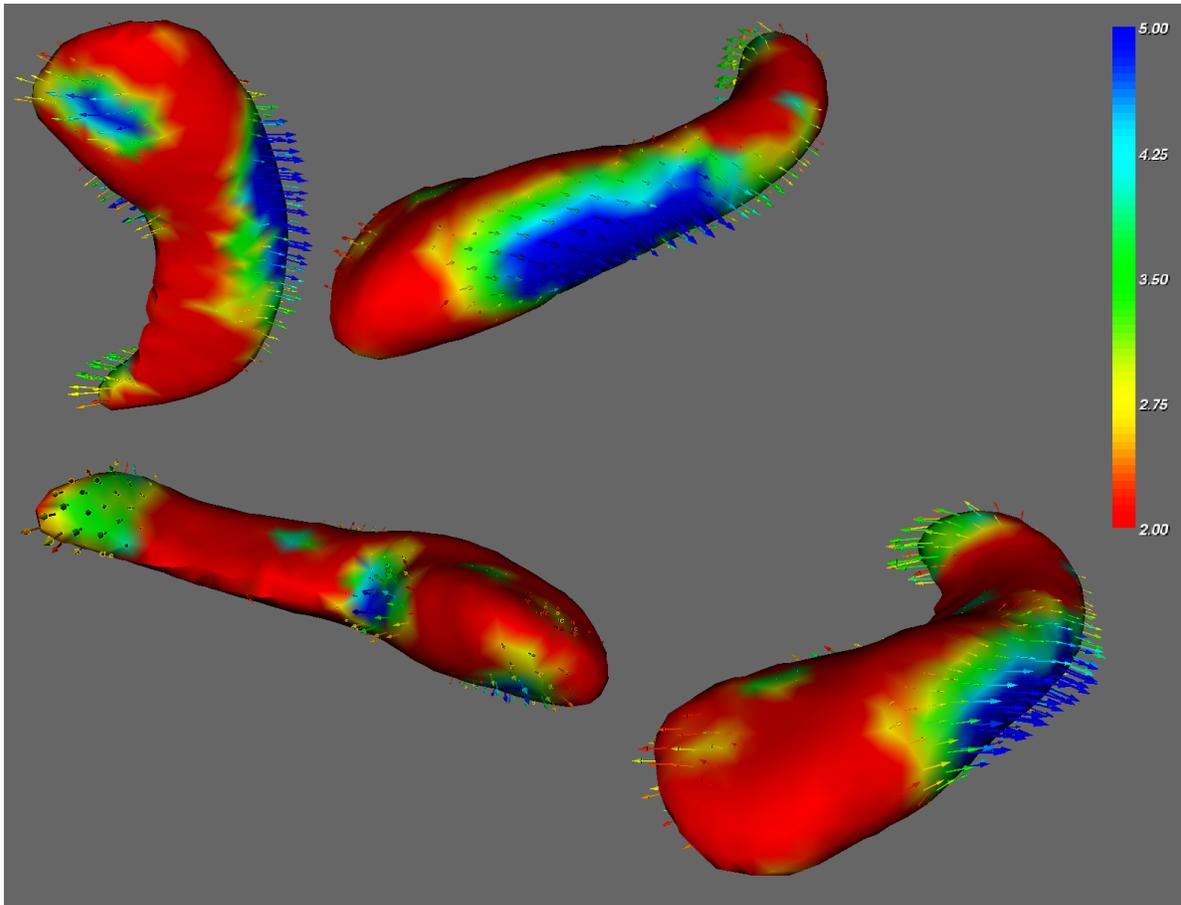


Figure 6.8: Results of hippocampal surface analysis comparing malrotated left and normal left hippocampal volumes

Crucially, when performing statistical voxel comparisons in imaging studies it is important to ensure a correction is applied to account for multiple comparisons (MCC). At the time of writing this thesis FIRST offers a single option for multiple comparison correction. This was to apply a false discovery rate correction (FDR). The results for the FDR corrected surface rendering are shown in figure 6.9 on the following page.

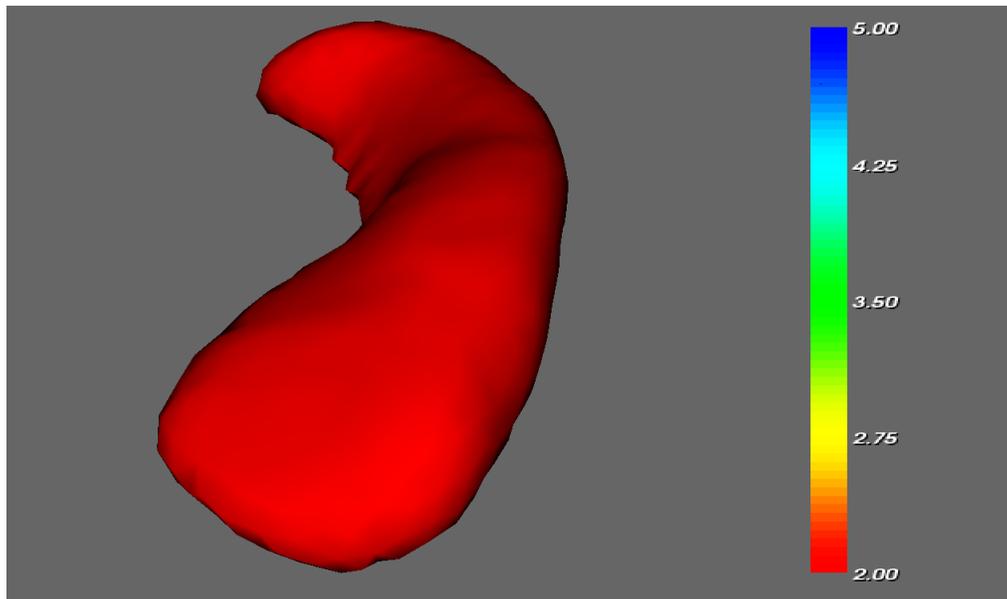


Figure 6.9: FDR corrected results of hippocampal surface analysis comparing malrotated left and normal left hippocampal volumes

Following the FDR MCC no statistically significant volume differences were observed between the malrotated and normal groups. This could have been because the volume differences between the two groups were too subtle to be detected by this method, or that there were too few numbers of subjects in each group to enable the appropriate level of confidence in detecting the volume differences. Further refinements to the MCC correction method might improve the likelihood of detecting significant volume differences. One way in which this could be done would be to refine the hypothesis of where in the hippocampus the volume change is to be expected. A mask could then be applied to remove parts of the volume which were not of interest, this would reduce the total number of statistical comparisons that had to be done.

6.8.4 Results of amygdala vertex analysis

6.8.4.1 Surface analysis of amygdala volumes: amygdala volumes adjacent to malrotated hippocampi vs. amygdala volumes adjacent to normal hippocampi

In a similar manner to the hippocampal vertex analysis, analysis of the amygdala volumes was performed. The aim here was to compare the surface of the amygdala volumes for normal subjects with normal hippocampi and normal subjects with malrotated hippocampi. As for

the case of the hippocampal analysis, the analysis consisted of 78 amygdala volumes in total, 26 volumes where the adjacent left sided hippocampus was malrotated and 52 amygdala volumes adjacent to normal hippocampi. The results of the analysis are shown in figure 6.10 on the following page. The amygdala results are shown in the context of the hippocampus as it was difficult to interpret the amygdala orientation when the volume is presented in isolation.

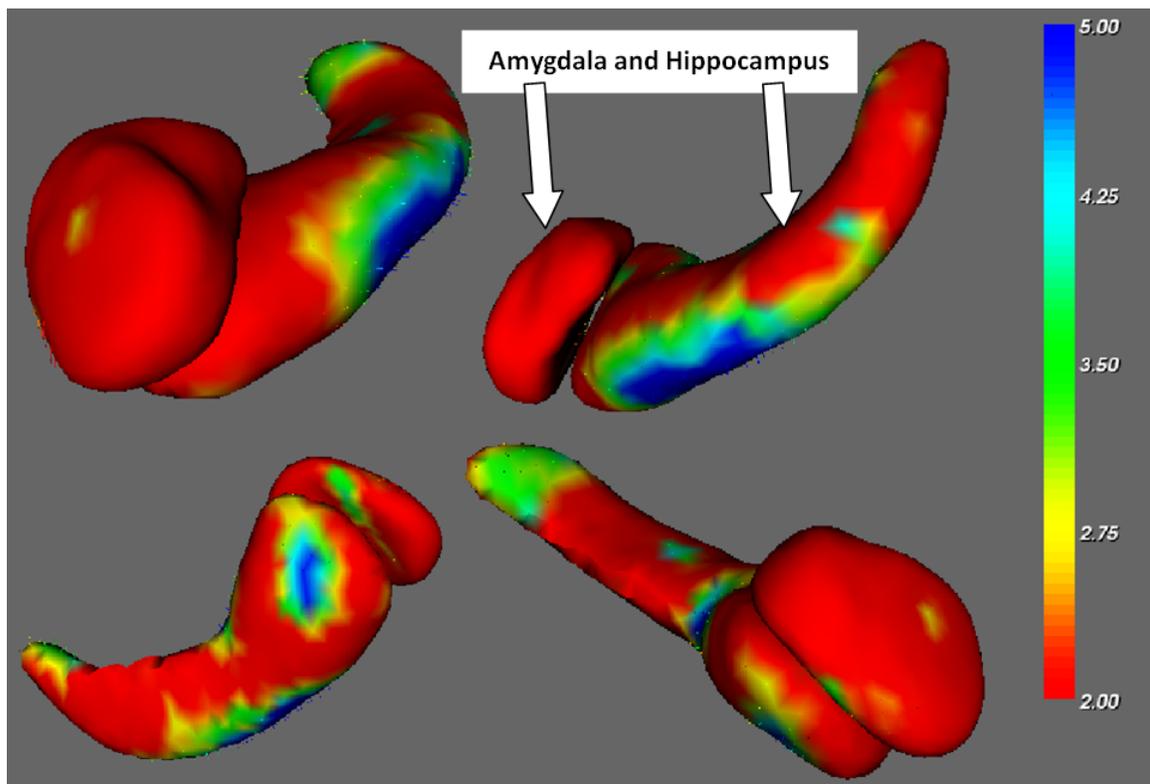


Figure 6.10: Results of amygdala and hippocampal surface analysis comparing volumes from malrotated and normal hippocampal groups

As with the case of the hippocampal volume analysis, it was important to perform a multiple comparison correction for the amygdala vertex analysis. The FDR correction was done and the results of the FDR correction are shown in figure 6.11 on the following page.

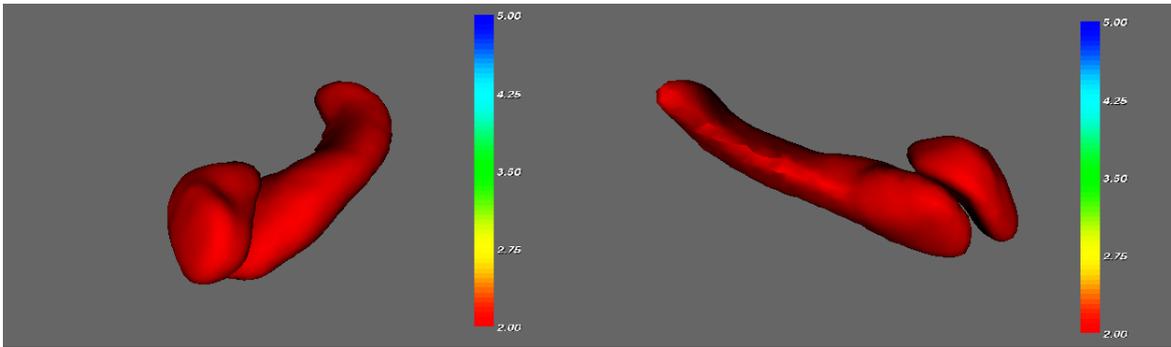


Figure 6.11: MCC corrected results of amygdala and hippocampal surface analysis comparing malrotated and normal left-sided amygdala and hippocampal volumes

From figure 6.11, following the FDR multiple comparison correction there were no statistically significant regions of difference on the left amygdala or left hippocampal volumes between the malrotated and normal groups.

6.9 Discussion of radiology outcomes

MRI imaging of the hippocampus provides sufficient contrast and resolution to begin to identify some of internal structure and subfields of the hippocampus [251]. Histology of the hippocampus [180] has shown five hippocampal layers and three dentate layers which are well correlated at 7 Tesla [346]. The internal blurring of the hippocampus was one criteria that has been used to assess IHI, however in our dataset, none of the subjects had blurring of the internal structure and this lack of perceived blurring may be due to the better resolution provided by the 3T MRI scanner. Alternatively, it could be argued that where blurring has been observed in patients with IHI that its presence might be related to temporal lobe epilepsy.

The association between hippocampal abnormalities and epilepsy is well documented in the literature. Hippocampal sclerosis is a frequent finding in patients with temporal lobe epilepsy [117]. A link between IHI and epilepsy has already been suggested in the literature, however the exact prevalence and the nature of a causal relationship, if any exists, is still unknown. Incomplete hippocampal Inversion (IHI) has also been well recognized in patients with temporal lobe epilepsy [279]. There are various terms used in the literature for hippocampal malrotation such as hippocampal malrotation (HIMAL) [32], incomplete hippocampal inversion [27] and round shaped hippocampus [63]. Thus there appears to be a need to try

and reach a consensus over the best terminology to use, the author of this thesis and his radiological colleagues are in favour of the continued use of the IHI notation as it refers to both the origins and current observation of this type of hippocampal formation.

The reported frequency of hippocampal malrotation varies in epileptic patients from 6% to 30% [32, 279, 26]. Baulac et al [34] showed 1.7% of partial epilepsy patients had hippocampal malrotation. Some studies have shown higher frequency of hippocampal malrotation in patients with developmental disorders [28, 340, 31]. Hippocampal abnormalities are found in cortical developmental malformation, however no specific pattern of hippocampal abnormality has been found. The frequency of hippocampal shape and rotation abnormality in cortical malformation has been estimated as 29.5% and was more frequently associated with periventricular heterotopia [243]. In recent published series, Barsi et al showed evidence of IHI in 6% (32/527) of patients with a suspicion of epilepsy and suggested that IHI was probably not the basic cause of epilepsy [32]. Contrary to this belief, Gamss et al studied prevalence of IHI in patients referred for conditions other than seizures and concluded that IHI was a rare finding in a population without seizures and that it was likely to be a pathological entity [131]. Similarly, a multicentre study of children with febrile status epilepticus (The FEBSTAT study), examined hippocampal abnormalities on MR imaging. They found all IHI on the left side in 23% (12/53) of cases and concluded that IHI was a pathological finding because it was more commonly observed in children with febrile seizures. They suggested follow up imaging to determine whether IHI predisposes the children to temporal lobe epilepsy or mesial temporal sclerosis [194].

Hippocampal malrotation has been described previously in a seizure free healthy population. Bronen and Cheugh studied 29 normal subject's hippocampi and showed that 20% (12/58) had oval shaped hippocampi [62]. Bajic et al described the term incomplete hippocampal inversion (a round or pyramidal shaped hippocampus and a vertical collateral sulcus) and found 19% of normal population without epilepsy or other development anomaly had incomplete hippocampal inversion [26]. Furthermore they found statistically significant differences in incomplete hippocampal inversion between the population with and without seizures (18% versus 30%, $p < 0.05$). The frequency of incomplete hippocampal inversion was higher in epileptic syndrome but they found no statistically significant difference between the patients with incomplete hippocampal inversion in temporal lobe epilepsy versus the control group.

In our study of normal healthy volunteers we found IHI in 37% of normal subjects. A similar criteria to that which has been used previously to assess IHI was adopted [32]. The main criteria on which IHI was identified was based upon observation of a round shaped hippocampus and vertical collateral sulcus.

6.10 Discussion of outcomes of the hippocampal volume and surface analysis

Here we have applied high resolution MRI and novel, quantitative image analysis methods to interrogate the volume and morphology of hippocampal malrotation in a normal population. The aim was to investigate hippocampal malrotation in a population of healthy, normal volunteers in greater detail than had been performed previously to gain as much information about the hippocampus and its surrounding environment as possible.

We set out to assess whether shape differences observed radiologically in malrotation would be reflected in the gross hippocampal volume differences and if there were volume differences, was it possible to localise these to particular subfields within the hippocampus. Having compared malrotated and normal hippocampal volumes, corrected for ICV, the malrotated hippocampal volumes were smaller than the normal volumes. This difference was statistically significant with a p-value of 0.006. We also compared the raw malrotated hippocampal volumes, all from the left hemisphere, with the right hippocampal volumes in the same subjects. While the malrotated volumes were smaller on average, the difference between the right and the left hippocampi was not statistically significant. This trend for smaller left hippocampi was in keeping with the broader literature on normal hippocampal volumes as was discussed in chapters 2 and 5 [232].

These results raise some interesting points, not least about choosing the appropriate methodology to assess hippocampal volumes in malrotation. While hippocampal malrotation remains a somewhat contentious issue in neuroradiology, there is perhaps broader agreement that regardless of its prevalence and the clinical implications of this anomalous feature that it is a developmental feature. Bajic et al investigated incomplete hippocampal inversion in non-epileptic populations and found 18% had IHI. The same group of researchers went on to assess IHI at different stages in the gestational period and found that prior to 24 weeks gestation, the hippocampus was not completely inverted in around 50% of cases but following 25 gestational weeks the incidence of IHI was comparable to that of an adult population. Thus, any failure of the hippocampus to completely invert was likely to be the result of a phenomenon occurring early in development [27, 25]. The issue with now discussing malrotation in a developmental context is whether when assessing malrotated and normal hippocampal volumes should one be corrected for ICV or not? One might speculate that if IHI has occurred as a result of some developmental 'insult' that perhaps this would also have an effect on the global intracranial volume. On the other hand, we have assessed ICV independently of volume and

observed that the ICV in the IHI group was larger on average than in the normal group, confounding the statistically significant ICV corrected volume difference observed between these groups.

When comparing hippocampal volumes between the malrotated left and normal right sided volumes there was a slight trend as seen in the literature for left hippocampal volumes to be smaller than right hippocampal volumes in healthy volunteers though this was not a statistically significant difference.

In further work it would be interesting to apply the quantitative methods discussed here to the hippocampi of epilepsy patients with IHI. As hippocampal sclerosis has been a widely reported feature in epilepsy it could be the case that IHI alone does not play a role in epilepsy, but IHI combined with a small hippocampus may well be important. That is, perhaps hippocampi need only be smaller to have a causative role in TLE and that IHI is simply a normal variant that on occasion is superimposed upon hippocampal sclerosis.

6.11 Discussion of outcomes of the amygdala volume and surface analysis

When comparing amygdala volumes between the malrotated left and normal right sided volumes one may wish to reflect on the slight trend seen in the literature for left hippocampal volumes to be smaller than right hippocampal volumes in healthy volunteers though this is typically not a difference which reaches statistical significance. No study to date to the best knowledge of the author has performed a quantitative assessment of amygdala volumes in cases of IHI. Here we found no difference between volume or morphology of the amygdala ipsilateral to IHI hippocampi when compared to amygdala volumes ipsilateral to normal hippocampi. This would imply that the developmental cause or causes of IHI do not extend to affect the development of the amygdala in gross anatomical terms.

6.12 Conclusion from radiology outcomes

IHI was observed in healthy individuals yet it is not clear whether IHI is a pathological entity or a normal developmental variant of no clinical significance though it does appear as though there is a broad spectrum of variations seen within normal hippocampal development. We conclude that IHI is likely to be a feature of normal healthy population and is much more common than has been previously reported [32, 279, 26]. IHI was overwhelmingly lateralised

to the left hippocampus, and this was in agreement with what was observed in the FEBSTAT study, though the data in the FEBSTAT study was in patients with febrile seizures [194]. Thus we have observed a relatively high incidence of IHI in a normal population and we therefore believe that the appearance of IHI is not likely to be a causal factor in patients with epilepsy.

6.13 Conclusions from hippocampal and amygdala volume segmentation

As per the optimised technique discussed in chapters 2 and 5, automated segmentations of the malrotated and normal hippocampi were augmented by manual refinements to provide the resultant hippocampal volumes. Malrotated volumes, being a predominantly left-sided phenomenon, were compared with normal left-sided hippocampal volumes. The left malrotated volumes were found to be smaller than normal hippocampal volumes ($p=0.006$) after correcting for differences in ICV. However, on closer inspection ICV differences were found to be significantly different, with ICV being approximately 5% larger on average in the IHI group than the normal group. This confounds the ICV corrected hippocampal volume results, and the reason that the ICV should be different between the groups remains a point for further investigation.

Malrotated left hippocampi volumes were compared to normal right hippocampal volumes. There was no statistically significant volume difference between these two groups ($p=0.19$).

Automatic segmentation of the amygdala was augmented by manual refinements to the resultant volumes. Volumes of the left amygdala, adjacent to malrotated hippocampi were compared to volumes of the left amygdala adjacent to normal hippocampi. There was no difference between the two groups of amygdala volumes after correcting for differences in ICV.

Similarly to the case for the hippocampus, left amygdala volumes adjacent to malrotated hippocampi were compared to right-sided amygdala volumes adjacent to normal hippocampi. There was no statistically significant difference between the two mean volume measurements ($p=0.08$).

To compliment the hippocampal volume assessment of IHI and normal subjects a novel surface analysis method was applied [278]. This allowed volumes differences to be compared at the level of the voxel. In a model which included age and ICV as covariates and following

an FDR correction there were no regions of statistically significant volume differences between the malrotated and normal left hippocampal volumes. This same process was applied to compare the amygdala volumes but again, no regions of statistically significant volume difference were found.

Presented here, for the first time, was a comprehensive and qualitative assessment of hippocampal and amygdala volumes in IHI, completed to complement the observations made by experienced neuroradiologists. At first there appeared to be evidence that ICV corrected IHI hippocampal volumes were smaller than normal hippocampal volumes but this was later confounded by group differences in ICV. Further work is necessary to reconcile the issue surrounding the ICV correction to hippocampal volumes, but overall, the hippocampal segmentation methodology developed in earlier chapters of this worked well within this context.

The work performed in this chapter is currently under discussion to decide if a single manuscript covering the radiological and quantitative methodological aspects of this work should be produced, or whether two separate manuscripts would better cover both of these topics. Once this decision has been made the manuscript(s) will be submitted to a peer reviewed journal.

Chapter 7

CA4/dentate hippocampal subfield volumes before and after treatment with an anti-inflammatory drug

7.1 Abstract

There is a growing body of evidence to suggest a link between inflammation, stress and depression. It has been posited that neurogenesis is a mechanism by which anti-depressants may work. It has also been reported that some anti-inflammatory drugs have been observed to have anti-depressant qualities. Here a pilot study of Rheumatoid arthritis (RA) patients were assessed before and after treatment with an anti-inflammatory drug. This small pilot study was used to establish neuroimaging techniques in non-depressed patients undergoing treatment with anti-TNF-alpha medication to test for evidence of the following primary hypothesis: that anti-TNF-alpha therapy would reduce circulating TNF-alpha and hence modify the activity of the serotonin transporter (SERT), resulting in a reduction of its binding potential as measured by SPECT. Since structural MRI scans of each patient were also being acquired to aid the SPECT image analysis this allowed a second, exploratory hypothesis to be investigated: that by reducing TNF-alpha the therapy would result in an increase in neurogenesis and that this would be reflected by an increased volume in the dentate gyrus of the hippocampus. This work has been published as an article in a peer reviewed journal [74].

By exploiting the image quality of 3T MRI the volume of the CA4/dentate subfield of the

hippocampus was assessed before and after treatment. Two methodologies were developed and implemented. In the first method a protocol was produced to manually segment the CA4/dentate region of the hippocampus from coronal T2-weighted FSE images. Given that few studies have assessed hippocampal subfields, an assessment of study power was conducted and based on the variances observed in the CA4/dentate volume measurements. Sample size estimations were made for given levels of study power to inform future work.

In the second method, the data were pre-processed with the N3 inhomogeneity algorithm before the DARTEL-VBM image processing pipeline was applied, this method was described previously in this thesis (chapter 3) however, due to the small number of subjects involved in this pilot study, statistical nonparametric mapping was applied in the final statistical interpretation of the VBM data. Following an FDR correction, a single GM voxel in the hippocampus was deemed to be statistically significant. This was suggestive of a small GM volume increase following anti-inflammatory treatment, however this result is severely limited by the small number of subjects involved in the study. Further work on a larger set of subjects and related studies, perhaps best performed in animals, to validate the relationship of neuroimaging markers such as volume, diffusion and perfusion to neurogenesis would lend greater confidence to results observed in in-vivo, human studies. Until such work is performed it would be difficult to state with confidence that volumetric differences observed in adult human are attributable to neurogenesis.

As stated above, any results suggestive of hippocampal volumetric differences will require further work on larger sample sizes to validate this initial tentative result. A further, similar study, building on the pilot work presented here has recently begun at the author's institution.

7.2 Introduction

Due to the inherent complexity of the hippocampus it is intuitive to try and investigate this structure by reducing it into its component parts. Until fairly recently it was not possible to even consider attempting to assess hippocampal subfields in-vivo. However, with the advent of ultra-high field MRI systems ($\geq 3T$), the improved visualisation of the hippocampus provides new opportunities to assess this most complex of structures.

Depression in the medically ill is five to fifteen times more prevalent than in the normal population [261]. Mood disorders can have a detrimental effect on the course of medical illness. Clinical associations between medical illness and major depressive disorder are not solely attributable to illness-induced physiological distress. Rather, there is a growing body

of evidence implicating biological mechanisms in a bi-directional link between mood disorders and many medical conditions [351, 206]. The focus of this work was to underpin the neurobiological relationship between the autoimmune disorder Rheumatoid arthritis (RA) and depression [160]. To achieve this a pilot study was performed to assess anatomical changes in the brains of RA patients using MRI. Measures related to dopamine and serotonin levels in the caudate and putamen were also determined from Single Photon Emission Computed Tomography (SPECT) imaging, but these were not considered as part of this thesis.

The subfield of the hippocampus of most interest to the field of mood disorder research is known as the dentate gyrus. The dentate surrounds the CA4 subfield at the centre of the hippocampus. The importance of this subfield zone is that neurogenesis, stimulated by anti-depressant and environmental interventions, has been demonstrated to occur here [209, 112, 307]. Therefore, initial measures on hippocampal subfield volumes will focus on the CA4 and dentate gyrus.

At the time of conducting the work for this thesis there were few studies which have aimed to assess hippocampal subfield volumes in-vivo from MRI data. Here subfield volumes were assessed not only by means of manual delineation but also by means of voxel based morphometry using the latest image processing techniques to increase the likelihood of detecting true volumetric differences, if they exist. The assessment of hippocampal subfield volumes in RA patients following anti-inflammatory treatment is also a novel aspect of this work. However, this chapter will focus on the methodological aspects of the work and not on the biological mechanisms underpinning the study.

7.3 Literature review

7.3.1 Review of subfield volume methodology

One of the first studies to assess, in-vivo, hippocampal subfields using MRI was the study by Mueller et al [257]. Mueller assessed forty-two healthy controls and 3 patients with Alzheimer's disease (AD). Imaging data was acquired using a coronal T2-weighted FSE sequence on a 4T MRI scanner. The entorhinal cortex (ERC), cornu ammonis 1 (CA1), CA2, CA3/4 and dentate gyrus were all outlined. Mueller et al's method was to manually outline these subfields but only in the hippocampal body i.e. subfields were not segmented in the hippocampal head or tail. Mueller et al used anatomical landmarks as a guide to define the boundaries between adjacent subfield zones. By this method Mueller et al found a significant correlation between age and the volume of CA1 in the normal controls. This was most

significant in the seventh decade of life. In AD patients (N=3) the CA1 and subiculum were smaller than in healthy-controls. While Mueller et al's study was satisfactory up to a point, there are clearly some limitations.

By only assessing the subfield zones within the hippocampal body, both the head and tail of the hippocampus are being ignored. One might argue that these areas are at least of equal or greater importance than the body of the hippocampus and therefore should also have been segmented. It is likely that, at least in the case of the hippocampal head, perhaps the subfield zones could not be reliably defined, and in the case of the hippocampal tail, perhaps the same could be said of defining its posterior boundary. With 24 x 2mm slices giving an anterior to posterior coverage of 48mm, this was sufficient coverage to obtain images from the hippocampal head to its tail. Another limitation with this study was the fact that using anatomical landmarks to define subfield cellular zones is inherently flawed. As the subfield boundaries themselves cannot be seen on MRI there is little, at present, that can be done to solve this problem, thus the approach of using landmarks, while flawed, is a reasonable approximation to start assessing subfields. By this method the aim was to build a protocol which although it might not be accurate, could be precise and reproducible, thus allowing comparisons to be made within and between subject groups. This was a further refinement to assessing whole hippocampal volumes, a topic which itself is still struggling for consensus and reliable measurements and therefore at this relatively early stage in subfield segmentation attempts to reach consensus should be welcomed [185]. It was hoped that the work undertaken in this thesis will contribute to the small but growing body of literature around hippocampal subfield measurements and segmentation methods.

Building on their earlier work, Mueller et al performed further work on more normal subjects and AD patients [254]. Later they also assessed the subfield volumes of TLE patients [252]. In the first of these later works Mueller et al assessed the relationship between the major lipoprotein transporter Apolipoprotein E (Apo E) and subfield volume in patients with AD. They found that AD patients with ApoE had smaller CA3/dentate than AD patients without ApoE. They argue that this implies that lipoprotein has a regionally selective effect within the hippocampus [254]. The second of these works assessing patients with TLE found ipsilateral atrophy in the CA1, in CA3/dentate zones in patients with mesial temporal sclerosis TLE. In 17% of cases no gross hippocampal atrophy was observed (i.e. normal hippocampal volume) yet subfield atrophy was detected implying that the subfield measurements can provide additional information over and above gross hippocampal volume measurements [252]. An extension of this method was implemented on T2-weighted FSE data acquired coronally using a 7T MRI scanner (Chupin et al, Proc ISMRM 17 2009, 49). In this case the subfield

areas of the entire hippocampal formation were manually segmented. This included differentiation between white matter regions such as the alveus and fimbria as well as the larger grey matter CA1-4 subfields. This has been the most detailed subfield description of the hippocampus to-date and crucially includes the head and tail of the hippocampus so as to assess the subfields throughout its entire length.

An alternative approach to Mueller et al in trying to further refine the study of hippocampal volumes using MRI was to segment the hippocampus into its head, body and tail as was done by Malykhin et al [212]. They used 1.5T T1-weighted 3D data which while giving superior resolution in three planes to T2-weighting imaging, the contrast between subfield regions was poorer thus making internal differentiation of the hippocampal subfields difficult. However, larger segmented regions such as the head, body and tail were segmented using anatomical landmarks and subsequently assessed. Malykhin et al presented normal ranges for his volumes for the amygdala, and hippocampal head, body and tail and showed that these volumes could be reliably reproduced. This was again an improvement and refinement from the many studies which have assessed whole hippocampal volumes, but it was still a step away from being able to quantify the volumes of the subfield regions themselves as a result of the poorer contrast of the internal structure of the hippocampus on the T1-weighted imaging.

A further approach to localise regions of volume difference within hippocampus, which was similar in some respects to that which was performed by Malykhin et al, was to perform surface analysis on the whole hippocampal volume. This method was implemented by Wang et al to study the hippocampal surface in patients with very mild dementia of the Alzheimer's type [362]. This built upon earlier work by the same group who had previously defined subfield zones for the CA1, subiculum, combined CA2-CA4 and dentate regions [97]. Thus in both these works a hippocampal template was created. In the study by Wang et al, the hippocampus was manually outlined and segregated into further subfields CA1, CA2, CA3, CA4 and the gyrus dentatus. Thus the various zones on the surface of the hippocampal template were defined. The individual hippocampal volumes for each of the study patients were then mapped to the template in two steps. First, the template scans and target images were coarsely aligned using landmarks previously defined on the template and on all the individuals scans. Second, the precise mapping of the template image to each of the target images was defined by a high-dimensional diffeomorphic (i.e. invertible, continuous and differentiable) transformation. Movements of the voxels in the template MR image during these transformations were constrained by assigning the matrix of voxels with the physical properties of a viscous fluid. Thus the regions of volume difference between the subject

groups could be visualized on a surface rendering of the hippocampus which has been further differentiated into various subfield regions. By this method Wang et al found that when comparing very mild DAT (Dementia of the Alzheimer's type) with non-demented subjects, the mild DAT patients had inward deformities of the hippocampal surface in proximity to the CA1 subfield and subiculum [362].

The major benefit of the methodology of Wang et al was in producing a template of hippocampal subfield regions which subsequent images could be warped to [362]. In practical terms this would be likely to reduce the analysis time as it was not necessary to segment each individual hippocampus into the various subfields. Perhaps in argument against a method like this would be the fact that the result is dependent upon the success and reliability of various image normalisation steps, though measures could be put in place to assess the quality of these methods. Moreover, this method is certainly more complex to implement and an appropriate visualisation scheme would also need to be developed. Finally, methods based on normalisation to a template typically, while being able to highlight local regions of volume difference, cannot provide absolute volume differences. Depending on the underlying question a researcher or clinician hopes to answer this may or may not be important and perhaps in most cases identifying regions of relative change would be sufficient.

Thus there are various methods which can be applied to examine the subfields of the hippocampus, all of which have their own fundamental limitations and benefits. For the purposes of this chapter where we aim to define normal volumes of particular subfield regions an approach more similar to that of Mueller et al and Chupin et al will be implemented. A surface based method for localising regions of hippocampal volume difference has been described elsewhere in this thesis (chapter 6) [278].

7.3.2 Review of RA, depression and their relevance to the hippocampus

RA is the most common inflammatory joint disease in the UK being a major cause of disability, morbidity and mortality. The worldwide prevalence of RA is 0.7-1.0% of the adult population and increases with age. Depressive symptoms are significantly more common in those with RA than in the general population, being associated with higher pain, higher fatigue and reduced quality of life metrics. Conservative estimates indicate that Major Depressive Disorder (MDD) affects between 13% and 17% of patients with RA. However, prevalence exceeding 40% has been reported with 11% of hospital RA out-patients experiencing suicidal

ideation. MDD is an independent risk factor for both work disability and mortality in those with RA [147].

The etiology of MDD in the presence of a medical co-morbidity is unknown, although a variety of hypotheses have been proposed. Of direct relevance to the link between RA and depression is the proposal that pro-inflammatory cytokines, released during tissue damage in the body, provoke changes in brain structure and function that lead to the development of co-morbid MDD [294]. The imbalance between pro- and anti-inflammatory cytokine activity is central to the pathogenesis of RA [231].

The mechanisms whereby cytokines induce the symptoms of depression are unknown. Cytokines can directly modulate pathways implicated in the etiology and treatment of depression, including the hypothalamus-pituitary-adrenal axis, monoaminergic neurotransmitter systems and hippocampal neurogenesis [161].

Whereas the former has received some prior attention, the latter two have not been explored directly in human disease despite persuasive pre-clinical data. It has been shown that both the density and the activity of the serotonin (5-HT) transporter (SERT) are increased by proinflammatory cytokines, leading to an increase in 5HT uptake from the synapse, thus decreasing 5HT transmission [249][379]. SERT itself is, of course, a key target site for selective serotonin re-uptake inhibitor antidepressants (SSRI's).

In addition to action at SERT, preclinical studies also suggest that cytokines can impair neurogenesis in the brain [352, 242, 36]. While a pathogenic role for reduced hippocampal neurogenesis in depression is unclear, there is now considerable experimental evidence that the generation of new neurons in the dentate gyrus of the hippocampus is enhanced by antidepressant treatment [75]. Changes in volume within the hippocampus, as measured by MRI, are associated with depression and the use of anti-depressants [72, 358, 100].

Tumour necrosis factor alpha (TNF-alpha) is a pro-inflammatory cytokine produced by macrophages that increases the immune response and amplifies inflammation. The anti-inflammatory drug used in this study was a soluble fusion protein that mimics the inhibitor effects of naturally occurring soluble TNF-alpha receptors but with a much extended half-life in blunting the immune response. This small pilot study was used to establish neuroimaging techniques in non-depressed patients undergoing treatment with anti-TNF-alpha medication (adalimumab) to test for evidence of the following primary hypothesis: that anti-TNF-alpha therapy would reduce circulating TNF-alpha and hence modify the activity of the SERT, resulting in a reduction of its binding potential as measured by SPECT. Since structural MRI

scans of each patient were being acquired to aid the SPECT image analysis this allowed a second, exploratory hypothesis to be investigated: that by reducing TNF-alpha the therapy would result in an increase in neurogenesis and that this would be reflected by an increased volume in the dentate gyrus of the hippocampus. This work has been published as an article in a peer reviewed journal [74].

Whereas in chapters 2, 3 and 4, socioeconomic status, and in particular deprivation was considered to be the inflammatory agent, leading to stress, in this chapter pathological RA was considered as the inflammatory agent, but the goal of developing the methodology to assess volumetric changes within the hippocampus remains the same.

7.4 Study aim

7.4.1 Primary aim

The primary aim of this chapter of the thesis as to determine if manual segmentation of the CA4/dentate hippocampal subfield could be reliably measured using data acquired from a 3T MRI system?

7.4.2 Secondary aims

Following implementation of the manual delineation protocol for the CA4/dentate for this pilot study, study power and sample size calculations will be performed to inform future work.

7.4.3 Tertiary aim

By using DARTEL-VBM and statistical non-parametric mapping we aimed to assess whether it would be possible to detect an increase in focal GM volumes between the two groups, pre and post anti-inflammatory treatment?

7.4.4 Ethics

This study was approved by the local Research Ethics Committee and by the Administration of Radioactive Substances Advisory Committee (ARSAC).

7.4.5 Subjects

Six patients with seropositive RA who met the eligibility criteria of the British Society for Rheumatology Guidelines for treatment for anti-TNF-alpha treatment were recruited from the Centre for Rheumatic Diseases at Glasgow Royal Infirmary. Inclusion criteria included: seropositive RA meeting the American College of Rheumatology (ACR) diagnostic criteria. Patients were already scheduled to receive an anti-inflammatory drug indicated by active RA and meeting the British Society for Rheumatology Guidelines for treatment.

7.4.6 Exclusion criteria

The exclusion criteria included: Non-rheumatoid arthritis; contraindications to anti-inflammatory drug; patients on antidepressant medication; cerebrovascular risk factors or a previous stroke; documented head trauma or neurodegenerative disorders; other Diagnostic and Statistical manual of mental disorders (DCM) Axis 1 psychiatric diagnoses; psychotic symptoms; alcohol and/or substance misuse; currently taking any other psychotropic medication; thyroid disorder and contraindications for MRI imaging.

