THE UNIVERSITY
of EDINBURGH

Structural Brain Imaging and Cognitive Function in Individuals at High Familial Risk of Mood Disorders

Martina Papmeyer

Doctor of Philosophy

University of Edinburgh

2014

DECLARATION

I hereby declare that this thesis has been composed by myself and that the work presented within is my own. The participants of the Bipolar Family Study were recruited by Prof. Andrew M. McIntosh and associated clinicians and the data were collected by research members of the Division of Psychiatry of the University of Edinburgh. This work has not been submitted for any other degree or professional qualification.

_____

- Martina Papmeyer -

PUBLICATION

The following article is based on the work presented in Chapter 4:

Papmeyer, M., Giles, S., Sussmann, J.E., Kielty, S., Stewart, T., Lawrie S.M., Whalley, H.C. & McIntosh, A.M. (in press). Cortical thickness in individuals at high familial risk of mood disorders as they develop major depressive disorder. Biol Psych.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all participants who took part in the Bipolar Family Study. Without their support, this research project would not have been feasible and we would not be able to gain new insights in the pathogenesis of mood disorders.

Furthermore, I would like to thank my supervisors Prof. Andrew M. McIntosh, Prof. Stephen M. Lawrie and Dr. Heather C. Whalley for their ongoing support and guidance during all stages of my Ph.D. Andrew's passion for science has always inspired me and his cheerfulness and open-mindedness has not only made this Ph.D. a very pleasurable experience, but has also helped me to become a more independent scientist.

I would also like to thank everyone who contributed to this research project in various ways. I am especially grateful for Stephen Giles's support with various IT issues and for Jess Sussmann's extremely well-organised data files. Moreover, I would like to thank my colleagues Dr. Emma Sprooten and Dr. Betty Tijms for having such a pleasurable time in our shared office. I would also like to thank the National Health Service (NHS) Research Scotland for their financial support towards this research project through the Scottish Mental Health Research Network and the Medical Research Council (MRC) for funding my Ph.D.

Last but not least, I would like to thank my family and friends for their ongoing emotional support and love during the three years of my Ph.D. I feel extremely blessed for having my husband Marcus Papmeyer by my side who went through his own Ph.D. experiences together with me. His passion for science always helped me to stay focussed and motivated and his humour, love and warmth made sure that I enjoyed mostly the ups of the Ph.D. and not so much experienced the downs.

ABSTRACT

Bipolar disorder (BD) and major depressive disorder (MDD) are characterised by a fundamental disturbance of mood, with strong support for overlapping causal pathways. Structural brain and neurocognitive abnormalities have been associated with mood disorders, but it is unknown whether these reflect early adverse effects predisposing to mood disorders or emerge as a consequence of illness onset.

The Bipolar Family Study is well-suited to examine the origin of structural brain and neuropsychological abnormalities in mood disorders further. The volumes of subcortical brain regions, cortical thickness and surface area measures of frontal and temporal regions of interest and neuropsychological performance over a two-year time interval was compared at baseline and longitudinally between three groups: young individuals at high risk of mood disorders who subsequently developed MDD during the follow-up period (HR-MDD), individuals at high risk of mood disorders who remained well (HR-well), and healthy control subjects (HC).

The longitudinal analysis of cortical thickness revealed significant group effects for the right parahippocampal and right fusiform gyrus. Cortical thickness in both of these brain regions across the two time points was reduced in both high-risk groups relative to controls, with the HR-MDD group displaying a thinner parahippocampus gyrus than the HR-well group. Moreover, a significant interaction effect was observed for the left inferior frontal and left precentral gyrus. The HR-well subjects had progressive thickness reductions in these brain regions relative to controls, while the HR-MDD group showed cortical thickening of these areas. Finally, longitudinal analyses of neuropsychological performance revealed a significant group effect for long delay verbal memory and extradimensional set-shifting performance. Reduced neurocognitive performance during both tasks across the two time points was found in the HR-well group relative to controls,

with the HR-MDD group displaying decreased extradimensional set-shifting abilities as compared to the HC group only.

These findings indicate, that reduced left parahippocampal and fusiform thickness constitute a familial trait marker for vulnerability to mood disorders and may thus form potential neuroanatomic endophenotypes. Particularly strong thickness reductions of the parahippocampal gyrus appear be linked to an onset of MDD. Moreover, progressive thickness reductions in the left inferior frontal and precentral gyrus in early adulthood form a familial trait marker for vulnerability to mood disorders, potentially reflecting early neurodegenerative processes. By contrast, an absence of cortical thinning of these brain regions in early adulthood appears to be linked to the onset of MDD, potentially reflecting a lack or delay of normal synaptic pruning processes. Reduced long delay verbal memory and extradimensional set-shifting performance across time constitute a familial trait marker for vulnerability to mood disorders, likely representing disturbances of normal brain development predisposing to illness. These findings advance our understanding of the origin of structural brain and neurocognitive abnormalities in mood disorders.

ABBREVIATIONS

5HTT	Serotonin transporter gene
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BD	Bipolar disorder
BD-I	Bipolar I disorder
BD-II	Bipolar II disorder
BDNF	Brain-derived neurotrophic factor gene
BIRC	Brain Imaging Research Centre
CACNA1C	Calcium channel, voltage-dependent, L type, alpha 1 subunit gene
CANTAB	Cambridge Neuropsychological Test Automated Battery
COMT	Catechol-O-Methyltransferase gene
CPT	Continuous Performance Test
CVLT	California Verbal Learning Test
DAOA	see G72
DGKH	Diacylglycerol kinase eta gene
DISC1	Disrupted in Schizophrenia 1 gene
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, fourth edition
DSST	Digit Symbol Substituition Test
FDR	False-Discovery Rate

fMRI	Functional magnetic resonance imaging
G72	D-amino acid oxidase activator gene
GWAS	Genome-wide association study
HAM-D	Hamilton Depression Rating Scale
HC	Healthy control subjects
HR-MDD	Individuals at high familial risk of mood disorders who developed MDD during a two-year time period
HR-well	Individuals at high familial risk of mood disorders who remained well during a two-year time period
IED	Intra-/Extradimensional Set Shifting Task
MAOA	Monoamine oxidase-A gene
MDD	Major depressive disorder
MPRAGE	Magnetisation Prepared Rapid Gradient Echo
MRI	Magnetic resonance imaging
MYO5B	Myosin VB gene
NART	National Adult Reading Test
NCAN	Neurocan gene
NHS	National Health Services
ODZ4	Odd Oz/ten-m homolog 4 gene
OPCRIT	Operational Criteria Symptom Checklist
PALB2	Partner and localizer of BRCA2 gene
RF	Resonance frequency
ROI	Region of interest

RVLT	Rey Verbal Learning Test
SCID	Structured Clinical Interview for DSM-IV Axis-I Disorders
SDMT	Symbol Digit Modalities Test
SKAP1	Src kinase associated phosphoprotein 1 gene
SLC6A4	Solute carrier family 6 member 4 gene
SLC6A15	Solute carrier family 6 member 15 gene
VBM	Voxel-based morphometry
WCST	Wisconsin Card Sorting Test
YMRS	Young Mania Rating Scale