7.4.7 Image acquisition protocol

MR images were acquired using the body transmit coil and an 8-channel head receiver coil on a 3 Tesla GE Signa scanner (GE Medical Systems, Milwaukee, Wisconsin, USA). In order to estimate volumetric change in the CA4/dentate gyrus of the hippocampus high resolution T2-weighted FSE coronal slices of the hippocampus were acquired. These were planned perpendicular to the long axis of the hippocampus, see figure 7.1 on the following page. The image parameters for the acquisition were as follows: TR = 4075ms, TE = 115ms, FOV = 200mm x 200mm, matrix = 512 x 512, bandwidth = 31.25kHz, echo train length = 24. The in-plane pixel size was 0.39mm x 0.39mm, slice thickness was 4mm with a slice gap of 1mm. To guide the SPECT region of interest (ROI) analysis and for MRI DARTEL-VBM analysis a 3D T1-weighted IR-FSPGR dataset was acquired. The image acquisition parameters for this acquisition were: TR = 6.8ms, TE = 1.4ms, TI = 500ms, flip angle of 12°, FOV = 260mm, matrix = 320 x 224, PFOV = 70%, bandwidth = 31.25. Thus, the voxel size was 0.81mm x 1.16mm x 1.00mm. MRI was performed up to 14 days before the commencement of anti-inflammatory treatment and within 4 days after the last dose of the treatment regime.

The image acquisition protocol began with a basic three plane localiser scan followed by the 3D T1-weighted IR-FSPGR sequence. The sagittal T1-weighted FLAIR and T2-weighted

FSE PROPELLER sequences were essentially enhanced localiser scans to provide good visualisation of the hippocampus in their respective orientations. The sagittal T1-weighted FLAIR was a short scan of 1minute 32seconds and had the following imaging parameters: TR = 2500ms, TE = 8.9ms, TI = 920ms, FOV = 240mm x 240mm, matrix = 320 x 224, bandwidth = 31.25kHz, slice thickness = 5mm and a slice gap of 2mm. The T2-weighted FSE PROPELLER sequence was particularly useful as it provided high quality images resistant to the effects of patient motion [285]. The T2-weighted FSE PROPELLER has the following imaging parameters: TR = 5000ms, TE = 109ms, FOV = 240mm x 240mm, matrix 512 x 512, bandwidth = 50kHz, slice thickness 4mm and no slice gap. The T2-weighted FSE sequences had high in-plane resolution with excellent grey and white matter contrast. It was these acquisitions that were later used to produce the subfield volume measures.

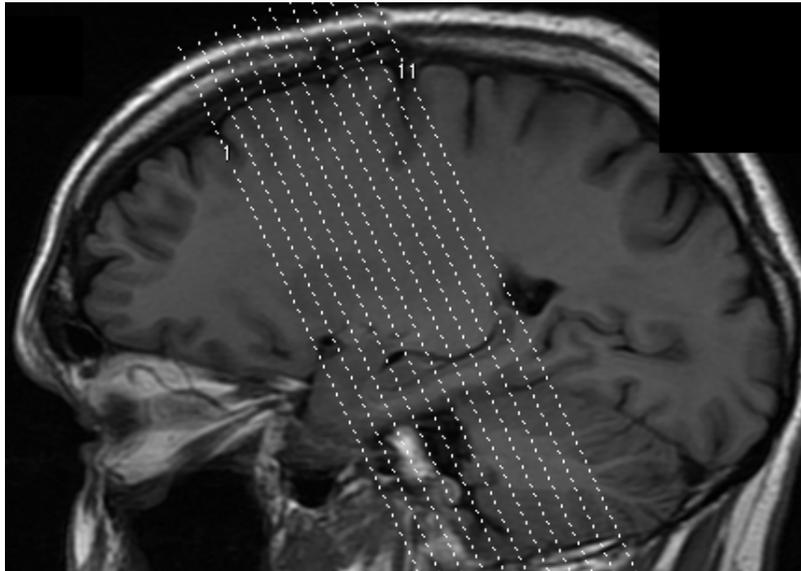


Figure 7.1: Planning coronal slices along the long axis of the hippocampus

7.4.8 Analysis protocol for the outlining of the dentate/CA4 hippocampal subfield

7.4.8.1 Manual outlining of subfields of the hippocampus

Manual outlining of the dentate gyrus was performed using a method detailed by Mueller et al (2006) by an operator (JM) blinded to the identity of the patients and the order of the scans.

The dentate/CA4 region was outlined on the first coronal slice beyond the hippocampal head and then on subsequent posterior slices. Internal landmarks based on other hippocampal subfields such as CA1 and CA2 also guided the manual outlining procedure. The manual outlining was performed using the Medical Image Processing, Analysis and Visualisation tool MIPAV v4.0.0 (<http://mipav.cit.nih.gov/>).

Figure 7.2 below illustrates two independent attempts by the same operator of manual outlining the dentate/CA4 region within the hippocampal body.

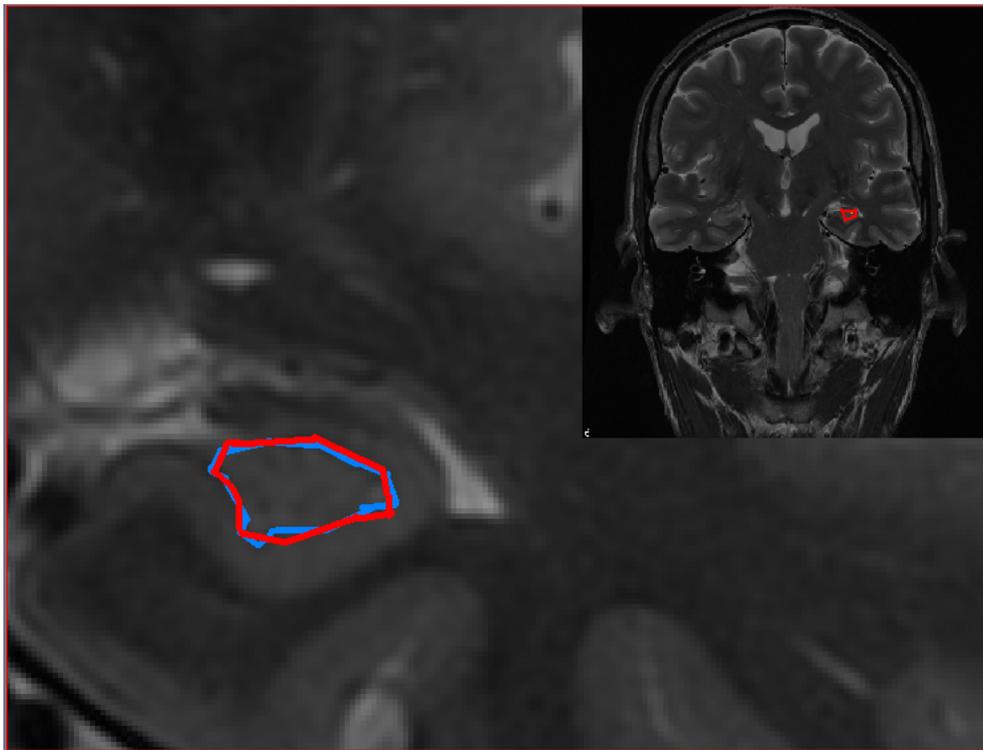


Figure 7.2: Two, independent manual delineations of the dentate/CA4 region of the hippocampal body

Figure 7.2 shows, in the top, right corner of the image, the relative position and size of the dentate with respect to the whole brain. The larger part of the image shows the zoomed-in image of the dentate/CA4. The two independent manual delineations of the dentate are shown in red and blue and broadly overlap.

7.4.9 Analysis of data by DARTEL-VBM

DARTEL-VBM analysis is performed using the 3D T1-weighted IR-FSPGR data. The DARTEL-VBM procedure here has been extensively described elsewhere in this thesis (see chapter 3). To briefly summarise the process. The T1-weighted 3D image sets were roughly aligned to each other and were preprocessed using the N3 bias correction algorithm. Initial normalization parameters were then established and the data was segmented into either GM, WM or CSF tissue types. Following this, GM and WM study specific templates were created, the original segmentations were then warped to the template and the subsequent modulated images were then smoothed prior to statistical analysis being applied.

7.4.10 Statistical design and analysis

While the processing steps applied here are in broad terms the same as that which was applied previously in this thesis (Chapter 3), there are a number of differences as a result of the study numbers and design. Previously there were two distinct groups of 21 subjects, here the data from 5 subjects was acquired longitudinally. Thus, a paired t-test design was set-up. As the data are paired, independence of the data cannot be assumed. Given the small numbers, as well as applying the default SPM statistical analysis to the data, more appropriate nonparametric statistics were also computed, this was done using the nonparametric mapping (NPM) tool developed by Chris Rorden [301]. Another tool for performing statistical nonparametric mapping is the SNPM plugin for SPM [265].

The nonparametric approach used in NPM, SNPM and also within FSL for VBM analysis use permutation based methods. Nonparametric permutation methods are a means of addressing the problem of multiple comparisons while requiring only a minimal set of assumptions. Historically, the major drawback to permutation based methods was the increased computational demands, and therefore processing time was increased. Holmes et al developed a nonparametric alternative to the random fields approach using permutation test theory [163]. This approach was simple, relies on minimal assumptions, deals with multiple comparisons, and can be applied in circumstances where the parametric assumptions are violated. In some circumstances the nonparametric permutation approach may even outperform parametric approach [17, 144, 152, 164].

The crux of permutation theory is as follows. If we consider a single subject performing an fMRI experiment, this single subject is scanned in an 'active' / 'rest' block design. Considering a single voxel, if there is truly no difference between the active and rest states then we would

be considerably surprised if the signal during the ‘active’ state exceeded that of the ‘rest’ state and thus we would be inclined to believe that there was genuine evidence for activation at that voxel. Permutation tests simply quantify the element of “surprise” in terms of probability, allowing significance tests with p-values to be obtained. Thus, in situations where there is no experimental effect, i.e. no correlation between the task and the measured signal, then the labelling of the experimental condition is arbitrary. That is, for instance, the group labels can be swapped and the same result would be obtained. This applies for not just group affiliations but also for covariates of interest. Given the null hypothesis that the labels are arbitrary, the significance of a statistic expressing the experimental effect can then be assessed by comparison with the distribution of values obtained when the labels are permuted. The justification for exchanging the labels comes from weak distributional assumptions, or by appeal to the randomisation scheme used in designing the experiment. Tests based on the idea that the initial conditions are randomized are known as randomization tests or re-randomization tests.

This approach is in contrast to the parametric GLM based approach using random field theory described in chapter 3.

7.5 Results

7.5.1 Data quality: Intra-rater variability

To assess the intraoperator reliability of measuring the CA4/dentate subfield by the above protocol, two repeat measures were made on both the right and left CA4/dentate of hippocampi from 3 subjects before and after anti-inflammatory treatment (12 CA4/dentate volume measures in total). The aim here was to assess the variability in the measurement, this enabled power calculations to be done to inform further work.

The mean standard deviation for the repeat measurements was 55mm^3 , the standard error of the mean was 23mm^3 , the 95% confidence interval was 57mm^3 for the left CA4/dentate and similarly the mean standard deviation for the repeat measurements was 56mm^3 , the standard error of the mean was 23mm^3 , the 95% confidence interval was 57mm^3 for the right CA4/dentate. These measures of the variability are summarised in the table 7.1 on the following page.

The ICC (2,1) i.e. the ICC (agreement) between repeat measurements was 0.81 for the right

Subfield	Mean Volume	Std Dev	n	SE	CoV	95% CI
L CA4/Dentate	355	55	6	23	7	57
R CA4/Dentate	321	56	6	23	7	57

Table 7.1: Summary Statistics for repeat measures of right and left CA4/ dentate

CA4/dentate and 0.81 for the left CA4/dentate subfield as calculated by a two-way random effects ANOVA for absolute agreement using SPSS v18.

7.5.2 Data quality: power calculations on identifying CA4/dentate subfield volume differences

As we expect that further volume measurements on hippocampal subfields are likely to be performed in future it was worthwhile assessing what level of study power might be achievable.

If we take the mean volume of the CA4/dentate to be $\sim 350\text{mm}^3$, and we somewhat speculatively say that a significant change in volume would be a 10% change, i.e. 35mm^3 , combined with the standard deviation in the measurements from the pilot data being 55mm^3 , the following power estimations were made for various projected subject numbers. The results of the power calculation are shown in figure 7.3 below.

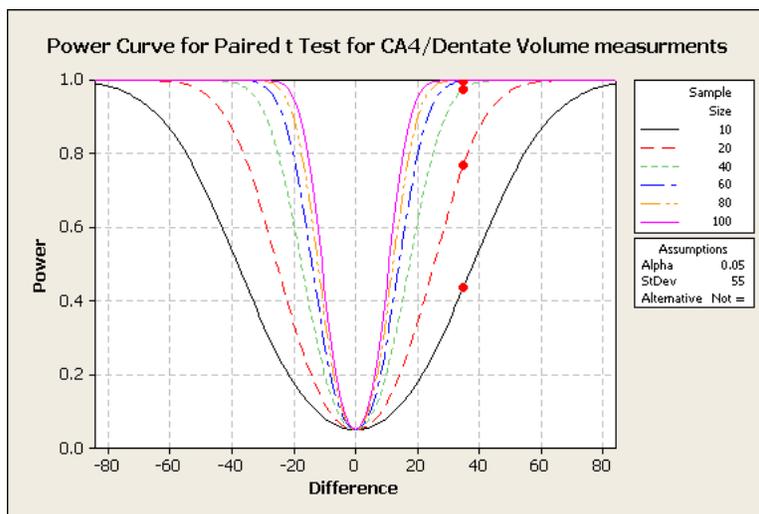


Figure 7.3: Power calculation estimation for Paired-t-test CA4/dentate subfield volume measurements

Thus to obtain a power in the design of approximately 0.90 in identifying a 10% volume difference in the CA4/dentate approximately 30 subjects would be required.

Broadening the context of the power calculation, comparing two groups rather than a paired design, the resultant power estimations are shown in figure 7.4 below.

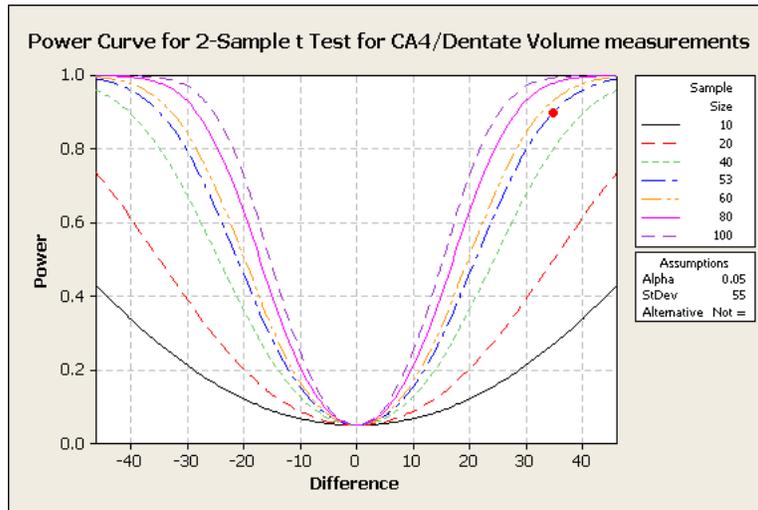


Figure 7.4: Power calculation estimation for 2 sample t-test design of CA4/dentate subfield volume measurements

Thus, for this statistical design, to achieve a power of 0.90 in detecting a volume difference of 10% (35mm^3) 53 subjects in each group would be needed. Thus, as the design was not a paired design, more subjects are required to achieve a similar level of power.

The power calculations performed here were set-up to assess the case where volumes were unequal i.e. where it was uncertain as to whether one will observe a volume increase or a volume decrease. For the case where a more specific hypothesis is generated such that testing is being done for a 10% volume decrease, the estimated number of subjects to obtain a power of 0.90 was 30 for the paired t-test design and 43 for the two sample unpaired t-test design. That is, when it is known that either a volume decrease or a volume increase will be found, fewer subjects may be required to achieve equivalent levels of study power as when the direction of volume change is not known.

7.5.3 Establishing the mean and variance of dentate/CA4 volumes before and after a course of anti-inflammatory drug administration

Of the six patients in this pilot study, five had MR imaging. One patient was unable to be scanned due to claustrophobia. Repeat manual segmentations determined the mean volume for the CA4/dentate and 95% confidence interval for the mean. There was no significant group difference in the CA4/dentate volumes between the pre and post anti-inflammatory treatment using this analysis method. The mean CA4/dentate volume from the pre and post treatment with the 95% confidence intervals are shown in figure 7.5 below.

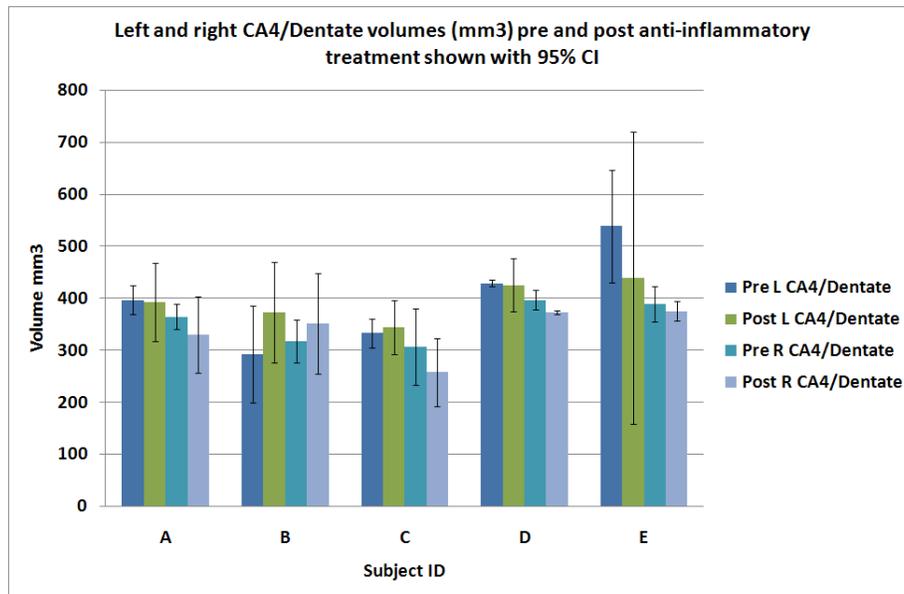


Figure 7.5: Results of CA4/dentate volumes (mm³) before and after anti-inflammatory treatment

The study was set-up in a paired design where measurements were made on the same subjects before and after therapy, where the therapeutic agent in this case was an anti-inflammatory medicine. As a volume of a region of the brain was measured, this measurement may be confounded by additional factors, namely subject age and ICV. That is, in the broadest of terms, it may be expected that older subjects will have correspondingly smaller brain structures and similarly, subjects with larger skulls will have correspondingly larger volumes of brain tissue. These issues are discussed in greater depth elsewhere in the thesis (chapters

2 and 5) but are mentioned here simply to highlight the fact that a paired analysis was performed where the measurements were made within the space of approximately 10 weeks and therefore age effects on the measurements were negligible. Similarly, each subject's ICV measurement did not change throughout the duration of the study and thus as with age this will not influence the subfield measurements made prior to and following drug treatment.

The data corresponding to the figure 7.5 above are shown below in table 7.2.

Subject	Pre L		Post L		Pre R		Post R	
	CA4/Dentate	CI	CA4/Dentate	CI	CA4/Dentate	CI	CA4/Dentate	CI
A	397	28	393	75	365	24	331	73
B	293	93	373	96	317	41	351	97
C	333	28	344	52	307	74	258	65
D	429	6	426	51	397	19	373	4
E	539	109	440	281	389	34	376	19

Table 7.2: Left and right mean CA4/dentate volumes pre and post a course of anti-inflammatory drug. 95% Confidence Intervals (CI) are also shown.

The mean pre and post treatment, left CA4/dentate subfield volumes were 398mm³ and 395mm³ respectively. The mean pre and post treatment, right CA4/dentate subfield volumes were 355mm³ and 338mm³ respectively.

7.5.4 Volumes with age

As stated above, as we have a paired design where images were acquired in a relatively small space of time, the age of the subjects was not a major concern. However, if one were to assess an unpaired group of subjects, for example, assessing hippocampal volumes in two groups being treated with different anti-inflammatory drugs, one would need to take steps to ensure the subjects were age matched. Even if such steps were taken to ensure subjects were age matched it is still worth understanding how the hippocampal subfield volumes are affected by age.

In this study we have several subjects ranging in age from 44 to 67 years old (529 to 809 months old). For interest, pre-treatment, ICV corrected, CA4/dentate volumes were plotted against age, see figure 7.6 on the following page.

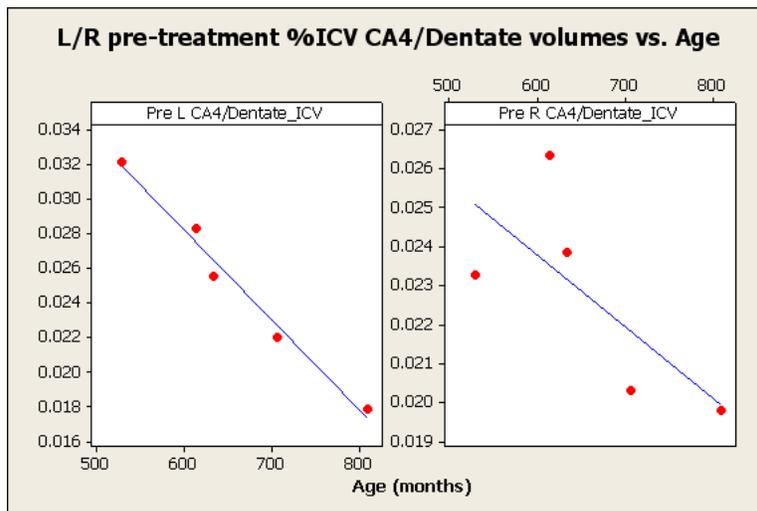


Figure 7.6: Left and right, pre-treatment, ICV corrected CA4/dentate volumes vs. Age (months)

The Pearson correlation value for ICV corrected CA4/dentate volume against age was -0.99 at a statistical significance of $p=0.001$ for the left hippocampus and a Pearson correlation value of -0.72 at a statistical significance of $p=0.078$ for the right hippocampus.

7.5.5 Results of DARTEL-VBM to assess hippocampal volume differences

Having compared manual segmentation volumes of CA4/dentate pre and post anti-inflammatory treatment, GM volume differences across the whole brain were then compared at the level of the voxel using the DARTEL-VBM method. The T1-weighted 3D images were used for this analysis, the DARTEL-VBM processing pipeline has been described in more detail in chapter 3.

The paired t-test design was set-up within SPM and is shown in figure 7.7 on the following page.

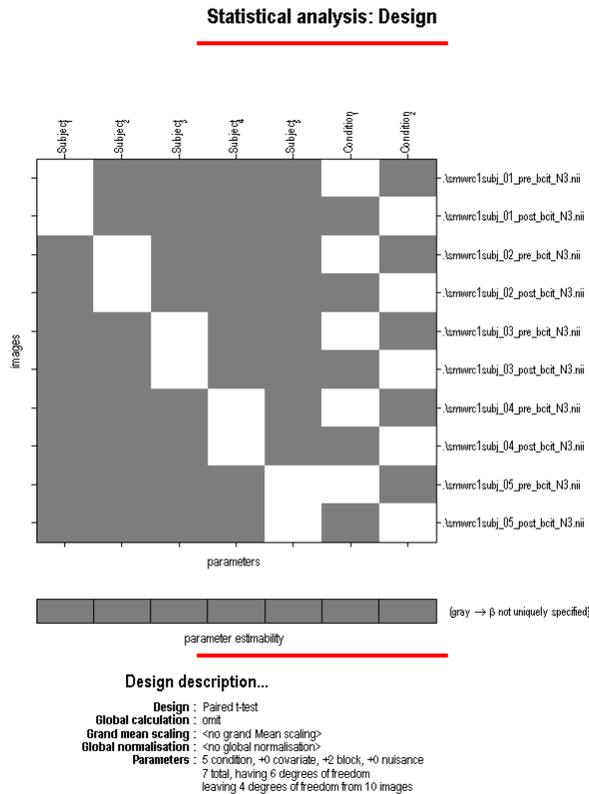


Figure 7.7: SPM paired t-test design to compare pre and post anti-inflammatory treatment brain by VBM

The results of the F-test to compare the groups was initially performed to assess for any statistical differences in GM volume. The results of the F-test are shown in figure 7.8 on the following page.

Note that the statistics shown here were uncorrected for multiple comparisons with an F-value of 0.001. As the aim here was to assess whether or not there was an increased GM volume in the hippocampus following application of the anti-inflammatory drug, then a t-test was performed to identify regions of increased volume in the later scans. The results of the t-test are shown in figure 7.9 on the following pages.

As shown in figure 7.9, there was a statistically significant region of increased GM volume

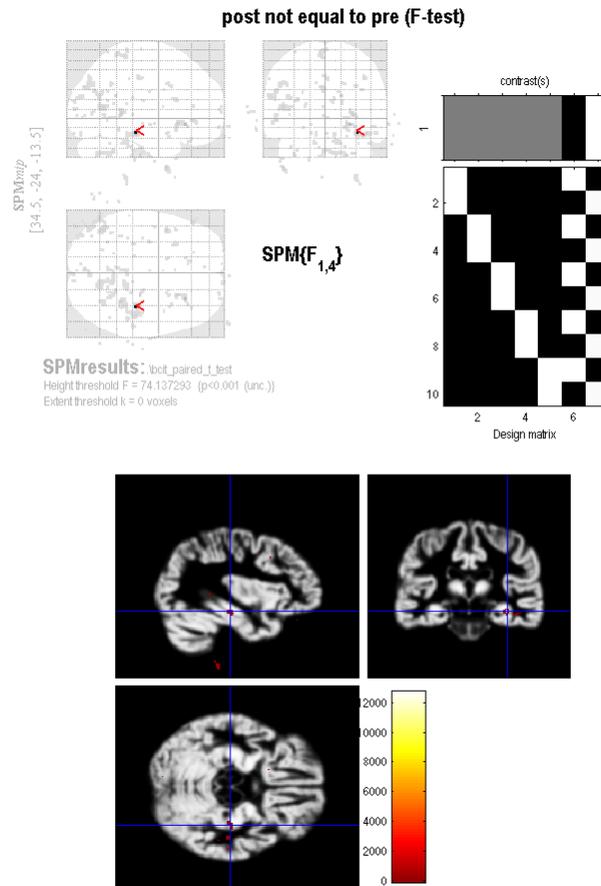


Figure 7.8: SPM F-test results comparing pre and post treatment groups

within the hippocampus. Note that these results have not been corrected for multiple comparisons and are thresholded with a p-value of 0.001. Running a cluster corrected statistical approach centered on the cluster in the hippocampus, the FDR p-value for the hippocampus was 0.055 i.e. just below a level that would be deemed statistically significant.

The same paired t-test model was also created in the nonparametric statistics toolbox (NPM). The results of the post vs. pre treatment data are shown in figure 7.10 on the following page.

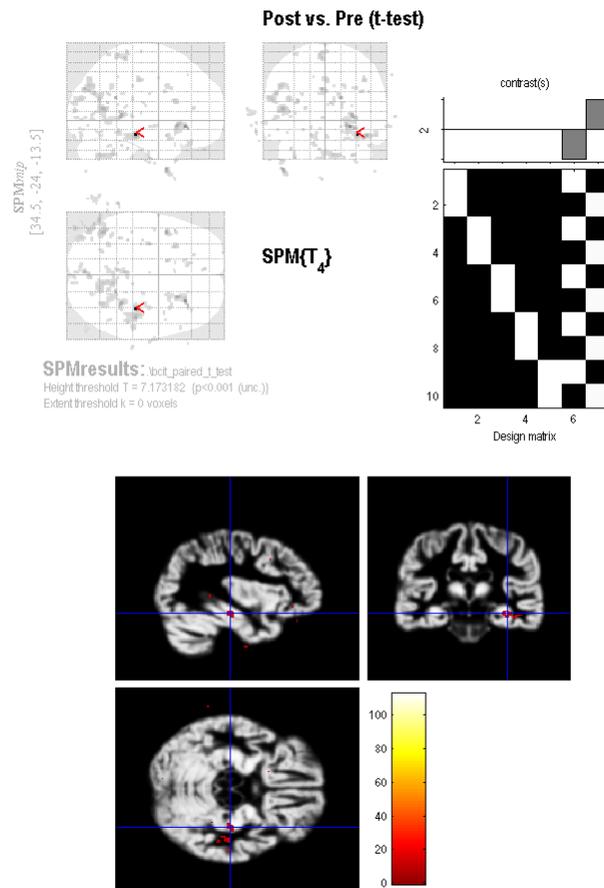


Figure 7.9: SPM Paired t-test results, post treatment vs. pre treatment

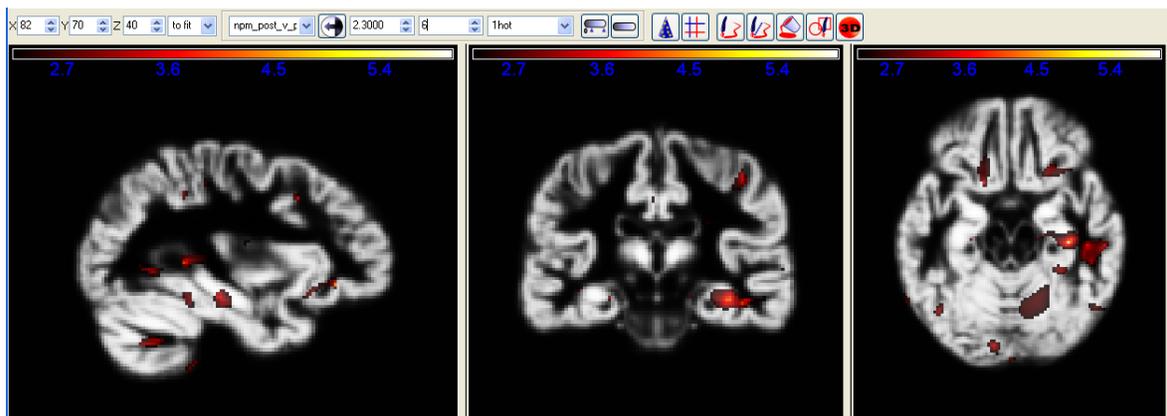


Figure 7.10: NPM paired t-test (uncorrected) results, post treatment vs. pre treatment

The results in 7.10 are the uncorrected statistics from the NPM tool which mirror that of the SPM results shown in figure 7.9. The NPM FDR corrected statistics are shown in figure

7.11 below.

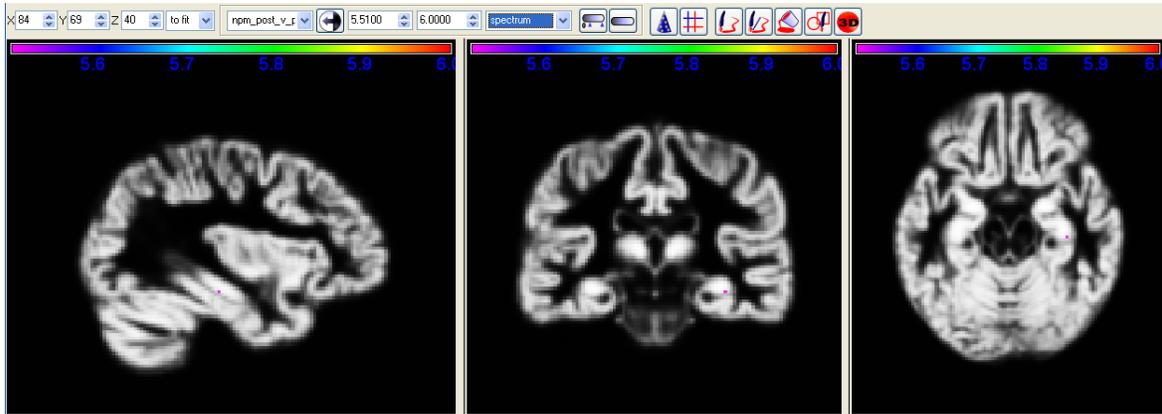


Figure 7.11: NPM FDR corrected paired t-test results, post treatment vs. per treatment

The NPM results in figure 7.11 are the FDR corrected statistics. The only voxel to survive the FDR 0.05 Z-threshold (5.510), was a single voxel in the hippocampus, more precisely around the CA1 region of the hippocampal body. Thus, while tentative given the small numbers of subjects involved in this pilot study, there results are suggestive of a GM volume increase in the CA1 subfield of the hippocampus following anti-inflammatory treatment.

7.6 Discussion

7.6.1 Intra-operator variability

Repeat manual outlining measures of the CA4/dentate indicate that the intraoperator correlation coefficient were 0.81 for both the right and left subfield regions. The 95% CI was 57mm^3 for both right and left subfield regions based around mean subfield volumes of 355mm^3 and 321mm^3 respectively. These values are in keeping with values from the literature on similar regions of the brain [257].

To assess the intraoperator reliability of measuring the CA4/dentate subfield by the above protocol, two repeat measures were made on both the right and left CA4/dentate of hippocampi from 3 subjects before and after anti-inflammatory treatment (12 CA4/dentate volume measures in total). The Intraclass correlation coefficient for consistency intra-operator measurements was determined. The average intraclass correlation coefficients for consistency were 0.88 and 0.87 for the left and right CA4/dentate region respectively. As the method

applied here was most similar to that which was applied by Mueller et al, the ICC values from this study for operator 1 were 0.98 and 0.99 for the left and right CA4/CA3/dentate subfield respectively and for rater 2 were 0.86 and 0.96 for the left and right CA4/CA3/dentate subfield respectively. Thus, while the ICC values obtained in the work done in this thesis were slightly lower than achieved by Mueller [251], they are still of a reasonable standard where typically ICC values above 0.8 are typically deemed acceptable. The lower values obtained by the second rater in the Mueller et al study were to be expected given that the second rater is likely to be less familiar with the segmentation protocol. The implication for further work was to ensure operators performing manual outlining are suitably trained and that the performance of each operator was assessed as well as being compared.

7.6.2 Data quality: Power calculations on CA4/dentate subfield volumes

At the time of performing this thesis there was very little information in the literature regarding power calculations for hippocampal subfield volumes. Based on the manual segmentation protocol of the CA4/dentate region this issue was addressed. Based on projected volume differences of 10% between groups (unpaired design) and changes within groups (paired design), given the variance calculated from the data acquired in this study, the estimated number of subjects required in each group to achieve a study power of 0.90 was 43 and 30 respectively. In the context of medical literature such numbers are modest. Imaging studies generally have fewer numbers, dictated perhaps by expense rather than desired power. It could be argued that the expectation of achieving a 10% volume difference is perhaps too optimistic and clearly as the expected difference between groups decreases, the number of subjects must increase to maintain the same level of power.

7.6.3 CA4/dentate volumes before and after a course of anti-inflammatory drug treatment

A clear protocol was implemented to outline the CA4/dentate hippocampal subfield, this was based primarily on the work previously performed by Mueller et al [251]. Following manual delineation of the CA4/dentate on T2-weighted FSE coronal MRI images pre and post treatment volumes were compared. The five subjects studied were in the age range 44-67. The mean pre and post treatment, left CA4/dentate subfield volumes were 398mm³ and 395mm³ respectively. The mean pre and post treatment, right CA4/dentate subfield volumes were 355mm³ and 338mm³ respectively.

It was difficult to compare these volumes with other values in the literature due to differences in the delineation protocol and in the way ICV corrections were applied. However, the ICV corrected values performed in this study were reasonably similar to the initial ICV corrected values calculated by Mueller et al [251]. ICV corrected CA4/dentate values here ranged from 0.018% - 0.032% (or 182 - 320, when using Mueller et al's units), where Mueller et al's normalised values ranged from approximately 110 to 210 with 42 subjects covering a wide range of ages (21-85). Thus the range of the values in this study appear slightly larger when compared to Mueller et al's results. This may be because the delineating protocol created by the author for this study continued to outline the CA4/dentate regions on more posterior slices than Mueller et al did with the aim of achieving a more comprehensive coverage of the CA4/dentate within the hippocampus. That is, Mueller et al's protocol outlined the CA4/dentate on the first slice after the hippocampal head and then only on a further two slices. Furthermore, they used thinner slices (2mm) to acquire the hippocampal images and thus each voxel on Mueller et al's images had a volume of 0.4mm^3 , whereas in the study described in this chapter the slices were 4mm thick with a 1mm slice gap, thus each voxel has a volume of 0.76mm^3 . Thus for a similar slice, different acquisition protocols may partly account for some of the variation between measurements, though an error of this type would only become important when the CA4/dentate failed to be delineated on a particular slice.

An alternative approach to measuring hippocampal subfields to that which was employed here and that which was employed by Mueller et al would be to delineate the subfields based on a T1-weighted 3D MRI dataset as was done by Malykhin et al [212]. The measurements made in this study by Malykhin et al covered the hippocampus more completely, however, they only subdivided the hippocampus into its head, body and tail regions and so the volumes may not be considered representative of the underlying subfield areas. The T1-weighted 3D dataset offers the advantage of improved through-plane resolution over what is typically achievable with T2-weighted FSE sequences, however, the sequence has inferior GM to WM contrast necessary for clearer visualisation and delineation of the hippocampal subfields. Similar approaches have been implemented by Csernansky et al, Wang et al and Apostolova et al [98, 362, 13]. In these studies subfields were defined across the whole hippocampus based upon boundary positions as defined by an atlas and thus the location of focal volume differences may be visualised on the average hippocampal template.

7.6.4 CA4/dentate volumes with age

Though it was not explicitly the aim of this work to assess subfield volumes with age, it was interesting to note that there was a negative correlation between age and CA4/dentate vol-

ume. That is, as subjects become older, their CA4/dentate volumes tend to decrease. This on its own was not surprising given that it is well established that gross hippocampal volume declines with age [110], moreover recent evidence indicates that this process accelerates beyond the 6th decade of life [297]. However, little work has been performed assessing which subfields, if any, are preferentially affected during aging. This has potential consequences when considering, for example, the progression of Alzheimer's disease and other conditions of pathological aging [85]. Therefore, a clear understanding of how aging affects subfield volumes was indeed worth considering. While the numbers here are small and within the context of the paired study it was not of great significance, it is noteworthy that the CA4/dentate subfield volume declines with increasing age.

7.6.5 Results of DARTEL-VBM comparing hippocampal GM volume before and after anti-inflammatory treatment

Using a novel image normalisation technique, a unique study specific template was created. Using this template, regions of GM volume difference were then assessed between the pre and post anti-inflammatory treatment groups. Given the small numbers present in this pilot study, it was deemed more appropriate to use a nonparametric permutation method to control the false positive rate dictated by multiple comparison problem. Following an FDR correction to the data (0.05), an area suggestive of a GM volume increase was observed in the right hippocampus with a Z-score of 5.51 following treatment of the anti-inflammatory medication. While these results are presented tentatively given the small numbers available in this study, the fact that a volume difference was observed in the apriori region of interest is reassuring. Further work on a larger sample size to try and replicate and reinforce this finding is warranted.

Potentially, if this finding could be verified in a larger sample size, it would be of great significance as it would show in adult humans, in-vivo, that a focal volume increase had occurred in the hippocampus following anti-inflammatory treatment which it has been posited may have anti-depressant qualities via neurogenesis in the dentate gyrus. However, future human studies would benefit greatly from animal work which could be used to establish clear correlations between measures of neurogenesis from histology and MRI outcomes measures.

7.6.6 Comparison with other studies, limitations of the manual outlining procedure and further work

A fundamental limitation of the method employed here was that the definition of the delin-

eated region was partly based upon slice position and not on an anatomical boundary. An improved procedure would be to delineate the CA4/dentate subfield based on anatomical boundaries. As discussed above, the protocol itself is limited by excluding those parts of the CA4/dentate in the hippocampal head. Inclusion of this region may be critical to detecting volumetric changes in the head of the hippocampus as a result of disease or effect of treatment, thus the author would advocate a more complete approach. However, it is worth highlighting that reliable delineation of the hippocampal head may well be considerably more challenging than delineating the CA4/dentate body as a result of the complex, undulating nature of the hippocampal head.

A further improvement to the protocol implemented in this study would be to include delineation of additional hippocampal subfields. While initially we were only interested in the CA4/dentate subfield it may be informative to assess other, neighbouring subfields. Mueller et al [251] did include delineation of other subfields namely: CA1, CA2, subiculum and entorhinal cortex as well as white matter subfields such as the alveus and fimbria. Depending on the study hypothesis it may well be worth extending the analysis of these subfield volumes. In chapter 8, the subfield measurement protocol will be extended to incorporate many of the subfield regions discussed above.

In assessing the reliability of the volume delineation measurements, an improved method would have been to employ some measure of how well the volumes themselves overlap such as the dice metric. This would have enabled a more complete comparison of the repeat hippocampal volumes to be made as the location of the hippocampus and not just the volume are considered. Dice metrics of an extended protocol for hippocampal subfield segmentation will be presented in chapter 8.

Over and above the practical and technical limitations of this study there are some more fundamental issues that underpin the aims of this work that ought to be addressed. From the DARTEL-VBM data there was indeed evidence to suggest that anti-inflammatory treatment may have induced a small increase in hippocampal volume. One possible mechanism by which this may have occurred is by neurogenesis. However, whether or not neurogenesis would actually lead to a detectable difference in volume on an MRI scan and post processing as described above has yet to be validated. Unless one was able to access and assess post-mortem data and correlate neuroimaging volumetric findings with changes with histology, the only other method would be to confirm the methodological process in animal. That is, a study should be performed where it is known that neurogenesis will be induced in the animal, high resolution longitudinal MRI data should be acquired, the pre- and post neurogenesis data should be assessed to see if volumetric changes were detectable. Until this link between the

histology and neuroimaging dataset has been confirmed, human, adult, in-vivo studies will suffer from the doubt that the small and subtle volumetric changes which are being detected are not the result of real changes in the underlying tissue but rather are a subsequence of the methodology. If such a study was to be performed it would be useful to assess the relationship between other neuroimaging markers such as T1, T2, diffusion and perfusion against neurogenesis, again to build confidence in subsequent methodologies applied in-vivo, in adult humans.

7.7 Conclusion

A protocol was established for measuring the CA4/dentate subfield of the hippocampus. Repeat measures were made and it was established that the intra-operator variability was of an acceptable level. CA4/dentate volumes tended to decrease with age. There was no significant difference in CA4/dentate volume following a course of anti-inflammatory medication as measured by manual segmentation on T2-weighted FSE coronal MRI images. Given the variance of the manual measurements of the CA4/dentate region, sample size calculations were estimated based on a desired study power of 0.90 for detecting a projected volume difference of 10%. These calculations determined that a paired study would require 30 subjects in each group and an unpaired design would require 43 subjects in each group in order to satisfy the desired study power.

To complement the manual volume measurements made of the CA4/dentate region DARTEL-VBM using permutation based analysis was performed. This method detected an area suggestive of a GM volume increase in the right hippocampus following anti-inflammatory treatment (FDR corrected, $Z = 5.51$). This finding needs further validation in a larger dataset and would benefit from a complementary study in animals to validate the relationship between neuroimaging markers and known neurogenesis, thus providing greater assurance in the methodology.

Regardless of the underlying biological mechanisms at work and despite the small numbers of subjects sampled in this pilot study, tentative evidence has been observed for volumetric differences between the pre and post treatment groups. The methodology appears to be robust, including the use of DARTEL-VBM and the application of permutation based statistics to VBM. Further work on larger sample sizes is strongly recommended and the author of this thesis is currently involved with such work.

Chapter 8

Hippocampal subfield volumes in normal subjects using 3T MRI

8.1 Abstract

In the earlier chapters of this thesis methods to assess the volume of the whole hippocampus were presented (chapters 2 and 5). Further methods were presented to localise regions of volume difference between two subject groups in the hippocampus by a surface based method (chapter 6) and then by producing a protocol to manually outline a specific subfield of the hippocampus, absolute volume measures were assessed (chapter 7). Here we extend the approach presented in chapter 7 to include a manual delineation protocol for a comprehensive set of hippocampal subfields throughout the entire length of the hippocampus. This is one of the first attempts to segment the entire hippocampus into its subfields using 3T MRI and as such, it was important to assess the quality of the measurement procedure. This included inter and intra operator variability, the dice metric to assess how well volumes overlap, as well as the subfield volumes and the variability within a range of normal subjects. Furthermore, given the subfield volumes and the variability in these measurements, power and sample size calculations were also estimated to inform further work. Subfield volumes were also compared to age and sex of the normal volunteers.

Of the thirty-nine normal, healthy volunteers discussed in chapter 5, seventeen subjects were randomly selected for subfield volume analysis. The subjects were scanned using 3T MRI with coronal T2-weighted FSE images planned perpendicular to the long axis of the hippocampus. A detailed manual segmentation protocol was created to guide two independent operators to

measure the hippocampal subfield volumes. Repeat measures were made by a single operator for intra-operator variability and inter-operator variability was also assessed. The Intra-rater results proved reasonably successful where values compared well but were typically slightly poorer than similar attempts in the literature. This was likely to be the result of the additional complication of trying to segment subfields in the head and tail of the hippocampus where previous studies have focused only on the body of the hippocampus. Inter-rater agreement measures for subfield volumes were generally poorer than would be acceptable if full exchangeability of the data between the raters was necessary. This would indicate that further refinements to the manual segmentation protocol and more training of raters is needed before confidence in subfield volumes could be achieved across raters.

Normal ranges for hippocampal subfield volumes were established. Sample size calculations and Dice measures were also obtained. Broadly speaking grey matter CA1-CA3 and CA4/dentate subfields were measured more reproducibly and with less variability than other subfields, consequently these subfields had better Dice metrics. This implied that lower sample sizes would be necessary to extract statistically significant differences between groups for these regions. For example, for the CA1-CA3 and CA4/dentate it was determined that approximately 25-80 subjects per group would be necessary to achieve a study power of 0.9 for detecting a 10% volume difference. However, other grey matter subfield regions such as the entorhinal cortex (ERC) and white matter subfield regions such as the fimbria and alveus were less reproducible and as such would require greater subject numbers to achieve good study power by the method presented here. That is, approximately 130-200 subjects per group would be required to achieve a study power of 0.9 for detecting a 10% volume difference for these subfield regions. These are large numbers of subjects for imaging studies and thus while studies of this size are not impossible to perform they are quite rare and perhaps practically unrealistic. Future work should seek to improve the methodology to reduce the variability and improve the reproducibility in these measures. Such developments would benefit from consensus over the anatomical regions and boundaries of the various hippocampal subfields and it is hoped this work will contribute to the body of the peer reviewed literature in this area.

The work performed in this chapter, i.e. the creation of protocol for manually segmenting the subfield volumes throughout the whole hippocampus and the assessment of the quality of these measures is believed to be of interest to the wider scientific community. Therefore, this work forms the basis for a scientific manuscript currently in preparation for publication to a peer reviewed journal.

8.2 Introduction

As we have seen from discussions throughout this thesis, the hippocampus is an important and complicated structure with a number of its surfaces being curved, particularly at the hippocampal head, where the gyrus folds over on itself. While there are many studies of hippocampal volumes using MRI at 1.5T, there has been growing interest in recent years to exploit the potential of higher field MRI systems to not only explore the hippocampal structure as a whole, but to begin to investigate its substructures, or more precisely, the subfield regions.

There are several reasons why it was worthwhile to further assess the various subfields of the hippocampus. It is clear that the hippocampus is a mass of variable cell types, defining cellular regions or subfields which have distinct functions, all of which integrate into a complicated network of cellular activity. If one is to further interrogate these regions in either structural or functional studies, and importantly to begin to generate more specific hypotheses then one must understand the natural variability in these regions such that robust answers may be obtained. Only then can reliable comparisons within and between diseased populations be made.

An important development in recent years has been the discovery that neurogenesis, the process of new cell production, has been found to occur in the dentate gyrus area of the hippocampal head. It has been speculated that neurogenesis may be a possible mechanism by which anti-depressant medication acts and thus there was clear motivation to focus on the hippocampus, its subfields and in particular the dentate gyrus.

8.3 Literature review

The literature review of measuring hippocampal subfield volumes was conducted in chapter 7. The implementation of subfield measurements in chapter 7 was extended here to include a more comprehensive manual delineation of the subfields of the entire hippocampus, including the head, body and tail.

8.4 Aim and hypotheses

In this chapter a novel and comprehensive working protocol for determining the subfield volumes of the hippocampus will be presented. This will include a full description of subfield

segmentation on images from the hippocampal head, body and tail, an approach which goes beyond that which has been presented in the current literature. Subfield volumes will be presented with the focus on quantification of the variability in these structures to establish the appropriate power that would be necessary for further structural volume studies. An error analysis of the volumes will also be performed with intra and interoperator variability measures being presented and discussed. Right and left hippocampal subfield volumes will also be compared. Furthermore, the relationship of subfield volumes to age and ICV will also be explored.

8.5 Methodology

8.5.1 Ethics

The study was approved by the local NHS research ethics committee.

8.5.2 Volunteers

The subjects who provided the data for this chapter of the thesis were a subset of the subjects described in chapter 5 of this thesis. Therefore the inclusion and exclusion criteria for these subjects has been previously described in chapter 5. The subset of 17 volunteers were randomly selected from the larger group of 39 subjects. Given the resources available and the time taken to manually segment the hippocampus into its subfield components, the choice of 17 subjects, although somewhat arbitrary, was a compromise between obtaining a reasonable subject number to assess the subfield volume variability in a practically achievable timeframe. The mean age range of the subgroup was 21-59 years old, with a mean age of 36 and standard deviation of 12. 7 volunteers were male and 10 were female.

8.5.3 Imaging hardware

All MRI scans were acquired using a 3T GE Medical systems, Signa Excite HD system (Milwaukee, USA) with software version 12m5. Furthermore, a body transmit coil and an eight-channel phased array receive only head coil were used to acquire the imaging data.

8.5.4 Exclusion criteria

The subject exclusion criteria has previously been described in chapter 5. Briefly, subjects were healthy volunteers, with no history of head injury or neurological disorder. Subjects

were chosen such that they were suitable for MRI scanning and had no medical implants that were contraindicated with the MRI environment.

8.5.5 Data acquisition

The data acquisition protocol has been described previously in chapter 7. Briefly, the acquisition of the sagittal FLAIR T1-weighted images, allowed the subsequent T2-weighted propeller FSE to be planned parallel to the hippocampus. Then the high resolution T2-weighted FSE coronal acquisition was planned perpendicular to the long-axis of the hippocampus using the T1-weighted sagittal FLAIR to obtain the correct pitch for the slices. The T2-weighted propeller FSE was then used to check the anterior and posterior coverage of the hippocampus. Typically 11, 4mm T2-weighted FSE slices, with a 1mm slice gap, were acquired in the coronal plane.

8.5.6 Image analysis

8.5.6.1 Hippocampus: Manual segmentation of subfield

To-date several studies have aimed to segment the hippocampus into further subfields [257, 212, 362]. These approaches have exploited different types of information, namely, the use of anatomical landmarks, and the use of intensity differences within the hippocampus. These methods all suffer from the same limitation, that is, one must use MRI based intensities to define cellular subfields and due to the variance in hippocampal anatomy, the application of sometimes arbitrary landmarks can lead to inaccuracy. However, despite this perceived limitation, this is not to say one cannot develop a consistent and reproducible framework within which both cross-sectional and longitudinal studies can be performed. To be able to perform such work it is critical to have a clear and precise protocol that can be applied consistently. Previously similar issues to these have been addressed with respect to the hippocampus as a whole [185], this was a direct result of the large variability in the methodology that was being applied to assess the hippocampus. We suggest here that by defining a robust and comprehensive protocol for defining substructures of the hippocampus, that others may follow this protocol to avoid the issues that occurred within hippocampal volume studies. Failing that, we would hope others would criticise this protocol with a view to its refinement such that the optimum method of assessing the subfields of the in-vivo hippocampus using MRI may be determined. To aid with the challenge in answering the most fundamental question, that is, how does the MRI subfield regions compare with histology, we would encourage others to pursue this work.

8.5.6.2 Protocol for manually outlining hippocampal subfields

Regions of the hippocampus and related structures to be manually outlined on coronal T2-weighted images were: the alveus, fimbria, subiculum, CA1, CA2 and CA3. The white matter layer between CA1, CA2, CA3 and the dentate gyrus was also outlined as was the dentate gyrus/CA4 subfield and the entorhinal cortex. The protocol used to define these subfield regions is described below and exploits a mixture of anatomical landmarks and image intensities to achieve subfield segmentation.

8.5.7 Description of hippocampal subfields

8.5.7.1 Entorhinal cortex

The entorhinal cortex (ERC) is the grey matter layers ranging from the parahippocampal gyrus to the subiculum. Similarly to the subiculum this cellular field has been defined by fundamental cellular histology. Here we define the lateral boundary of the ERC as the base of the rhinal sulcus and then as we progress posteriorly through the slices the base of the rhinal or collateral sulcus will be used, with the medial border being used to approximate the border with the subiculum, defined as being the most medial aspect of the hippocampal formation and temporal cortex. When viewing coronal slices of the hippocampus the ERC lies anteriorly, inferiorly and medially to the hippocampal head. Moving posteriorly from the most anterior slice, the ERC will begin to be outlined on the first slice on which the hippocampal head is present, this will continue until the slice on which the white matter of the Parahippocampal gyrus protrudes to the medial border between ERC and the subiculum.

On figure 8.1 on the following page point A indicates the border between the rhinal sulcus and the ERC. Point B indicates the medial and superior boundary between the ERC and subiculum. Point C indicates the apex of the hippocampal head which we use as a landmark denoting the lateral boundary between the subiculum and CA1. The image on the right is an exact match for the image on the left where the image on the right shows the subfield segmentations.

8.5.7.2 The subiculum

The subiculum is the grey matter area of the hippocampal gyrus that extends from the most medial aspect of the hippocampus (i.e the ERC) to the boundary of CA1. These

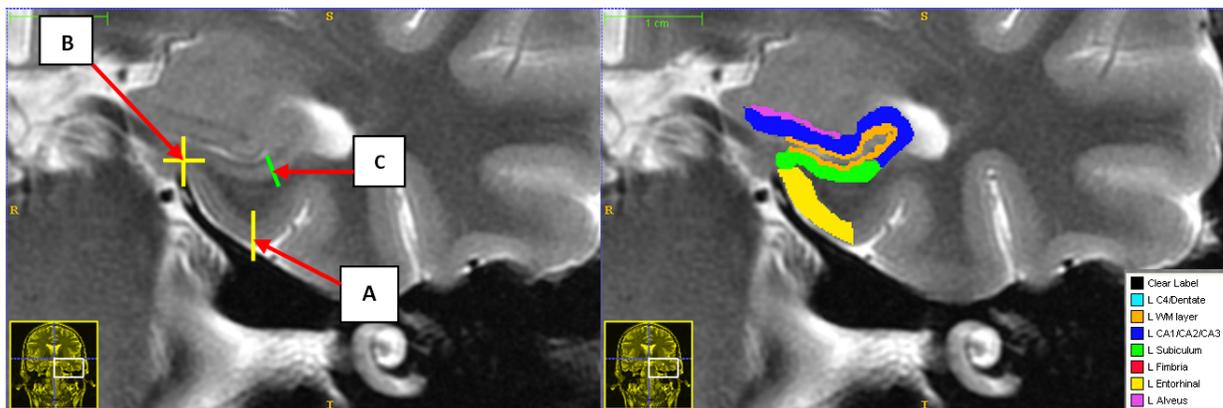


Figure 8.1: Delineation of hippocampal subfields at the level of the hippocampal head.

regions are defined at a cellular level based on differing cell types, such a definition was not possible using MRI at present and as such the subfields were only approximately defined by using anatomical landmarks. For the purposes of reproducible manual outlining we adopt the description used by Mueller et al [251], where the lateral boundary of the subiculum was defined directly inferiorly to the medial aspect of CA4. This was at the level of the hippocampal body. While Mueller et al describes how to define the subiculum on slices on the hippocampal body, there was no description of how to outline this subfield for the hippocampal head and tail. From the Duvernoy atlas [114], at the hippocampal head, the subiculum is shown to extend from the ERC to the lateral apex of the hippocampus. As we move posteriorly through a second slice within the hippocampal head, the subiculum recedes to a level directly superior to the middle of the CA4/dentate region. See figure 8.2 on the following page for an imaging representation. As we reach the level of the hippocampal body, Mueller et al's description becomes valid and the boundary of the subiculum/CA1 becomes the boundary directly inferiorly to the medial aspect of the CA4/dentate.

On figure 8.2 point A indicates the border between the collateral sulcus (on anterior slices the rhinal cortex was used) and the ERC. Point B indicates the medial and superior boundary between the ERC and subiculum. Point C indicates the landmark directly inferior to the central hippocampal body which denotes the lateral boundary between the subiculum and CA1. Point D is a line running tangentially to the CA4/dentate region and acts as a landmark between this area and the CA3 component of the CA1-CA3 subfield. The image on the right is an exact match for the image on the left where the image on the right shows the subfield segmentations.

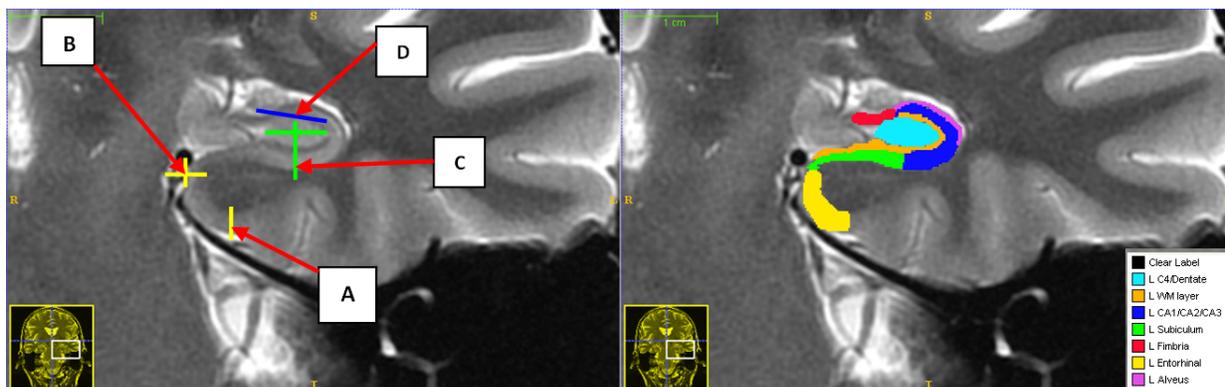


Figure 8.2: Delineation of hippocampal subfields at the anterior aspect of the hippocampal body.

8.5.7.3 Grey matter of CA1, CA2, and CA3

CA1, CA2 and CA3 are observed as distinct regions by histology and define different cellular subfields around the inferior, medial and superior aspects of the hippocampal head, body and tail. This subfield region lies between the subiculum and the CA4 subfield as we follow the naturally inward curvature of the hippocampus. The CA1 – CA3 region follows from the boundary of the subiculum until the entrance to the central portion of the hippocampal head and the CA4/ dentate, where the boundary between CA3 and CA4 was the line running tangentially to the CA4/dentate region. At the anterior and superior aspect of the hippocampal head, the CA1, lined by the alveus protrudes medially. This protrusion is responsible for giving the 'seahorse' its head. As we move to more posterior slices the body of the hippocampus becomes thinner and the most medial aspects of the CA1-CA3 subfield on the superior portion of the hippocampus was the boundary with the CA4 subfield.

8.5.7.4 The dentate gyrus and CA4

The central portion of the hippocampus consists of the dentate gyrus and the CA4 subfield region, this region projects through the entire length the hippocampus. The CA4/dentate region is perhaps the most easily definable region of the hippocampus being that it is almost entirely enclosed within a WM layer, though it can be more difficult to define within the hippocampal head due to the undulating superior digitations of CA1-CA3. Thus, within the hippocampal head, the dentate/CA4 region was outlined as the region that was visibly encompassed within the WM layer and distinct from CA1-CA3 subfield. It is worth noting at this point that the volume of CA4 cellular subfield accounts for the majority of the volume in this combined space. See figure 8.3 below.

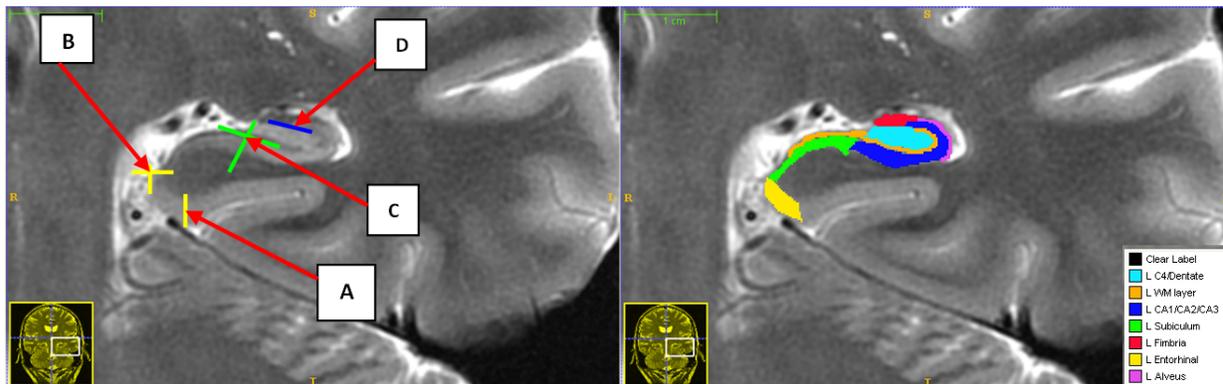


Figure 8.3: Delineation of hippocampal subfields at the central portion of the hippocampal body.

On figure 8.3 point A indicates the border between the collateral sulcus and the ERC. Point B indicates the medial and superior boundary between the ERC and subiculum. Point C indicates the landmark directly inferior to the most medial aspect of the hippocampal body (i.e. the medial aspect of the CA4/dentate subfield) which denotes the boundary between the subiculum and CA1. Point D is a line running tangentially to the CA4/dentate region and acts as a landmark between this area and the CA3 component of the CA1-CA3 subfield. The image on the right is an exact match for the image on the left where the image on the right shows the subfield segmentations

8.5.7.5 White matter layer between CA1, CA2, CA3 and the dentate gyrus

In the body and tail of the hippocampus there is a thin layer of white matter between the internal dentate gyrus and the external CA1, CA2 and CA3 subfields. This appears as a dark band, hypointense region on the T2-weighted FSE acquisition. More precisely this region is the stratum moleculare of the CA layer but we will consider it simply as a WM layer separating the CA1-3 and the dentate gyrus. This region will be manually delineated largely based on its lower intensity, however in the hippocampal head it was also necessary to define a medial border for the WM layer which on the inferior aspect of the hippocampal head straddles the subiculum. Thus the WM layer straddling the subiculum in the hippocampal head will be outlined as far as it is visible moving medially, but no further than the inferior-medial border of the subiculum itself. In some cases where image quality was poor then it

became difficult to define the WM layer as this layer is relatively thin. Similarly, it also suffered badly from through-plane partial volume effects.

8.5.7.6 Alveus

The alveus is a thin white matter layer lying predominately on the superior surface of the hippocampus although it does project anteriorly and inferiorly providing a boundary between the hippocampal head and the amygdala. On T2-weighted FSE images the alveus appears as a dark layer between the brighter intensity grey matter regions CA1 – CA3 and the much brighter CSF surrounding the anterior portion of the hippocampal head. Similarly the alveus forms a distinct layer moving laterally and also posteriorly, between CA1-CA3 and the CSF in the lateral ventricles. On the lateral surface of the hippocampus the alveus acts as a boundary between the CA1 and CSF and continues until it joins the larger mass of WM of the parahippocampal gyrus. Within the hippocampal body the alveus runs along the surface of the CA1-CA3 subfield until it reaches the fimbria, a broader region of WM, this is typically approximate to the position directly superior to the point where CA3 ends and enters the CA4/dentate complex. Thus, the major factor in defining the 'end' of the alveus and thus the 'beginning' of the fimbria was the point at the broadening of the superficial WM layer above CA3. See figure 8.4 on the following page. If image quality was poor then it became difficult in some cases to define the relatively thin layer of the alveus. Similarly, it also suffered badly from through-plane partial volume effects. This can be particularly bad as a result of the fact the alveus lies adjacent to the CSF.

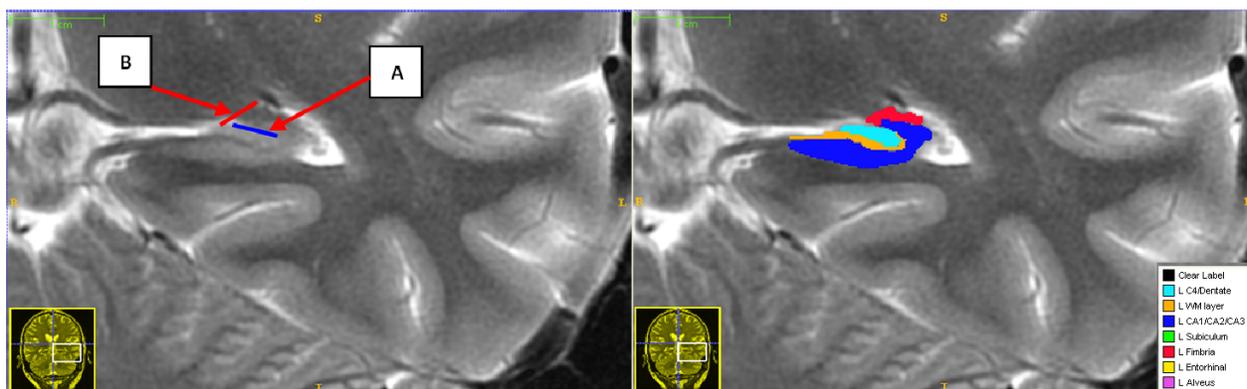


Figure 8.4: Delineation of hippocampal subfields at the hippocampal tail.

On figure 8.4 point A is a line running tangentially to the CA4/dentate region and acts as a landmark between this area and the CA3 component of the CA1-CA3 subfield. Point B acts

as the medial boundary of the fimbria. It is the line running superiorly as the tangent to the most medial aspect of the CA4/dentate subfield. The image on the right is an exact match for the image on the left where the image on the right shows the subfield segmentations

8.5.7.7 The fimbria

The fimbria, from the Latin for fringe, is the white matter bundle lying at the superior medial aspect of the hippocampus. The fimbria projects posteriorly from the posterior edge of the hippocampal head to the tail of the hippocampus before joining the fornix. As it is a white matter bundle the fimbria appears darker grey on T2-weighted FSE images, relative to the brighter grey GM regions. While the intensity of the alveus and fimbria are similar, the fimbria does appear slightly darker than the alveus and as discussed above, the fimbria is a broader region of WM than the thin superficial alveus. Furthermore, these structures may be distinguished by their position, where the alveus lies between the grey matter CA1-CA3 subfield and the CSF on the lateral and superior aspects of the hippocampal surface. This is in contrast to the fimbria, which is thicker in size and is not positioned between the CA1-CA3 and CSF, but more at the superior-medial aspect of the hippocampal body. Moving posteriorly from the body of the hippocampus to the tail, it becomes more difficult to define the medial border for the fimbria. The fimbria itself becomes thinner as we reach the tail of the hippocampus where it joins the fornix. We define the medial boundary of the fimbria on the most posterior slice of the hippocampus as the point no further than the line running tangentially to the medial border of the CA4/dentate crosses the fimbria. See point B on figure 8.4.

8.5.7.8 Order of delineation

The order in which the author found it most straightforward and beneficial to delineate the subfields was as follows. CA4/dentate, ERC, subiculum, CA1-CA3, WM layer, alveus and then the fimbria. This typically involved starting to outline subfield regions on a slice in the hippocampal body as shown in figure 8.5 above. Further subfields were then segmented in the hippocampal head and tail. The subfield segmentations were then reviewed and corrected as necessary to satisfy the protocol.

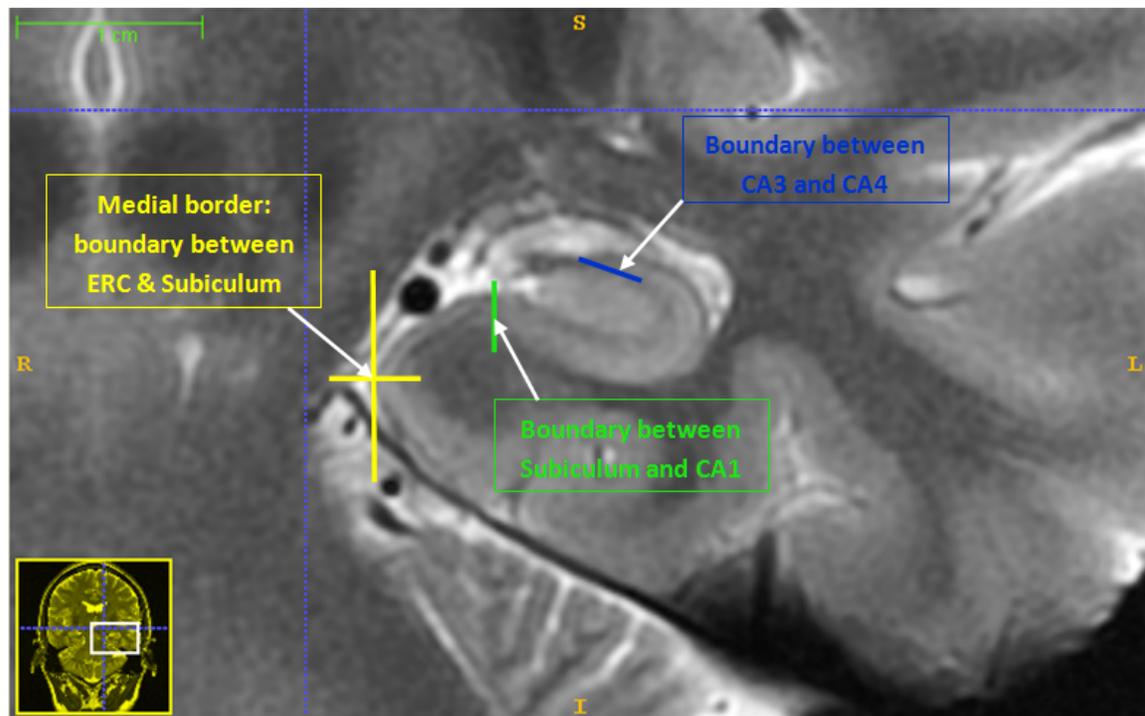


Figure 8.5: Hippocampal substructures: markers delineating subfield boundaries

8.5.7.9 Implementation of manual segmentation protocol

Manual segmentation of the subfields of the hippocampus was performed using the ITKsnap software (v1.8, Paul Yushkevich, www.ITKsnap.com). Labels were created for the subfield regions to be segmented. To ensure consistency in windowing the images, the auto windowing option within the ITKsnap display options was enforced. The MRI images were converted from dicom to nifti format using Chris Rorden's dcm2nii GUI tool which is bundled as part of the MRICron software download (<http://www.cabiatl.com/mricron/>). The nifti images were then uploaded into ITKsnap and the subfields were manually outlined as per the protocol detailed above. For the outlining the polygon tool was used initially for C4/dentate, ERC, subiculum and CA1-CA3 subfields after which the paint tool with a round shape and size of 2 or 3 pixels was used to define the WM layer, alveus and fimbria. Finally, the paint tool was used to refine the segmentation of all subfields. The left hippocampal subfields were outlined first, followed by the right hippocampal subfields. The resulting image segmentations were saved as an image mesh (.vtk format) and the volume and statistics for the segmented subfields were extracted. In total the segmentation of all the subfields within an individual hippocampus took between 20-30mins. This was comparable to the time it takes to manually segment the whole hippocampus on a 3D T1-weighted image.

8.6 Results of normal subjects hippocampal subfield data

8.6.1 Test normality of subfield volumes

The distribution of all subfield volumes were tested for normality. Using the Andersen-Darling test, applied using the Minitab v16 statistical package, the null hypothesis that the data are all from a normal distribution was not rejected for any of the subfield volumes. Thus the subfield volume data may be treated as being part of a normal distribution and therefore, it was appropriate to use normal statistics to interrogate the dataset.

8.6.2 Normal ranges for raw subfield volumes

Following the manual outlining procedure the mean volume and 95% confidence limits (CI's) were determined for both the R and L hippocampal subfields. The results are shown in table 8.1 on the following page.

Subfield	Mean Volume (mm ³)	Std Dev	Std Error	CoV (%)	95% CI	Confidence Limits	
						Mean - CI	Mean + CI
L CA4/Dentate	585	106	26	18	54	531	640
R CA4/Dentate	556	109	26	20	56	499	612
L WM layer	394	64	16	16	33	331	394
R WM layer	403	83	20	21	42	360	445
L CA1, CA2, CA3	1345	163	40	12	83	1260	1429
R CA1, CA2, CA3	1408	138	33	10	70	1337	1479
L Subiculum	390	104	25	27	53	336	443
R Subiculum	338	77	19	23	39	299	378
L Fimbria	107	30	7	28	15	92	123
R Fimbria	105	28	7	27	14	91	119
L Entorhinal cortex	634	189	46	30	96	597	791
R Entorhinal cortex	589	136	33	23	69	519	659
L Alveus	161	46	11	29	23	134	185
R Alveus	195	65	16	33	33	163	292

Table 8.1: Mean volume, 95% CI and lower and upper CI limits

In addition to the standard deviation, standard error and confidence intervals the coefficient of variation (CoV) was also included.

$$CoV = (StdError/Mean) * 100 \quad (8.1)$$

The CoV was of benefit as it expresses the relationship between the standard deviation and the mean as a percentage. Thus it is an error measurement which considers the size of the measurement itself i.e. it is a relative error measurement. This makes it useful when comparing errors across different measurements and in particular in this case, for comparing the error values across different subfields.

This data is also shown graphically in figure 8.6 on the following page.

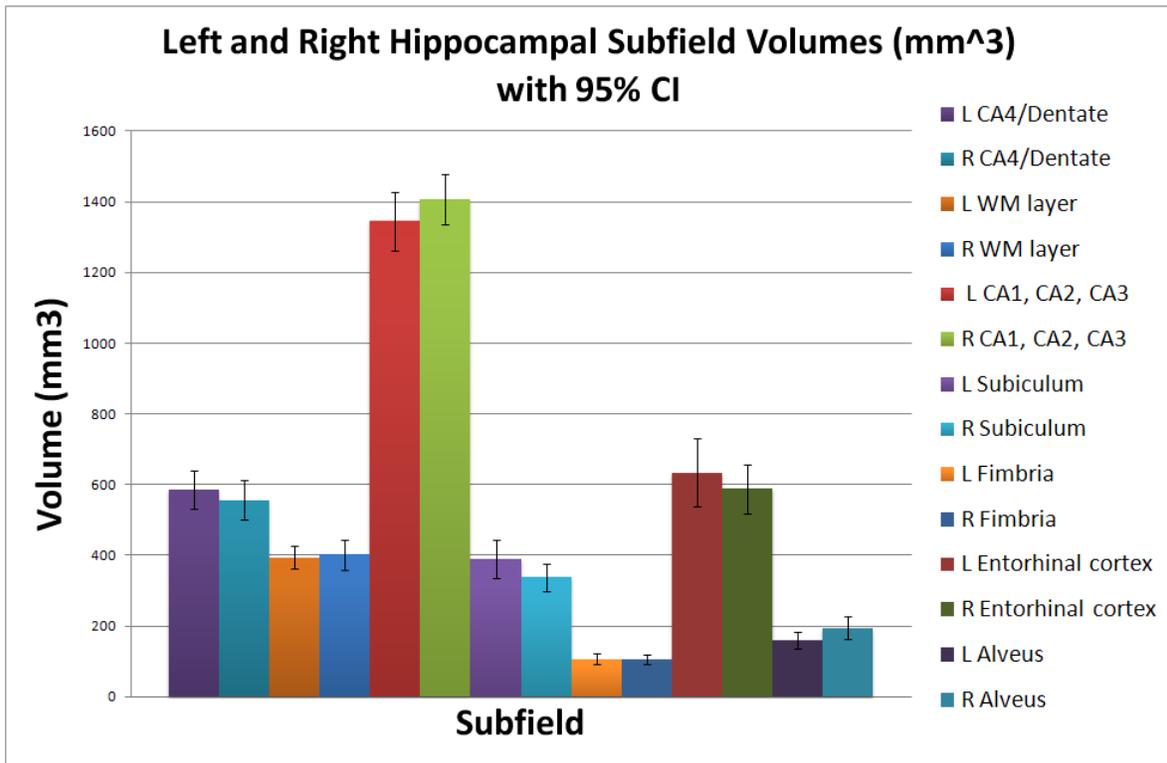


Figure 8.6: Normal hippocampal subfield volumes and 95% CI's

8.6.3 Assess correlation between raw subfield volumes and ICV

The left and right subfield volumes plotted against ICV are shown in figure 8.7 on the pages that follow.