TABLE OF CONTENTS

Declaration ..... **i**

Publication ..... **ii**

Acknowledgement ..... **iii**

Abstract ..... **iv**

Abbreviations ..... **vi**

List of Figures ..... **xiv**

List of Tables ..... **xv**

Chapter 1 Background **1**

1.1 Clinical characteristics of mood disorders 1

1.2 Epidemiology of mood disorders 6

1.3 Aetiology of mood disorders 8

1.3.1 Genetic susceptibility factors 9

1.3.1.1 Family and twin studies 9

1.3.1.2 Molecular genetic studies 10

1.3.2 Environmental susceptibility factors 15

1.3.3 Gene-environment interplay 15

1.4 The endophenotype concept 16

1.5 Structural grey matter abnormalities in mood disorders 17

1.5.1 Structural grey matter abnormalities in affected patients 18

1.5.2 Structural grey matter abnormalities in unaffected relatives 21

1.5.3 The importance of studying grey matter pathology in mood disorders 23

1.6 Neurocognitive functioning in mood disorders 26

1.6.1 Neurocognitive tasks commonly employed in mood disorder research 26

1.6.2 Neurocognitive functioning in affected patients 33

1.6.3 Neurocognitive functioning in unaffected relatives 35

1.7 Summary and aims 36

Chapter 2 Methods 39

2.1 The Bipolar Family Study 39

2.1.1 Participants 39

2.1.2 Clinical assessment 41

2.1.3 Neuropsychological assessment 42

2.1.4 Magnetic resonance imaging acquisition and pre-processing 42

2.2 Principles of structural magnetic resonance brain imaging 43

2.2.1 Magnetic resonance imaging 43

2.2.2 Processing structural brain images with FreeSurfer 45

2.2.2.1 The volume-based stream 49

2.2.2.2 The surface-based stream 52

2.3 Principles of neuropsychological assessment 58

2.3.1 The Intra-/Extradimensional Set Shifting Task 59

2.3.2 The California Verbal Learning Test 64

2.3.3 The Digit Span 66

2.4 Statistical analyses 67

Chapter 3 Subcortical volumes in individuals at high familial risk of mood disorders 70

3.1 Introduction 70

3.2 Methods 72

3.2.1 Participants and clinical assessment 72

3.2.2 Magnetic resonance imaging and processing 73

3.2.3 Statistical analysis 75

3.3 Results 76

3.3.1 Cross-sectional analysis 76

3.3.1.1 Group differences at baseline 76

3.3.1.2 Correlation analysis 76

3.3.1.3 Analysis of potential confounders 77

3.3.2 Longitudinal analysis 79

3.3.2.1 Group differences over time 79
3.3.2.2 Correlation analysis 79
3.3.2.3 Analysis of potential confounders 80
3.4 Discussion 85

Chapter 4 Cortical thickness in individuals at high familial risk of mood disorders 90
4.1 Introduction 90
4.2 Methods 93
4.2.1 Participants and clinical assessment 93
4.2.2 Magnetic resonance imaging and processing 93
4.2.3 Statistical analysis 93
4.3 Results 95
4.3.1 Cross-sectional analysis 95
4.3.1.1 Group differences at baseline 95
4.3.1.2 Correlation analysis 95
4.3.1.3 Analysis of potential confounders 96
4.3.2 Longitudinal analysis 99
4.3.2.1 Group differences over time 99
4.3.2.2 Correlation analysis 100
4.3.2.3 Analysis of potential confounders 101
4.4 Discussion 107

Chapter 5 Cortical surface area in individuals at high familial risk of mood disorders 112
5.1 Introduction 112
5.2 Methods 114
5.2.1 Participants and clinical assessment 114
5.2.2 Magnetic resonance imaging and processing 114
5.2.3 Statistical analysis 115

5.3 Results ... 116
5.3.1 Cross-sectional analysis ... 116
5.3.1.1 Group differences at baseline ... 116
5.3.1.2 Correlation analysis ... 116
5.3.1.3 Analysis of potential confounders ... 117
5.3.2 Longitudinal analysis ... 119
5.3.2.1 Group differences over time ... 119
5.3.2.2 Correlation analysis ... 120
5.3.2.3 Analysis of potential confounders ... 121
5.4 Discussion ... 126

Chapter 6 Neurocognitive performance in individuals at high familial risk of mood disorders ... 130
6.1 Introduction ... 130
6.2 Methods ... 132
6.2.1 Participants and clinical assessment ... 132
6.2.2 Neuropsychological assessment ... 132
6.2.3 Statistical analysis ... 134
6.3 Results ... 135
6.3.1 Cross-sectional analysis ... 135
6.3.1.1 Group differences at baseline ... 135
6.3.1.2 Correlation analysis ... 135
6.3.1.3 Analysis of potential confounders ... 136
6.3.2 Longitudinal analysis ... 138
6.3.2.1 Group differences over time ... 138
6.3.2.2 Correlation analysis ... 139
6.3.2.3 Analysis of potential confounders ... 139
6.4 Discussion ... 146

Chapter 7 General conclusions and implications for the future..... 150

7.1 Summary of main findings 150

7.2 Methodological remarks and limitations..... 155

7.3 Implications for the future..... 159

7.4 General conclusion 161

References 162

LIST OF FIGURES

Figure 2.1 Three-dimensional representation of the subcortical ROI extracted with the volume-based stream of FreeSurfer **50**

Figure 2.2 Visual illustration of a FreeSurfer processed brain scan before (A) and after (B) removal of non-brain tissue **51**

Figure 2.3 Coronal (A) and horizontal (B) visual illustration of volume-based segmented brain regions **52**

Figure 2.4 Visual representation of the cortical ROIs **55**

Figure 2.5 Illustration of the segmentation of white matter (A) and the subsequent separation of the two hemispheres (B) **57**

Figure 2.6 Visual illustration of the white (shown in yellow) and the pial (shown in red) surfaces, overlaid onto the original brain scan **58**

Figure 2.7 The IED task **63**

Figure 4.1 Significant group-by-time interactions (a, b) and group effects (c, d) of longitudinal cortical thickness analyses **103**

Figure 6.1 Significant group effects (a, b) of longitudinal neurocognitive performance analyses **142**

LIST OF TABLES

Table 1.1 DSM-IV criteria for a major depressive episode **2**

Table 1.2 DSM-IV criteria for a manic episode **4**

Table 1.3 DSM-IV criteria for a mixed episode **4**

Table 1.4 DSM-IV criteria for a hypomanic episode **5**

Table 3.1 Demographic and clinical characteristics **74**

Table 3.2 Cross-sectional analysis of regional subcortical volumes **77**

Table 3.3 Correlation between regional subcortical volumes and HAM-D sum scores **78**

Table 3.4 Cross-sectional analysis of regional subcortical volumes in unrelated subjects **78**

Table 3.5 Longitudinal analysis of regional subcortical volumes **81**

Table 3.6 Correlation between changes in regional subcortical volumes and changes in HAM-D sum scores **82**

Table 3.7 Longitudinal analysis of regional subcortical volumes in unrelated subjects **83**

Table 3.8 Longitudinal analysis of regional subcortical volumes in unmedicated HR-MDD subjects **84**

Table 4.1 Cross-sectional analysis of cortical thickness **97**

Table 4.2 Correlation between cortical thickness and HAM-D sum score **98**

Table 4.3 Cross-sectional analysis of cortical thickness in unrelated subjects **99**

Table 4.4 Longitudinal analysis of cortical thickness **102**

Table 4.5 Correlation between changes in cortical thickness and changes in HAM-D sum score **104**