Pearson correlation values were obtained to determine whether or not a linear correlation

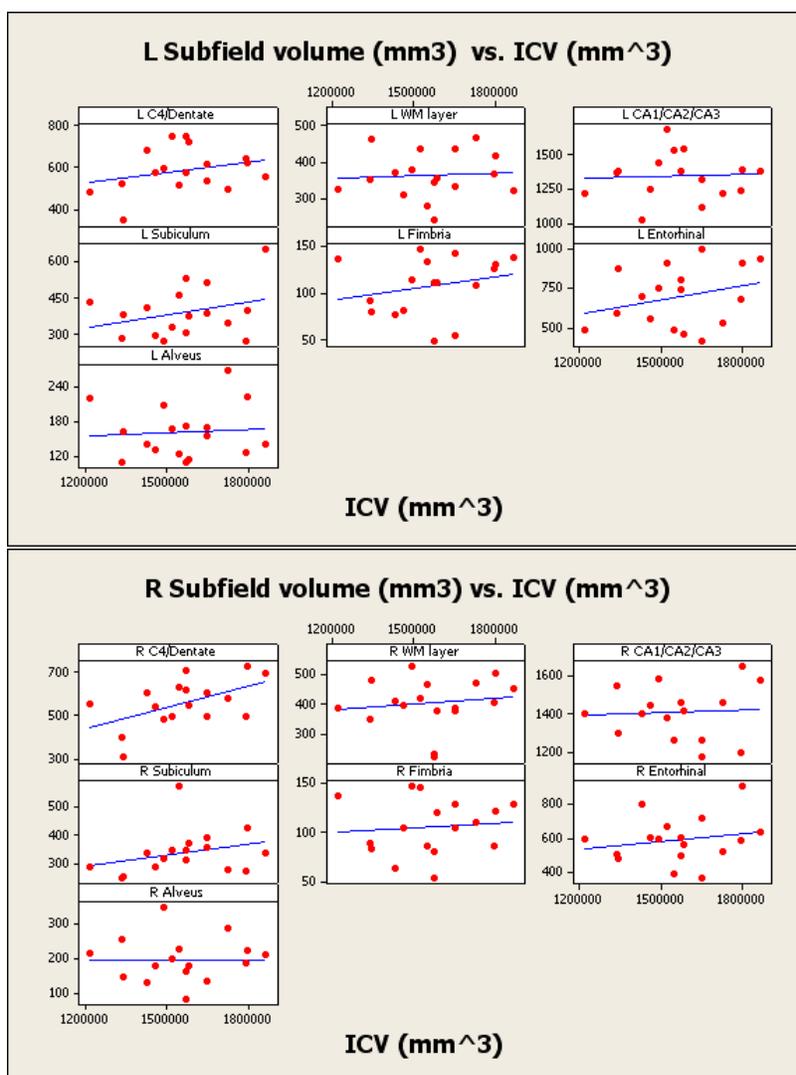


Figure 8.7: Scatterplot of left and right hippocampal subfield volumes vs. ICV

existed between each subfield volume and the ICV. This was necessary to determine whether it was appropriate to correct subfield volumes for ICV when presenting normal ranges and for subsequent analysis in comparing subfield volumes across different subject groups. The strongest correlation of any of the subfield regions with ICV was the right CA4/dentate region, with an r -value of 0.54 ($p=0.03$). In the regression model this equated to ICV accounting for 24.7% (R^2 -adjusted value) of the variance in the volume of the right CA4/dentate region. Several other subfield volumes exhibited modest correlations ($r>0.2$) with ICV, including the left CA4/dentate, left and right subiculum, left fimbria and left/right entorhinal cortex, though none of these correlations were deemed statistically significant ($p<0.05$). The remaining subfield volumes exhibited little or no correlation with ICV (r -value <0.2). For both the modest and low correlation subfield volume vs. ICV cases the corresponding R^2 -adjusted

values were all very low ($<3\%$) as would be expected given the low Pearson r -values.

8.6.4 Assess correlation between raw subfield volumes and age

The left and right subfield volumes plotted against age are shown in figure 8.8 on the pages that follow.

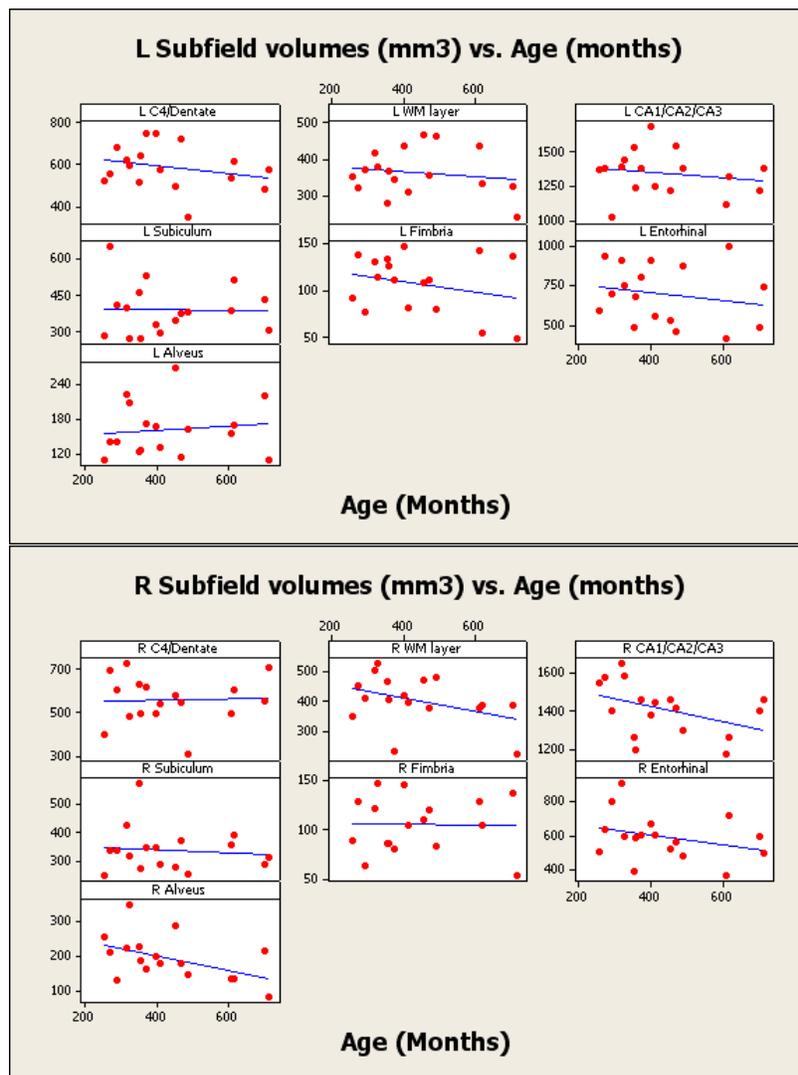


Figure 8.8: Scatterplot of left and right hippocampal subfield volumes vs. Age

Pearson correlations were obtained to determine whether or not a linear correlation existed between each subfield volume and subject age. This was necessary to determine whether it was appropriate to correct subfield volumes for age when presenting normal ranges and

for subsequent analysis in comparing subfield volumes across different subject groups. The strongest linear correlations observed were for the right WM layer ($r = -0.4$, $p = 0.11$); right CA1, CA2 ($r = -0.42$, $p = 0.10$) and CA3 and the right alveus ($r = -0.48$, $p = 0.05$). Thus only the right alveus volume was deemed to have a statistically significant correlation with age. Several other modest correlations existed ($0.2 < r < 0.4$) for the left CA4/dentate; left fimbria and right entorhinal cortex but none were deemed statistically significant. The remaining subfield volumes typically expressed weak negative correlations with age, none of which were deemed statistically significant. In a regression model for subfield volumes with age as a predictor, the R^2 -adjusted value was greatest for the right alveus at 18.2%; R CA1,CA2,CA3 at 11.7% and right WM layer at 10.5%. Both age and ICV explained little ($<3\%$) of the variance for the remaining subfield volumes.

8.6.5 Left vs. right subfield volumes

As discussed in chapters 2 and 5, there has been a trend for the right hippocampus to be smaller than the left hippocampus in normal volunteers. Thus it follows that if we partition the hippocampus into its subfield components we may be able to identify if a specific subfield was responsible for the perceived lateralised volume difference or perhaps establish that the lateralised volume difference was the result of the sum of smaller volume reductions distributed across the entire volume of the whole hippocampus. A two sample t-test was performed to compare the left and right hippocampal subfield volumes. Table 8.1 shown previously (section 8.6.2) summarises the results. From table 8.1 there were no statistically significant differences between right and left hippocampal subfield volumes.

8.6.6 Male vs. Female

The results in table 8.2 below compares the female and male raw hippocampal subfield volumes.

The results for comparing the female and male raw hippocampal subfield volumes was that, for all but one subfield, there was no statistically significant differences between the volumes. The exception was the fimbria where there was a general trend for the female subfield volume to be slightly smaller, this bordered on being statistically significant. To further investigate

Female vs. Male Subfield volumes (mm³), Based on F = 20, M = 14 Hippocampi					
Subfield Structure	Female Mean	Male Mean	Female Std Dev	Male Std Dev	p-value
CA4/Dentate	560	586	119	90	0.48
WM layer	381	386	76	78	0.86
CA1, CA2, CA3	1374	1380	149	163	0.91
Subiculum	342	395	72	114	0.14
Fimbria	99	117	30	23	0.05
Entorhinal cortex	671	600	163	179	0.25
Alveus	175	183	63	51	0.65

Table 8.2: Comparison of female and male hippocampal subfield volumes

the subfield volumes between men and women the ICV corrected volumes were determined, see table 8.3 on the following page.

Female vs. Male Subfield volumes corrected for ICV (as a %ICV), Based on F = 20, M = 14 Hippocampi					
Subfield	Female Mean	Male Mean	Female Std Dev	Male Std Dev	p-value
CA4/Dentate	0.037	0.036	0.006	0.006	0.83
WM layer	0.025	0.024	0.006	0.005	0.39
CA1, CA2, CA3	0.092	0.086	0.014	0.012	0.21
Subiculum	0.023	0.024	0.005	0.007	0.45
Fimbria	0.007	0.007	0.002	0.001	0.33
Entorhinal cortex	0.044	0.037	0.010	0.011	0.07
Alveus	0.011	0.011	0.005	0.003	0.76

Table 8.3: Comparison of female and male hippocampal subfield volumes corrected for ICV

After correcting the subfield volumes for ICV the volumes for females and males were compared. The difference in the fimbria volume between men and women no longer bordered on being statistically significant, there was no significant difference in volume between the groups. With the exception of the subiculum subfield the female ICV corrected volumes tended to be slightly larger than the male volumes. However, none of these ICV corrected subfield volume differences were statistically significant between the male and female groups.

8.6.7 Intra-rater variability and the Dice overlap metric

The concept of ICC and Dice overlapping metric were introduced in this chapter 2. The ICC intra-rater agreement and Dice overlap metric were calculated and the results are presented in table 8.4 below and figure 8.9 on the following page.

	Subfield Structure													
	CA4/ D	CA4/ D	WM layer	WM layer	CA1/ CA2/ CA3	CA1/ CA2/ CA3	Sub	Sub	Fim	Fim	Ent	Ent	Al	Al
	L	R	L	R	L	R	L	R	L	R	L	R	L	R
ICC (agreement)	0.79	0.85	0.59	0.63	0.83	0.77	0.94	0.55	0.61	0.57	0.84	0.79	0.75	0.62
Dice	0.83	0.79	0.60	0.56	0.77	0.74	0.70	0.64	0.72	0.62	0.81	0.76	0.60	0.58
Where CA4/D = CA4/Dentate; WM = white matter layer; Sub = Subiculum; Fim = Fimbria; Ent = Entorhinal cortex and Al = Alveus														

Table 8.4: Subfield ICC (agreement) and Dice metric

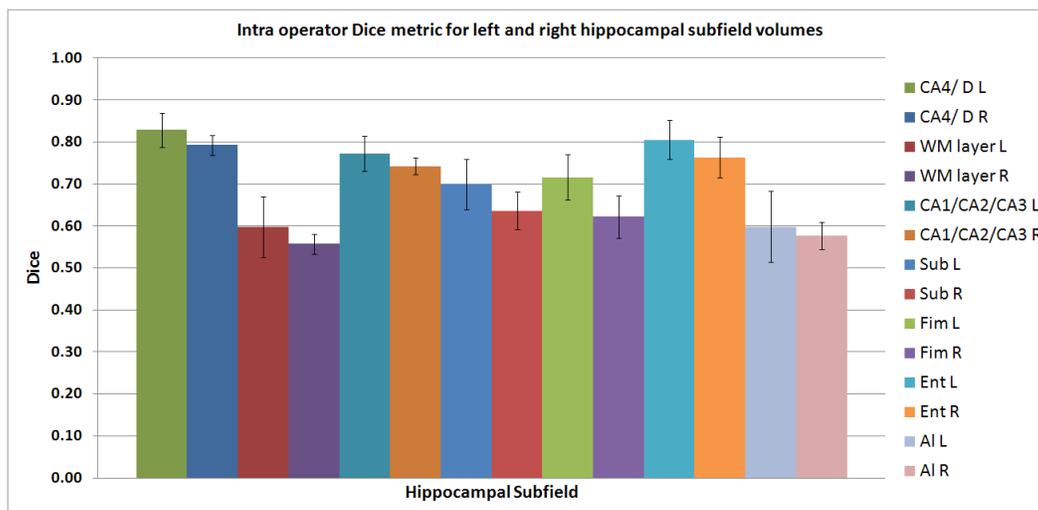


Figure 8.9: Intraoperator Dice for hippocampal subfield volumes

As can be seen from table 8.4 and figure 8.9, the largest Dice measures were observed for the CA4/dentate, ERC and CA1/CA2/CA3. Poorer Dice measures were observed for the subiculum, fimbria, alveus and WM layer.

8.6.8 Inter-rater variability and Dice overlap metric

To test the reproducibility of the manual subfield segmentation protocol across different raters, the Dice metric for all the subfields for 11 subjects were compared between the two raters. The results of the inter-rater Dice metric measurements are shown in figure 8.10 on the following page.

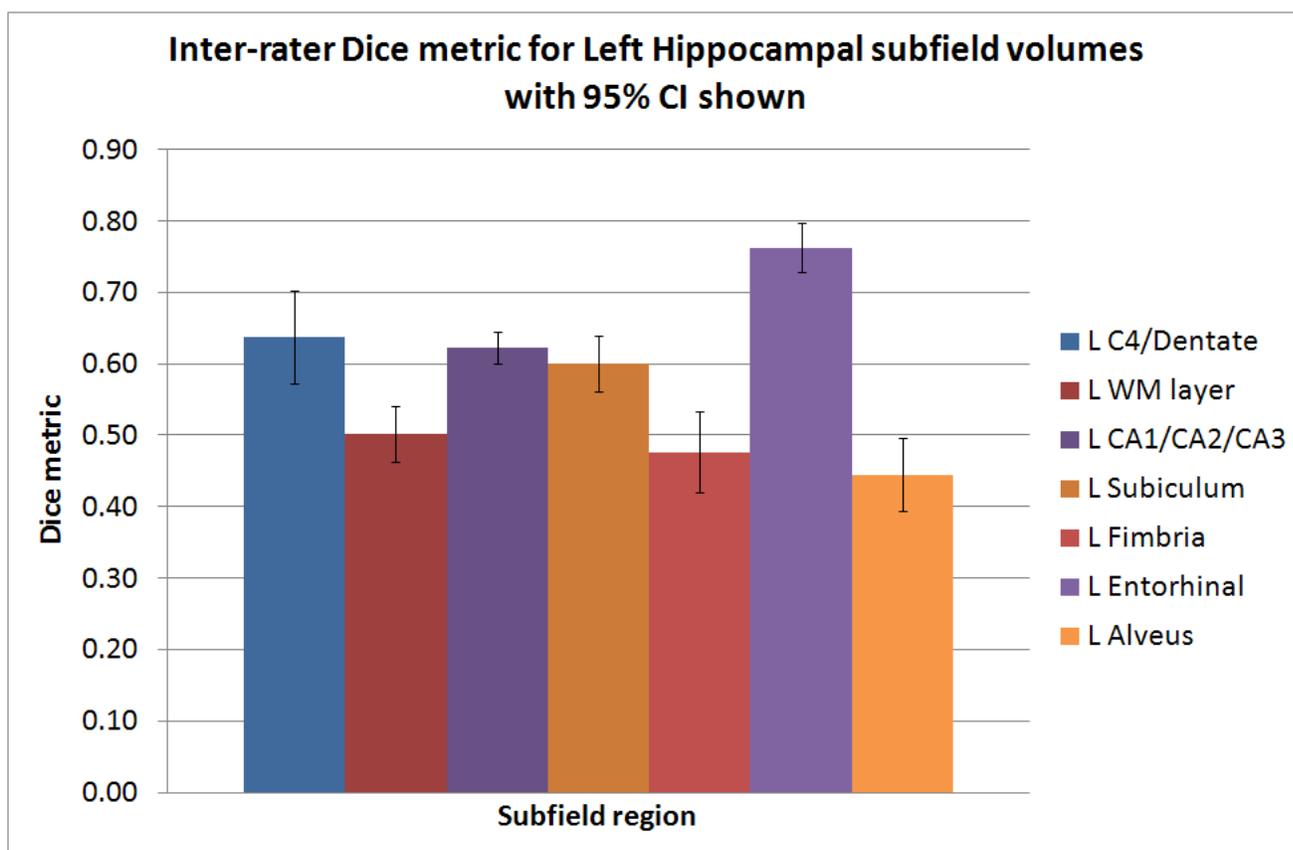


Figure 8.10: Inter-rater Dice for left hippocampal subfield volumes

The corresponding Dice metric data from figure 8.10 is shown in table 8.5 below. As per the intra-rater assessment, ICC agreement measures are also included. As a point for interest and discussion the ICC (consistency) measure between the two raters are also shown in table 8.5.

Subfield Structure: Inter-rater Dice and ICC agreement and consistency measures							
	CA4/ Dentate	WM layer	CA1/CA2/CA3	Subiculum	Fimbria	Entorhinal	Alveus
	L	L	L	L	L	L	L
ICC (agreement)	0.34	0.32	0.32	0.87	0.24	0.63	0.34
ICC (consistency)	0.73	0.64	0.47	0.86	0.46	0.78	0.60
Dice	0.64	0.50	0.62	0.60	0.43	0.76	0.44

Table 8.5: Inter-rater Dice metric and ICC agreement and consistency measures

From figure 8.10 and table 8.5 Dice metric values ranged from 0.43 for the fimbria to 0.76 for the entorhinal cortex. As would be expected the Dice metric measures for inter-rater agreement were poorer for all subfields than the intra-rater Dice metric values. Similarly, the ICC (agreement) statistics were also considerably worse for the inter-rater case than for the intra-rater case. In general the inter-rater ICC (agreement) values were disappointing, with the only subfield that could be deemed to have been recreated successfully was the subiculum with an ICC (agreement) of 0.87, though even for this subfield the inter-rater Dice was still quite poor at 0.60. Perhaps the most disappointing result was the ICC (agreement) value of 0.34 for the CA4/dentate subfield. This was the primary subfield region of interest and it was anticipated that this would have been one of the 'easier' regions to achieve conformance across raters, though this was not reflected in the ICC (agreement) results. The inter-rater Dice of 0.64 for the CA4/dentate offers some reason for optimism where given further refinements to the protocol and training of the raters, better inter-rater agreement may be achievable.

8.6.9 Power calculations for corrected subfield volumes

Having established the mean volume and standard deviation for the different subfields, power calculations were then done to estimate how many subjects, given an estimated group difference and standard deviations, would be necessary to obtain results of sufficient power. For the purposes of this discussion we focused on two scenarios, firstly the number of subjects needed to detect a subfield volume difference between two groups of 10% with a power of 0.9. Secondly, for 17 subjects in each of two groups we aimed to estimate the volume difference for each subfield that would be necessary to obtain a study power value of 0.9. In order to cover the most general case, the power calculation was done by assuming the groups under test were not equal, and that no preference to the direction of the volume difference between

the groups existed, i.e. there was uncertainty as to whether a volume increase or decrease was expected.

8.6.9.1 CA4/dentate power calculation

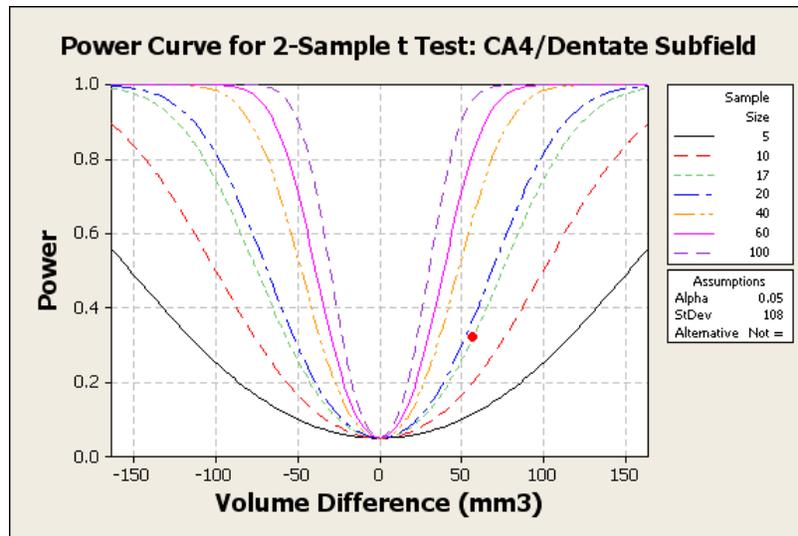


Figure 8.11: Power calculation for CA4/dentate hippocampal subfield

For the CA4/dentate subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.32. If the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 77 individuals in each group. See figure 8.11 above.

As discussed above the power calculation plots were presented for the case where there was uncertainty over whether a volume decrease or increase was expected between the groups. If we consider, for example, that we expect a CA4/dentate subfield volume decrease in a particular group when compared to another, the resultant power calculation is shown in figure 8.12 below.

For the CA4/dentate subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.32. If the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 63 individuals in each group.

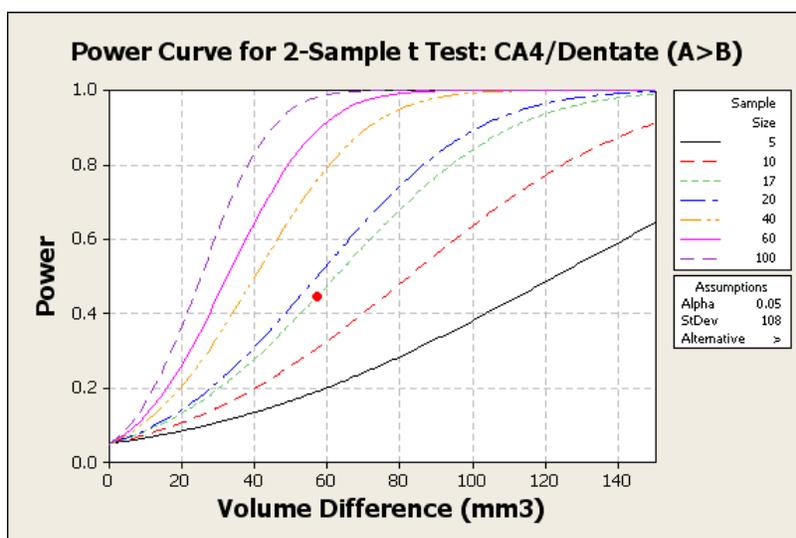


Figure 8.12: Power calculation for the CA4/dentate subfield

Thus, comparing the power calculation for the directional and non-directional cases, the case where a clear directional hypothesis was made improves the power of the design by between 5-10%. While this was a modest increase it serves to illustrate the importance of defining a clear hypothesis.

8.6.9.2 WM layer power calculation

For the WM layer subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.33, see figure 8.13 above. If the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 66 individuals in each group.

8.6.9.3 CA1/CA2/CA3 power calculation

For the CA1/CA2/CA3 subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.73, see figure 8.14 on the previous page. If the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 26 individuals in each group.

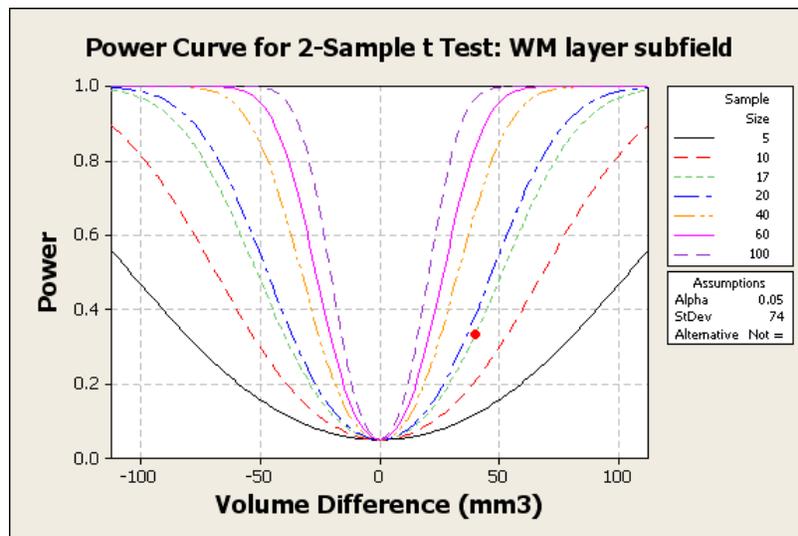


Figure 8.13: Power calculation for WM layer hippocampal subfield

8.6.9.4 Subiculum power calculation

For the subiculum subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.20, see figure 8.15 above. If the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 133 individuals in each group.

8.6.9.5 Fimbria power calculation

For the fimbria subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.19, as shown in figure 8.16 above. If the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 191 individuals in each group.

8.6.9.6 ERC power calculation

For the entorhinal cortex subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.18, see figure 8.17. If

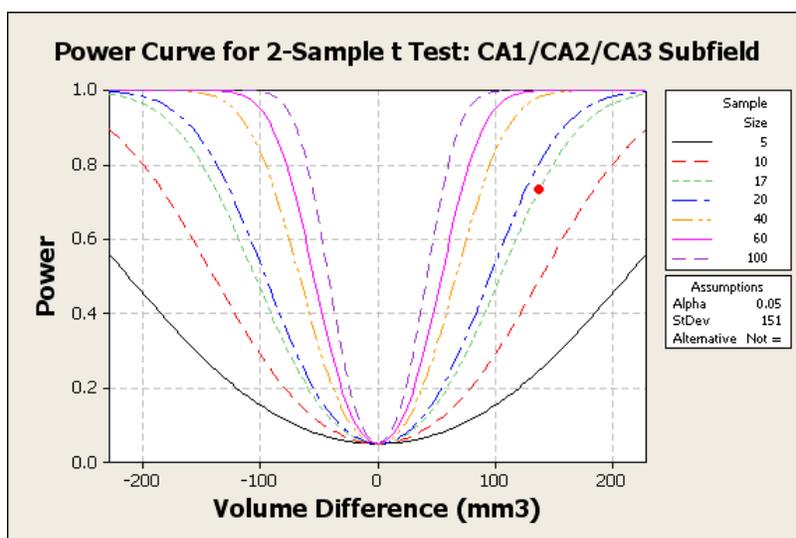


Figure 8.14: Power calculation for CA1, CA2,CA3 hippocampal subfield

the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 146 individuals in each group.

8.6.9.7 Alveus power calculation

For the alveus subfield, with a sample size of 17 in each group, for a 10% difference in volume between two study groups the power was determined to be 0.15, see figure 8.18 above. If the desired power for a study was 0.9, to elicit the same 10% volume difference between the groups would require approximately 198 individuals in each group.

8.7 Discussion

8.7.1 Mean Volumes and Normal ranges

For normal volunteers in the age range 21 to 59 (mean 36) the normal range based on the mean subfield volumes (mm^3) and $\pm 95\%$ confidence intervals were shown in table 8.1. Based upon the coefficient of variation (CoV) of each of the subfield measurements, i.e. the relative errors involved, the smallest CoV's, in order of increasing CoV, were for the CA1/CA2/CA3, WM layer and CA4/dentate with CoV's of less than 20%. Other subfields such as the subiculum, entorhinal cortex, fimbria and alveus had higher CoV's in the range 23% -33% implying that

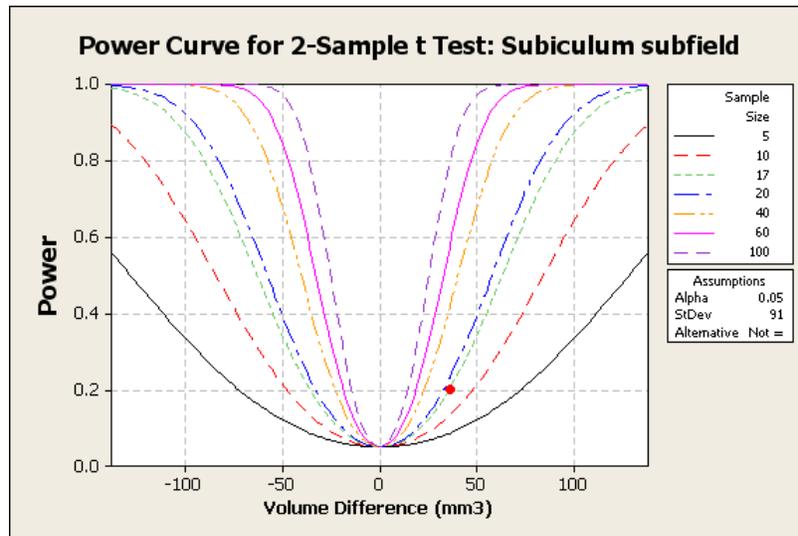


Figure 8.15: Power calculation for subiculum hippocampal subfield

the variance in the measurement of these structures was larger as a proportion of their mean compared to the CA1/CA2/CA3, WM layer and CA4/dentate.

The recent work using a 4T MRI system by Mueller et al stated values for hippocampal subfield volumes and standard deviations (mm³) in a normal control population as: ERC = 191 ± 54 ; subiculum = 191 ± 38 ; 326 ± 48 and CA3-dentate = 226 ± 41 [256]. There were some differences between the protocol implemented by the author of this thesis and Mueller et al, this may account for large portions of the error between the results. For example, the ERC values as measured by Mueller et al were less than a third of the values measured in this study. It was likely that this was mainly due to the number of slices on which the ERC itself was outlined. Throughout this thesis, the ERC was outlined in such a way as to depend on anatomical landmarks, rather than using slice position, this was intuitively a better approach.

It was extremely difficult to compare volume measurements across studies because of differences in protocols. This may be because protocol descriptions are incomplete or lacking in detail. A common observation when reviewing the literature was for images of hippocampal segmentations to be presented for the body of the hippocampus but not for head and tail. In fact, the segmentation protocols for studies such as Yushkevich et al [375], do not attempt to distinguish subfields in the hippocampal head or tail, these regions are simply defined as the hippocampal head and tail. This was perhaps because the aim of this work was to develop an automated algorithm, and therefore it was worth only segmenting the subfields in the more distinct hippocampal body. In any case, if the subfield volumes are not extended into these

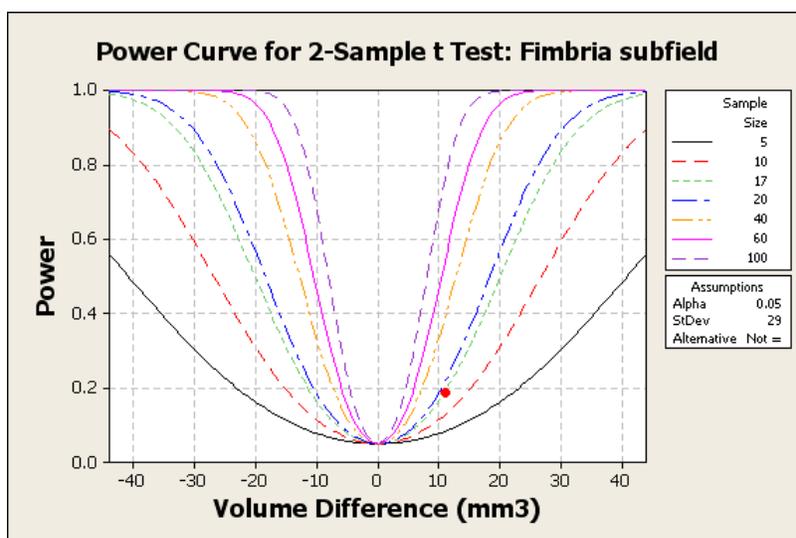


Figure 8.16: Power calculation for fimbria hippocampal subfield

regions then the subfield volumes measured in studies will be underestimated when compared to studies where subfield regions have been extended into the head and tail regions as has been the case in this thesis. On the other side of the argument of course is the fact that as one attempts to delineate subfield volumes into these more difficult regions then the error in these measurements will likely increase, that is, while the accuracy of the subfield volume measures may increase, the precision of the measurements was likely to decrease.

Recalling also that there are two fundamentally different approaches to assessing subfield differences, one being the method where, typically on coronal T2-weighted FSE images, subfield locations are estimated by anatomical landmarks and intensity values, and the other, where typically T1-weighted 3D images are used, surface models are applied and inferences on subfield location are made based on landmarks around these surfaces. Thus these approaches partition the hippocampus into its subfields in different ways.

The early works on subfield volume estimation by Csernansky et al [98] and then from the same group by Wang et al [363], cited deviations from the mean for regions of volume difference on the hippocampal surface rather than in distinct subfields. This being where the hippocampus was partitioned into three zones (lateral zone \equiv CA1, inferior-medial zone \equiv subiculum and superior zone \equiv CA2/CA3/CA4/dentate). While this method can be used to compare the volume differences on the surface of groups, it does not give an indication of the underlying absolute subfield volume. A similar approach, also assessing differences between groups in hippocampal surface volumes was performed by Apostolova et al [13]. As the underlying methodology was different the results of our study cannot be compared to the

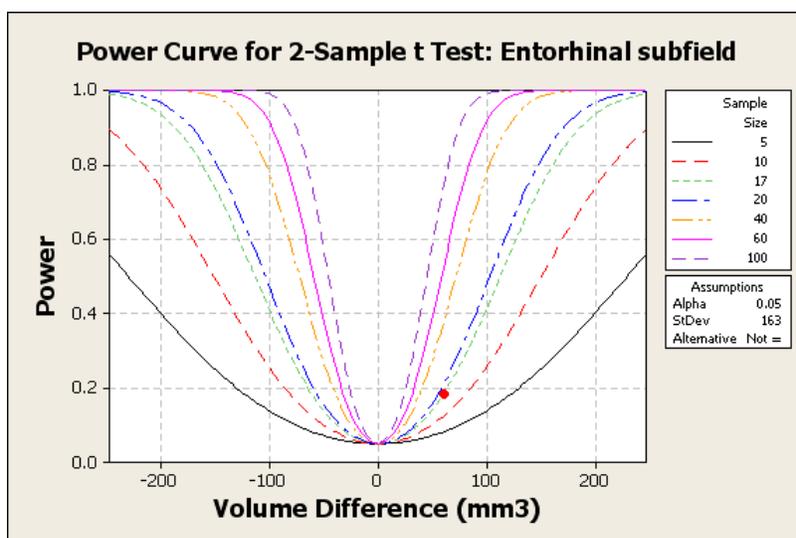


Figure 8.17: Power calculation for entorhinal cortex subfield

results of these works.

Similarly to the studies described above, the approach by Malykhin et al [212], using a 1.5T MRI system and T1-weighted 3D images, was to segment the hippocampal into its head, body and tail. Once again, because of the differences in methodology it was difficult to compare results from this study with the results of our work. Malykhin et al revisited these ideas more recently and by using a 4.7T and T2-weighted FSE images the hippocampal head, body and tail were further subdivided into the dentate, subiculum and CA1/CA2/CA3 subfields [214]. Subfield volumes normalised for ICV were determined. However, again the regions differ from those segmented in our study and thus it was difficult to compare volumes directly.

The first study to have a methodology similar to that performed in this study was the work by Mueller et al [251]. Indeed Mueller's subsequent works typically refer back to this seminal paper describing the reliability of the measurements [255, 258, 253]. Mueller's work was done using a 4T scanner and all volumes are cited after correction for ICV i.e. the raw volumes themselves were not published. Furthermore, the protocol applied by Mueller et al does not segment subfields in the hippocampal head and tail and therefore there will be substantial discrepancies between the volumes produced in these studies from those produced in the study presented here.

Other studies published more recently where subfield volumes have been quantified include Wang et al [365] where subfields have been assessed in patients with PTSD, Yassa et al [373] where the aim was to improve localisation of fMRI activation within the hippocampus of patients with Mild Cognitive Impairment (MCI). Neylan et al [263] assessed CA3/dentate

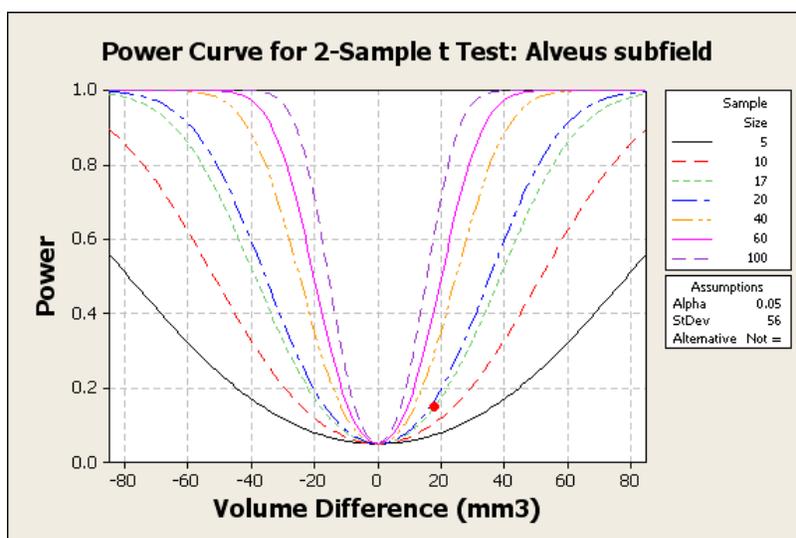


Figure 8.18: Power calculation for alveus hippocampal subfield

volumes based on the methodology described previously by Wang et al [365] in patients with insomnia. As illustrated by these studies, as some groups of researchers have now developed and assessed the reliability of their protocols for assessing the subfield volumes, they are beginning now to have greater confidence in asking research questions about hippocampal subfields in various diseases.

What was noticeable throughout the various studies conducted to date was that in this still emerging field with respect to the broader topic of hippocampal volumes, the emphasis has been on assessing the methodology, reliability, automation, and differences in surface volumes from a group mean. The pursuit of absolute volumes of subfield regions does not feature heavily. Thus the emphasis very much seems to be on reliability and reproducibility rather than accuracy. Indeed, few would lay claim to have an automated method of measuring the whole hippocampal volume with sufficient accuracy and reliability that, for example might be used routinely in clinical practice or in research where one would be sufficiently confident as to not have to manually check and edit the automatically produced segmentations. Therefore, in many respects it is understandable why reliably delineating or segmenting hippocampal subfields still has a long way to go and many improvements still have to be made. Indeed, greater issues face subfield segmentation than currently face whole hippocampal segmentation, with the fundamental issue being the fact that many of the true subfields and subfield boundaries themselves, for example the CA1-CA2 boundary, cannot be visualised. However, until then, one must aim to produce as clear and reproducible protocols as possible.

8.7.2 Subfield volumes vs. ICV

In general subfield volumes expressed modest, little or no correlation with ICV. Subfields such as the CA4/dentate, entorhinal cortex and subiculum expressed modest correlations, with the CA1/CA2/CA3, fimbria, WM layer and alveus having little or no correlation with ICV.

Pearson correlations were obtained to determine whether or not a linear correlation existed between each subfield volumes and the ICV. This was necessary to determine whether or not it would be most appropriate to correct subfield volumes for ICV when presenting normal ranges and for subsequent analysis in comparing subfield volumes across different subject groups. The strongest correlation of any of the subfield regions with ICV was the right CA4/dentate region, with an r -value of 0.54 ($p=0.03$). In the regression model this equated to ICV accounting for 24.7% (R^2 -adjusted value) of the variance in the volume of the R CA4/dentate region. Several other subfield volumes exhibited modest correlations ($r>0.2$) with ICV, including the L CA4/dentate, L/R subiculum, L fimbria and L/R entorhinal cortex, though none of these correlations were deemed statistically significant ($p<0.05$). The remaining subfield volumes exhibited little or no correlation with ICV (r -value <0.2). For both the modest and low correlation subfield volume vs. ICV cases the corresponding R^2 -adjusted value were all very low ($<3\%$) as one would expect given the low Pearson r -values.