Table 4.6 Longitudinal analysis of cortical thickness in unrelated subjects **105**

Table 4.7 Longitudinal analysis of cortical thickness in unmedicated subjects **106**

Table 5.1 Cross-sectional analysis of cortical surface area **117**

Table 5.2 Correlation of cortical surface area and HAM-D sum score **118**

Table 5.3 Cross-sectional analysis of cortical surface area in unrelated subjects **119**

Table 5.4 Longitudinal analysis of cortical surface area **122**

Table 5.5 Correlation between changes in surface area and changes in HAM-D sum score **123**

Table 5.6 Longitudinal analysis of cortical surface area in unrelated subjects **124**

Table 5.7 Longitudinal analysis of cortical surface area in unmedicated HR-MDD subjects **125**

Table 6.1 Neuropsychological performance parameters **133**

Table 6.2 Cross-sectional analysis of neurocognitive performance **136**

Table 6.3 Correlation between neurocognitive performance measures and HAM-D sum scores **137**

Table 6.4 Cross-sectional analysis of neurocognitive performance in unrelated subjects **137**

Table 6.5 Longitudinal analysis of neurocognitive performance **141**

Table 6.6 Correlation between changes in neurocognitive performance and changes in HAM-D sum scores **143**

Table 6.7 Longitudinal analysis of neurocognitive performance in unrelated subjects **144**

Table 6.8 Longitudinal analysis of neurocognitive performance in unmedicated HR-MDD subjects **145**

Chapter 1

Background

1.1 Clinical characteristics of mood disorders

The diagnostic term ‘mood disorders’ refers to a range of conditions in which the most prominent symptom is a fundamental disturbance in the person’s mood to either depression or elation (World Health Organization, 2004). Based on whether an episode of mania or hypomania has ever been present, mood disorders are divided into depressive disorders or bipolar disorders (BD) as outlined in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994).

There are three main types of depressive disorders. The first is major depressive disorder (MDD). A diagnosis of MDD requires the presence of at least a single major depressive episode. Such an episode consists of a broad range of symptoms that are experienced for most of the time, nearly every day, for at least two weeks and represent a significant change from previous functioning. The core feature of a major depressive episode is either a depressed mood or anhedonia, although both can occur simultaneously. Furthermore, the presence of at least four (or three if both depressed mood and anhedonia are present) of the following symptoms are required for a diagnosis: change in appetite or sleep patterns (insomnia or hypersomnia), psychomotor agitation or retardation, fatigue, feelings of worthlessness or guilt, poor concentration, recurrent thoughts of death and suicide. The symptoms cause significant distress or impairment in social, occupational or other areas of functioning to the person concerned (American Psychiatric Association, 1994). The precise criteria for a major depressive episode according to DSM-IV are provided in Table 1.1.

Table 1.1 DSM-IV criteria for a major depressive episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.
Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.
   (1) Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful). Note: In children and adolescents, can be irritable mood
   (2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others)
   (3) Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. In children, consider failure to make expected weight gains
   (4) Insomnia or hypersomnia nearly every day
   (5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
   (6) Fatigue or loss of energy nearly every day
   (7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
   (8) Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
   (9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide
B. The symptoms do not meet criteria for a Mixed Episode.
C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to the direct physiological effects of a substance or a general medical condition.
E. The symptoms are not better accounted for by bereavement.

Other subtypes of depressive disorders include dysthymic disorder and depressive disorder not otherwise specified. Given that these two conditions are not the focus of this thesis, only a brief description of them is provided. Dysthymic disorder is a milder but more chronic type of depression. Criteria include the presence of depressed mood for the majority of time during a time period of at least two years. Moreover, two or more of the following symptoms are evident when the person is depressed: change in appetite or sleep patterns, fatigue, reduced self-esteem, poor concentration or decision making abilities, feelings of hopelessness or pessimistic thoughts. Depressive symptoms that do not meet criteria for any mental disorder outlined in DSM-IV can be classified as depressive disorder not otherwise

specified. This category applies for example to cases in which fewer than the required five depressive symptoms for a diagnosis of MDD are present for at least two weeks (American Psychiatric Association, 1994).

The main features of BD are intermittent episodes of mania or hypomania, commonly interlaced with episodes of depression. Several types of BD can be distinguished, including bipolar I disorder (BD-I) and bipolar II disorder (BD-II). A diagnosis of BD-I is made when at least a single manic or mixed episode, with or without a depressive episode, has ever occurred. A manic episode is defined as a period of excessively elevated, expansive or irritable mood that lasts for one week or longer. At least three (or four if mood is only irritable) of the following symptoms are required to meet criteria for a manic episode: increased feelings of self-esteem or grandiosity, less need for sleep, increased talkativeness, enhanced distractibility, increased goal-directed activity or psychomotor agitation, exorbitant involvement in pleasurable activities that have a high potential for painful consequences (such as buying sprees or risky business investments). A mixed episode by contrast concerns the presence of symptoms that fulfil criteria for both a manic episode and a major depressive episode (except for duration) over a period of at least one week. The mood disturbance is causing marked impairment in social or occupational areas of function (American Psychiatric Association, 1994). The detailed criteria for a manic episode and a mixed episode are provided in Tables 1.2 and 1.3.

Table 1.2 DSM-IV criteria for a manic episode

A. A distinct period of abnormally and persistently elevated, expansive or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary).
B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:
(1) Inflated self-esteem or grandiosity
(2) Decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
(3) More talkative than usual or pressure to keep talking
(4) Flight of ideas or subjective experience than thoughts are racing
(5) Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
(6) Increase in goal-directed activity (at work, at school, or sexually) or psychomotor agitation)
(7) Excessive involvement in pleasurable activities that have a high potential for painful consequences
C. The symptoms do not meet criteria for a Mixed Episode.
D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there a psychotic features.
E. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication or other treatment) or a general medical condition (e.g., hyperthyroidism).

Table 1.3 DSM-IV criteria for a mixed episode

A. The criteria are met both for a Manic Episode and for a Major Depressive Episode (except for duration) nearly every day during a 1-week period.
B. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.
C. The symptoms are not due to the direct physiological effects of a substance or a general medical condition.

A diagnosis of BD-II is made when at least one major depressive episode and at least one hypomanic episode has occurred. Hypomanic episodes are milder forms of manic episodes and are characterised as a period of persistently elevated, expansive or irritable mood that lasts for at least four days. At least three (or four if mood is only irritable) of the symptoms of a manic episode are required to meet criteria for a hypomanic episode. In contrast to a manic episode however, the symptoms do not cause significant distress or impairment in social, occupational or

other areas of functioning (American Psychiatric Association, 1994). The precise criteria for a hypomanic episode are outlined in Table 1.4.

Table 1.4 DSM-IV criteria for a hypomanic episode

A.	A distinct period of persistently elevated, expansive or irritable mood, lasting throughout at least 4 days, that is clearly different from the usual non-depressed mood.
B.	During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree: (1) Inflated self-esteem or grandiosity (2) Decreased need for sleep (e.g., feels rested after only 3 hours of sleep) (3) More talkative than usual or pressure to keep talking (4) Flight of ideas or subjective experience that thoughts are racing (5) Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli) (6) Increase in goal-directed activity (at work, at school, or sexually) or psychomotor agitation (7) Excessive involvement in pleasurable activities that have a high potential for painful consequences
C.	The episode is associated with an unequivocal change in functioning that is uncharacteristic of the person when not symptomatic.
D.	The disturbance in mood and the change in functioning are observable by others.
E.	The mood disturbance is not severe enough to cause marked impairment in social or occupational functioning, or to necessitate hospitalization, and there are no psychotic features.
F.	The symptoms are not due to the direct physiological effects of a substance or a general medical condition.