As has been shown in chapters 2 and 5 of this thesis and in some studies before there is a clear correlation between whole hippocampal volume and ICV. Therefore, one would expect intuitively that it follows that there must be a correlation between hippocampal subfields and ICV, if not across all subfield regions then preferentially within specific subfields. Despite this natural argument correlations between subfield volumes and ICV have been more difficult to confirm. From the work done in this study the largest correlations between volume and ICV were determined to be in the CA4/dentate, entorhinal and subiculum region. After the CA1/CA2/CA3 subfield these subfields are amongst the larger subfields that were measured and thus perhaps the fact these regions all show some correlation with age would, if summed up, account for the correlation observed between ICV and the whole hippocampus.

However, we are not unique in struggling to establish clear relationships between subfield volumes and ICV. Mueller et al [251] while investigating age effects on subfield volumes was unable to show any significant correlation between subfield volumes and ICV. Despite this the subfield volumes were still normalised to account for ICV and plotted against age as would be in keeping with the convention used for whole hippocampal volume analysis. This

correction for ICV also persists through the rest of Mueller et al's work. Possible explanations why Mueller was unable to show any relationship with subfield volume might be because the subfield regions such as the hippocampal head and tail which were not segmented may be more preferentially related to ICV than the regions within the body of the hippocampus. Or perhaps more likely, volume changes are distributed across all subfields, perhaps to a greater and lesser extent, but for these lesser regions the subfield measurements themselves are not being made with sufficient accuracy to elicit the expected correlation with ICV. Regardless, as the topic of hippocampal subfield measurements is still progressing we feel it is helpful to express both the raw volumes measured as well as the ICV corrected volumes. Expressing raw volumes is also useful as different authors may choose to correct subfield volumes for ICV in different ways and thus inspecting raw volumes makes comparing protocols and methodologies more accessible.

8.7.3 Subfield volumes vs. age

In general subfield volumes expressed modest negative, little or no correlation with age.

Pearson correlations were obtained to determine whether or not a linear correlation existed between each subfield volume and subject age. This was necessary to determine whether or not it would be most appropriate to correct subfield volumes for age when presenting normal ranges and for subsequent analysis in comparing subfield volumes across different subject groups. The strongest linear correlations observed were for the right WM layer ($r = -0.4$, $p = 0.11$); right CA1, CA2 and CA3 ($r = -0.42$, $p = 0.10$) and the right alveus ($r = -0.48$, $p = 0.05$). Only the right alveus volume was deemed to have a statistically significant correlation with age. Several other modest correlations existed ($0.2 < r < 0.4$) for the left CA4/dentate; left fimbria and right entorhinal cortex but none were deemed statistically significant. The remaining subfield volumes typically expressed weak negative correlations with age, none of which were deemed statistically significant. In a regression model for subfield volumes with age as a predictor, the R^2 -adjusted value was greatest for the right alveus at 18.2%; right CA1, CA2, CA3 at 11.7% and right WM layer at 10.5%. Both age and ICV explained little (<3%) of the variance for the remaining subfield volumes.

As for the relationship with ICV and hippocampal volume, it is well known that increasing age is correlated with decreasing hippocampal volume, this is discussed in chapters 2 and 5. Estimates for the strength of the correlation between age and hippocampal volume may vary and more recent studies have shown that the relationship between these factors was not linear but changes with age where a more rapid decline in volume was observed in subjects of

60 years or older [297]. As the correlation with aging and hippocampal volumes in younger adults may reasonably be assumed to be linear then this may not have a great impact when assessing and comparing hippocampal volumes in diseases such as psychosis. However, non-linear, accelerating changes in older adults may potentially have massive implications when trying to compare patients with diseases of pathological aging such as Alzheimer's. Therefore, given the complexity involved, particularly in older age groups, with whole hippocampal volumes it is likely that the relationship with subfield volumes and age is going to be equally if not more complex. A comprehensive assessment of how subfield volumes differ with age, particularly in older age groups should lead to an improved understanding of the normal and pathological age affects on the hippocampus and could potentially lead to more accurate and precise measurements being made to aid the detection, monitoring and treatment of such diseases. Given that people are living longer in most of the developed world, the incidence of diseases of pathological aging are likely to increase and therefore public, government and scientific interest to improve the management and treatment of these patients is also likely to increase. Indeed, work to assess hippocampal subfields volumes in these areas is already underway [13, 259, 12].

The results from our work showed that the volume of several subfields being correlated with age, the strongest of these correlations being that of the right WM layer, right CA1/CA2/CA3 and right alveus. In contrast, the work of Mueller et al [251] assessing age effects on subfield volumes only established a significant relationship between the CA1 subfield and age. As discussed above in the Mueller et al study ICV did not explain any of the variance of subfield volume where only age explained an appreciable portion of the variance in the CA1 subfield volume. Interestingly, Mueller et al noted an increased volume loss in older subjects (> 60 years old) similar to that which was observed by Raz et al [297], though their attempt to fit a quadratic term in the regression model failed to improve the quality of the fit. In our study subfield volumes were only delineated in subjects between the ages of 21 and 59 and therefore it was not possible to investigate the relationship between the effects of aging on subfield volumes in a more aged population.

8.7.3.1 Left vs. right subfield volumes

There were no statistically significant differences between right and left hippocampal subfield volumes. Volumes for left and right whole hippocampi and subfield volumes are often not treated separately [251]. In some instances this may be perfectly appropriate. However, as discussed in chapter 5, there is a trend for normal subjects to have slightly smaller left hippocampi, and for the purposes of our work here we are assessing normal subjects, future

work will aim to explore patients with psychiatric disorders. In these circumstances one may wish to assess the relationships between subfield volumes and cognitive scores. As brain function in some circumstances may be lateralised it would be most appropriate to retain the distinction between the right and left subfield volumes to allow the most specific hypotheses to be addressed.

8.7.4 Male vs. Female

The results for comparing the female and male raw hippocampal subfield volumes are that there was, with one exception, no statistically significant difference between the volumes. The exception was the entorhinal cortex where there was a general trend for the female raw subfield volumes to be slightly smaller, but given the amount of variance present in the measurements, and the modest number of subjects sampled, these differences were not statistically significant.

After correcting for ICV the female and male subfield volumes were compared. With the exception of the subiculum subfield the female ICV corrected volumes tended to be slightly larger than the male volumes. However, none of these corrected subfield volumes were statistically significant.

Given the small number of subjects involved in this analysis, a full treatment comparing the male and female subfield volumes was not possible.

8.7.5 Intra-rater agreement and Dice metric

The ICC agreement measures for the various subfields were calculated and the values are shown in table 8.4. As stated above, due to differing protocols it was difficult to exactly compare subfield volumes across studies, the same can be said for comparing the ICC agreement values. The early work on subfield volume difference estimation by Csernansky et al [98] and then from the same group of Wang et al [363] cited a comparison of 80% between manual outlining and their automated transformation with ICC values comparing manually outlined zones and the template zones were for the three zones: lateral zone (CA1) = 0.97, inferior-medial zone (subiculum) = 0.97, and superior zone (CA2/CA3/CA4/dentate) = 0.90. These results compare the ICC for different methodologies i.e. manually vs. template and do not therefore compare intra-rater values and thus we cannot compare the ICC agreement values for the study described in this chapter with these results.

Malykhin et al [212] also used T1-weighted 3D images to segment the hippocampal volumes into its head, body and tail. As in this instance the segmentation was performed and repeated by two raters. Measurements of intra and inter-rater variability were obtained for these three segmented regions. Intra-rater values were: hippocampal head: 0.92; body 0.93 and tail 0.88. While these values are higher (better) than those obtained in our study, it is important to remember that the regions assessed in Malykhin et al's study are larger and more homogeneous. Therefore, while the results are not directly comparable they are informative as to what levels of intra-rater agreement one should be aiming to achieve for hippocampal subfields. Malykhin et al [214] revisited hippocampal segmentation more recently using 4.7T MRI and T2-weighted FSE images and refined their approach from assessing the hippocampal head, body and tail to assessing subfields. By this method they achieved very good intra-rater ICC's for the left and right (L/R) hippocampal subfield volumes of: CA1/CA2/CA3 (L/R) = 0.95/0.96; DG (L/R) = 0.96, 0.96; and subiculum (L/R) = 0.97/0.98.

The first studies to have a methodology similar to that performed in this chapter were the work by Mueller et al [251]. The intra-rater ICC for the hippocampal subfields were as follows, for rater 1: left/ right ERC = 0.83/1, left/right subiculum = 0.99/0.99, left/right CA1 = 0.99/0.99, left/right CA2 = 0.96/0.77 and left/right CA3/CA4/dentate = 0.98/0.99. For rater 2: left/ right ERC = 0.67/0.98, left/right subiculum = 0.91/0.66, left/right CA1 = 0.98/0.87, left/right CA2 = 0.81/0.78 and left/right CA3/CA4/dentate = 0.86/0.96.

The intra-rater of Mueller et al are generally good (>0.80) with rater 2 performing slightly worse with the occasional (left ERC, right subiculum) ICC < 0.7 [251]. Such high levels of agreement may be attributable to several factors. The protocol itself was clearly defined within the body of the hippocampus. The fact that both raters have good intra-rater results indicates the second rater adopted the protocol well. On the downside, for the most part the subfields within the hippocampal head and tail were not segmented. From one perspective this may be considered to raise reliability in that more precise measurements can be made but it adds difficult considerations such as: where does the hippocampal head end and the body start, and similarly, where does the body end and the tail start? Neglecting to segment these regions, while increasing precision of the measurement inevitably reduces the accuracy of the subfield volume. It might be argued that any such discrepancy in accuracy would be likely to express itself as a systematic error, however, this would imply that one has evidence to suggest that volumes from the body correlate strongly with volumes in the head and tail, while this might be the case in normal subjects, it would be less likely to argue this point in diseased states where more localised volume differences, preferential to specific subfields may be affected.

To summarise, while the intra-rater ICC of Muller et al's work are better than demonstrated in our study, this may partly be because a simpler protocol focusing on only the hippocampal body has been applied. Despite that which has been discussed above it would be unfair to suggest that Mueller's better intra-rater ICC values are simply down to how the hippocampal subfields were defined. Mueller et al's acquisition protocol used a 4T scanner, it is likely this had some additional benefit in image quality, furthermore, Mueller et al's used slice thicknesses of 2mm without a gap, compared to the 4mm slice thickness and 1mm slice gap used in the study in this thesis. This would significantly improve the segmentation on each individual slice, mainly by reducing partial volume effects. Of course, applying thinner slices means more slices and therefore longer scan times would be required to achieve the same coverage of the hippocampus. Similarly, for more slices, the time taken to manually segment the subfields will also increase.

In the recent work using a 4T Bruker system by Yushkevich et al developed a 'Nearly automatic segmentation of hippocampal subfields in in-vivo focal T2-weighted MRI' [375]. Intra-rater ICC for various subfields were determined. The approach adapted by Yushkevich et al was more akin to that which had previously been done by Mueller et al with the main body of the hippocampus being segmented into the various subfields, with the head and tail of the hippocampus being distinct regions. The ERC, parahippocampal gyrus and subiculum were also segmented. The intra-rater ICC results for these regions were: CA1 = 0.92, CA2 = 0.73, CA3 = 0.81, DG = 0.93, head = 0.93, tail = 0.93, subiculum = 0.88, ERC = 0.86, PHG = 0.86. These ICC agreement values are high, with all of them being greater than 0.80. For the most part these values are comparable or better than the ICC values estimate for our work. However, again, it is worth highlighting the fact that the protocols differ in that the head and tail were treated as single regions and were not segmented into subfields. In addition, a point which improves the complexion of the results is that in the work by Yushkevich et al [375] the thin layers and small subfield regions such as the fimbria, alveus and WM layer were not segmented but were either ignored, or absorbed into a neighbouring subfield regions. This will artificially boost the volume of the subfield volumes measured and while this may lead to more reproducible results, this may be at the expense of anatomical accuracy. That being said, Yushkevich et al's work is laudable, and the values achieved are excellent for what is a 'nearly' automated method.

To compliment the intra-rater ICC agreement measures, the intra-rater Dice metric was also calculated. The largest Dice measures were observed for the CA4/dentate (L = 0.83/R = 0.79), entorhinal cortex (L = 0.81/R = 0.76) and CA1/CA2/CA3 (L = 0.77/R = 0.74), with the subiculum (L = 0.70/R = 0.64) , fimbria (L = 0.72/R = 0.60) , alveus (L = 0.60/R =

0.58) and WM layer ($L = 0.60/R = 0.56$) having poorer Dice measures.

Dice measures are often overlooked in the literature yet they provide an important metric over and above a simple comparison between volumes. The Dice may be calculated simply using the ITKsnap program and convert3d plug-in (<http://www.ITKsnap.org/pmwiki/pmwiki.php>) [377]. The software was developed by Paul Yushkevich who amongst other things has published works assessing the hippocampus and its subfields, including recent works assessing the ex-vivo hippocampus at ultra-high field strengths [376]. Even more recently Yushkevich et al have developed a 'nearly' automatic segmentation of hippocampal subfields in in-vivo focal T2-weighted MRI [375]. In this study Yushkevich et al also use the Dice metric, not to compare intra or inter-rater reliability of manual outlining, but to compare the manual and automated outlining methods themselves i.e. treated as different raters. In the ex-vivo study, where a 9.4T Varian magnet was used, the following intra operator Dice metrics were determined: CA1 = 0.88, CA2/CA3 = 0.70, dentate gyrus (Hilus) = 0.76, dentate Gyrus (Stratum Moleculare (SM)) = 0.66, stratum radiatum (SR) + stratum lacunosum-moleculare (SLM) + vestigial hippocampal sulcus (VHS) = 0.56, Stratum oriens (SO) + pyramidal cell layer (PCL) = 0.89, dentate gyrus (DG) = 0.83, hippocampus (H) = 0.92. While some of the subfields were defined in a slightly different manner one can see that in some cases the Dice metrics between this study and our own study are not vastly different. More generally, the ex-vivo segmentations have higher Dice metrics as one would expect but even at this ultra-high field strength the Dice metrics are lowest (worst) when manually outlining the thin layer subfields such as the combined subfield of the stratum radiatum (SR) + stratum lacunosum-moleculare (SLM) + vestigial hippocampal sulcus (VHS). In this thesis poorer Dice metrics were also recorded for the white matter layer between the CA4/dentate and the CA1/CA2/CA3 and these subfields remain a challenge to be delineated reliably.

8.7.6 Inter-rater agreement and Dice metric

Table 8.5 show the Inter-rater ICC measures of agreement and Dice metric. The ICC (agreement) statistics were considerably worse for the inter-rater case than for the intra-rater case with values ranging from 0.24 for the fimbria to 0.87 for the subiculum. The Dice metric values looked slightly better on average than the ICC (agreement) measures with values ranging from 0.43 for the fimbria to 0.76 for the entorhinal cortex. As would be expected the Dice metric measures for inter-rater agreement were poorer for all subfields than the intra-rater Dice metric values. In general the inter-rater ICC (agreement) values were disappointing, with the only subfield that could be deemed to have been recreated successfully was the subiculum with an ICC (agreement) of 0.87, though even for this subfield the inter-rater

Dice was still quite poor at 0.60. Perhaps the most disappointing result was the ICC (agreement) value of 0.34 for the CA4/dentate subfield. This was the primary subfield region of interest and it was anticipated that this would have been one of the 'easier' regions to achieve conformance across raters, though this was not reflected in the ICC (agreement) results. The inter-rater Dice of 0.64 for the CA4/dentate offers some reason for optimism where given further refinements to the protocol and training of the raters, better inter-rater agreement may be achievable.

Mueller et al using 4T MRI and a T2-weighted FSE acquisition reported inter-rater ICC values for the: left/ right ERC = 0.61/0.88, left/right subiculum = 0.78/0.39, left/right CA1 = 0.93/0.91, left/right CA2 = 0.86/0.68 and left/right CA3/CA4/dentate = 0.88/0.89 [251]. While these values are superior to the values reported in this chapter, it is worth reiterating the point that the values were obtained for manual segmentation within only the body of the hippocampus and not the head and tail. Also, Mueller et al did not include the fimbria and alveus in their segmentation protocol. Mueller et al did have some success with their protocol demonstrating for some subfields inter-rater agreement of 0.80 and above, with some values for the CA1 subfield of greater than 0.90. However, there was evidence that they too struggled with the manual segmentation of some subfields with values of 0.68, 0.61 and 0.39 for the right CA2, left ERC and right subiculum respectively.

A further technical point on the work performed by Mueller et al is that that it was not clear exactly how the ICC values were calculated, that is, the software used was not cited nor was the underlying ICC method referenced. They do state that a two-way ANOVA with random effects was used, but this could either be a measure of agreement or consistency and as we have shown from table 8.5 these results can be quite different. Therefore it is difficult to have absolute confidence in comparing the results of this chapter with that of Mueller et al's study.

A further study which reported inter-rater agreement measures for subfield regions was performed by Malykhin et al, though they only partitioned the hippocampus into its head, body and tail [212]. They did this using a 1.5T magnet with a T1-weighted 3D acquisition. They reported inter-rater ICC values of 0.95 for the hippocampal head, 0.83 for the hippocampal body and 0.95 for the hippocampal tail. These are very high levels of agreement. These are superior to what was determined by the work presented in this chapter but as the regions measured differ so significantly it is difficult to compare the results. However, the improved resolution of the T1-weighted acquisition sequence by minimising partial volume effects between slices might be part of the reason the ICC agreement values of Malykhin et al were so high. Malykhin et al state which software and method was used to calculate ICC agreement

and cite the original reference, they make it clear that it was ICC agreement that was used and not the ICC consistency measures.

For the purposes of research conducted in this chapter, the expectation might be that two operators could share a dataset, perhaps manually segmenting half of the total number of subjects each. Ideally then the data would be combined and treated as if it had been manually segmented by a single rater, this is what is known as full exchangeability of the data. For this to be done, a high level of confidence in the agreement of the manual segmentations of the raters must be obtained, this is what the ICC (agreement) measures and here the values were lower than a level that would be considered satisfactory.

For the assessment of inter-rater agreement, the ICC (consistency) measure was also calculated and the values are shown in table 8.5. Strictly speaking this is not a statistic that is appropriate or necessary to meet the aim of the inter-rater assessment presented here i.e. the aim of determining whether or not the data from the different raters are exchangeable. If the aim had been to determine whether the subfield volumes were measured consistently across raters i.e. systematic differences in performance were not important, then this would have been the correct statistic to use. From table 8.5 the ICC (consistency) measure can be seen to be as much twice as high as the ICC (agreement) measure. That is, the consistency for subfield segmentation between raters was much better than the absolute agreement. The point of discussing these statistics together is to emphasize the importance of choosing the correct ICC statistic when handling data such as this, and to make clear which statistic has been used when presenting the data. This is not always the case in the peer review literature.

Taking the ICC (agreement) and Dice metric measures together the results were generally poor. Further work would be necessary to ensure confidence in the manual segmentation protocol, this may be achievable by refinements to the manual segmentation protocol or from greater effort in training the manual segmentation operators to improve inter-rater agreement. Of course, any improvements which can be made on a more fundamental level to the underlying methodology such as from improved data acquisition resolution or contrast and improvements in post-processing are likely to improve both the intra and inter rater agreement levels. Indeed, such improvements to the methodology are more likely to facilitate larger improvements to inter and intra rater agreement than perhaps the incremental progress that might be achieved from refinements to the existing manual segmentation protocol.

8.7.7 Subfield power and sample size calculations and their implications

Power calculations are rarely discussed explicitly within the body of literature covering brain imaging structures and subfield segmentation. Such a treatment of the data can be highly informative and therefore it is somewhat disappointing this metric is not published more often. Perhaps the reason power calculations are not often cited is that provided a paper states the variance in the measurement and the number of subjects studied, the reader can then derive the study power. Short of obtaining pilot data for oneself this would often be the approach one might use to establish estimates of study power that are necessary for grant applications i.e. examine the existing literature value for the variance involved from previous studies with similar methodologies. The latter point of ensuring that a similar methodology is used is critical as the variance and therefore power of a study are inextricably linked to the way in which the study was performed. This is perhaps a further reason why many publications do not discuss study power as they might see this value as something which is unique to their own methodological approach and study site. One other, less worthy reason the power associated with studies is often not cited is that the power may often be lower than the authors initial expectation. There may be many genuine reasons for this, for example, the expectation for the measurable difference between two study groups was not realised, however, the author of this thesis would argue here that highlighting poor study power is an important part of informing the reader of the limitations of a scientific work. A final, more charitable, point is that power is often not cited, regardless of whether the power of the study is high or low, so thus perhaps it is simply a metric many do not deem necessary to publish by deciding that it would not be of interest to the reader.

For the purpose of this study and in particular the measurement of hippocampal subfields, the associated power for each subfield measure was determined and will be discussed. The author had no great expectation of what the resultant power would or should be. The purpose of considering the study power was simply to quantify what its value was for each subfield. There are several reasons why in this context a detailed consideration of power was important. To date, few studies of hippocampal subfields have considered the power in their measurements, thus a fuller treatment would be informative to the community in general. More locally, it was desirable to establish the power in the applied methodology such that further work may be informed by these results. This was necessary to, for example, determine how many subjects would be necessary to elicit 'say' a 10% volume difference between groups with 'good' power. Given the manually intensive nature of the work with subfields taking ~ 30mins to segment, clearly if one believed 100 subjects were necessary in each group to

detect a statistically significant difference then one ought to consider the 200 hours of manual labour necessary to complete the work. Furthermore, we wish to establish how the power may differ across the different hippocampal subfields, perhaps it would be the case that some subfields will have better power than others.

The results of power calculation estimations based on 17 subjects in a given study group, where a 10% volume difference between the groups was anticipated, are shown in table 8.6 on the following page.

Power calculation on Hippocampal Subfields based on 17 subjects, where 10% volume differences are anticipated					
Subfield	Sample size	Mean	Std Dev	Volume Difference	Power
CA4/Dentate	17	530	108	10%	0.32
WM layer	17	356	74	10%	0.33
CA1, CA2, CA3	17	1278	151	10%	0.73
Subiculum	17	338	91	10%	0.20
Fimbria	17	99	29	10%	0.19
Entorhinal cortex	17	596	163	10%	0.18
Alveus	17	166	56	10%	0.15

Table 8.6: Power calculation estimations for hippocampal subfields based on 17 subjects with the anticipation of a 10% volume difference between the groups

The power calculation estimations clearly show that with 17 subjects there was little likelihood that one could have confidence in detecting a 10% subfield volume difference between two groups. With the exception of the CA1/CA2/CA3 subfield (power =0.73) all the power estimations were below 0.33 i.e. below 33%. It is worth pointing out that the fact the CA1/CA2/CA3 has the best estimation of power and that this was not a great surprise given that the CoV for this subfield was low as a result of the relatively low standard deviation compared to the mean segmented volume.

While it is worth pointing out that the power estimations for a relatively low number of subjects are generally poor, it is more constructive to discuss how many subjects would indeed be necessary to have a higher likelihood of having confidence in statistically significant subfield volume differences between groups. Thus the sample size results based on the desire to obtain a power of 0.9 for an estimated 10% volume difference are shown in table 8.7 below.

Sample size estimation based on desired study power of 0.9, where 10% volume differences are anticipated					
Subfield	Desired Power	Mean	Std Dev	Volume Difference	Sample size
CA4/Dentate	0.9	530	108	10%	77
WM layer	0.9	356	74	10%	66
CA1, CA2, CA3	0.9	1278	151	10%	26
Subiculum	0.9	338	91	10%	133
Fimbria	0.9	99	29	10%	191
Entorhinal cortex	0.9	596	163	10%	146
Alveus	0.9	166	56	10%	198

Table 8.7: Sample size estimation for hippocampal subfields based on the anticipation of a 10% volume difference where a study power of 0.9 is desired

Given the results in table 8.7, it was unsurprising that the subfield for which the fewest number of subjects would be required is the CA1/CA2/CA3, where 26 subjects in each group would achieve 90% power. Beyond that the WM layer with 66 subjects and CA4/dentate subfield with 77 subjects may be considered reasonable sample numbers to achieve this same, high level of power. However, for the remaining subfield regions where more than 133 subjects would be required in each group, this would present a significantly greater challenge to recruit, acquire, analyse and detect statistically significant volume differences in these regions. Note that for both the power calculation and sample size estimation that these values would have been slightly higher and lower respectively if the direction of the 10% group difference was known.

8.7.8 Further work and future advancements

As with all aspects of scientific development, improvements can occur on several fronts. In MRI structure and subfield segmentation the two main areas where developments are likely to occur are the data acquisition and image analysis. If we include MRI hardware developments as part of the acquisition process, clearly as demonstrated in some of the high and ultra-high field studies discussed above there were benefits from the greater signal-noise at higher field strengths.

In the area of pulse sequence development, the two main methodologies discussed in this chapter, one being the surface analysis of T1-weighted 3D images and the other being manual

outlining of subfields on T2-weighted FSE images, exploit different benefits from these two types of sequences. These are the superior three-plane resolution of the T1-weighted 3D images and the superior grey matter to white matter contrast of the T2-weighted FSE images for better internal hippocampal visualisation.

In an ideal world, having a T2-weighted 3D sequence with high in-plane resolution and optimal GM/WM contrast would be of great benefit. Such sequences have been developed but to the knowledge of the author at this time they remain to be assessed in the context of hippocampal subfield measurements. Further technical advancements might include exploiting parallel imaging techniques such that images can be acquired more quickly, or with better quality for the same length of scan time. This may occur as head coil technology progresses from 8 channel to 16, 32 and perhaps even 128 channels in the future. Moreover, the widespread introduction of compressed sensing methods into MRI pulse sequence will likely enable a further leap forward in reducing imaging time and/or improving image quality [200]. Of course, such developments often come as a trade-off and higher field strengths may present challenges with image susceptibility. Similarly, increasing the number of coil elements may increase the inhomogeneity or bias across the field of view, though to counter these issues one might expect that correction algorithms for both of these issues will advance in parallel.

As well as hardware developments, the post processing methods have a large part to play in measuring subfield volumes. Yushkevich et al recently published on a 'nearly automatic segmentation of hippocampal subfields', this is an impressive achievement and demonstrated good agreement (Dice) between manual and automatic segmentations for several of the hippocampal subfield [375], though as with Mueller et al's work, the subfield segmentations did not extent into the hippocampal head and tail. Future developments will likely aim to expand into these areas and to continue to improve and develop methodologies for segmenting the body of the hippocampus.

Discussing the need to segment subfields into the hippocampal head and tail leads nicely to highlighting what is probably the most fundamental issue with hippocampal subfield segmentation. That is, at present we are unable to visualise the true extent of the subfields and more specifically the boundaries between the subfields themselves. A significant improvement in resolution and in contrast would be required for this to be achievable in-vivo in a practically realistic timeframe. Ex-vivo studies [369, 77, 121] and work on formalin fixed post-mortem brains [376] show that clearer differentiation between the internal grey matter and white matter structures is possible, however, a method of distinguishing between cellular subfields such as CA1, CA2 and CA3 other than using macroscopic anatomical landmarks

still remains elusive. Perhaps integrating other MR techniques such as diffusion imaging might offer additional benefit in contrasting subfield regions as was done in the ex-vivo work by Shepherd et al [322]. Finally as mentioned in chapter 7, if in-vivo imaging markers are to be determined for fundamental, cellular processes such as neurogenesis much improvement in image quality will likely be necessary. Perhaps if correlates between imaging measures and these processes could be established through animal work then in-vivo studies might begin to be possible.

8.8 Conclusion

A novel and comprehensive protocol for the manual delineation of hippocampal subfields, throughout the entire length of the hippocampus has been presented. This is one of the few studies to date which has attempted to assess the absolute subfield volumes throughout the entire length of the hippocampus and to assess the intra and inter rater reproducibility of these measurements. This protocol was implemented in 17 healthy volunteers. The mean volume and normal ranges were determined for the hippocampal subfields and repeat intra and inter operator measures were made to assess the validity of the segmentation protocol.

The strength of the linear relationship between hippocampal subfield volumes and ICV was assessed and in general there was little or no correlation between these parameters. Correlations between subfield volumes and subject age were also assessed and generally, there were modest negative correlations. This would imply that there is a gradual decrease in subfield volumes with age. This mirrors that which has been observed for whole hippocampal volumes and was intuitively what was expected.

Of the 17 subjects where subfield volumes were determined repeat measures were made on 11 subjects by the first operator (the author of this thesis) to assess intra-rater reliability. A second operator also performed subfield measurements on 11 subjects and these results were compared to that of the first operator to assess inter-rater reliability. Intraclass correlation coefficients for absolute agreement and Dice metrics were calculated in both circumstances. Intra-rater ICC (agreement) and Dice measures were good for some subfields, namely: the CA4/dentate, CA1-CA3 and entorhinal cortex with values of approximately 0.80 or better. Other subfield segmentations agreed less well. In general the inter-rater results for both ICC (agreement) and Dice measures were poor. This implies that more work on refining the manual segmentation protocol method and training would be necessary before the protocol may be used across different raters with confidence.

Given that measuring hippocampal subfield volumes is a relatively new and novel type of measurement it was deemed necessary to perform an assessment of study power and sample size calculations. It was estimated that in order to achieve a study power of 90%, when detecting a 10% volume difference in the CA4/dentate subfield region, 77 subjects would be required in each group when a unpaired t-test design was being used to assess the data. Smaller sample numbers of 26 and 66 gave equivalent power for the CA1-CA3 and WM layer subfields respectively, as a result of the smaller variation measured for these regions. The GM subfields of the ERC and subiculum and WM subfields of the alveus and fimbria were found to require much greater numbers (> 130 subjects) to achieve equivalent power by using the same method.

Subfield volumes were also compared between male and females and between right and left hemispheres. In both instances, there were no significant differences between these groups of data.

The work performed in this chapter, i.e. the creation of protocol for manually segmenting the subfield volumes throughout the whole hippocampus and the assessment of the quality of these measures is believed to be of interest to the wider scientific community. Therefore, this work forms the basis for a scientific manuscript currently in preparation for publication to a peer reviewed journal.

Chapter 9

Final conclusions

9.1 Conclusions and original contributions of this thesis

The aim of this thesis was to use 3T MRI and the latest developments in image processing software to assess a range of measures of the hippocampus. This included measures of hippocampal volumes, metabolites, morphology and subfield volumes. These tools were used to assess several datasets. Chapters one to four of the thesis focused on developing the methods necessary to study two groups of twenty-one subjects each with distinct socioeconomic status i.e a more deprived group and a more affluent group. Chapter five was concerned with establishing normal ranges for hippocampal volumes in normal, healthy volunteers (thirty-seven subjects in total). Chapter six combined these two datasets to provide a quantitative assessment of incomplete hippocampal inversion in normal volunteers. In chapter seven a protocol for assessing hippocampal subfields was implemented on a small pilot dataset of five subjects before and after anti-inflammatory therapy. Finally, the subfield volumetry protocol introduced in chapter 7 was extended to include subfield volumes in the hippocampal head and tail. Seventeen normal, healthy subjects were assessed.

In chapter 2, a trend was observed for manually segmented left hippocampal volumes to be smaller in the more deprived socioeconomic group than in a more affluent group. Right hippocampal volumes also tended to be smaller but this difference was less than was observed in the left hippocampus. In both circumstances, volumes were compared while covarying for age and ICV. Post-hoc correlations between the inflammatory markers: IL-6 and cortisol and hippocampal volumes found no significant linear relationship between these measures. It was suggested that a larger scale, more refined assessment of inflammatory and metabolic

makers would be necessary to illicit the moderating and mediating factors between SES and its relationship to hippocampal volume.

From the data that was acquired sample sizes were also estimated for desired levels of study power, this will inform further work in this area. Based on the mean differences between the differing SES groups and the variance in the hippocampal volume measurements, it was estimated that in order to achieve a study power of 0.9, approximately 57 subjects would be required in each of the two study groups.

The time taken to manually segment hippocampal volumes was significant and could be restrictive to future work, particularly if large scale studies are to be performed. Therefore, a comprehensive comparison between the automated method of hippocampal volumetry and the manual segmentation procedure was done. The agreement between the manual and automated method was reasonable with an overlapping volume Dice metric of 0.81. These values were comparable to what has been presented in the literature for other automated methods as well as the FSL FIRST algorithm used in this thesis. Despite the modest success of the automated method it was observed that it can occasionally fail to give accurate results. Therefore it was recommended that in future, to maximise quality while minimising the time penalty for analysis, that automated segmentation should be performed, followed by a manual inspection and correction of the segmented volume. This approach was adopted in chapters 5 and 6 of this thesis to assess hippocampal volumes and malrotation in a normal group of healthy volunteers. Furthermore, regions of difference between the manual and automated methods were visualised on an average brain template to show areas of frequent error. Few studies have provided such insight but this is important as it highlights where the automated segmentation method most commonly fail such that future improvements may be implemented in a more focused and localised manner.

To compliment the work performed in chapter 2, in chapter 3 the VBM methodology was developed to assess whether or not grey matter volume differences existed between groups of poorer and more affluent subjects. This was a novel opportunity to assess brain volumes differences from two differing SES groups from a community based sample.

Two VBM approaches were implemented: optimised VBM (the older, more traditional approach) and DARTEL-VBM which has been reported to improve image normalisation. The DARTEL-VBM method was found to be an improvement over the optimised VBM method. Regions suggestive of grey matter volume reductions were found in both the left and right hippocampi of the poorer group when compared to the more affluent group. The GLM model to compare the groups included age, ICV and alcohol consumption as covariates.

It has been shown previously that increases in inflammatory markers such as IL-6 and cortisol correlate with decreasing hippocampal volumes in healthy volunteers [217, 291]. This had been posited as supporting evidence for hippocampal volume reductions being the result of chronic environmental stress factors. Both IL-6 and cortisol measures were included in a VBM model of multiple regression but these factors were found not to correlate with hippocampal volume. However, cortisol measures have been shown to be notoriously variable and therefore much greater confidence would be needed in this measurement before one could reliably assess correlations between brain volume measures and cortisol levels [109].

During a post-hoc VBM analysis of the rest of the brain, the most significant region suggestive of volume reduction in the poorer group was found to be the grey matter of the cerebellum. This is an interesting finding as cerebellar volume reduction has been observed in subjects suffering from PTSD which it has been proposed may result in brain volume reduction by chronic stress, a similar mechanism which has been proposed in subjects from deprived backgrounds [104].

While regions suggestive of volume reductions in a group with lower socioeconomic status have been observed in important brain regions such as the hippocampus and cerebellum, the mechanisms by which these reductions occur are still unclear. It is hoped the work here will add to the body of knowledge in this area and to guide further, larger studies, with more specific goals around eliciting the mediating and moderating factors between SES and volumetric changes in the brain.

Chapter 4 moved away from focusing on hippocampal volumes by exploiting another aspect of MR technology to perform MR spectroscopy (MRS). Using 3T MRS hippocampal metabolites were measured in two groups of differing socioeconomic status from a community based sample. We believe this is the first time this has been done. These were the same two groups studied in chapters 2 and 3 and it was hoped that the MRS metabolite measures might be a useful addition to the volumetric assessment.

No significant difference was observed in the NAA/Cr and Cho/Cr ratios between the more deprived and more affluent groups. The motivation behind assessing hippocampal metabolites originated from the fact that the hippocampus may be preferentially affected in subjects exposed to chronic stress, where in this case we considered social deprivation as the stressor. To further investigate this proposed mechanism, the hippocampal metabolites were correlated against cortisol and interleukin-6, two markers for inflammatory processes.

There were no significant correlations between NAA/Cr or Cho/Cr and cortisol. On removal of an outlying IL-6 value, a statistically significant correlation was observed between the

right hippocampal NAA/Cr ratio and IL-6 ($p = 0.005$, $r = 0.53$). Therefore, in general these results did not support the theory that inflammatory processes might affect hippocampal metabolism, although the correlation between right hippocampal NAA/Cr and IL-6 warrants replication in a larger sample size.

Further to the primary aims of chapter 4, hippocampal metabolites were correlated with age and hippocampal volume. No significant correlations were observed between hippocampal metabolite ratios and age or between hippocampal metabolite ratios and ICV corrected hippocampal volumes. Perhaps the reason hippocampal metabolites were not observed to correlate with age, was due to the limited age range of the population studied where other studies that included an assessment of more elderly subjects showed decreasing NAA/Cr with age [315].

This was the first time hippocampal metabolites have been assessed using 3T MRI between two groups of differing SES. While a group difference between hippocampal metabolites was not detected, it is hoped that the metabolite ratio and variance in those measurements will inform future work as to what is likely to be achievable using 3T MRI.

Chapter 5 returned to assessing hippocampal volumes where the aim was to use 3T MRI to establish a normal range for hippocampal volumes in healthy volunteers. The T1-weighted 3D FSPGR data was processed using an automatic segmentation method augmented by manual editing. Left and right hippocampal volumes were segmented for a group of 39 normal, healthy volunteers. The subjects ranged in ages from 19 to 64 years old. The relationships between hippocampal volumes, ICV and age were investigated where it was found that hippocampal volumes were larger with increasing ICV and smaller with increasing age. The ICV results were in keeping with what has already been observed in the peer reviewed literature though the conflicting reports on the effect of aging on the hippocampus make results difficult to compare with any degree confidence. Given the relationships between hippocampal volumes, age and ICV a multiple regression model was created to determine an equation to describe the size of the hippocampus for a given ICV of a subject of known age. Thus the normal range for hippocampal volumes, for subjects aged 19 to 64 was determined. That is, the mean and standard deviation of the left and right hippocampal volumes were $3421\text{mm}^3 \pm 399\text{mm}^3$ and $3487\text{mm}^3 \pm 431\text{mm}^3$ respectively.

In further comparisons between males and females, after correcting for ICV, no significant difference was observed between the hippocampal volumes across the sexes. In considering laterality differences the results obtained in chapter 5 reflected that which have been reported

elsewhere in the literature. That is, that was a trend for the left hippocampal volume to be slightly smaller than the right.

Thus, a methodology was produced and a normal database of hippocampal volumes for 3T MRI has been established for subjects in the age range between 19 and 64. This normative data may prove useful in future for further comparisons with psychosocial study groups or diseased populations, though its utility will be limited by the lack of a comprehensive assessment of IQ or education level of the normal volunteers which may be responsible for observed differences in hippocampal volumes. Therefore, the acquisition of the normative database does not negate the need for appropriate, study specific, control groups to be acquired in future work.

In chapter 6 the methodology developed in chapters 2 and 5 was implemented and extended to assess a query of more clinical importance than has been discussed elsewhere in this thesis. That is, is Incomplete Hippocampal Inversion a rare occurrence or simply a normal variant of the hippocampal formation. To investigate this problem the frequency of IHI in a normal population was assessed. A quantitative approach to assessing the hippocampus and amygdala was implemented to augment the radiological observations. This is the first time hippocampal malrotation has been assessed in such a comprehensive way.

In our study of normal healthy volunteers we found IHI in 37% of normal subjects. A similar criteria to that which has been used previously to assess IHI was adopted [32]. The main criteria on which IHI was identified was based upon observation of a round shaped hippocampus and vertical collateral sulcus. IHI was observed in healthy individuals yet it is not clear whether IHI is a pathological entity or a normal developmental variant of no clinical significance though it does appear as though there is a broad spectrum of variations seen within normal hippocampal development. We conclude that IHI is likely to be a feature of normal healthy population and is much more common than has been previously reported [32, 279, 26]. IHI was overwhelmingly lateralised to the left hippocampus, and this was in agreement with what was observed in a previous work [194]. Thus we have observed a relatively high incidence of IHI in a normal population and we therefore believe that it is likely that the appearance of IHI is not a causal factor in patients with epilepsy.