Apart from BD-I and BD-II, other subtypes of BD include cyclothymia or BD not otherwise specified. Given that these two conditions are not the focus of this thesis, only a brief description of them is provided. Cyclothymia is characterised by the presence of hypomanic episodes with periods of major depressive symptoms that do not meet criteria for MDD over a period of at least two years. A diagnosis of BD not otherwise specified can be made if there are symptoms of BD that do not fall within one of the other established subtypes. For example, an individual may experience intermittent manic and depressive symptoms that only last for a few days or may experience recurrent hypomanic episodes in the absence of any depressive symptoms (American Psychiatric Association, 1994).

1.2 Epidemiology of mood disorders

With an estimated lifetime prevalence of more than 20%, mood disorders are one of the most common mental disorders worldwide and a leading cause of disability (Kessler et al., 2005). Although classified together as mood disorders, MDD and BD are heterogeneous with respect to their epidemiological characteristics. MDD is far more common in the population than BD, with lifetime prevalence estimates being about 16% for MDD and about 2% for BD-I and BD-II as a whole (Kessler et al., 2003; Merikangas et al., 2007).

The median age at onset of MDD is about 29-32 years but only 21-25 years for BD (Kessler, et al., 2005; Perlis, Brown, Baker, & Nierenberg, 2006). Earlier ages at onset of MDD have been associated with a more severe course of illness, including greater symptom severity, more lifetime episodes of depression and suicide attempts, increased social and occupational impairment, poorer quality of life and greater comorbidity as compared to individuals with later age onset (Zisook et al., 2007). Similarly, early onset BD has been linked to particularly severe clinical features and a worse course of illness that is characterised by more lifetime psychotic symptoms, more rapid cycling of episodes, increased suicide attempts, greater comorbidity and poorer response to lithium pharmacotherapy (Carter, Mundo, Parikh, & Kennedy, 2003; Schurhoff et al., 2000; Suominen et al., 2007). While woman suffer approximately twice as frequently from MDD than men, both genders appear to be equally often affected by BD (Piccinelli & Wilkinson, 2000; Viana & Andrade, 2012).

Relatively few studies have examined the course of illness of mood disorders across lifetime, although attempts have been made to draw conclusions about their persistence by comparing 12-month and lifetime prevalence rates (Kessler, Merikangas, & Wang, 2007). The ratios indicate a higher persistence of BD-I (63%) than MDD (40%) (Kessler, et al., 2007). These estimates are broadly in line with retrospective and prospective assessments in community and clinical samples

that point towards mood disorders being mainly episodically chronic-recurrent in nature. For MDD, the recurrence rates of a major depressive episode after an initial recovery have been estimated to be between 38% and 78% (Angst & Merikangas, 1997; Mueller et al., 1999; van Weel-Baumgarten, van den Bosch, van den Hoogen, & Zitman, 1998). However, findings also highlight that the majority of MDD patients recover within a time frame of one year from a major depressive episode whilst only a minority of individuals experiences no recovery or remission of symptoms and their condition can take a chronic course for years (Richards, 2011). When compared to MDD, the clinical course of BD-I is generally characterised by a higher but shorter number of episodes and cycles (Angst & Preisig, 1995; Perlis, et al., 2006). In a prospective longitudinal study, 93% of the 82 BD-I patients experienced at least one affective episode within a 10-year time period following their intake, with the average number of episodes being three, lasting on average ten weeks (Judd et al., 2003).

Comorbidity in mood disorders is generally high. In particular, anxiety disorders, substance use disorders and impulse control disorders occur frequently within the context of MDD and BD-I. About 60% of MDD patients and 87% of BD patients suffer from a comorbid lifetime anxiety disorder, most commonly generalised anxiety disorder or phobic disorder (Kessler, et al., 2003; Kessler, et al., 2007). MDD and BD have been associated with high rates of lifetime substance use disorders of 24% and 60%, respectively (Kessler, et al., 2003; Kessler, et al., 2007). Moreover, impulse control disorders such as intermittent explosive disorder, pathological gambling, bulimia nervosa, conduct disorder, oppositional defiant disorder or antisocial personality disorder occur in approximately 32% of MDD patients and 72% of BD individuals throughout their life (Kessler, et al., 2003; Kessler, et al., 2007). Comorbidity has also been established between mood disorders and physical conditions such as cardiovascular disorders, diabetes mellitus or respiratory syndromes (Kupfer, 2005; Richards, 2011). Moreover, the presence of a comorbid disorder has also been associated with greater severity of

symptoms, lower treatment response rates, greater social and occupational impairment (Richards, 2011; N. M. Simon et al., 2004).

Mood disorders are highly disabling conditions and provide a substantial social and economic burden to society (G. E. Simon, 2003). They cause marked social and functional impairment that ultimately results in a decreased quality of life (Papakostas et al., 2004; G. E. Simon, Bauer, Ludman, Operskalski, & Unutzer, 2007) and have been associated with highly increased morbidity and mortality rates. The risk of completed suicide is about 20 times higher among both BD and MDD inpatients than for the general population (Holma et al., 2010; Osby, Brandt, Correia, Ekbom, & Sparen, 2001).

The effective treatment of mood disorders involves pharmacotherapy or psychotherapy, either as monotherapy or in combination. Antidepressants are the most commonly prescribed medication for MDD (Nemeroff & Owens, 2002), while mood stabilizers such a lithium are the most frequently applied pharmacological treatment for BD (Blanco, Laje, Olfson, Marcus, & Pincus, 2002). This discrepancy in pharmacological treatment approaches together with the heterogeneity of symptoms and epidemiological characteristics suggests a partly distinct underlying pathology of MDD and BD. However, research on the aetiology of mood disorders also points towards an at least partly shared biological basis of the two – an important topic that will be discussed in the next paragraph.

1.3 Aetiology of mood disorders

Environmental and genetic factors both play an important role in the aetiology of mood disorders and recent findings suggest that their interaction is of relevance, too (Lau & Eley, 2010). Since human brain development is influenced by continuing complex interactions of genetic and environmental influences, genetic and environmental factors that have been found to be associated with an onset of mood disorders will be discussed in this paragraph. It should be noted that

relatively little is known about the relative impact of genes and environment on grey matter while the brain is actively developing (Lenroot & Giedd, 2008). Nevertheless, this paragraph aims at describing genetic and environmental factors that have been associated with mood disorders and are thus likely to impact on grey matter pathology commonly observed in mood disorders.

1.3.1 Genetic susceptibility factors

The first indication that genes may contribute towards the pathogenesis of mood disorders initially derived from the observation that mood disorders frequently run in families (Lau & Eley, 2010). Many different approaches have since been applied to examine the role of genetic factors in the aetiology of mood disorders, including family and twin study designs as well as molecular genetic analyses such as linkage and association approaches.

1.3.1.1 Family and twin studies

Family studies have shown that mood disorders aggregate within families and as such provided the first line of evidence to suggest that they may be heritable (Lau & Eley, 2010). Familial aggregation cannot, however, differentiate shared genetic from shared environmental influences, but family studies nevertheless provide very important information. It has now been established that first-degree relatives of MDD patients have a three-fold enhanced risk of developing MDD and that they are also three times more likely to suffer from BD as compared to the general population (Smoller & Gardner-Schuster, 2007; Sullivan, Neale, & Kendler, 2000). By contrast, first-degree relatives of BD patients have a 10-fold excess risk of BD as compared to the general population, and a 3-fold increased risk of MDD (Smoller & Finn, 2003). Given that MDD is more prevalent in the population than BD, the overall risk of developing MDD in first-degree BD relatives is about twice

as high as developing BD (Smoller & Gardner-Schuster, 2007). Importantly, these findings provide a first line of evidence towards pathogenic overlap between the two conditions.

To estimate the relative contributions of genetic and environmental factors in the aetiology of mood disorders, twin study designs provide a powerful tool. In a twin study, the concordance rates for a given phenotype (such as MDD or BD) are compared between monozygotic twins, who share about 100% of their genetic makeup, and dizygotic twins, who share about 50% of their genes. Since pair members of both monozygotic and dizygotic twins are considered to be exposed to similar family, social, and cultural environmental influences, a comparison of the concordance rates for a given phenotype between both groups of twins enables one to estimate the extent to which genetic factors play a role in the pathogenesis of the condition. Twin studies have consistently demonstrated that there is a strong genetic component of susceptibility to mood disorders, with heritability estimates ranging from 31-42% for MDD and from 60-85% for BD (Barnett & Smoller, 2009; Sullivan, et al., 2000). These findings illustrate that there is a strong genetic component in the aetiology of mood disorders and that the heritability of BD exceeds that of MDD.

1.3.1.2 Molecular genetic studies

Based on the high heritability estimates from twin studies, it became apparent that genes play an important role in the aetiology of mood disorders. However, a large magnitude of heritability does not provide any insights into the genetic architecture of a disorder such as the specific genes contributing to its development, their number and magnitude of effect or their mode of inheritance (Smoller & Gardner-Schuster, 2007). In an attempt to identify susceptibility genes or chromosome loci which house candidate genes for mood disorders, many molecular genetic studies have been conducted.

The two commonly applied approaches in molecular genetic research are linkage and association analyses. Linkage studies are carried out on families with at least two family members sharing a specific phenotype such as BD or MDD. This approach examines if genetic markers co-segregate within families. Association analyses by contrast compare the frequency of genetic markers in non-related individuals who do or do not carry the phenotype. A genetic marker is said to be associated with the phenotype if it occurs significantly more often in individuals with a certain phenotype than in individuals without the phenotype. It has now been established that genetic liability for MDD and BD arises from the effects of multiple susceptibility genes that are each of small effect and are neither necessary nor sufficient for the onset of the disorder (Lau & Eley, 2010).

Several genome-wide linkage analyses for MDD have been carried out, reporting relatively inconsistent findings across various chromosomic regions (Abkevich et al., 2003; Camp et al., 2005; Holmans et al., 2007; Zubenko et al., 2003). The potentially most important finding is a genome-wide linkage to chromosome 3p25-26 that has been found in two independent study samples (Breen et al., 2011; Pergadia et al., 2011). The linked region contains the gene *GRM7* which is known to encode a protein for the metabotropic glutamate receptor 7 but also contains many other genes (S. P. Hamilton, 2011).

Association studies are grouped into candidate gene association approaches and genome-wide association studies (GWAS). While candidate gene studies examine whether a specific a priori selected gene is associated with a disorder, GWAS assess the whole genome for association without having to specify candidate genes in a hypothesis-free approach. Candidate genes can be selected based on their proposed involvement in biological mechanisms underlying the disorder (biological candidate gene) or based on their genomic location that has already been linked to the disorder in previous research (positional candidate gene). This approach suffers several limitations. Firstly, the selection of candidate genes is restricted by the current knowledge of the disorder's biology or insights from

linkage or cytogenetic studies. Secondly, the likelihood of false-positive results is highly increased due to the limited prior evidence for an involvement of the gene in the disorder. Thirdly, most studies have been underpowered to robustly detect associations with genes that have small effect sizes, resulting in a disproportionally high number of false-negative findings (Smoller & Gardner-Schuster, 2007). The latter limitation also applies to GWAS although their sample sizes tend to increase more and more.

Hundreds of putative candidate genes have been investigated in association analyses of MDD (see Bosker et al., 2011 for a summary of candidate genes). Many of the genes studied are implicated in serotonergic neurotransmission, coding for serotonergic receptors or serotonergic transporters since abnormalities of the serotonergic system are considered to underlie the pathophysiology of MDD (Levinson, 2006). However, the majority of positive findings could not be replicated and have not been supported by meta-analyses or GWAS (Anguelova, Benkelfat, & Turecki, 2003; Lau & Eley, 2010). More recently, genes involved in neuroplasticity have become the focus of attention as it has been hypothesized that MDD is caused by stress-induced neurotoxic effects that damage cells of the hippocampus, thereby mediating several symptoms of depression (Levinson, 2006). Genes involved in the expression of growth and survival-promoting factors such as the *BDNF* gene which encodes the brain derived neurotrophic factor have accordingly been studied, with inconsistent results (Lau & Eley, 2010).

Nine GWAS have been published to date (Aragam, Wang, & Pan, 2011; Kohli et al., 2011; Lewis et al., 2010; Muglia et al., 2010; Rietschel et al., 2010; Shi et al., 2011; Shyn et al., 2011; Sullivan et al., 2009; Wray et al., 2012) but only one of these detected a genome-wide association with MDD, namely for the neuron-specific neutral amino acid transporter *SLC6A15* gene (Kohli, et al., 2011). Although being the largest GWAS of MDD as yet, a recent GWAS mega-analysis failed to identify susceptibility variants on a genome-wide supported level of significance (Ripke et al., 2013), and the same applies to a recently conducted

GWAS of depressive symptoms that included nearly 35,000 individuals (Hek et al., 2013). These results imply that the current sample sizes are still underpowered to detect small genetic effects that are typical for complex diseases such as MDD (Ripke, et al., 2013). It has also been argued that the phenotypic diversity of MDD, its underlying genetic heterogeneity and the influence of environmental factors towards liability to depression represent major obstacles for the identification of susceptibility genes (Kohli, et al., 2011).

Over the last decades, many linkage studies have been conducted on BD but results have been inconclusive (Craddock & Sklar, 2013). An analysis combining eleven genome-wide linkage scans of BD has identified genome-wide significance for linkage on chromosome 6q for BD-I (McQueen et al., 2005). This region has also been identified in an earlier genome-wide linkage analysis (Dick et al., 2003) but not in two meta-analyses of BD linkage analyses (Badner & Gershon, 2002; Segurado et al., 2003). The power to detect significant results is much greater when analysing combined study data than conducting meta-analyses which may explain the inconsistencies in findings. One advantage of the linkage approach is that when single large families are studied, it is possible to detect rare family-specific genetic variants that are involved in BD. This strategy has for example led to the identification of the disrupted in schizophrenia 1 (*DISC1*) locus on chromosome 1q24 in a large Scottish pedigree in which translocation of the gene co-segregates with psychiatric disorders such as BD and schizophrenia (St Clair et al., 1990). It has generally been difficult thus far to identify specific genes that contribute towards the development of BD on the basis of linkage findings. This phenomenon derives at least partly from the fact that the majority of linkage areas contain multiple plausible candidate genes which hampers the search (Nurnberger, 2012).