For the quantitative assessment of IHI the optimised technique discussed in chapters 2 and 5 was implemented. Automated segmentation of the malrotated and normal hippocampi were augmented by manual refinements to provide the resultant hippocampal volumes. Malrotated volumes, being a predominantly left-sided phenomenon, were compared with normal left-sided hippocampal volumes. The left malrotated volumes were found to be smaller than

normal hippocampal volumes ($p=0.006$) after correcting for differences in ICV. However, on closer inspection ICV differences were found to be significantly different, with ICV being approximately 5% larger on average in the IHI group than the normal group. This confounds the ICV corrected hippocampal volume results, and the reason that the ICV should be different between the groups remains a point for further investigation.

Malrotated left hippocampi volumes were compared to normal right hippocampal volumes. There was no statistically significant volume difference between these two groups ($p=0.19$).

Automatic segmentation of the amygdala was augmented by manual refinements to the resultant volumes. Volumes of the left amygdala, adjacent to malrotated hippocampi were compared to volumes of the left amygdala adjacent to normal hippocampi. There was no difference between the two groups of amygdala volumes after correcting for differences in ICV.

Similarly to the case for the hippocampus, left amygdala volumes adjacent to malrotated hippocampi were compared to right sided amygdala volumes adjacent to normal hippocampi. There was no statistically significant difference between the two mean volume measurements ($p=0.08$).

To compliment the hippocampal volume assessment of IHI and normal subjects a novel surface analysis method was applied [278]. This allowed volumes differences to be compared at the level of the voxel. In a model which included age and ICV as covariates and following an FDR correction there were no regions of statistically significant volume differences between the malrotated and normal left hippocampal volumes. This same process was applied to compare the amygdala volumes but again, no regions of statistically significant volume difference were found.

Presented in chapter 6, for the first time, was a comprehensive and qualitative assessment of hippocampal and amygdala volumes in IHI, completed to complement the observations made by experienced neuroradiologists. At first there appeared to be evidence that ICV corrected IHI hippocampal volumes were smaller than normal hippocampal volumes but this was later confounded by group differences in ICV. Further work is necessary to reconcile the issue surrounding the ICV correction to hippocampal volumes, but overall, the hippocampal segmentation methodology developed in earlier chapters of this worked well within this context.

The work performed in chapter 6 is currently under discussion to decide if a single manuscript covering the radiological and quantitative methodological aspects of this work should be

produced, or whether two separate manuscripts would better cover both of these topics. Once this decision has been made the manuscript(s) will be submitted to a peer reviewed journal.

In chapter 7 volumetric analysis of the hippocampus continued though we aimed to exploit the power of 3T MRI to assess hippocampal subfields. A protocol was established for measuring the CA4/dentate subfield of the hippocampus. This was the subfield of primary interest for our research. Repeat measures were made and it was established that the intra-operator variability was of an acceptable level. In the few subjects assessed here it was observed that CA4/dentate volumes tended to decrease with age. There was no significant difference in CA4/dentate volume following a course of anti-inflammatory medication as measured by manual segmentation on T2-weighted FSE coronal MRI images. Given the variance of the manual measurements of the CA4/dentate region, sample size calculations were estimated based on a desired study power of 0.90 for detecting a projected volume difference of 10%. These calculations determined that a paired study would require 30 subjects in each group and an unpaired design would require 43 subjects in each group in order to achieve the desired study power.

To complement the manual volume measurements made of the CA4/dentate region DARTEL-VBM using permutation based analysis was performed. This method detected a GM volume increase in the right hippocampus following anti-inflammatory treatment (FDR corrected, $Z = 5.51$). This finding needs further validation in a larger dataset and would benefit from a complementary study in animals to validate the relationship between neuroimaging markers and known neurogenesis, thus providing greater assurance in the methodology.

Regardless of the underlying biological mechanisms at work and despite the small numbers of subjects sampled in this pilot study, tentative evidence has been observed for volumetric differences between the pre and post treatment groups. The methodology appears to be robust, including the use of DARTEL-VBM and the application of permutation based statistics to VBM. Further work on larger sample sizes is strongly recommended and the author of this thesis is currently involved with such work

In chapter 8, the subfield protocol introduced in chapter 7 was extended to become a novel, comprehensive protocol for the manual delineation of hippocampal subfields throughout the entire length of the hippocampus. This is one of the few studies to date which has attempted to assess the absolute subfield volumes throughout the entire length of the hippocampus and to assess the intra and inter rater reproducibility of these measurements. This protocol was implemented in 17 normal volunteers. The mean volume and normal ranges were determined

for the hippocampal subfields and repeat intra and inter operator measures were made to assess the validity of the segmentation protocol.

The strength of the linear relationship between hippocampal subfield volumes and ICV was assessed and in general there was little or no correlation between these parameters. Correlations between subfield volumes and subject age were also assessed and generally, there were modest negative correlations. This would imply that there is a gradual decrease in subfield volumes with age. This mirrors that which has been observed for whole hippocampal volumes and was intuitively what was expected.

Of the 17 subjects where subfield volumes were determined repeat measures were made on 11 subjects by the first operator (the author of this thesis) to assess intra-rater reliability. A second operator also performed subfield measurements on 11 subjects and these results were compared to that of the first operator to assess inter-rater reliability. Intraclass correlation coefficients for absolute agreement and Dice metrics were calculated in both circumstances. Intra-rater ICC (agreement) and Dice measures were good for some subfields, namely: the CA4/dentate, CA1-CA3 and entorhinal cortex with values of approximately 0.80 or better. Other subfield segmentations agreed less well. In general the inter-rater results for both ICC (agreement) and Dice measures were poor. This implies that more work on refining the manual segmentation protocol method and training would be necessary before the protocol may be used across different raters with confidence.

Given that measuring hippocampal subfield volumes is a relatively new and novel type of measurement it was deemed necessary to perform an assessment of study power and sample size calculations. It was estimated that in order to achieve a study power of 90%, when detecting a 10% volume difference in the CA4/dentate subfield region, 77 subjects would be required in each group when an unpaired t-test design was being used to assess the data. Smaller sample numbers of 26 and 66 gave equivalent power for the CA1-CA3 and WM layer subfields respectively, as a result of the smaller variation measured for these regions. The GM subfields of the ERC and subiculum and WM subfields of the alveus and fimbria were found to require much greater numbers (> 130 subjects) to achieve equivalent power by using the same method.

Subfield volumes were also compared between male and females and between right and left hemispheres. In both instances, there were no significant differences between these groups of data.

The work performed in chapter 8, i.e. the creation of protocol for manually segmenting the subfield volumes throughout the whole hippocampus and the assessment of the quality of

these measures is believed to be of interest to the wider scientific community. Therefore, this work forms the basis for a scientific manuscript currently in preparation for publication to a peer reviewed journal.

9.2 Limitations of this thesis, methodological issues and future work

The utility of automated methods for determining hippocampal volumes has been widely covered in the peer reviewed literature though advancements in these methods such as that of the FIRST algorithm used in this thesis occur fairly infrequently. It is fair to say that there is no perfect method of automatic segmentation of the hippocampus. Treating the manual segmentation of the hippocampus as the gold standard automated methods have achieved Dice metrics of greater than 0.80 [123, 278]. There is some evidence to suggest the accuracy of hippocampal segmentation has improved over the past ten years, with values observed in studies using 3T being similar, and typically larger, than what was observed in earlier studies using 1.5T magnets [232]. There has also been efforts to try and reach a consensus over manual hippocampal segmentation which is of importance given it is the gold standard for comparing automated methods to [185]. Ultimately though, as the hippocampus is connected at various aspects to other areas of the brain, then there is always going to be some degree of subjectivity in identifying the boundary between these regions. Even if a consensus is reached for how to segment the hippocampus, subjectivity is likely to still exist between different raters interpretation of the underlying protocol. However, we can hope and expect that as image quality continues to improve so will the accuracy and reproducibility of hippocampal segmentation, both by manual and automated methods.

The limitations of VBM have been well documented in the peer reviewed literature. This includes the limitation of the VBM methodology to be applied in situations where images were not well aligned [50]. A further limitation was that cluster statistics were not recommended for VBM as a result of variations in smoothness across images, the so-called non-stationary problem [21]. However, in chapter 3, both of these points were addressed by refinements to the VBM methodology. Firstly, an improved normalisation algorithm (DARTEL) was used to bring all the study images into a common template space [20] and secondly, a non-stationary correction was applied when performing the final group comparison statistics [159, 158]. Thus in chapter 3, the major limitations which have been cited about the VBM methodology were largely addressed. Despite the confidence in using these aspects of the VBM methodology it was always considered best practice to visually inspect the output of any gross automated

image processing steps to ensure any erratic errors in the processing pipeline were detected. However following good pre-processing of the data, such errors were rare.

A further point regarding VBM is that it does not easily provide a means to determine absolute volume differences between groups. Also, it does not give information over what aspect of the GM is changing, i.e. it is non-specific to whether the cortical thickness or surface area has changed and these parameters do not necessarily change together in a linear way. Without going into a full debate over the relative merits of both VBM and cortical thickness techniques such as that provided by the Freesurfer software (<http://surfer.nmr.mgh.harvard.edu/>) suffice it to say that techniques are not mutually exclusive and many would argue that both techniques can offer useful, complementary information [168, 41, 186, 89]. As the primary aim of this thesis was to assess the hippocampus only the DARTEL-VBM methodology was considered in chapter 3. However, future work assessing the cortical areas of the brain between the poorer and more affluent groups, including analysis of cortical thickness and surface area measurements will be performed to compliment the VBM results.

A final technical point regarding DARTEL-VBM is that DARTEL is more computationally intensive than previous normalisation methods implemented within SPM and as such takes a considerably longer time to run to completion. This may be restrictive, particularly for very large studies, though this would likely be easily overcome with modest improvements to the computing infrastructure of a given laboratory.

One limitation to implementing MRS analysis performed in chapter 4 was that on the GE Signa HD system the MRS voxel had to be planned parallel to the orientation of the planning image acquisition. That is, it was not possible to plan oblique voxels by rotating the voxel on the MR console. This was particularly restrictive when attempting to acquire spectra from the hippocampus as it typically lies on a steep angle within the temporal lobe. Future developments in MR scanning technology are likely to resolve this issue by enabling voxels to be defined on oblique angles.

Another limitation of the MRS acquisition in the hippocampus was the poor signal to noise that was achieved. In the majority of cases the data on the NAA, Cho and Cr was interpretable but for some of the voxels the spectra were too poor in quality to be considered so. Implementing higher order shimming over the region of interest may have improved the spectral quality. More adjustment could have been made to the voxel size to ensure it was entirely encompassed within the hippocampus, though while this may avoid voxel contamination from CSF, the SNR of the signal will be reduced in smaller voxels. In an ideal world much better spatial resolution with improved SNR would be achievable, allowing MRSI acquisitions to be

performed which could show variations in metabolites within the hippocampus. To achieve such improvements SNR improvements of at least around an order of magnitude are likely to be necessary. At present it is not obvious how such gain in spectral quality may be achieved.

Future developments which may allow MRS Imaging (MRSI) to be performed in acceptable time frames include greater exploitation of parallel imaging techniques as greater numbers of receiver coil elements are used, indeed, 32-channel devices are already being used [380]. Moreover, the introduction of compressed sensing techniques to greatly reduce the acquisition time (or improve quality if scanning time is maintained) are also likely to offer benefits to MRSI. Compressed sensing MRSI acquisitions have already been acquired in animals, where in dynamic nuclear polarisation (DNP) experiments, rapid acquisition schemes are even more beneficial [166, 167]. Therefore it is likely improvements in MRS will come on several fronts, with developments in receiver hardware and pulse sequences combining to enable faster acquisitions times or improved SNR.

A criticism that could be leveled at all of the work in chapters 2-4 was that the subjects were separated into their two socioeconomic groups based on existing socioeconomic measures. That is, a subjects current SES may not reflect their SES as a child and given the large impact of environmental component at the developmental stage, this is a confound which might have been worth avoiding. However, those subjects who have deviated to a different SES are also of great interest as they have rebuked the trend of their peers. What makes them special? Such questions are likely to be addressed in further work.

Various technological advances which may occur over the next several years include the potential for faster imaging by further exploitation of multi-element coils, and improved pulse sequences. However, technological advances may only be useful if they are framed within plausible biological models. Future work to develop a clearer biological framework and a more comprehensive investigation of metabolic and inflammatory markers may be more informative. However, it could also have been the case that as we were dealing with an otherwise healthy population, the anticipated metabolic changes in the hippocampus simply do not exist as a result of the posited inflammatory pathway. Further studies in animals may be useful for interrogating the underlying metabolism as a result of stress and inflammatory processes. Moreover, identification of valid neuroimaging markers from animal studies should also be performed before attempting to translate the work into human studies.

The subfield analysis performed in chapter 7 and 8 was limited by a number of factors, the fundamental limitation being the fact that using anatomical landmarks to define subfield cellular zones is inherently inaccurate. As the subfield boundaries themselves cannot be seen

on MRI there is little, at present, that can be done to solve this problem. Thus the approach of using landmarks, while limited in terms of accuracy, is a reasonable approximation to start assessing subfields. By this method the aim was to build a protocol which although it might not be accurate, could be precise and reproducible, thus allowing comparisons to be made within and between subject groups. It was hoped that the work undertaken in this thesis will contribute to the small but growing body of literature around hippocampal subfield measurements and segmentation methods.

Over and above the practical and technical limitations of this study there are some more fundamental issues that underpin the aims of this work that ought to be addressed. From the DARTEL-VBM data there was evidence to suggest that anti-inflammatory treatment may have induced a small increase in hippocampal volume. One possible mechanism by which this may have occurred is by neurogenesis. However, whether or not neurogenesis would actually lead to a detectable volume difference on an MRI scan and post processing as described above has yet to be validated. Unless one was able to access and assess post-mortem data and correlate neuroimaging volumetric findings with changes with histology, the only other method would be to confirm the methodological process in animals. That is, a study should be performed where it is known that neurogenesis will be induced in the animal. High resolution longitudinal MRI data should be acquired and the pre- and post neurogenesis data assessed to see if volumetric changes were detectable. Until this link between the histology and neuroimaging dataset has been confirmed, human, adult, in-vivo studies will suffer from the doubt that the small and subtle volumetric changes which are being detected are not the result of real changes in the underlying tissue but rather are a subsequence of the methodology. If such a study was to be performed it would be useful to assess the relationship between other neuroimaging markers such as T1, T2, diffusion and perfusion against neurogenesis, again to build confidence in subsequent methodologies applied in-vivo, in adult humans.

As with all aspects of scientific development, improvements can occur on several fronts. In MRI structure and subfield segmentation the two main areas where developments are likely to occur are the data acquisition and image analysis. If we include MRI hardware developments as part of the acquisition process, clearly as demonstrated in some of the high and ultra-high field studies there are benefits from the greater signal-noise at higher field strengths.

In the area of pulse sequence development, the two main methodologies applied for volume analysis through this thesis, one being the surface analysis of T1-weighted 3D images and the other being manual outlining of subfields on T2-weighted FSE images, exploit different benefits from these two types of sequences. These are the superior three-plane resolution of

the T1-weighted FSPGR images and the superior grey matter to white matter contrast of the T2-weighted FSE images for better internal hippocampal visualisation.

In an ideal world, having a T2-weighted 3D sequence with high in-plane resolution and optimal GM/WM contrast would be of great benefit. Such sequences have been developed but to the knowledge of the author at this time they remain to be assessed in the context of hippocampal subfield measurements. Further technical advancements might include exploiting parallel imaging techniques such that images can be acquired more quickly, or with better quality for the same length of scan time. This may occur as head coil technology progresses from 8 channel to 16, 32 and perhaps even 128 channels in the future. Moreover, the widespread introduction of compressed sensing methods into MRI pulse sequence will likely enable a further leap forward in reducing imaging time and/or improving image quality [200]. Of course, such developments often come as a trade-off and higher field strengths may present challenges with image susceptibility. Similarly, increasing the number of coil elements may increase the inhomogeneity or bias across the field of view, though to counter these issues one might expect that correction algorithms for both of these issues will advance in parallel.

As well as hardware developments, the post processing methods have a large part to play in measuring subfield volumes. Yushkevich et al recently published on a 'nearly automatic segmentation of hippocampal subfields', this is an impressive achievement and demonstrated good agreement (Dice) between manual and automatic segmentations for several of the hippocampal subfield [375], though as with Mueller et al's work [251], the subfield segmentations did not extend into the hippocampal head and tail. Future developments will likely aim to expand into these areas and to continue to improve and develop methodologies for segmenting the body of the hippocampus. A reliable method of automatically segmenting the entire hippocampus into its subfields would be the ultimate goal, but this remains some way off. Moreover, even if a such a method was to be produced and proven to be reliable, the issue of the underlying accuracy may still prove to be a fundamental problem.

9.3 Summary of findings and contribution of this thesis

The aim of project was to develop and assess the methodologies involved in measuring whole hippocampal volumes, hippocampal metabolites and hippocampal subfield volumes using 3T MRI. For whole hippocampal volumes a comprehensive manual protocol was developed and manual tracings were compared to a recently developed automated segmentation method. This assessment was performed in two groups of subjects of differing socioeconomic status.

While the manual and automated methods compared well, with dice overlap metrics of 0.81, manual inspection of the outcome of the automated segmentation procedure did on occasion require manual refinement as a result of gross errors. This methodology of automated segmentation and manual correction was later implemented to assess a group of normal volunteers in chapter 5 of this thesis.

In chapter 3 a different methodology was employed to assess for focal differences in GM volume between the lower SES and more affluent groups. The voxel-based morphometry method has evolved over recent years and therefore results from different processing pipelines were compared. While both processing pipeline which were assessed included additional inhomogeneity corrections, the normalization method where one pipeline utilised the DARTEL normalization tool. Including DARTEL in the VBM processing pipeline improved the outcome where greater confidence in the SPM maps was observed. In contrasting the two SES groups and following a small volume correction on the whole brain VBM data, results were suggestive of a focal GM decrease in the hippocampus of the more deprived group but further work on a larger sample size would be necessary to confirm this finding.

In chapter 4 we aimed to measure hippocampal metabolite concentrations using 3T magnetic resonance spectroscopy. The hippocampus is a difficult region to assess using MRS as a result of its uneven shape and its location in the brain near regions air spaces and also the fact that CSF surrounds the hippocampus at many of its aspects. A protocol was developed which it was hoped would achieve good quality spectra. Hippocampal metabolites were assessed in the same subjects of differing SES as were used in chapters 2 and 3. Of the 42 subjects assessed, good quality MRS spectra was achieved in 73% (31/42) subjects. On the remaining data of suitable quality NAA/Cr and Cho/Cr ratios were compared between the two differing SES groups. The author believes this may be the first time metabolite ratios have the hippocampus have been assessed in humans using 3T MRI in groups of differing SES.

In chapter 5 the protocol developed in chapter 2 for assessment of hippocampal volumes was implemented on a group of normal volunteers. The aim was to obtain an estimation of hippocampal values in healthy volunteers. Hippocampal volumes over the age range of 19 to 64 were measured. While these values provide a baseline for the methodology developed to assess hippocampal volumes the future use of this normative database is severely limited as a result of the lack of control for the education or intellectual capacity of the volunteers in this normative group.

In chapter 6 the normal hippocampal data and the two groups of differing SES were combined

to form a larger group of normal subjects such that the incidence of Incomplete hippocampal inversion could be assessed. Following the reporting of the scans a similar incidence rate for IHI was observed across the three groups, this check was done to ensure there was no group bias to the incidence of IHI. From the radiological assessment of IHI, a incidence rate of 37% for IHI was observed. By implementing and then extending the methodology of hippocampal volumes developed in chapter 2 a comprehensive quantitative assessment of hippocampal and amygdala volumes was performed, this included a comparison between IHI whole hippocampal volumes and normal hippocampal volumes and also a surface analysis of the same groups. No significant regions of volume difference were observed between these two groups by the surface analysis method. The author believes this is the first quantitative assessment of hippocampal and amygdala volumes in subjects with IHI.

In chapter 7 a protocol was developed and implemented to assess the volume of the CA4/dentate subfield of the hippocampus. These volumes were assessed in two small groups of patient with rheumatoid arthritis. The intra-operator variability was assessed and was found to be reasonable with ICC (consistency) measures of 0.88 and 0.87 for the left and right CA4/dentate volume respectively. The small numbers of subjects in the two groups meant a reliable comparison was not possible.

Building on the work in chapter 7 a more comprehensive protocol for assessing hippocampal subfields was developed in chapter 8 and was implemented in a group of healthy subjects. Given the novelty of using the 3T MRI system and the development of a hippocampal subfield segmentation protocol it was possible following estimation of the mean volumes and variances of the various subfields to estimate power and sample sizes that would be necessary in future work if subfield volumes were to be assessed. The detailed subfield segmentation protocol will be written up for publication in the peer reviewed literature. It is hoped that this will stimulate debate on the measurement of hippocampal subfields with a view to trying to establish a consensus over the method by which hippocampal subfield segmentation should be performed. As studies continue to use differing protocols this limits the interpretability of any results across the literature. Given the difficulties which have been observed as a result of the variation in whole hippocampal segmentation protocols it would be advantageous for the academic community to reach consensus over subfield measurements to improve comparisons across studies. The author believes the work in the area of human hippocampal subfield measurement described in the final chapter of this thesis will prove to be the greatest contribution this thesis provides to the scientific literature.

Bibliography

- [1] J. Acosta, G.B. Williams, J.M.S. Pereira, G. Pengas, and P.J. Nestor. The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. *Neuroimage*, 39(4):1654–1665, Feb 2008.
- [2] N. Adler, A. Singh-Manoux, J. Schwartz, J. Stewart, K. Matthews, and M. G. Marmot. Social status and health: a comparison of british civil servants in whitehall-ii with european- and african-americans in cardia. *Soc Sci Med*, 66(5):1034–1045, Mar 2008.
- [3] N. E. Adler, T. Boyce, M. A. Chesney, S. Cohen, S. Folkman, R. L. Kahn, and S. L. Syme. Socioeconomic status and health. the challenge of the gradient. *Am Psychol*, 49(1):15–24, Jan 1994.
- [4] N. E. Adler, W. T. Boyce, M. A. Chesney, S. Folkman, and S. L. Syme. Socioeconomic inequalities in health. no easy solution. *JAMA*, 269(24):3140–3145, 1993.
- [5] N. E. Adler and J. Stewart. Health disparities across the lifespan: meaning, methods, and mechanisms. *Ann N Y Acad Sci*, 1186:5–23, Feb 2010.
- [6] I. Agartz, R. Momenan, R. R. Rawlings, M. J. Kerich, and D. W. Hommer. Hippocampal volume in patients with alcohol dependence. *Arch Gen Psychiatry*, 56(4):356–363, Apr 1999.
- [7] G. Albouy, V. Sterpenich, E. Balteau, G. Vandewalle, M. Desseilles, T. Dang-Vu, A. Darsaud, P. Ruby, P.H. Luppi, C. Degueldre, et al. Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron*, 58(2):261–272, 2008.
- [8] J. Altman and G.D. Das. Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *The Journal of comparative neurology*, 124(3):319–335, 1965.

- [9] J.A. Amat, R. Bansal, R. Whiteman, R. Haggerty, J. Royal, and B. S. Peterson. Correlates of intellectual ability with morphology of the hippocampus and amygdala in healthy adults. *Brain Cogn*, 66(2):105–114, Mar 2008.
- [10] N. C. Andreasen, M. Flaum, V. Swayze, D. S. O’Leary, R. Alliger, G. Cohen, J. Ehrhardt, and W. T. Yuh. Intelligence and brain structure in normal individuals. *Am J Psychiatry*, 150(1):130–134, Jan 1993.
- [11] E. Angelie, A. Bonmartin, A. Boudraa, P. M. Gonnaud, J. J. Mallet, and D. Sappey-Mariner. Regional differences and metabolic changes in normal aging of the human brain: proton mr spectroscopic imaging study. *AJNR Am J Neuroradiol*, 22(1):119–127, Jan 2001.
- [12] L. G. Apostolova, L. Mosconi, P. M. Thompson, A. E. Green, K. S. Hwang, A. Ramirez, R. Mistur, W. H. Tsui, and M. J. deLeon. Subregional hippocampal atrophy predicts alzheimer’s dementia in the cognitively normal. *Neurobiol Aging*, 31(7):1077–1088, Jul 2010.
- [13] L.G. Apostolova, I. D. Dinov, R. A. Dutton, K. M. Hayashi, A. W. Toga, J. L. Cummings, and P. M. Thompson. 3d comparison of hippocampal atrophy in amnesic mild cognitive impairment and alzheimer’s disease. *Brain*, 129(Pt 11):2867–2873, Nov 2006.
- [14] L.G. Apostolova, R. A. Dutton, I. D. Dinov, K. M. Hayashi, A. W. Toga, J. L. Cummings, and P. M. Thompson. Conversion of mild cognitive impairment to alzheimer disease predicted by hippocampal atrophy maps. *Arch Neurol*, 63(5):693–699, May 2006.
- [15] D. Araújo, A. C. Santos, T. R. Velasco, L. Wichert-Ana, V. C. Terra-Bustamante, V. Alexandre, C. G. Carlotti, J. A. Assirati, H. R. Machado, R. Walz, J. P. Leite, and A. C. Sakamoto. Volumetric evidence of bilateral damage in unilateral mesial temporal lobe epilepsy. *Epilepsia*, 47(8):1354–1359, Aug 2006.
- [16] B. A. Ardekani, S. Guckemus, A. Bachman, M.J. Hoptman, M. Wojtaszek, and J. Nierenberg. Quantitative comparison of algorithms for inter-subject registration of 3d volumetric brain mri scans. *J Neurosci Methods*, 142(1):67–76, Mar 2005.
- [17] S. Arndt, T. Cizadlo, N. C. Andreasen, D. Heckel, S. Gold, and D. S. O’Leary. Tests for comparing images based on randomization and permutation methods. *J Cereb Blood Flow Metab*, 16(6):1271–1279, Nov 1996.

- [18] B. B. Arnetz, S. O. Brenner, L. Levi, R. Hjelm, I. L. Petterson, J. Wasserman, B. Petrini, P. Eneroth, A. Kallner, and R. Kvetnansky. Neuroendocrine and immunologic effects of unemployment and job insecurity. *Psychother Psychosom*, 55(2-4):76–80, 1991.
- [19] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-euclidean framework for statistics on diffeomorphisms. *Med Image Comput Comput Assist Interv*, 9(Pt 1):924–931, 2006.
- [20] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, Oct 2007.
- [21] J. Ashburner and K. J. Friston. Voxel-based morphometry—the methods. *Neuroimage*, 11(6 Pt 1):805–821, Jun 2000.
- [22] J. Ashburner and K. J. Friston. Why voxel-based morphometry should be used. *Neuroimage*, 14(6):1238–1243, Dec 2001.
- [23] J. Ashburner and K. J. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, Jul 2005.
- [24] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*, 12(1):26–41, Feb 2008.
- [25] D. Bajic, U. Ewald, and R. Raininko. Hippocampal development at gestation weeks 23 to 36. an ultrasound study on preterm neonates. *Neuroradiology*, 52(6):489–494, Jun 2010.
- [26] D. Bajic, E. Kumlien, P. Mattsson, S. Lundberg, C. Wang, and R. Raininko. Incomplete hippocampal inversion-is there a relation to epilepsy? *Eur Radiol*, 19(10):2544–2550, Oct 2009.
- [27] D. Bajic, C. Wang, E. Kumlien, P. Mattsson, S. Lundberg, O. Eeg-Olofsson, and R. Raininko. Incomplete inversion of the hippocampus—a common developmental anomaly. *Eur Radiol*, 18(1):138–142, Jan 2008.
- [28] L. L. Baker and A. J. Barkovich. The large temporal horn: Mr analysis in developmental brain anomalies versus hydrocephalus. *AJNR Am J Neuroradiol*, 13(1):115–122, 1992.
- [29] P. B. Barker, D. O. Hearshen, and M. D. Boska. Single-voxel proton mrs of the human brain at 1.5t and 3.0t. *Magn Reson Med*, 45(5):765–769, May 2001.

- [30] P. B. Barker, B. J. Soher, S. J. Blackband, J. C. Chatham, V. P. Mathews, and R. N. Bryan. Quantitation of proton nmr spectra of the human brain using tissue water as an internal concentration reference. *NMR Biomed*, 6(1):89–94, 1993.
- [31] A.J. Barkovich. *Pediatric neuroimaging*. Lippincott Williams & Wilkins, 2005.
- [32] P. Barsi, J. Kenéz, D. Solymosi, A. Kulin, P. Halász, G. Rásonyi, J. Janszky, A. Kalóczkai, G. Barcs, M. Neuwirth, E. Paraicz, Z. Siegler, M. Morvai, J. Jerney, M. Kassay, and A. Altmann. Hippocampal malrotation with normal corpus callosum: a new entity? *Neuroradiology*, 42(5):339–345, May 2000.
- [33] P. M Bauer, J. L. Hanson, R. K. Pierson, R. J. Davidson, and S. D. Pollak. Cerebellar volume and cognitive functioning in children who experienced early deprivation. *Biol Psychiatry*, 66(12):1100–1106, Dec 2009.
- [34] M. Baulac, N. De Grissac, D. Hasboun, C. Oppenheim, C. Adam, A. Arzimanoglou, F. Semah, S. Lehericy, S. Clémenceau, and B. Berger. Hippocampal developmental changes in patients with partial epilepsy: magnetic resonance imaging and clinical aspects. *Ann Neurol*, 44(2):223–233, Aug 1998.
- [35] C. E. Bearden, J C. Soares, A. D. Klunder, M. Nicoletti, N. Dierschke, K. M. Hayashi, K. L. Narr, P. Brambilla, R. B. Sassi, D. Axelson, N. Ryan, B. Birmaher, and P. M. Thompson. Three-dimensional mapping of hippocampal anatomy in adolescents with bipolar disorder. *J Am Acad Child Adolesc Psychiatry*, 47(5):515–525, May 2008.
- [36] R. D. Beck, C. Wasserfall, G. K. Ha, J. D. Cushman, Z. Huang, M. A. Atkinson, and J. M. Petitto. Changes in hippocampal il-15, related cytokines, and neurogenesis in il-2 deficient mice. *Brain Res*, 1041(2):223–230, Apr 2005.
- [37] S. Becker and J.M. Wojtowicz. A model of hippocampal neurogenesis in memory and mood disorders. *Trends in cognitive sciences*, 11(2):70–76, 2007.
- [38] O. Bengochea and L. M. Gonzalo. Effect of chronic alcoholism on the human hippocampus. *Histol Histopathol*, 5(3):349–357, Jul 1990.
- [39] T. P. Beresford, D B. Arciniegas, J. Alfers, L. Clapp, B. Martin, Y. Du, D. Liu, D. Shen, and C. Davatzikos. Hippocampus volume loss due to chronic heavy drinking. *Alcohol Clin Exp Res*, 30(11):1866–1870, Nov 2006.
- [40] L. Bergouignan, M. Chupin, Y. Czechowska, S. Kinkingnéhun, C. Lemogne, G. LeBastard, M. Lepage, L. Garnero, O. Colliot, and P. Fossati. Can voxel based morphometry,

- manual segmentation and automated segmentation equally detect hippocampal volume differences in acute depression? *Neuroimage*, 45(1):29–37, Mar 2009.
- [41] P. Bermudez, J. P. Lerch, A. C. Evans, and R. J. Zatorre. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex*, 19(7):1583–1596, Jul 2009.
- [42] C.M. Bird and N. Burgess. The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience*, 9(3):182–194, 2008.
- [43] D. L. Birken and W. H. Oldendorf. N-acetyl-l-aspartic acid: a literature review of a compound prominent in 1h-nmr spectroscopic studies of brain. *Neurosci Biobehav Rev*, 13(1):23–31, 1989.
- [44] C. Blair and A. Diamond. Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure. *Dev Psychopathol*, 20(3):899–911, 2008.
- [45] J. M. Bland and D. G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476):307–310, Feb 1986.
- [46] S. Blüml, K. J. Seymour, and B. D. Ross. Developmental changes in choline- and ethanolamine-containing compounds measured with proton-decoupled (31)p mrs in in vivo human brain. *Magn Reson Med*, 42(4):643–654, Oct 1999.
- [47] W. Block, F. Träber, O. Widdern, M. Metten, H. Schild, W. Maier, A. Zobel, and F. Jessen. Proton mr spectroscopy of the hippocampus at 3 t in patients with unipolar major depressive disorder: correlates and predictors of treatment response. *Int J Neuropsychopharmacol*, 12(3):415–422, Apr 2009.
- [48] K. M. Bohnert and N. Breslau. Stability of psychiatric outcomes of low birth weight: a longitudinal investigation. *Arch Gen Psychiatry*, 65(9):1080–1086, Sep 2008.
- [49] O. Bonne, M. Vythilingam, M. Inagaki, S. Wood, A. Neumeister, A. C. Nugent, J. Snow, D. A. Luckenbaugh, E. E. Bain, W. C. Drevets, and D. S. Charney. Reduced posterior hippocampal volume in posttraumatic stress disorder. *J Clin Psychiatry*, 69(7):1087–1091, Jul 2008.
- [50] F. L. Bookstein. "voxel-based morphometry" should not be used with imperfectly registered images. *Neuroimage*, 14(6):1454–1462, Dec 2001.

- [51] R. G. Boyes, J. L. Gunter, C. Frost, A. L. Janke, T. Yeatman, D. L. G. Hill, M. A. Bernstein, P. M. Thompson, M. W. Weiner, N. Schuff, G. E. Alexander, R. J. Killiany, C. DeCarli, C. R. Jack, and N. C. Fox. Intensity non-uniformity correction using n3 on 3-t scanners with multichannel phased array coils. *Neuroimage*, 39(4):1752–1762, Feb 2008.
- [52] H. Braak. On the structure of the human archicortex. *Cell and Tissue Research*, 152(3):349–383, 1974.
- [53] H. Braak. Architectonics of the human telencephalic cortex. 1980.
- [54] R. H. Bradley and R. F. Corwyn. Socioeconomic status and child development. *Annu Rev Psychol*, 53:371–399, 2002.
- [55] J. Brandtstädter, B. Baltes-Götz, C. Kirschbaum, and D. Hellhammer. Developmental and personality correlates of adrenocortical activity as indexed by salivary cortisol: observations in the age range of 35 to 65 years. *J Psychosom Res*, 35(2-3):173–185, 1991.
- [56] P. A. Braveman, C. Cubbin, S. Egerter, S. Chideya, K. S. Marchi, M. Metzler, and S. Posner. Socioeconomic status in health research: one size does not fit all. *JAMA*, 294(22):2879–2888, Dec 2005.
- [57] J. D. Bremner, P. Randall, T. M. Scott, R. A. Bronen, J. P. Seibyl, S. M. Southwick, R. C. Delaney, G. McCarthy, D. S. Charney, and R. B. Innis. Mri-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. *Am J Psychiatry*, 152(7):973–981, Jul 1995.
- [58] J. D. Bremner, P. Randall, E. Vermetten, L. Staib, R. A. Bronen, C. Mazure, S. Capelli, G. McCarthy, R. B. Innis, and D. S. Charney. Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse—a preliminary report. *Biol Psychiatry*, 41(1):23–32, Jan 1997.
- [59] T. Breyer, I. Wanke, S. Maderwald, F. G. Woermann, O. Kraff, J. M. Theysohn, A. Ebner, M. Forsting, M. E. Ladd, and M. Schlamann. Imaging of patients with hippocampal sclerosis at 7 tesla: initial results. *Acad Radiol*, 17(4):421–426, Apr 2010.
- [60] R. S. Briellmann, A. Syngeniotes, and G. D. Jackson. Comparison of hippocampal volumetry at 1.5 tesla and at 3 tesla. *Epilepsia*, 42(8):1021–1024, Aug 2001.

- [61] R. A. Bronen and G. Cheung. Mri of the normal hippocampus. *Magn Reson Imaging*, 9(4):497–500, 1991.
- [62] R. A. Bronen and G. Cheung. Mri of the temporal lobe: normal variations, with special reference toward epilepsy. *Magn Reson Imaging*, 9(4):501–507, 1991.
- [63] R. A. Bronen, G. Cheung, J. T. Charles, J. H. Kim, D. D. Spencer, S. S. Spencer, G. Sze, and G. McCarthy. Imaging findings in hippocampal sclerosis: correlation with pathology. *AJNR Am J Neuroradiol*, 12(5):933–940, 1991.
- [64] J. Brooks-Gunn and G. J. Duncan. The effects of poverty on children. *Future Child*, 7(2):55–71, 1997.
- [65] E. S. Brown, D. J. Woolston, A. Frol, L. Bobadilla, D. A. Khan, M. Hanczyc, A. J. Rush, J. Fleckenstein, E. Babcock, and C. M. Cullum. Hippocampal volume, spectroscopy, cognition, and mood in patients receiving corticosteroid therapy. *Biol Psychiatry*, 55(5):538–545, Mar 2004.
- [66] S. Brulatout, P. Méric, I. Loubinoux, J. Borredon, J. L. Corrèze, P. Roucher, B. Gillet, G. Bérenger, J. C. Beloeil, B. Tiffon, J. Mispelter, and J. Seylaz. A one-dimensional (proton and phosphorus) and two-dimensional (proton) in vivo nmr spectroscopic study of reversible global cerebral ischemia. *J Neurochem*, 66(6):2491–2499, Jun 1996.
- [67] T.W. Buchanan and W.R. Lovallo. Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology*, 26(3):307–317, 2001.
- [68] J. C. Buckner, E. Mezzacappa, and W. R. Beardslee. Characteristics of resilient youths living in poverty: the role of self-regulatory processes. *Dev Psychopathol*, 15(1):139–162, 2003.
- [69] C. Buss, C. Lord, M. Wadiwalla, D. H. Hellhammer, S. J. Lupien, M. J. Meaney, and J. C. Pruessner. Maternal care modulates the relationship between prenatal risk and hippocampal volume in women but not in men. *J Neurosci*, 27(10):2592–2595, Mar 2007.
- [70] C. Caldji, J. Diorio, and M. J. Meaney. Variations in maternal care alter gaba(a) receptor subunit expression in brain regions associated with fear. *Neuropsychopharmacology*, 28(11):1950–1959, Nov 2003.