For BD, candidate gene analyses have predominantly focussed on the dopaminergic, serotonergic and noradrenergic systems as these are targeted by pharmacological treatment approaches of the disorder. In particular, genes encoding the serotonin transporter (*5HTT*), monoamine oxidase A (*MAOA*) and

catechol-O-methyl-transferase (*COMT*) have been extensively studied (Craddock & Sklar, 2009). Meta-analyses support the association of variation in the gene *G72* (also called *DAOA* for d-amino acid oxidase activator) which deactivates NMDA receptors, the gene encoding 5,10 methylenetetrahydrofolate reductase (*MTHFR*) and the serotonin transporter gene *SLC6A4* with BD (Smoller & Gardner-Schuster, 2007). Moreover, a reasonably consistent association has also been established for *BDNF* (Nurnberger, 2012). GWAS have identified a genome-wide significant association with the genes calcium channel, voltage-dependent, L type, alpha 1C subunit (*CACNA1C*), *ODZ4* and the neurocan gene *NCAN* (Cichon et al., 2011; P. Sklar et al., 2011). Calcium channels regulate neuronal excitability and are involved in long-term potentiation and synaptic plasticity, while the *ODZ4* gene is potentially involved in cell surface signalling and neuronal pathfinding (Sullivan, Daly, & O'Donovan, 2012). *NCAN* encodes neurocan, an extracellular matrix glycoprotein, which is considered to play an important role in cell adhesion and migration (Cichon, et al., 2011). Several other putative susceptibility genes for BD have been identified through GWAS that did not reach genome-wide significance, including *DGKH* (Baum et al., 2008), *MYO5B* (P. Sklar et al., 2008), *SKAP1* (Ferreira et al., 2008), *PALB2* (The Wellcome Trust Case Control Consortium, 2007) and many more.

It has recently been established that part of the genetic risk factors for MDD and BD are associated with a wide range of psychiatric disorders, indicating that there is a certain amount of pathogenic overlap between different conditions. In particular, variation in calcium-channel activity genes seems to play a role in the aetiology of many psychiatric disorders, including MDD and BD (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013).

1.3.2 Environmental susceptibility factors

While genetic factors undoubtedly play an important role in the aetiology of mood disorders, several environmental risk factors have also been identified. It has been estimated that about 35% - 63% of the variance in liability to MDD and about 21% - 38% of the variance in liability to BD can be accounted to individual-specific environmental factors that are not shared by family members (Foley, Neale, & Kendler, 1998; Kendler, Pedersen, Neale, & Mathe, 1995; Lichtenstein et al., 2009; Sullivan, et al., 2000). The contribution of shared environmental effects to the aetiology of mood disorders by contrast appears to be minimal (Smoller & Finn, 2003; Sullivan, et al., 2000). The latter finding does not indicate that shared environmental influences such as parental rearing style or poverty are irrelevant, but rather suggests that it may be important how an individual interacts with these environmental factors or how these factors influence an individual across developmental stages (Sullivan, et al., 2000).

Identified individual-specific environmental risk factors for MDD include disturbed parent-child relationships, sexual abuse, divorce, financial difficulties, low optimism, stressful life events, exposure to traumatic events, premature parental loss, substance misuse, predisposing personality traits, social difficulties and low social support (Kendler & Gardner, 2001; Kendler, Gardner, & Prescott, 2002, 2006). Similarly, susceptibility to BD has been associated with childhood abuse, disturbed parent-child relationships, childbirth and stressful life events (Alloy et al., 2005; Tsuchiya, Byrne, & Mortensen, 2003).

1.3.3 Gene-environment interplay

Although genetic and environmental risk factors for major psychiatric disorders appear to be easily distinguishable, there is a complex gene-environment interplay which complicates our understanding of their aetiology (Lau & Eley, 2010). On the one hand, genetic factors can influence the exposure to environmental conditions –

this relationship is also known as gene-environment correlation. For example, research has shown that there is a genetic influence on the likelihood of stressful life events so that genetic vulnerability can lead to a higher probability of stressful life events which in turn feedback and increase disease susceptibility (Kendler, 2001). On the other hand, genetic factors can influence the responses to environmental events and environmental events can attenuate genetic risk factors – a relationship called gene-environment interaction. For example, genetic differences can influence the likelihood that exposure to a stressful life event results in pathology (Nugent, Tyrka, Carpenter, & Price, 2011).

It has been speculated that the exposure to specific environmental experiences may explain why one disorder or another develops in vulnerable individuals (Kendler, Prescott, Myers, & Neale, 2003). The future of research on the aetiology of mood disorders clearly rests on investigating the underlying genetic and environmental susceptibility factors and their interplay to determine how these combine to produce the observed clinical symptoms specific to each disorder.

1.4 The endophenotype concept

Our understanding of the pathophysiological mechanisms underlying mood disorders is still limited. As highlighted in the previous paragraph, genetic factors play an important role in the aetiology of mood disorders but the specific susceptibility genes for MDD and BD remain largely unknown. Identifying genes that contribute to disease vulnerability holds the potential to obtain important insights into the causes of mood disorders that may help to develop novel diagnostic and therapeutic strategies (Glahn, Thompson, & Blangero, 2007). It is plausible that the heterogeneity of the clinical presentation of MDD and BD as well as their multifactorial and polygenic origin complicates the search for disease-related genes. Therefore, the concept of endophenotypes has been introduced to research on complex psychiatric diseases.

Endophenotypes represent disease-associated traits that are more closely linked to the underlying genetics than the clinical phenotype itself (Gottesman & Gould, 2003). They are measurable components that are not visible to the unaided eye and lie along the causal chain from genetic and environmental influences to their behavioural manifestations as symptoms (Peterson & Weissman, 2011). Accordingly, the rationale for identifying endophenotypes for psychiatric disorders is to detect more successfully susceptibility genes causing the disease, thereby providing new insights into the causal biological mechanisms. Gottesman and Gould (2003) developed the following criteria to identify endophenotypes: They have to be (a) associated with the illness in the population, (b) heritable, (c) primarily state-independent, (d) co-segregating with illness within families, and (e) found in unaffected relatives at a higher rate than in the general population.

An endophenotype can be neurophysiological, biochemical, endocrinological, neuroanatomical or neurocognitve in nature (Gottesman & Gould, 2003). In the following paragraphs 1.5 and 1.6, potential structural grey matter neuroanatomical and neurocognitive endophenotypes for MDD and BD are discussed. The major focus is placed on Gottesman and Gould's (2003) criterion 1 and 5, that is evidence towards neuroanatomical and cognitive pathology in BD and MDD patients as well as their unaffected relatives.