- [71] H.A. Cameron and R.D.G. McKay. Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. *The Journal of comparative neurology*, 435(4):406–417, 2001.
- [72] S. Campbell, M. Marriott, C. Nahmias, and G. M. MacQueen. Lower hippocampal volume in patients suffering from depression: a meta-analysis. *Am J Psychiatry*, 161(4):598–607, Apr 2004.
- [73] O. T. Carmichael, H. A. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, and Y. Liu. Atlas-based hippocampus segmentation in alzheimer’s disease and mild cognitive impairment. *Neuroimage*, 27(4):979–990, Oct 2005.
- [74] J. Cavanagh, C. Paterson, J. McLean, S. Pimlott, M. McDonald, J. Patterson, D. Wyper, and I. McInnes. Tumour necrosis factor blockade mediates altered serotonin transporter availability in rheumatoid arthritis: a clinical, proof-of-concept study. *Ann Rheum Dis*, 69(6):1251–1252, Jun 2010.
- [75] J. Cavanagh, J. Patterson, S. Pimlott, D. Dewar, J. Eersels, M. F. Dempsey, and D. Wyper. Serotonin transporter residual availability during long-term antidepressant therapy does not differentiate responder and nonresponder unipolar patients. *Biol Psychiatry*, 59(4):301–308, Feb 2006.
- [76] A. Cerbone, FR Patacchioli, and AG Sadile. A neurogenetic and morphogenetic approach to hippocampal functions based on individual differences and neurobehavioral covariations. *Behavioural brain research*, 55(1):1–16, 1993.
- [77] D. W. Chakeres, C. D. S. Whitaker, R. A. Dashner, D. W. Scharre, D. Q. Beversdorf, A. Raychaudhury, and P. Schmalbrock. High-resolution 8 tesla imaging of the formalin-fixed normal human hippocampus. *Clin Anat*, 18(2):88–91, Mar 2005.
- [78] J. R. Challis, D. Sloboda, S. G. Matthews, A. Holloway, N. Alfaidy, F. A. Patel, W. Whittle, M. Fraser, T. J. Moss, and J. Newnham. The fetal placental hypothalamic-pituitary-adrenal (hpa) axis, parturition and post natal health. *Mol Cell Endocrinol*, 185(1-2):135–144, Dec 2001.
- [79] F. A. Champagne. Epigenetic mechanisms and the transgenerational effects of maternal care. *Front Neuroendocrinol*, 29(3):386–397, Jun 2008.
- [80] F. A. Champagne and M. J. Meaney. Stress during gestation alters postpartum maternal care and the development of the offspring in a rodent model. *Biol Psychiatry*, 59(12):1227–1235, Jun 2006.

- [81] L. Chang, T. Ernst, D. Osborn, W. Seltzer, M. Leonido-Yee, and R. E. Poland. Proton spectroscopy in myotonic dystrophy: correlations with ctg repeats. *Arch Neurol*, 55(3):305–311, Mar 1998.
- [82] S. Chen, W. Xia, L. Li, J. Liu, Z. He, Z. Zhang, L. Yan, J. Zhang, and D. Hu. Gray matter density reduction in the insula in fire survivors with posttraumatic stress disorder: a voxel-based morphometric study. *Psychiatry Res*, 146(1):65–72, Jan 2006.
- [83] Z. Cho, J. Han, S. Hwang, D. Kim, K. Kim, N. Kim, S. J. Kim, J. Chi, C. Park, and Y. Kim. Quantitative analysis of the hippocampus using images obtained from 7.0 t mri. *Neuroimage*, 49(3):2134–2140, Feb 2010.
- [84] Z. Cho, Y. Kim, J. Han, N. Kim, S. Hwang, S. Kim, and S. Cho. Altered t2* relaxation time of the hippocampus in major depressive disorder: implications of ultra-high field magnetic resonance imaging. *J Psychiatr Res*, 44(14):881–886, Oct 2010.
- [85] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehericy, H. Benali, L. Garnero, O. Colliot, and Alzheimer’s Disease Neuroimaging Initiative. Fully automatic hippocampus segmentation and classification in alzheimer’s disease and mild cognitive impairment applied on data from adni. *Hippocampus*, 19(6):579–587, Jun 2009.
- [86] D. Cicchetti and S. L. Toth. Child maltreatment. *Annu Rev Clin Psychol*, 1:409–438, 2005.
- [87] S. Cohen, W. J. Doyle, and A. Baum. Socioeconomic status is associated with stress hormones. *Psychosom Med*, 68(3):414–420, 2006.
- [88] S. Cohen, J. E. Schwartz, E. Epel, C. Kirschbaum, S. Sidney, and T. Seeman. Socioeconomic status, race, and diurnal cortisol decline in the coronary artery risk development in young adults (cardia) study. *Psychosom Med*, 68(1):41–50, 2006.
- [89] S. J. Colloby, M. J. Firbank, A. Vasudev, S. W. Parry, A. J. Thomas, and J. T. O’Brien. Cortical thickness and vbm-dartel in late-life depression. *J Affect Disord*, 133(1-2):158–164, Sep 2011.
- [90] B. Condon, R. Grant, D. Hadley, and A. Lawrence. Brain and intracranial cavity volumes: in vivo determination by mri. *Acta Neurol Scand*, 78(5):387–393, Nov 1988.
- [91] R. D. Conger and M. B. Donnellan. An interactionist perspective on the socioeconomic context of human development. *Annu Rev Psychol*, 58:175–199, 2007.

- [92] J. D. Coplan, M. W. Andrews, L. A. Rosenblum, M. J. Owens, S. Friedman, J. M. Gorman, and C. B. Nemeroff. Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult nonhuman primates exposed to early-life stressors: implications for the pathophysiology of mood and anxiety disorders. *Proc Natl Acad Sci U S A*, 93(4):1619–1623, Feb 1996.
- [93] J. D. Coplan, S. J. Mathew, C. G. Abdallah, X. Mao, J. G. Kral, E. L. P. Smith, L. A. Rosenblum, T. D. Perera, A. J. Dwork, P. R. Hof, J. M. Gorman, and D. C. Shungu. Early-life stress and neurometabolites of the hippocampus. *Brain Res*, 1358:191–199, Oct 2010.
- [94] J. T. Coyle. The nagging question of the function of n-acetylaspartylglutamate. *Neurobiol Dis*, 4(3-4):231–238, 1997.
- [95] F. I. M. Craik and T. A. Salthouse. *The handbook of aging and cognition*. Routledge, 2000.
- [96] J. R. Crawford, K. M. Allan, R. H. Cochrane, and D. M. Parker. Assessing the validity of nart-estimated premorbid iqs in the individual case. *Br J Clin Psychol*, 29 (Pt 4):435–436, Nov 1990.
- [97] J. G. Csernansky, L. Wang, J. Swank, J. P. Miller, M. Gado, D. McKeel, M. I. Miller, and J. C. Morris. Preclinical detection of alzheimer’s disease: hippocampal shape and volume predict dementia onset in the elderly. *Neuroimage*, 25(3):783–792, Apr 2005.
- [98] J. G. Csernansky, L. Wang, J. Swank, J. P. Miller, M. Gado, D. McKeel, M. I. Miller, and J. C. Morris. Preclinical detection of alzheimer’s disease: hippocampal shape and volume predict dementia onset in the elderly. *Neuroimage*, 25(3):783–792, Apr 2005.
- [99] A. Curtatolo, P. D Arcangelo, A. Lino, and A. Brancati. Distribution of n-acetyl-aspartic and n-acetyl-aspartyl-glutamic acids in nervous tissue. *J Neurochem*, 12:339–342, Apr 1965.
- [100] B. Czéh and P. J. Lucassen. What causes the hippocampal volume decrease in depression? are neurogenesis, glial changes and apoptosis implicated? *Eur Arch Psychiatry Clin Neurosci*, 257(5):250–260, Aug 2007.
- [101] B. Czéh, T. Michaelis, T. Watanabe, J. Frahm, G. de Biurrun, M. van Kampen, A. Bartolomucci, and E. Fuchs. Stress-induced changes in cerebral metabolites, hippocampal volume, and cell proliferation are prevented by antidepressant treatment with tianeptine. *Proc Natl Acad Sci U S A*, 98(22):12796–12801, Oct 2001.

- [102] A.M. Dam. The density of neurons in the human hippocampus. *Neuropathology and Applied Neurobiology*, 5(4):249–264, 1979.
- [103] P. D’Arcangelo and A. Brancati. Distribution of n-acetylaspartate, n-acetylaspartylglutamate, free glutamate and aspartate following complete mesencephalic transection in rat neuraxis. *Neurosci Lett*, 114(1):82–88, Jun 1990.
- [104] M. D. DeBellis and M. Kuchibhatla. Cerebellar volumes in pediatric maltreatment-related posttraumatic stress disorder. *Biol Psychiatry*, 60(7):697–703, Oct 2006.
- [105] S. A. Decker. Salivary cortisol and social status among dominican men. *Horm Behav*, 38(1):29–38, Aug 2000.
- [106] C. C. Dickey, R. W. McCarley, M. L. Xu, L. J. Seidman, M. M. Voglmaier, M. A. Niznikiewicz, E. Connor, and M. E. Shenton. Mri abnormalities of the hippocampus and cavum septi pellucidi in females with schizotypal personality disorder. *Schizophr Res*, 89(1-3):49–58, Jan 2007.
- [107] B. P. Dohrenwend. The role of adversity and stress in psychopathology: some evidence and its implications for theory and research. *J Health Soc Behav*, 41(1):1–19, Mar 2000.
- [108] J. B. Dowd and N. Goldman. Do biomarkers of stress mediate the relation between socioeconomic status and health? *J Epidemiol Community Health*, 60(7):633–639, Jul 2006.
- [109] J. B. Dowd, A. M. Simanek, and A. E. Aiello. Socio-economic status, cortisol and allostatic load: a review of the literature. *Int J Epidemiol*, 38(5):1297–1309, Oct 2009.
- [110] I. Driscoll, D. A Hamilton, H. Petropoulos, R. A. Yeo, W. M. Brooks, R. N. Baumgartner, and R. J. Sutherland. The aging hippocampus: cognitive, biochemical and structural findings. *Cereb Cortex*, 13(12):1344–1351, Dec 2003.
- [111] R. S. Duman, J. Malberg, and J. Thome. Neural plasticity to stress and antidepressant treatment. *Biol Psychiatry*, 46(9):1181–1191, Nov 1999.
- [112] R.S. Duman. Depression: a case of neuronal life and death? *Biological psychiatry*, 56(3):140–145, 2004.
- [113] G. J. Duncan, C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, H. Sexton, K. Duckworth, and C. Japel. School readiness and later achievement. *Dev Psychol*, 43(6):1428–1446, Nov 2007.

- [114] H M. Duvernoy. *Human Hippocampus*. Springer-Verlag Berlin and Heidelberg GMBH & CO. KG, 2004.
- [115] M. Ereciska and I. A. Silver. Atp and brain function. *J Cereb Blood Flow Metab*, 9(1):2–19, Feb 1989.
- [116] P. S. Eriksson, E. Perfilieva, T. Björk-Eriksson, A. M. Alborn, C. Nordborg, D. A. Peterson, and F. H. Gage. Neurogenesis in the adult human hippocampus. *Nat Med*, 4(11):1313–1317, Nov 1998.
- [117] S. H. Eriksson, M. Thom, P. A. Bartlett, M. R. Symms, A. W. McEvoy, S. M. Sisodiya, and J. S. Duncan. Propeller mri visualizes detailed pathology of hippocampal sclerosis. *Epilepsia*, 49(1):33–39, Jan 2008.
- [118] G. W. Evans. The environment of childhood poverty. *Am Psychol*, 59(2):77–92, 2004.
- [119] G. W. Evans and K. English. The environment of poverty: multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child Dev*, 73(4):1238–1248, 2002.
- [120] G.W. Evans and P. Kim. Childhood poverty and health. *Psychological Science*, 18(11):953, 2007.
- [121] G. M. Fatterpekar, T. P. Naidich, B. N. Delman, J. G. Aguinaldo, S. H. Gultekin, C. C. Sherwood, P. R. Hof, B. P. Drayer, and Z. A. Fayad. Cytoarchitecture of the human cerebral cortex: Mr microscopy of excised specimens at 9.4 tesla. *AJNR Am J Neuroradiol*, 23(8):1313–1321, Sep 2002.
- [122] MB First, M. Gibbon, RL Spitzer, JBW Williams, and L.S. Benjamin. User’s guide for the structured clinical interview for dsm-iv axis ii personality disorders. *Washington, DC: American Psychiatric Press. Grilo CM, McGlashan TH, Quinlan DM, Walker ML, Greenfeld D, Edell WS (1998): Frequency of personality disorders in two age cohorts of psychiatric inpatients. American Journal of Psychiatry*, 155:140–142, 1997.
- [123] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van. der. Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, Jan 2002.
- [124] B. Fischl, D. H. Salat, A. J. W. van. der. Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale. Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, 23 Suppl 1:S69–S84, 2004.

- [125] P. A. Fisher, M. R. Gunnar, P. Chamberlain, and J. B. Reid. Preventive intervention for maltreated preschool children: impact on children's behavior, neuroendocrine activity, and foster parent functioning. *J Am Acad Child Adolesc Psychiatry*, 39(11):1356–1364, Nov 2000.
- [126] L.A. Flashman, N.C. Andreasen, M. Flaum, II Swayze, and W. Victor. Intelligence and regional brain volumes in normal controls. *Intelligence*, 25(3):149–160, 1997.
- [127] N. Forget-Dubois, G. Dionne, J. Lemelin, D. Pérusse, R. E. Tremblay, and M. Boivin. Early child language mediates the relation between home environment and school readiness. *Child Dev*, 80(3):736–749, 2009.
- [128] D. Francis, J. Diorio, D. Liu, and M. J. Meaney. Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science*, 286(5442):1155–1158, Nov 1999.
- [129] S. D. Friedman. Comment on "magnetic resonance spectroscopy identifies neural progenitor cells in the live human brain". *Science*, 321(5889):640; author reply 640, Aug 2008.
- [130] K. J. Friston, D. E. Glaser, R. N A Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and bayesian inference in neuroimaging: applications. *Neuroimage*, 16(2):484–512, Jun 2002.
- [131] R. P. Gamss, S. E. Slasky, J. A. Bello, T. S. Miller, and S. Shinnar. Prevalence of hippocampal malrotation in a population without seizures. *AJNR Am J Neuroradiol*, 30(8):1571–1573, Sep 2009.
- [132] O. Gersten. Neuroendocrine biomarkers, social relations, and the cumulative costs of stress in taiwan. *Soc Sci Med*, 66(3):507–19; discussion 520–35, Feb 2008.
- [133] P. J. Gianaros, J. R. Jennings, L. K. Sheu, P. J. Greer, L. H. Kuller, and K. A. Matthews. Prospective reports of chronic life stress predict decreased grey matter volume in the hippocampus. *Neuroimage*, 35(2):795–803, Apr 2007.
- [134] P. J. Gianaros, J. R. Jennings, L. K. Sheu, P. J. Greer, L. H. Kuller, and K. A. Matthews. Prospective reports of chronic life stress predict decreased grey matter volume in the hippocampus. *Neuroimage*, 35(2):795–803, Apr 2007.
- [135] J. N. Giedd, J. E. Schmitt, and M. C. Neale. Structural brain magnetic resonance imaging of pediatric twins. *Hum Brain Mapp*, 28(6):474–481, Jun 2007.

- [136] V. Glover and T. G. O'Connor. Effects of antenatal stress and anxiety: Implications for development and psychiatry. *Br J Psychiatry*, 180:389–391, May 2002.
- [137] D. P. Goldberg, R. Gater, N. Sartorius, T. B. Ustun, M. Piccinelli, O. Gureje, and C. Rutter. The validity of two versions of the ghq in the who study of mental illness in general health care. *Psychol Med*, 27(1):191–197, Jan 1997.
- [138] C. D. Good, J. Ashburner, and R. S. Frackowiak. Computational neuroanatomy: new perspectives for neuroradiology. *Rev Neurol (Paris)*, 157(8-9 Pt 1):797–806, Sep 2001.
- [139] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. Henson, K. J. Friston, and R. S. Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage*, 14(1 Pt 1):21–36, Jul 2001.
- [140] E. Goodman, B. S. McEwen, B. Huang, L. M. Dolan, and N. E. Adler. Social inequalities in biomarkers of cardiovascular risk in adolescence. *Psychosom Med*, 67(1):9–15, 2005.
- [141] E. Gould, B. S. McEwen, P. Tanapat, L. A. Galea, and E. Fuchs. Neurogenesis in the dentate gyrus of the adult tree shrew is regulated by psychosocial stress and nmda receptor activation. *J Neurosci*, 17(7):2492–2498, Apr 1997.
- [142] E. Gould, P. Tanapat, T. Rydel, and N. Hastings. Regulation of hippocampal neurogenesis in adulthood. *Biol Psychiatry*, 48(8):715–720, Oct 2000.
- [143] V. Govindaraju, K. Young, and A. A. Maudsley. Proton nmr chemical shifts and coupling constants for brain metabolites. *NMR Biomed*, 13(3):129–153, May 2000.
- [144] T. J. Grabowski, R. J. Frank, C. K. Brown, H. Damasio, L. L. Ponto, G. L. Watkins, and R. D. Hichwa. Reliability of pet activation across statistical methods, subject groups, and sample sizes. *Hum Brain Mapp*, 4(1):23–46, 1996.
- [145] W. S. Grolnick, S. T. Gurland, W. DeCoursey, and K. Jacob. Antecedents and consequences of mothers' autonomy support: an experimental investigation. *Dev Psychol*, 38(1):143–155, Jan 2002.
- [146] M. R Gunnar, P. A. Fisher, Stress The Early Experience, and Prevention Network. Bringing basic research on early experience and stress neurobiology to bear on preventive interventions for neglected and maltreated children. *Dev Psychopathol*, 18(3):651–677, 2006.

- [147] G. Gupta. Depression and suicidal ideation in dermatology patients with acne, alopecia areata, atopic dermatitis and psoriasis. *British Journal of Dermatology*, 139(5):846–850, 1998.
- [148] T. V. Gurvits, M. E. Shenton, H. Hokama, H. Ohta, N. B. Lasko, M. W. Gilbertson, S. P. Orr, R. Kikinis, F. A. Jolesz, R. W. McCarley, and R. K. Pitman. Magnetic resonance imaging study of hippocampal volume in chronic, combat-related posttraumatic stress disorder. *Biol Psychiatry*, 40(11):1091–1099, Dec 1996.
- [149] D. A. Hackman and M. J. Farah. Socioeconomic status and the developing brain. *Trends Cogn Sci*, 13(2):65–73, Feb 2009.
- [150] D. A. Hackman, M. J. Farah, and M. J. Meaney. Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nat Rev Neurosci*, 11(9):651–659, Sep 2010.
- [151] R. J Haier, R. E. Jung, R. A. Yeo, K. Head, and M. T. Alkire. Structural brain variation and general intelligence. *Neuroimage*, 23(1):425–433, Sep 2004.
- [152] M. Halber, K. Herholz, K. Wienhard, G. Pawlik, and W. D. Heiss. Performance of a randomization test for single-subject (15)o-water pet activation studies. *J Cereb Blood Flow Metab*, 17(10):1033–1039, Oct 1997.
- [153] M. Hamilton. A rating scale for depression. *J Neurol Neurosurg Psychiatry*, 23:56–62, Feb 1960.
- [154] A. Hammers, R. Heckemann, M. J. Koepp, J. S. Duncan, Jo. V. Hajnal, D. Rueckert, and P. Aljabar. Automatic detection and quantification of hippocampal atrophy on mri in temporal lobe epilepsy: a proof-of-principle study. *Neuroimage*, 36(1):38–47, May 2007.
- [155] M. Hanamiya, Y. Korogi, S. Kakeda, N. Ohnari, K. Kamada, J. Moriya, T. Sato, M. Kitajima, N. Akamatsu, and S. Tsuji. Partial loss of hippocampal striation in medial temporal lobe epilepsy: pilot evaluation with high-spatial-resolution t2-weighted mr imaging at 3.0 t. *Radiology*, 251(3):873–881, Jun 2009.
- [156] J. L. Hanson, A. Chandra, B. L. Wolfe, and S. D. Pollak. Association between income and the hippocampus. *PLoS One*, 6(5):e18712, 2011.
- [157] M. K. Harte, S. B. Powell, L. M. Reynolds, N. R. Swerdlow, M. A. Geyer, and G. P. Reynolds. Reduced n-acetylaspartate in the temporal cortex of rats reared in isolation. *Biol Psychiatry*, 56(4):296–299, Aug 2004.

- [158] S. Hayasaka and T. E. Nichols. Combining voxel intensity and cluster extent with permutation test framework. *Neuroimage*, 23(1):54–63, Sep 2004.
- [159] S. Hayasaka, K. L. Phan, I. Liberzon, K. J. Worsley, and T. E. Nichols. Nonstationary cluster-size inference with random field and permutation methods. *Neuroimage*, 22(2):676–687, Jun 2004.
- [160] S. Hayley, M. O. Poulter, Z. Merali, and H. Anisman. The pathogenesis of clinical depression: stressor- and cytokine-induced alterations of neuroplasticity. *Neuroscience*, 135(3):659–678, 2005.
- [161] S. Hayley, MO Poulter, Z. Merali, and H. Anisman. The pathogenesis of clinical depression: stressor-and cytokine-induced alterations of neuroplasticity. *Neuroscience*, 135(3):659–678, 2005.
- [162] J. C. Hoch, M. W. Maciejewski, and M. R. Gryk. Comment on "magnetic resonance spectroscopy identifies neural progenitor cells in the live human brain". *Science*, 321(5889):640; author reply 640, Aug 2008.
- [163] A. P. Holmes, R. C. Blair, J. D. Watson, and I. Ford. Nonparametric analysis of statistic images from functional mapping experiments. *J Cereb Blood Flow Metab*, 16(1):7–22, Jan 1996.
- [164] A. P. Holmes, J. D. Watson, and T. E. Nichols. Holmes and watson on 'sherlock'. *J Cereb Blood Flow Metab*, 18(6):697–698, Jun 1998.
- [165] A. House, P. Knapp, J. Bamford, and A. Vail. Mortality at 12 and 24 months after stroke may be associated with depressive symptoms at 1 month. *Stroke*, 32(3):696–701, Mar 2001.
- [166] K. Hu, C. Song, H. Yu, T. M. Swager, and Robert G Griffin. High-frequency dynamic nuclear polarization using biradicals: a multifrequency epr lineshape analysis. *J Chem Phys*, 128(5):052302, Feb 2008.
- [167] S. Hu, M. Lustig, A. Balakrishnan, P. E. Z. Larson, R. Bok, J. Kurhanewicz, S. J. Nelson, A. Goga, J. M. Pauly, and D. B. Vigneron. 3d compressed sensing for highly accelerated hyperpolarized (^{13}C) mrsi with in vivo applications to transgenic mouse models of cancer. *Magn Reson Med*, 63(2):312–321, Feb 2010.
- [168] C. Hutton, B. Draganski, J. Ashburner, and N. Weiskopf. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *Neuroimage*, 48(2):371–380, Nov 2009.

- [169] B. L. Jacobs, H. Praag, and F. H. Gage. Adult brain neurogenesis and psychiatry: a novel theory of depression. *Mol Psychiatry*, 5(3):262–269, May 2000.
- [170] J. F. A. Jansen, J. D. Gearhart, and J. W. M. Bulte. Comment on "magnetic resonance spectroscopy identifies neural progenitor cells in the live human brain". *Science*, 321(5889):640;author reply 640, Aug 2008.
- [171] B. J. M. H. Jefferis, C. Power, and C. Hertzman. Birth weight, childhood socioeconomic environment, and cognitive development in the 1958 british birth cohort study. *BMJ*, 325(7359):305, Aug 2002.
- [172] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, Oct 2002.
- [173] M. Jenkinson and S. Smith. A global optimisation method for robust affine registration of brain images. *Med Image Anal*, 5(2):143–156, Jun 2001.
- [174] C. Jeukens, M. Vlooswijk, H. Majoie, M. de Krom, A. P. Aldenkamp, P. A. M. Hofman, J. F. A. Jansen, and W. H. Backes. Hippocampal mri volumetry at 3 tesla: reliability and practical guidance. *Invest Radiol*, 44(9):509–517, Sep 2009.
- [175] A. Karl and A. Werner. The use of proton magnetic resonance spectroscopy in ptsd research—meta-analyses of findings and methodological review. *Neurosci Biobehav Rev*, 34(1):7–22, Jan 2010.
- [176] J. Keller, L. Shen, R. G. Gomez, A. Garrett, H. B. Solvason, A. Reiss, and A. F. Schatzberg. Hippocampal and amygdalar volumes in psychotic and nonpsychotic unipolar depression. *Am J Psychiatry*, 165(7):872–880, Jul 2008.
- [177] S. Kelly, C. Hertzman, and M. Daniels. Searching for the biological pathways between stress and health. *Annu Rev Public Health*, 18:437–462, 1997.
- [178] G. Kempermann, H. G. Kuhn, and F. H. Gage. More hippocampal neurons in adult mice living in an enriched environment. *Nature*, 386(6624):493–495, Apr 1997.
- [179] A. Khiat, Z. Yared, C. Bard, A. Lacroix, and Y. Boulanger. Long-term brain metabolic alterations in exogenous cushing's syndrome as monitored by proton magnetic resonance spectroscopy. *Brain Res*, 911(2):134–140, Aug 2001.

- [180] E. L. Kier, J. H. Kim, R. K. Fulbright, and R. A. Bronen. Embryology of the human fetal hippocampus: Mr imaging, anatomy, and histology. *AJNR Am J Neuroradiol*, 18(3):525–532, Mar 1997.
- [181] C. Kirschbaum and D. H. Hellhammer. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. *Psychoneuroendocrinology*, 19(4):313–333, 1994.
- [182] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802, Jul 2009.
- [183] W. E. Klunk, C. Xu, K. Panchalingam, R. J. McClure, and J. W. Pettegrew. Quantitative 1h and 31p mrs of pca extracts of postmortem alzheimer’s disease brain. *Neurobiol Aging*, 17(3):349–357, 1996.
- [184] K. J. Koller, R. Zaczek, and J. T. Coyle. N-acetyl-aspartyl-glutamate: regional levels in rat brain and the effects of brain lesions as determined by a new hplc method. *J Neurochem*, 43(4):1136–1142, Oct 1984.
- [185] C. Konrad, T. Ukas, C. Nebel, V. Arolt, A. W. Toga, and K. L. Narr. Defining the human hippocampus in cerebral magnetic resonance images—an overview of current segmentation protocols. *Neuroimage*, 47(4):1185–1195, Oct 2009.
- [186] P. Cédric M P Koolschijn, Neeltje E M van Haren, Hugo G Schnack, Joost Janssen, Hilleke E Hulshoff Pol, and René S Kahn. Cortical thickness and voxel-based morphometry in depressed elderly. *Eur Neuropsychopharmacol*, Mar 2010.
- [187] R. Kreis, T. Ernst, and B. D. Ross. Development of the human brain: in vivo quantification of metabolite and water content with proton magnetic resonance spectroscopy. *Magn Reson Med*, 30(4):424–437, Oct 1993.
- [188] N. Krieger, D. R. Williams, and N. E. Moss. Measuring social class in us public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health*, 18:341–378, 1997.
- [189] H. G. Kuhn, H. Dickinson-Anson, and F. H. Gage. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J Neurosci*, 16(6):2027–2033, Mar 1996.

- [190] S. R. Kunz-Ebrecht, C. Kirschbaum, M. Marmot, and A. Steptoe. Differences in cortisol awakening response on work days and weekends in women and men from the whitehall ii cohort. *Psychoneuroendocrinology*, 29(4):516–528, May 2004.
- [191] L. J. Lengua. The contribution of emotionality and self-regulation to the understanding of children’s response to multiple risk. *Child Dev*, 73(1):144–161.
- [192] B. Leuner, Y. Kozorovitskiy, C. G. Gross, and E. Gould. Diminished adult neurogenesis in the marmoset brain precedes old age. *Proc Natl Acad Sci U S A*, 104(43):17169–17173, Oct 2007.
- [193] A. Levine, O. Zagoory-Sharon, R. Feldman, J. G. Lewis, and A. Weller. Measuring cortisol in human psychobiological studies. *Physiol Behav*, 90(1):43–53, Jan 2007.
- [194] DV Lewis, S. Chan, JA Bello, et al. HIMAL is a malformation that predisposes to prolonged febrile seizures: data from the FEBSTAT Study. *Epilepsia*, 47(Suppl 4):16, 2006.
- [195] Y. Li, S. Ga, Y. Huo, S. Li, and X. Gao. Characteristics of hippocampal volumes in healthy chinese from mri. *Neurol Res*, 29(8):803–806, Dec 2007.
- [196] A. J. Lloyd, I. N. Ferrier, R. Barber, A. Gholkar, A. H Young, and J. T. O’Brien. Hippocampal volume change in depression: late- and early-onset illness compared. *Br J Psychiatry*, 184:488–495, Jun 2004.
- [197] A. Louissaint, S. Rao, C. Leventhal, and S. A. Goldman. Coordinated interaction of neurogenesis and angiogenesis in the adult songbird brain. *Neuron*, 34(6):945–960, Jun 2002.
- [198] S. J. Lupien, S. King, M. J. Meaney, and B. S. McEwen. Child’s stress hormone levels correlate with mother’s socioeconomic status and depressive state. *Biol Psychiatry*, 48(10):976–980, Nov 2000.
- [199] S.J. Lupien, S. King, MJ Meaney, and BS McEwen. Can poverty get under your skin? basal cortisol levels and cognitive function in children from low and high socioeconomic status. *Development and Psychopathology*, 13(3):653–676, 2001.
- [200] M. Lustig, D. Donoho, and J. M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magn Reson Med*, 58(6):1182–1195, Dec 2007.

- [201] S. Maccari, M. Darnaudery, S. Morley-Fletcher, A. R. Zuena, C. Cinque, and O. Van Reeth. Prenatal stress and long-term consequences: implications of glucocorticoid hormones. *Neurosci Biobehav Rev*, 27(1-2):119–127, 2003.
- [202] C. E. Mackay, J. A. Webb, P. R. Eldridge, D. W. Chadwick, G. H. Whitehouse, and N. Roberts. Quantitative magnetic resonance imaging in consecutive patients evaluated for surgical treatment of temporal lobe epilepsy. *Magn Reson Imaging*, 18(10):1187–1199, Dec 2000.
- [203] F. P. MacMaster, Y. Mirza, P. R. Szeszko, L. E. Kmiecik, P. C. Easter, S. P. Taormina, M. Lynch, M. Rose, G. J. Moore, and D. R. Rosenberg. Amygdala and hippocampal volumes in familial early onset major depressive disorder. *Biol Psychiatry*, 63(4):385–390, Feb 2008.
- [204] G. M. MacQueen, S. Campbell, B. S. McEwen, K. Macdonald, S. Amano, R. T. Joffe, C. Nahmias, and L. T. Young. Course of illness, hippocampal function, and hippocampal volume in major depression. *Proc Natl Acad Sci U S A*, 100(3):1387–1392, Feb 2003.
- [205] T. M. Madsen, A. Treschow, J. Bengzon, T. G. Bolwig, O. Lindvall, and A. Tingström. Increased neurogenesis in a model of electroconvulsive therapy. *Biol Psychiatry*, 47(12):1043–1049, Jun 2000.
- [206] M. Maes. The cytokine hypothesis of depression: inflammation, oxidative & nitrosative stress (io&ns) and leaky gut as new targets for adjunctive treatments in depression. *Neuro Endocrinol Lett*, 29(3):287–291, Jun 2008.
- [207] EA Maguire, RSJ Frackowiak, and CD Frith. Learning to find your way: a role for the human hippocampal formation. *Proceedings: Biological Sciences*, 263(1377):1745–1750, 1996.
- [208] J. E. Malberg and L. E. Schechter. Increasing hippocampal neurogenesis: a novel mechanism for antidepressant drugs. *Curr Pharm Des*, 11(2):145–155, 2005.
- [209] J.E. Malberg, A.J. Eisch, E.J. Nestler, and R.S. Duman. Chronic antidepressant treatment increases neurogenesis in adult rat hippocampus. *Journal of Neuroscience*, 20(24):9104, 2000.
- [210] J. J. Maller, Z. J. Daskalakis, R. H. S. Thomson, M. Daigle, M. S. Barr, and P. B. Fitzgerald. Hippocampal volumetrics in treatment-resistant depression and schizophrenia: The devil’s in de-tail. *Hippocampus*, Sep 2010.

- [211] J. J. Maller, C. Réglade-Meslin, K. J. Anstey, and P. Sachdev. Sex and symmetry differences in hippocampal volumetrics: before and beyond the opening of the crus of the fornix. *Hippocampus*, 16(1):80–90, 2006.
- [212] N. V. Malykhin, T. P. Bouchard, C. J. Ogilvie, N. J. Coupland, P. Seres, and R. Camicioli. Three-dimensional volumetric analysis and reconstruction of amygdala and hippocampal head, body and tail. *Psychiatry Res*, 155(2):155–165, Jul 2007.
- [213] N. V. Malykhin, R. M. Lebel, N. J. Coupland, A. H. Wilman, and R. Carter. In vivo quantification of hippocampal subfields using 4.7 t fast spin echo imaging. *Neuroimage*, 49(2):1224–1230, Jan 2010.
- [214] N. V. Malykhin, R. M. Lebel, N. J. Coupland, A. H. Wilman, and R. Carter. In vivo quantification of hippocampal subfields using 4.7 t fast spin echo imaging. *Neuroimage*, 49(2):1224–1230, Jan 2010.
- [215] L. N. Manganas, X. Zhang, Y. Li, R. D. Hazel, S. D. Smith, M. E. Wagshul, F. Henn, H. Benveniste, P. M. Djuric, G. Enikolopov, and M. Maletic-Savatic. Magnetic resonance spectroscopy identifies neural progenitor cells in the live human brain. *Science*, 318(5852):980–985, Nov 2007.
- [216] A. L. Marsland, P. J. Gianaros, S. M. Abramowitch, S. B. Manuck, and A. R. Hariri. Interleukin-6 covaries inversely with hippocampal grey matter volume in middle-aged adults. *Biol Psychiatry*, 64(6):484–490, Sep 2008.
- [217] A. L. Marsland, P. J. Gianaros, A. A. Prather, J. R. Jennings, S. A. Neumann, and S. B. Manuck. Stimulated production of proinflammatory cytokines covaries inversely with heart rate variability. *Psychosom Med*, 69(8):709–716, Nov 2007.
- [218] A. S. Masten. Ordinary magic. resilience processes in development. *Am Psychol*, 56(3):227–238, Mar 2001.
- [219] S. G. Matthews and D. I. W. Phillips. Minireview: transgenerational inheritance of the stress response: a new frontier in stress research. *Endocrinology*, 151(1):7–13, Jan 2010.
- [220] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima, and

- B. Mazoyer. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). *Philos Trans R Soc Lond B Biol Sci*, 356(1412):1293–1322, Aug 2001.
- [221] M.A. McDaniel. Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence*, 33(4):337–346, 2005.
- [222] B. S. McEwen. The neurobiology of stress: from serendipity to clinical relevance. *Brain Res*, 886(1-2):172–189, Dec 2000.
- [223] B. S. McEwen. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiol Rev*, 87(3):873–904, Jul 2007.
- [224] B. S. McEwen, H. Cameron, H. M. Chao, E. Gould, A. M. Magarinos, Y. Watanabe, and C. S. Woolley. Adrenal steroids and plasticity of hippocampal neurons: toward an understanding of underlying cellular and molecular mechanisms. *Cell Mol Neurobiol*, 13(4):457–482, Aug 1993.
- [225] B. S. McEwen, P. G. Davis, B. Parsons, and D. W. Pfaff. The brain as a target for steroid hormone action. *Annu Rev Neurosci*, 2:65–112, 1979.
- [226] B. S. McEwen and P. J. Gianaros. Central role of the brain in stress and adaptation: links to socioeconomic status, health, and disease. *Ann N Y Acad Sci*, 1186:190–222, Feb 2010.
- [227] B. S. McEwen and P. J. Gianaros. Stress- and allostasis-induced brain plasticity. *Annu Rev Med*, 62:431–445, Feb 2011.
- [228] B. S. McEwen, J. M. Weiss, and L. S. Schwartz. Selective retention of corticosterone by limbic structures in rat brain. *Nature*, 220(5170):911–912, Nov 1968.
- [229] S. McGrath and R. Smith. Prediction of preterm delivery using plasma corticotrophin-releasing hormone and other biochemical variables. *Ann Med*, 34(1):28–36, 2002.
- [230] K. O. McGraw and S. P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1):30 – 46, 1996.
- [231] I.B. McInnes and G. Schett. Cytokines in the pathogenesis of rheumatoid arthritis. *Nature Reviews Immunology*, 7(6):429–442, 2007.
- [232] M. C. McKinnon, K. Yucel, A. Nazarov, and G. M. MacQueen. A meta-analysis examining clinical predictors of hippocampal volume in patients with major depressive disorder. *J Psychiatry Neurosci*, 34(1):41–54, Jan 2009.

- [233] V. C. McLoyd. Socioeconomic disadvantage and child development. *Am Psychol*, 53(2):185–204, Feb 1998.
- [234] M. J. Meaney, M. Szyf, and J. R. Seckl. Epigenetic mechanisms of perinatal programming of hypothalamic-pituitary-adrenal function and health. *Trends Mol Med*, 13(7):269–277, Jul 2007.
- [235] M. A. Mehta, N. I. Golembo, C. Nosarti, E. Colvert, A. Mota, S. C. R. Williams, M. Rutter, and E. J. S. Sonuga-Barke. Amygdala, hippocampal and corpus callosum size following severe early institutional deprivation: the english and romanian adoptees study pilot. *J Child Psychol Psychiatry*, 50(8):943–951, Aug 2009.
- [236] L. C. Meiners, A. van Gils, G. H. Jansen, G. de Kort, T. D. Witkamp, L. M. Ramos, J. Valk, R. M. Debets, A. C. van Huffelen, and C. W. van Veelen. Temporal lobe epilepsy: the various mr appearances of histologically proven mesial temporal sclerosis. *AJNR Am J Neuroradiol*, 15(8):1547–1555, Sep 1994.
- [237] E. Mervaala, J. Föhr, M. Könönen, M. Valkonen-Korhonen, P. Vainio, K. Partanen, J. Partanen, J. Tiihonen, H. Viinamäki, A. K. Karjalainen, and J. Lehtonen. Quantitative mri of the hippocampus and amygdala in severe depression. *Psychol Med*, 30(1):117–125, Jan 2000.
- [238] T. Michaelis, K. D. Merboldt, H. Bruhn, W. Hänicke, and J. Frahm. Absolute concentrations of metabolites in the adult human brain in vivo: quantification of localized proton mr spectra. *Radiology*, 187(1):219–227, Apr 1993.
- [239] M. W. Miller. Effect of early exposure to ethanol on the protein and dna contents of specific brain regions in the rat. *Brain Res*, 734(1-2):286–294, Sep 1996.
- [240] M. Miyake, Y. Kakimoto, and M. Sorimachi. A gas chromatographic method for the determination of n-acetyl-l-aspartic acid, n-acetyl-alpha-aspartylglutamic acid and beta-citryl-l-glutamic acid and their distributions in the brain and other organs of various species of animals. *J Neurochem*, 36(3):804–810, Mar 1981.
- [241] V. Mondelli, A. Cattaneo, M. B. Murri, M. DiForti, R. Handley, N. Hepgul, A. Miorrelli, S. Navari, A. S. Papadopoulos, K. J. Aitchison, C. Morgan, R. M. Murray, P. Dazzan, and C. M. Pariante. Stress and inflammation reduce brain-derived neurotrophic factor expression in first-episode psychosis: a pathway to smaller hippocampal volume. *J Clin Psychiatry*, May 2011.