1.5 Structural grey matter abnormalities in mood disorders

Our understanding of the pathophysiological mechanisms underlying mood disorders is still limited. However, the application of neuroimaging techniques has remarkably expanded our knowledge of the critical neural processes and brain regions associated with the disease. A large body of structural magnetic resonance imaging (MRI) studies in mood disorders has now been published, identifying several neuroanatomical changes in affected patients and their close unaffected relatives. The majority of structural MRI studies are based on volumetric

representations of the brain and assess the volume of a brain structure by summing up the number of its voxels or by quantifying the proportion of grey matter using a voxel-based morphometry (VBM) approach. More recently, it has become possible to assess the thickness and surface area of cortical brain regions in an automated fashion. Given that cortical grey matter volume is a composite of cortical thickness and surface area, with research suggesting that cortical grey matter volume is more closely related to surface area than to thickness (Winkler et al., 2009), it appears likely that volumetric analyses of brain structures capture mainly surface area but not thickness differences.

This paragraph focuses on volumetric brain abnormalities in patients with mood disorders and unaffected relatives because only very few studies have directly assessed surface area or cortical thickness in MDD or BD as yet. Accordingly, with no meta-analyses being published thus far, the validity of surface area or thickness findings remains largely unknown. The findings from cortical thickness and surface area studies will therefore be discussed in the context of Chapters 4 and 5. The review of volumetric structural brain findings in unaffected relatives of mood disorders patients is restricted to relatives of BD patients as this subgroup of high-risk participants is examined further in the analyses provided in this thesis.

1.5.1 Structural grey matter abnormalities in affected patients

Structural brain abnormalities have been most consistently detected in the limbic system, basal ganglia, frontal and temporal lobes of mood disorders patients (Arnone et al., 2009; Beyer & Krishnan, 2002; Bora, Fornito, Pantelis, & Yucel, 2012; Hallahan et al., 2011; Kempton, Geddes, Ettinger, Williams, & Grasby, 2008; Kempton et al., 2011; Konarski et al., 2008; Koolschijn, van Haren, Lensvelt-Mulders, Hulshoff Pol, & Kahn, 2009; Savitz & Drevets, 2009). In detail, a significant enlargement of the lateral ventricles has been repeatedly documented in both MDD and BD patients (Dahabra et al., 1998; Hauser et al., 2000; Morys et

al., 2003; Zipursky et al., 1997), and been confirmed by meta-analyses for both conditions (Arnone, et al., 2009; Kempton, et al., 2008; Kempton, et al., 2011; McDonald, Zanelli, et al., 2004) and a recent mega-analysis containing 321 BD patients (Hallahan, et al., 2011). This ventricular abnormality has been hypothesized to reflect medial temporal lobe, lateral prefrontal cortex or basal ganglia volume reductions (Savitz & Drevets, 2009). Other commonly replicated structural brain findings in mood disorders are basal ganglia volume abnormalities, including the caudate, putamen and globus pallidus (Beyer & Krishnan, 2002; Konarski, et al., 2008; Lorenzetti, Allen, Fornito, & Yucel, 2009). While studies investigating basal ganglia in BD have been inconsistent, reporting both volumetric increases (Arnone, et al., 2009; DelBello, Zimmerman, Mills, Getz, & Strakowski, 2004; Hallahan, et al., 2011; Strakowski et al., 2002; Wilke, Kowatch, DelBello, Mills, & Holland, 2004), decreases (Beyer et al., 2004), or no changes (Haznedar et al., 2005; Kempton, et al., 2008; McDonald, Zanelli, et al., 2004), recent meta-analyses of MDD have repeatedly demonstrated volume reductions of the caudate, putamen and globus pallidus (Arnone, McIntosh, Ebmeier, Munafo, & Anderson, 2012; Kempton, et al., 2011; Koolschijn, et al., 2009). It should be noted that the conflicting findings of basal ganglia abnormalities in BD patients are potentially caused by confounding effects of length of illness, age of onset and/or (anti-psychotic) medication (Savitz & Drevets, 2009). In MDD (Arnone, et al., 2012; Kempton, et al., 2011; Koolschijn, et al., 2009) but not BD patients (Arnone, et al., 2009; Hallahan, et al., 2011; Kempton, et al., 2008), the thalamus and hippocampus have been found to be reduced in meta-analyses or mega-analyses. Importantly, a significant difference between hippocampal volumes of MDD and BD patients has been detected, suggesting that decreased hippocampal volume is specific to MDD and might differentiate MDD from BD neuropathology (Kempton, et al., 2011). Moreover, grey matter reductions of the amygdala have been detected in a meta-analysis of VBM studies of MDD patients (Bora, et al., 2012) and a mega-analysis of first-episode BD subjects (Hallahan, et al., 2011). However, several other meta-analyses have not detected amygdala differences between affected patients and

control subjects (Arnone, et al., 2009; Arnone, et al., 2012; Kempton, et al., 2008; Koolschijn, et al., 2009; McDonald, Zanelli, et al., 2004), with findings suggesting particularly high heterogeneity of amygdala volume across studies (Kempton, et al., 2008; McDonald, Zanelli, et al., 2004) that may partially be linked to medication effects, at least in BD patients (Hallahan, et al., 2011). Interestingly, smaller amygdala volumes have been particularly observed in paediatric and adolescent subjects and larger volumes have been most frequently found in older age (Konarski, et al., 2008).

Strong evidence also points to frontal and temporal lobe differences between healthy control subjects and individuals with mood disorders. Grey matter volume reductions of the prefrontal cortex have been consistently observed in mood disorders (Arnone, et al., 2009; Koolschijn, et al., 2009), particularly in the orbitofrontal gyrus as shown by several MDD meta-analyses (Bora, et al., 2012; Kempton, et al., 2011; Koolschijn, et al., 2009) and some individual studies and one post-mortem analysis of BD patients (Cotter, Hudson, & Landau, 2005; Stanfield et al., 2009; Wilke, et al., 2004). Meta-analyses have also provided strong evidence towards volume reductions of the inferior, middle and superior frontal gyrus in MDD (Bora, et al., 2012; Du et al., 2012) and inferior frontal gyrus decreases in BD patients (Bora, Fornito, Yucel, & Pantelis, 2010; Selvaraj et al., 2012). There is also partial evidence for volume reductions of middle and superior frontal regions in BD (Lopez-Larson, DelBello, Zimmerman, Schwiers, & Strakowski, 2002; Lyoo et al., 2004). Moreover, anterior cingulate volume decreases have been found in MDD (Bora, et al., 2012; Koolschijn, et al., 2009) and BD meta-analyses (Bora, et al., 2010). Furthermore, grey matter reductions have been observed in the precentral gyrus of MDD patients (Bora, et al., 2012), with partial evidence towards volume decreases of this cortical brain region in BD-I patients (Lyoo, et al., 2004). For the temporal lobe, a meta-analysis of first-episode MDD patients has found volume reductions of the right superior temporal gyrus and the medial temporal lobe, particularly in the parahippocampal gyrus

(Bora, et al., 2012). Contrarily, a recent mega-analysis of BD patients identified significant temporal lobe volume increases that were most pronounced in the left hemisphere (Hallahan, et al., 2011), while a VBM-based meta-analysis detected a cluster of grey matter reductions in the right temporal cortex encompassing the superior temporal gyrus (Selvaraj, et al., 2012).