- [242] M. L. Monje, H. Toda, and T. D. Palmer. Inflammatory blockade restores adult hippocampal neurogenesis. *Science*, 302(5651):1760–1765, Dec 2003.
- [243] M. A. Montenegro, D. Kinay, F. Cendes, A. Bernasconi, N. Bernasconi, A. C. Coan, L. M. Li, M. M. Guerreiro, C. A M Guerreiro, I. Lopes-Cendes, E. Andermann, F. Dubeau, and F. Andermann. Patterns of hippocampal abnormalities in malformations of cortical development. *J Neurol Neurosurg Psychiatry*, 77(3):367–371, Mar 2006.
- [244] T. W. J. Moorhead, D. E. Job, M. D. Spencer, H. C. Whalley, E. C. Johnstone, and S. M. Lawrie. Empirical comparison of maximal voxel and non-isotropic adjusted cluster extent results in a voxel-based morphometry study of comorbid learning disability with schizophrenia. *Neuroimage*, 28(3):544–552, Nov 2005.
- [245] R. A. Morey, C. M. Petty, Y. Xu, J. P. Hayes, H. R. Wagner, D. V. Lewis, K. S. LaBar, M. Styner, and G. McCarthy. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage*, 45(3):855–866, Apr 2009.
- [246] R. A. Morey, C. M. Petty, Y. Xu, J. P. Hayes, H. R. Wagner, D. V. Lewis, K. S. LaBar, M. Styner, and G. McCarthy. Rebuttal to hasan and pedraza in comments and controversies: "improving the reliability of manual and automated methods for hippocampal and amygdala volume measurements". *Neuroimage*, 48(3):499–500, Nov 2009.
- [247] A. B. Morgan and S. O. Lilienfeld. A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clin Psychol Rev*, 20(1):113–136, Jan 2000.
- [248] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, C. Avedissian, S. K. Madsen, N. Parikshak, A. W. Toga, C. R. Jack, N. Schuff, M. W. Weiner, P. M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Automated mapping of hippocampal atrophy in 1-year repeat mri data from 490 subjects with alzheimer's disease, mild cognitive impairment, and elderly controls. *Neuroimage*, 45(1 Suppl):S3–15, Mar 2009.
- [249] R. Mossner, A. Heils, G. Stober, O. Okladnova, S. Daniel, and K.P. Lesch. Enhancement of serotonin transporter function by tumor necrosis factor alpha but not by interleukin-6. *Neurochemistry international*, 33(3):251–254, 1998.

- [250] A. Mouiha and S. Duchesne. Multi-decade hippocampal and amygdala volume analysis: equal variability and limited age effect. *Neurosci Lett*, 499(2):93–98, Jul 2011.
- [251] C. Mueller, J. Scorzin, R. Koenig, H. Urbach, R. Fimmers, J. Zentner, T. Lehmann, and J. Schramm. Comparison of manual tracing versus a semiautomatic radial measurement method in temporal lobe mri volumetry for pharmaco-resistant epilepsy. *Neuroradiology*, 49(3):189–201, Mar 2007.
- [252] S. G. Mueller, K. D. Laxer, J. Barakos, I. Cheong, P. Garcia, and M. W. Weiner. Subfield atrophy pattern in temporal lobe epilepsy with and without mesial sclerosis detected by high-resolution mri at 4 tesla: preliminary results. *Epilepsia*, 50(6):1474–1483, Jun 2009.
- [253] S. G. Mueller, K. D. Laxer, J. Barakos, I. Cheong, P. Garcia, and M. W. Weiner. Subfield atrophy pattern in temporal lobe epilepsy with and without mesial sclerosis detected by high-resolution mri at 4 tesla: preliminary results. *Epilepsia*, 50(6):1474–1483, Jun 2009.
- [254] S. G. Mueller, N. Schuff, S. Raptentsetsang, J. Elman, and M. W. Weiner. Selective effect of apo e4 on ca3 and dentate in normal aging and alzheimer’s disease using high resolution mri at 4 t. *Neuroimage*, 42(1):42–48, Aug 2008.
- [255] S. G. Mueller, N. Schuff, S. Raptentsetsang, J. Elman, and M. W. Weiner. Selective effect of apo e4 on ca3 and dentate in normal aging and alzheimer’s disease using high resolution mri at 4 t. *Neuroimage*, 42(1):42–48, Aug 2008.
- [256] S. G. Mueller, N. Schuff, K. Yaffe, C. Madison, B. Miller, and M. W. Weiner. Hippocampal atrophy patterns in mild cognitive impairment and alzheimer’s disease. *Hum Brain Mapp*, 31(9):1339–1347, Sep 2010.
- [257] S. G. Mueller, L. Stables, A. T. Du, N. Schuff, D. Truran, N. Cashdollar, and M. W. Weiner. Measurement of hippocampal subfields and age-related changes with high resolution mri at 4t. *Neurobiol Aging*, 28(5):719–726, May 2007.
- [258] S. G. Mueller and M. W. Weiner. Selective effect of age, apo e4, and alzheimer’s disease on hippocampal subfields. *Hippocampus*, 19(6):558–564, Jun 2009.
- [259] S. G. Mueller and M. W. Weiner. Selective effect of age, apo e4, and alzheimer’s disease on hippocampal subfields. *Hippocampus*, 19(6):558–564, Jun 2009.

- [260] M. S. Murmu, S. Salomon, Y. Biala, M. Weinstock, K. Braun, and J. Bock. Changes of spine density and dendritic complexity in the prefrontal cortex in offspring of mothers exposed to stress during pregnancy. *Eur J Neurosci*, 24(5):1477–1487, Sep 2006.
- [261] C.J.L. Murray and A.D. Lopez. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *The Lancet*, 349(9063):1436–1442, 1997.
- [262] A. Neumeister, S. Wood, O. Bonne, A. C. Nugent, D. A. Luckenbaugh, T. Young, E. E. Bain, D. S. Charney, and W. C. Drevets. Reduced hippocampal volume in unmedicated, remitted patients with major depression versus control subjects. *Biol Psychiatry*, 57(8):935–937, Apr 2005.
- [263] T. C. Neylan, S. G. Mueller, Z. Wang, T. J. Metzler, M. Lenoci, D. Truran, C. R. Marmar, M. W. Weiner, and N. Schuff. Insomnia severity is associated with a decreased volume of the ca3/dentate gyrus hippocampal subfield. *Biol Psychiatry*, 68(5):494–496, Sep 2010.
- [264] T. C. Neylan, N. Schuff, M. Lenoci, R. Yehuda, M. W. Weiner, and C. R. Marmar. Cortisol levels are positively correlated with hippocampal n-acetylaspartate. *Biol Psychiatry*, 54(10):1118–1121, Nov 2003.
- [265] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*, 15(1):1–25, Jan 2002.
- [266] J. T. O’Brien, A. Lloyd, I. McKeith, A. Gholkar, and N. Ferrier. A longitudinal study of hippocampal volume, cortisol levels, and cognition in older depressed subjects. *Am J Psychiatry*, 161(11):2081–2090, Nov 2004.
- [267] T. G. O’Connor, K. Deater-Deckard, D. Fulker, M. Rutter, and R. Plomin. Genotype-environment correlations in late childhood and early adolescence: antisocial behavioral problems and coercive parenting. *Dev Psychol*, 34(5):970–981, Sep 1998.
- [268] D. L. Olds, J. Eckenrode, C. R. Henderson, H. Kitzman, J. Powers, R. Cole, K. Sidora, P. Morris, L. M. Pettitt, and D. Luckey. Long-term effects of home visitation on maternal life course and child abuse and neglect. fifteen-year follow-up of a randomized trial. *JAMA*, 278(8):637–643, Aug 1997.
- [269] L. Ory-Lavollée, R. D. Blakely, and J. T. Coyle. Neurochemical and immunocytochemical studies on the distribution of n-acetyl-aspartylglutamate and n-acetyl-aspartate in rat spinal cord and some peripheral nervous tissues. *J Neurochem*, 48(3):895–899, Mar 1987.

- [270] H. Ozdemir and M. Rezaki. [general health questionnaire-12 for the detection of depression]. *Turk Psikiyatri Derg*, 18(1):13–21, 2007.
- [271] R. Palacios, A. Campo, K. Henningsen, M. Verhoye, D. Poot, J. Dijkstra, J. Van Audekerke, H. Benveniste, J. Sijbers, O. Wiborg, and A. Van der Linden. Magnetic resonance imaging and spectroscopy reveal differential hippocampal changes in anhedonic and resilient subtypes of the chronic mild stress rat model. *Biol Psychiatry*, 70(5):449–457, Sep 2011.
- [272] T. D. Palmer, A. R. Willhoite, and F. H. Gage. Vascular niche for adult hippocampal neurogenesis. *J Comp Neurol*, 425(4):479–494, Oct 2000.
- [273] J. W. Pan, J. T. Vaughan, R. I. Kuzniecky, G. M. Pohost, and H. P. Hetherington. High resolution neuroimaging at 4.1t. *Magn Reson Imaging*, 13(7):915–921, 1995.
- [274] J. Pantel, D. S. O’Leary, K. Cretsing, H. J. Bockholt, H. Keefe, V. A. Magnotta, and N. C. Andreasen. A new method for the in vivo volumetric measurement of the human hippocampus with high neuroanatomical accuracy. *Hippocampus*, 10(6):752–758, 2000.
- [275] H. R. Pardoe, G. S. Pell, D. F. Abbott, and G. D. Jackson. Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia*, 50(12):2586–2592, Dec 2009.
- [276] L. A. Passani, J. P. Vonsattel, R. E. Carter, and J. T. Coyle. N-acetylaspartylglutamate, n-acetylaspartate, and n-acetylated alpha-linked acidic dipeptidase in human brain and their alterations in huntington and alzheimer’s diseases. *Mol Chem Neuropathol*, 31(2):97–118, Jun 1997.
- [277] T. D. Patel and F. C. Zhou. Ontogeny of 5-ht1a receptor expression in the developing hippocampus. *Brain Res Dev Brain Res*, 157(1):42–57, Jun 2005.
- [278] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson. A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3):907–922, Jun 2011.
- [279] B. Peltier, Ph Hurtevent, G. Trehan, Ph Derambure, J-P. Pruvo, and G. Soto-Ares. [mri of hippocampal malformations in patients with intractable temporal lobe epilepsy]. *J Radiol*, 86(1):69–75, Jan 2005.
- [280] B. F. Pennington, P. A. Filipek, D. Lefly, N. Chhabildas, D. N. Kennedy, J. H. Simon, C. M. Filley, A. Galaburda, and J. C. DeFries. A twin mri study of size variations in human brain. *J Cogn Neurosci*, 12(1):223–232, Jan 2000.

- [281] A. C. Pereira, D. E. Huddleston, A. M. Brickman, A. A. Sosunov, R. Hen, G. M. McKhann, R. Sloan, F. H. Gage, T. R. Brown, and S. A. Small. An in vivo correlate of exercise-induced neurogenesis in the adult dentate gyrus. *Proc Natl Acad Sci U S A*, 104(13):5638–5643, Mar 2007.
- [282] J. M S Pereira, L. Xiong, J. Acosta-Cabronero, G. Pengas, G. B. Williams, and P. J. Nestor. Registration accuracy for vbm studies varies according to region and degenerative disease grouping. *Neuroimage*, 49(3):2205–2215, Feb 2010.
- [283] C. Van. Petten. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. *Neuropsychologia*, 42(10):1394–1413, 2004.
- [284] A. Pfefferbaum. Alcoholism damages the brain, but does moderate alcohol use? *Lancet Neurol*, 3(3):143–144, Mar 2004.
- [285] J. G. Pipe. Motion correction with propeller mri: application to head motion and free-breathing cardiac imaging. *Magn Reson Med*, 42(5):963–969, Nov 1999.
- [286] D. Posthuma, E. J. C. DeGeus, W. F. C. Baaré, H. E. H. Pol, R. S. Kahn, and D. I. Boomsma. The association between brain volume and intelligence is of genetic origin. *Nat Neurosci*, 5(2):83–84, Feb 2002.
- [287] P. J. Pouwels and J. Frahm. Differential distribution of naa and naag in human brain as determined by quantitative localized proton mrs. *NMR Biomed*, 10(2):73–78, Apr 1997.
- [288] P. J. Pouwels and J. Frahm. Regional metabolite concentrations in human brain as determined by quantitative localized proton mrs. *Magn Reson Med*, 39(1):53–60, Jan 1998.
- [289] S. W. Provencher. Automatic quantitation of localized in vivo 1h spectra with lcmoel. *NMR Biomed*, 14(4):260–264, Jun 2001.
- [290] V. Prudent, A. Kumar, S. Liu, G. Wiggins, D. Malaspina, and O. Gonen. Human hippocampal subfields in young adults at 7.0 t: feasibility of imaging. *Radiology*, 254(3):900–906, Mar 2010.
- [291] M. Pruessner, J. C. Pruessner, D. H. Hellhammer, G. B. Pike, and S. J. Lupien. The associations among hippocampal volume, cortisol reactivity, and memory performance in healthy young men. *Psychiatry Res*, 155(1):1–10, May 2007.

- [292] A. Qiu, W. D. Taylor, Z. Zhao, J. R. MacFall, M. I. Miller, C. R. Key, M. E. Payne, D. C. Steffens, and K. R. R. Krishnan. Apoe related hippocampal shape alteration in geriatric depression. *Neuroimage*, 44(3):620–626, Feb 2009.
- [293] R. Raininko and D. Bajic. "hippocampal malrotation": No real malrotation and not rare. *AJNR Am J Neuroradiol*, Jan 2010.
- [294] C. L. Raison, L. Capuron, and A. H. Miller. Cytokines sing the blues: inflammation and the pathogenesis of depression. *Trends Immunol*, 27(1):24–31, Jan 2006.
- [295] N. Ranjit, E. A. Young, and G. A. Kaplan. Material hardship alters the diurnal rhythm of salivary cortisol. *Int J Epidemiol*, 34(5):1138–1143, Oct 2005.
- [296] A. Rao, R. Chandrashekar, G. I. Sanchez-Ortiz, R. Mohiaddin, P. Aljabar, J. V. Hajnal, B. K. Puri, and D. Rueckert. Spatial transformation of motion and deformation fields using nonrigid registration. *IEEE Trans Med Imaging*, 23(9):1065–1076, Sep 2004.
- [297] N. Raz, P. Ghisletta, K. M. Rodrigue, K. M. Kennedy, and U. Lindenberger. Trajectories of brain aging in middle-aged and older adults: Regional and individual differences. *Neuroimage*, Mar 2010.
- [298] R. L. Repetti, S. E. Taylor, and T. E. Seeman. Risky families: family social environments and the mental and physical health of offspring. *Psychol Bull*, 128(2):330–366, Mar 2002.
- [299] K. M. Rodrigue, E. M. Haacke, and N. Raz. Differential effects of age and history of hypertension on regional brain volumes and iron. *Neuroimage*, 54(2):750–759, Jan 2011.
- [300] M. A. Rogers, K. Kasai, M. Koji, R. Fukuda, A. Iwanami, K. Nakagome, M. Fukuda, and N. Kato. Executive and prefrontal dysfunction in unipolar depression: a review of neuropsychological and imaging evidence. *Neurosci Res*, 50(1):1–11, Sep 2004.
- [301] C. Rorden, L. Bonilha, and T. E. Nichols. Rank-order versus mean based statistics for neuroimaging. *Neuroimage*, 35(4):1531–1537, May 2007.
- [302] L. Rosero-Bixby and W. H. Dow. Surprising ses gradients in mortality, health, and biomarkers in a latin american population of adults. *J Gerontol B Psychol Sci Soc Sci*, 64(1):105–117, Jan 2009.
- [303] R. Rosmond and P. Björntorp. Occupational status, cortisol secretory pattern, and visceral obesity in middle-aged men. *Obes Res*, 8(6):445–450, Sep 2000.

- [304] T. C. Roth and V. V. Pravosudov. Hippocampal volumes and neuron numbers increase along a gradient of environmental harshness: a large-scale comparison. *Proc Biol Sci*, 276(1656):401–405, Feb 2009.
- [305] T. M. Rudkin and D. L. Arnold. Proton magnetic resonance spectroscopy for the diagnosis and management of cerebral disorders. *Arch Neurol*, 56(8):919–926, Aug 1999.
- [306] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Trans Med Imaging*, 18(8):712–721, Aug 1999.
- [307] A. Sahay, M. R. Drew, and R. Hen. Dentate gyrus neurogenesis and depression. *Prog Brain Res*, 163:697–722, 2007.
- [308] R. M. Sapolsky. Glucocorticoid toxicity in the hippocampus: temporal aspects of neuronal vulnerability. *Brain Res*, 359(1-2):300–305, Dec 1985.
- [309] R. M. Sapolsky, M. J. Meaney, and B. S. McEwen. The development of the glucocorticoid receptor system in the rat limbic brain. iii. negative-feedback regulation. *Brain Res*, 350(1-2):169–173, Feb 1985.
- [310] R. M. Sapolsky, H. Uno, C. S. Rebert, and C. E. Finch. Hippocampal damage associated with prolonged glucocorticoid exposure in primates. *J Neurosci*, 10(9):2897–2902, Sep 1990.
- [311] A. Sarrieau, M. Dussaillant, F. Agid, D. Philibert, Y. Agid, and W. Rostene. Autoradiographic localization of glucocorticosteroid and progesterone binding sites in the human post-mortem brain. *J Steroid Biochem*, 25(5B):717–721, Nov 1986.
- [312] D. E. Saunders, F. A. Howe, A. van den Boogaart, J. R. Griffiths, and M. M. Brown. Aging of the adult human brain: in vivo quantitation of metabolite content with proton magnetic resonance spectroscopy. *J Magn Reson Imaging*, 9(5):711–716, May 1999.
- [313] W. Schlotz and D. I. W. Phillips. Fetal origins of mental health: evidence and mechanisms. *Brain Behav Immun*, 23(7):905–916, Oct 2009.
- [314] M. L. Schneider, C. F. Moore, G. W. Kraemer, A. D. Roberts, and O. T. DeJesus. The impact of prenatal stress, fetal alcohol exposure, or both on development: perspectives from a primate model. *Psychoneuroendocrinology*, 27(1-2):285–298, 2002.

- [315] N. Schuff, D. L. Amend, R. Knowlton, D. Norman, G. Fein, and M. W. Weiner. Age-related metabolite changes and volume loss in the hippocampus by magnetic resonance spectroscopy and imaging. *Neurobiol Aging*, 20(3):279–285, 1999.
- [316] N. Schuff, T. C. Neylan, S. Fox-Bosetti, M. Lenoci, K. W. Samuelson, C. Studholme, J. Kornak, C. R. Marmar, and M. W. Weiner. Abnormal n-acetylaspartate in hippocampus and anterior cingulate in posttraumatic stress disorder. *Psychiatry Res*, 162(2):147–157, Feb 2008.
- [317] N. Schuff, T. C. Neylan, M. A. Lenoci, A. T. Du, D. S. Weiss, C. R. Marmar, and M. W. Weiner. Decreased hippocampal n-acetylaspartate in the absence of atrophy in posttraumatic stress disorder. *Biol Psychiatry*, 50(12):952–959, Dec 2001.
- [318] C. M. Schumann, J. Hamstra, B. L. Goodlin-Jones, H. Kwon, A. L. Reiss, and D. G. Amaral. Hippocampal size positively correlates with verbal iq in male children. *Hippocampus*, 17(6):486–493, 2007.
- [319] J. E. Scorzin, S. Kaaden, C. M. Quesada, C. Müller, R. Fimmers, H. Urbach, and J. Schramm. Volume determination of amygdala and hippocampus at 1.5 and 3.0t mri in temporal lobe epilepsy. *Epilepsy Res*, 82(1):29–37, Nov 2008.
- [320] J. R. Seckl. Glucocorticoids, developmental 'programming' and the risk of affective dysfunction. *Prog Brain Res*, 167:17–34, 2008.
- [321] Y. I. Sheline, M. Sanghavi, M. A. Mintun, and M. H. Gado. Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *J Neurosci*, 19(12):5034–5043, Jun 1999.
- [322] T. M. Shepherd, E. Ozarslan, A. T. Yachnis, M. A. King, and S. J. Blackband. Diffusion tensor microscopy indicates the cytoarchitectural basis for diffusion anisotropy in the human hippocampus. *AJNR Am J Neuroradiol*, 28(5):958–964, May 2007.
- [323] P. E. Shrout and J. L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*, 86(2):420–428, Mar 1979.
- [324] J. Siegrist and M. Marmot. Health inequalities and the psychosocial environment—two scientific challenges. *Soc Sci Med*, 58(8):1463–1473, Apr 2004.
- [325] M. L. Simmons, C. G. Frondoza, and J. T. Coyle. Immunocytochemical localization of n-acetyl-aspartate with monoclonal antibodies. *Neuroscience*, 45(1):37–45, 1991.

- [326] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans Med Imaging*, 17(1):87–97, Feb 1998.
- [327] S. E. Smith, D. Lekiëffre, P. Sowinski, and B. S. Meldrum. Cerebroprotective effect of bw619c89 after focal or global cerebral ischaemia in the rat. *Neuroreport*, 4(12):1339–1342, Sep 1993.
- [328] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. DeLuca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. DeStefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23 Suppl 1:S208–S219, 2004.
- [329] B. J. Soher, R. E. Hurd, N. Sailasuta, and P. B. Barker. Quantitation of automated single-voxel proton mrs using cerebral water as an internal reference. *Magn Reson Med*, 36(3):335–339, Sep 1996.
- [330] H. S. Soininen, K. Partanen, A. Pitkänen, P. Vainio, T. Hänninen, M. Hallikainen, K. Koivisto, and P. J. Riekkinen. Volumetric mri analysis of the amygdala and the hippocampus in subjects with age-associated memory impairment: correlation to visual and verbal memory. *Neurology*, 44(9):1660–1668, Sep 1994.
- [331] N. Spencer, S. Bambang, S. Logan, and L. Gill. Socioeconomic status and birth weight: comparison of an area-based measure with the registrar general’s social class. *J Epidemiol Community Health*, 53(8):495–498, Aug 1999.
- [332] M. N. Starkman, S. S. Gebarski, S. Berent, and D. E. Scheingart. Hippocampal formation volume, memory dysfunction, and cortisol levels in patients with cushing’s syndrome. *Biol Psychiatry*, 32(9):756–765, Nov 1992.
- [333] M. N. Starkman, B. Giordani, S. S. Gebarski, S. Berent, M. A. Schork, and D. E. Scheingart. Decrease in cortisol reverses human hippocampal atrophy following treatment of cushing’s disease. *Biol Psychiatry*, 46(12):1595–1602, Dec 1999.
- [334] N. De Stefano, P. M. Matthews, and D. L. Arnold. Reversible decreases in n-acetylaspartate after acute brain injury. *Magn Reson Med*, 34(5):721–727, Nov 1995.
- [335] D. C. Steffens, C. E. Byrum, D. R. McQuoid, D. L. Greenberg, M. E. Payne, T. F. Blitchington, J. R. MacFall, and K. R. Krishnan. Hippocampal volume in geriatric depression. *Biol Psychiatry*, 48(4):301–309, Aug 2000.

- [336] M. B. Stein, C. Koverola, C. Hanna, M. G. Torchia, and B. McClarty. Hippocampal volume in women victimized by childhood sexual abuse. *Psychol Med*, 27(4):951–959, Jul 1997.
- [337] H. Stephan. Evolutionary trends in limbic structures. *Neuroscience & Biobehavioral Reviews*, 7(3):367–374, 1983.
- [338] A. Steptoe, L. Brydon, and S. Kunz-Ebrecht. Changes in financial strain over three years, ambulatory blood pressure, and cortisol responses to awakening. *Psychosom Med*, 67(2):281–287, 2005.
- [339] A. Steptoe, S. Kunz-Ebrecht, N. Owen, P. J. Feldman, G. Willemsen, C. Kirschbaum, and M. Marmot. Socioeconomic status and stress-related biological responses over the working day. *Psychosom Med*, 65(3):461–470, 2003.
- [340] P. Stiers, A. Fonteyne, H. Wouters, E. D’Agostino, S. Sunaert, and L. Lagae. Hippocampal malrotation in pediatric patients with epilepsy associated with complex prefrontal dysfunction. *Epilepsia*, Dec 2009.
- [341] R. S. Strauss. Adult functional outcome of those born small for gestational age: twenty-six-year follow-up of the 1970 british birth cohort. *JAMA*, 283(5):625–632, Feb 2000.
- [342] J. Tan, S. Bluml, T. Hoang, D. Dubowitz, G. Mevenkamp, and B. Ross. Lack of effect of oral choline supplement on the concentrations of choline metabolites in human brain. *Magn Reson Med*, 39(6):1005–1010, Jun 1998.
- [343] W. D. Taylor, D. C. Steffens, M. E. Payne, J. R. MacFall, D. A. Marchuk, I. K. Svenson, and K. R. R. Krishnan. Influence of serotonin transporter promoter region polymorphisms on hippocampal volumes in late-life depression. *Arch Gen Psychiatry*, 62(5):537–544, May 2005.
- [344] J. Thammaroj, C. Santosh, and JJ Bhattacharya. The hippocampus: modern imaging of its anatomy and pathology. *Practical Neurology*, 5(3):150, 2005.
- [345] J. M. Theysohn, O. Kraff, S. Maderwald, M. U. Schlamann, A. de Greiff, M. Forsting, S. C. Ladd, M. E. Ladd, and E. R. Gizewski. The human hippocampus at 7t in-vivo mri. *Hippocampus*, 19(1):1–7, Jan 2009.
- [346] B. P. Thomas, E. B. Welch, B. D. Niederhauser, W. O. Whetsell, A. W. Anderson, J. C. Gore, M. J. Avison, and J. L. Creasy. High-resolution 7t mri of the human hippocampus in vivo. *J Magn Reson Imaging*, 28(5):1266–1272, Nov 2008.

- [347] P. Tibbo, C. C. Hanstock, S. Asghar, P. Silverstone, and P. S. Allen. Proton magnetic resonance spectroscopy (1h-mrs) of the cerebellum in men with schizophrenia. *J Psychiatry Neurosci*, 25(5):509–512, Nov 2000.
- [348] G. Tsai and J. T. Coyle. N-acetylaspartate in neuropsychiatric disorders. *Prog Neurobiol*, 46(5):531–540, Aug 1995.
- [349] H. Uno, R. Tarara, J. G. Else, M. A. Suleman, and R. M. Sapolsky. Hippocampal damage associated with prolonged and fatal stress in primates. *J Neurosci*, 9(5):1705–1711, May 1989.
- [350] J. Urenjak, S. R. Williams, D. G. Gadian, and M. Noble. Specific expression of n-acetylaspartate in neurons, oligodendrocyte-type-2 astrocyte progenitors, and immature oligodendrocytes in vitro. *J Neurochem*, 59(1):55–61, Jul 1992.
- [351] V. Vaccarino, B. D. Johnson, D. S. Sheps, S. E. Reis, S. F. Kelsey, V. Bittner, T. Rutledge, L. J. Shaw, G. Sopko, C. N. B. Merz, Lung National Heart, and Blood Institute. Depression, inflammation, and incident cardiovascular disease in women with suspected coronary ischemia: the national heart, lung, and blood institute-sponsored wise study. *J Am Coll Cardiol*, 50(21):2044–2050, Nov 2007.
- [352] L. Vallières, I. L. Campbell, F. H. Gage, and P. E. Sawchenko. Reduced hippocampal neurogenesis in adult transgenic mice with chronic astrocytic production of interleukin-6. *J Neurosci*, 22(2):486–492, Jan 2002.
- [353] D. C. van den Boom. The influence of temperament and mothering on attachment and exploration: an experimental manipulation of sensitive responsiveness among lower-class mothers with irritable infants. *Child Dev*, 65(5):1457–1477, Oct 1994.
- [354] P. van. Eijndhoven, G. van. Wingen, K. van. Oijen, M. Rijpkema, B. Goraj, R. J. Verkes, R. O. Voshaar, G. Fernández, J. Buitelaar, and I. Tendolkar. Amygdala volume marks the acute state in the early course of depression. *Biol Psychiatry*, 65(9):812–818, May 2009.
- [355] H. van Praag, G. Kempermann, and F. H. Gage. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nat Neurosci*, 2(3):266–270, Mar 1999.
- [356] Y. N. Velupillai, C. J. Packard, G. D. Batty, V. Bezlyak, H. Burns, J. Cavanagh, K. Deans, I. Ford, A. McGinty, K. Millar, N. Sattar, P. Shiels, and C. Tannahill.

- Psychological, social and biological determinants of ill health (psobid): study protocol of a population-based study. *BMC Public Health*, 8:126, 2008.
- [357] M. N. Verbaten. Chronic effects of low to moderate alcohol consumption on structural and functional properties of the brain: beneficial or not? *Hum Psychopharmacol*, 24(3):199–205, Apr 2009.
- [358] P. Videbech and B. Ravnkilde. Hippocampal volume and depression: a meta-analysis of mri studies. *Am J Psychiatry*, 161(11):1957–1966, Nov 2004.
- [359] E. De Vita, D. L. Thomas, S. Roberts, H. G. Parkes, R. Turner, P. Kinchesh, K. Shmueli, T. A. Yousry, and R. J. Ordidge. High resolution mri of the brain at 4.7 tesla using fast spin echo imaging. *Br J Radiol*, 76(909):631–637, Sep 2003.
- [360] M. Vythilingam, E. Vermetten, G. M. Anderson, D. Luckenbaugh, E. R. Anderson, J. Snow, L. H. Staib, D. S. Charney, and J. D. Bremner. Hippocampal volume, memory, and cortisol status in major depressive disorder: effects of treatment. *Biol Psychiatry*, 56(2):101–112, Jul 2004.
- [361] G. D. Waiter, H. C. Fox, A. D. Murray, J. M. Starr, R. T. Staff, V. J. Bourne, L. J. Whalley, and I. J. Deary. Is retaining the youthful functional anatomy underlying speed of information processing a signature of successful cognitive ageing? an event-related fmri study of inspection time performance. *Neuroimage*, 41(2):581–595, Jun 2008.
- [362] L. Wang, J. P. Miller, M. H. Gado, D. W. McKeel, M. Rothermich, M. I. Miller, J. C. Morris, and J. G. Csernansky. Abnormalities of hippocampal surface structure in very mild dementia of the alzheimer type. *Neuroimage*, 30(1):52–60, Mar 2006.
- [363] L. Wang, J. P. Miller, M. H. Gado, D. W. McKeel, M. Rothermich, M. I. Miller, J. C. Morris, and J. G. Csernansky. Abnormalities of hippocampal surface structure in very mild dementia of the alzheimer type. *Neuroimage*, 30(1):52–60, Mar 2006.
- [364] Y. Wang and S. J. Li. Differentiation of metabolic concentrations between gray matter and white matter of human brain by in vivo 1h magnetic resonance spectroscopy. *Magn Reson Med*, 39(1):28–33, Jan 1998.
- [365] Z. Wang, T. C. Neylan, S. G. Mueller, M. Lenoci, D. Truran, C. R. Marmar, M. W. Weiner, and N. Schuff. Magnetic resonance imaging of hippocampal subfields in post-traumatic stress disorder. *Arch Gen Psychiatry*, 67(3):296–303, Mar 2010.

- [366] C. Watson, F. Andermann, P. Gloor, M. Jones-Gotman, T. Peters, A. Evans, A. Olivier, D. Melanson, and G. Leroux. Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. *Neurology*, 42(9):1743–1750, Sep 1992.
- [367] M. Weinstock. The long-term behavioural consequences of prenatal stress. *Neurosci Biobehav Rev*, 32(6):1073–1086, Aug 2008.
- [368] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. Hill, D. J. Hawkes, C. Studholme, J. B. Maintz, M. A. Viergever, G. Malandain, and R. P. Woods. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J Comput Assist Tomogr*, 21(4):554–566, 1997.
- [369] U. C. Wieshmann, M. R. Symms, J. P. Mottershead, D. G. MacManus, G. J. Barker, P. S. Tofts, T. Revesz, J. M. Stevens, and S. D. Shorvon. Hippocampal layers on high resolution magnetic resonance images: real or imaginary? *J Anat*, 195 (Pt 1):131–135, Jul 1999.
- [370] K. J. Worsley, M. Andermann, T. Koulis, D. MacDonald, and A. C. Evans. Detecting changes in nonisotropic images. *Hum Brain Mapp*, 8(2-3):98–101, 1999.
- [371] K. J. Worsley, A. C. Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for cbf activation studies in human brain. *J Cereb Blood Flow Metab*, 12(6):900–918, Nov 1992.
- [372] C. E. Wright and A. Steptoe. Subjective socioeconomic position, gender and cortisol responses to waking in an elderly population. *Psychoneuroendocrinology*, 30(6):582–590, Jul 2005.
- [373] M. A. Yassa, S. M. Stark, A. Bakker, M. S. Albert, M. Gallagher, and C. E. L. Stark. High-resolution structural and functional mri of hippocampal ca3 and dentate gyrus in patients with amnesic mild cognitive impairment. *Neuroimage*, 51(3):1242–1252, Jul 2010.
- [374] T. F. Yeh, Y. J. Lin, H. C. Lin, C. C. Huang, W. S. Hsieh, C. H. Lin, and C. H. Tsai. Outcomes at school age after postnatal dexamethasone therapy for lung disease of prematurity. *N Engl J Med*, 350(13):1304–1313, Mar 2004.

- [375] P. A. Yushkevich, B. B. Avants, S. R. Das, J. Pluta, M. Altinay, C. Craige, and Alzheimer's Disease Neuroimaging Initiative. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in adni 3 t mri data. *Neuroimage*, 50(2):434–445, Apr 2010.
- [376] P. A. Yushkevich, B. B. Avants, J. Pluta, S. Das, D. Minkoff, D. Mechanic-Hamilton, S. Glynn, S. Pickup, W. Liu, J. C. Gee, M. Grossman, and J. A. Detre. A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 t. *Neuroimage*, 44(2):385–398, Jan 2009.
- [377] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, Jul 2006.
- [378] W. Zheng, M. W. L. Chee, and V. Zagorodnov. Improvement of brain segmentation accuracy by optimizing non-uniformity correction using n3. *Neuroimage*, 48(1):73–83, Oct 2009.
- [379] C.B. Zhu, R.D. Blakely, and W.A. Hewlett. The proinflammatory cytokines interleukin-1beta and tumor necrosis factor-alpha activate serotonin transporters. *Neuropsychopharmacology*, 31(10):2121–2131, 2006.
- [380] H. Zhu, R. A. E. Edden, R. Ouwerkerk, and P. B. Barker. High resolution spectroscopic imaging of gaba at 3 tesla. *Magn Reson Med*, 65(3):603–609, Mar 2011.
- [381] M. E. Zimmerman, J. W. Pan, H. P. Hetherington, M. J. Katz, J. Verghese, H. Buschke, C. A. Derby, and R. B. Lipton. Hippocampal neurochemistry, neuromorphometry, and verbal memory in nondemented older adults. *Neurology*, 70(18):1594–1600, Apr 2008.

Appendix A

Permission to use images from *The Human Hippocampus. Functional Anatomy, Vascularization and Serial Sections with MRI* by Duvernoy, Henri can be seen in figures 9.1 and 9.2.

WG: Permission request

Page 1 of 2

WG: Permission request

Schmidt-Loeffler, Barbara, Springer DE [Barbara.Schmidt-Loeffler@springer.com]

Sent: 07 July 2011 08:56**To:** JOHN ROY MCLEAN**Attachments:** Request_for_permission_Hip~1.doc (38 KB)

The Human Hippocampus. Functional Anatomy, Vascularization and Serial Sections with MRI
 Duvernoy, Henri M.; 3rd ed., 2005
 Figures 1AB, 5AB, 7, 8, 10AB

Dear Mr. McLean,

With reference to your request (copy herewith) to re-use material on which Springer controls the copyright, our permission is granted free of charge, on the following conditions:

- if material in question appears with credit to another source, authorization from and reference to that source is required as well;
- permission is also obtained from the author (address is given on the imprint page or with the article);
- full credit is given to the publication in which the material was originally published by adding: With kind permission of Springer Science+Business Media

UMI ProQuest Information and Learning may supply copies of the master's thesis on demand.

An author may self-archive an author-created version of his thesis on his own website and his university's repository.

Permission free of charge does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

With kind regards,

—
Barbara Schmidt-Löffler
 Springer
 Rights and Permissions

—
 Tiergartenstr. 17 | 69121 Heidelberg | Germany
barbara.schmidt-loeffler@springer.com
www.springeronline.com
 —

Figure 9.1: Permission to use images from The Human Hippocampus

Courier

Page 1 of 1

Courrier

Henri DUVERNOY [henri.duvernoy@orange.fr]

Sent: 17 July 2011 15:59**To:** JOHN ROY MCLEAN

Dear Dr John McLean

I thank you for the use of the images from the book "The Human Hippocampus " There is no problem for all the figures as you want . I can help you if necessary .

I would try to do a new edition of the book but I am not sure to be able to do this book.

Sincerely yours

Henri Duvernoy

Figure 9.2: Permission to use images The Human Hippocampus

Appendix B

From Patenaude et al, 2011 [278] the equations implementing the MVGLM with Pillai's trace for FIRST's vertex analysis are shown below.

$$Y = X\beta + \varepsilon \quad (9.1)$$

$$\hat{\beta} = [X^T X]^{-1} X^T Y \quad (9.2)$$

$$E = [Y - X\hat{\beta}]^T [Y - X\hat{\beta}] \quad (9.3)$$

$$\hat{\beta}_H = [CX^T X C^T]^{-1} C X^T Y \quad (9.4)$$

$$H = [Y - X C^T \hat{\beta}_H]^T [Y - X C^T \hat{\beta}_H] * E \quad (9.5)$$

$$V = Tr(H(H + E)^{-1}) \quad (9.6)$$

$$F_{s(|q-d|+s), s(N_S - N_R - q + s)} = \frac{N_s - N_R - q + s}{|q - d| + s} \left(\frac{V}{s - V} \right) \quad (9.7)$$

Where $q = 3$ is the number of dimensionality of the vectors (3 coordinates each); Y is an $N_s * q$ matrix containing corresponding coordinates for one vertex taken from each of the N_S test subjects; X is the $N_S * N_R$ design matrix with N_R regressors; E is the sample error

covariance matrix; C is the contrast matrix such that XC^T is the reduced model, or set of confound regressors, for instance, this could be subject age, ICV or other variable; H is the effect covariance matrix; $s = \min(q, d)$; $d = \text{rank}(X) - \text{rank}(XC^T)$; V is Pillai's trace; and $F_{d1, d2}$ is an F-statistic with $d1$ and $d2$ degrees of freedom.

Appendix C

The poster used to recruit subjects to the study of the hippocampus in normal volunteer using 3T MRI is shown below.

MR Imaging of the Hippocampus in Normal Volunteers

Volunteers Wanted for brain imaging study

**We are looking for healthy people between 18-65
years of age to take part in a research project.**

We aim to find out how much variation there is in the
brains of healthy people.

Would you be willing to:-

- o Have a 3T MRI scan of
your brain,
- o Answer two
questionnaires about your
memory function and
mental health AND
- o Donate a blood sample for
genetic testing?



If you have no known history of neurological or
psychiatric disorders and would like more information
about this project, please contact:

Dr Jonathan Cavanagh, Consultant Psychiatrist on 0141 201
2482 OR
Dr Barrie Condon, Consultant Physicist on 0141 201 2127

Institute of Neurological Sciences,
Southern General Hospital.

Poster Version HCPPost1.0
Date: 15/11/05

Figure 9.3: Poster for recruitment of normal volunteers to the study of the hippocampus 3T MRI