Several of the brain regions altered in mood disorders are considered to participate in affect regulation and modulation – a core dysfunction in mood disorders. It has been proposed that the interplay of two neural systems is critical to facilitate affective processing. The ventral system, comprising the ventral prefrontal cortex, amygdala, insula, ventral striatum, thalamus, orbitofrontal gyrus and ventral anterior cingulate, is charged with the identification of the emotional significance of environmental stimuli, the generation of a subsequent affective state, and the production of an autonomic response. The dorsal system which includes the dorsolateral prefrontal cortex, the medial prefrontal cortex, the dorsal anterior cingulate and the hippocampus, is responsible for the effortful rather than automatic regulation of affective states (Phillips, Drevets, Rauch, & Lane, 2003a). Theoretically, dysfunction of specific components of the ventral or dorsal system or both could be underlying mood dysregulation in BD and MDD (Phillips, Drevets, Rauch, & Lane, 2003b).

1.5.2 Structural grey matter abnormalities in unaffected relatives

Only a few studies have examined whether structural grey matter abnormalities that are frequently observed in mood disorders patients also exist in unaffected close relatives of BD patients. Some studies did find volume reductions of the caudate in unaffected close relatives or twins of BD patients (McDonald, Bullmore, et al., 2004; McIntosh et al., 2004), while others reported volumetric increases of the caudate instead (Hajek et al., 2009b; Noga, Vladar, & Torrey, 2001), and several studies did not detect any volumetric abnormality in this brain region to be

associated with an increased liability to BD (Ladouceur et al., 2008; Matsuo et al., 2012; McIntosh et al., 2006). Volume reductions of the putamen have been associated with vulnerability to BD (McDonald, Bullmore, et al., 2004), but many studies, including region of interest (ROI) analyses, did not replicate this finding (Hajek, et al., 2009b; Matsuo, et al., 2012; McIntosh, et al., 2004; McIntosh, et al., 2006; Noga, et al., 2001). For the pallidum, only non-significant findings have been reported (Ladouceur, et al., 2008; McIntosh, et al., 2004; McIntosh, et al., 2006; Noga, et al., 2001). McIntosh and colleagues (2004) reported reduced volume of the bilateral thalamus in unaffected BD relatives, but others have not shown thalamus reductions to be associated with enhanced vulnerability to BD (McDonald, Bullmore, et al., 2004; McIntosh, et al., 2006). Ladouceur et al. (2008) reported increased hippocampal volumes in young healthy relatives of BD patients but the majority of research has not replicated this finding (Connor et al., 2004; Hajek et al., 2009a; Matsuo, et al., 2012; McDonald et al., 2006; McIntosh, et al., 2004; McIntosh, et al., 2006). Boccardi et al. (2010) found enlarged left amygdalae in their study cohort. However, others did not detect significant volumetric abnormalities of the amygdala or lateral ventricles in high-risk subjects (Hajek, et al., 2009a; Kieseppa et al., 2003; Matsuo, et al., 2012; McIntosh, et al., 2004; McIntosh, et al., 2006; Noga, et al., 2001). A recent ROI meta-analysis suggested indeed that there is no evidence for volumetric reductions of the amygdala, hippocampus or basal ganglia in healthy bipolar offspring (Fusar-Poli, Howes, Bechdolf, & Borgwardt, 2012).

A few studies have examined if structural grey matter abnormalities exist in cortical brain regions of unaffected BD relatives. Using a twin study design, van der Schot and colleagues (2010) found genetic risk of BD to be related to decreased grey matter volume in the right medial/dorsolateral frontal gyrus and precentral gyrus. McDonald et al. (2004) similarly reported liability to BD to be associated with medial frontal gyrus reductions but also found decreased grey matter in the anterior cingulate cortex of unaffected BD relatives. One study consisting of young

unaffected children and adolescents of BD patients observed a nominal trend towards middle frontal gyrus reductions (Ladouceur, et al., 2008), and another study reported larger right inferior frontal gyri in unaffected BD relatives (Hajek et al., 2013) but many other studies did not find any frontal or temporal grey matter abnormalities (Kempton et al., 2009; Kieseppa, et al., 2003; McIntosh, et al., 2004; McIntosh, et al., 2006; van der Schot et al., 2009). For the temporal lobe, only research by Ladouceur et al. (2008) found increased parahippocampal volumes in particularly young BD relatives.

In summary, there is large discrepancy in findings so that it remains largely unknown if any regional grey matter abnormality may serve as a neuroanatomic endophenotype for BD. Potential reasons for the observed heterogeneity of findings across studies are small samples sizes, small effect sizes of susceptibility genes, effects of age, the phenotypic heterogeneity of BD and the possible confounding effect of comorbid Axis-I disorders among both the unaffected relatives of BD patients as well as the BD patients themselves (Nery, Monkul, & Lafer, 2013).

1.5.3 The importance of studying grey matter pathology in mood disorders

Grey matter volume assessed with MRI is an indirect measure of a collection of various microscopic cellular elements, including neuronal cell bodies, axon terminals, dendrites, glial cells and blood vessels (Paus, 2005). Research has repeatedly shown that grey matter abnormalities commonly found in mood disorders are associated with histopathological abnormalities observed in postmortem analyses of MDD and BD patients. For example, postmortem studies have shown that the mean density and size of neurons in the orbitofrontal and dorsolateral prefrontal cortex is reduced in MDD patients (Rajkowska et al., 1999) and that the density of nonpyramidal neurons in the anterior cingulate is decreased in BD patients (Todtenkopf, Vincent, & Benes, 2005).

Grey matter abnormalities observed in mood disorders using MRI thus point towards pathological mechanisms in cellular structures and hold the potential to narrow the neurobiological basis of the disorder down, generate new hypotheses and stimulate research. For example, findings of grey matter reductions in mood disorders have stimulated the rise of the glutamate-induced excitotoxicity hypothesis and intensified research in this area (Savitz, Rauch, & Drevets, 2013). This theory holds that grey matter decreases in mood disorders result from a loss of neuropil which is associated with loss of glial cells, with each of these findings arising secondarily to glutamate-induced excitotoxicity (Price & Drevets, 2010).

Similarly, studying grey matter in individuals at high familial risk of mood disorders because of a close family history of the disorder enables one to examine whether grey matter pathology in specific brain areas emerges in the absence of disease as a consequence of shared genetic and environmental risk factors. This approach is particularly suitable to disentangle the pathogenesis of mood disorders as it provides important information about brain morphologic vulnerability factors that are associated with an increased risk of developing mood disorders and may help in the future to predict the onset of disease with high accuracy.

During early brain development, cortical grey matter volume increases as a function of thickening of the cortical mantle, expansion of the cortical surface area and gyrification of the cerebral cortex (Winkler, et al., 2009). In particular, grey matter volume in frontal and parietal brain regions increases until the ages of 10-12 years (Giedd et al., 1999). During adolescence and early adulthood, cortical grey matter volume decreases in a non-linear fashion across different brain regions. Reductions first appear in sensorimotor areas, followed by decreases in the frontal cortex, parietal cortex and finally in the temporal cortex (Gogtay et al., 2004). These grey matter reductions occurring around late adolescence and early adulthood have been linked to increased synaptic and neuronal pruning processes taking place at that time (Gogtay, et al., 2004). Interestingly, the onset of BD and MDD peaks around early adulthood when pruning processes are known to be

taking place. Studying grey matter abnormalities during early adolescence in individuals at high risk of mood disorders and affected patients can thus provide important new leads on locations of potentially abnormal brain development that are associated with enhanced disease vulnerability and the development of the disorder, respectively.

From a clinical perspective, studying grey matter holds the potential to identify structural brainmarkers for MDD and BD, thereby supporting the development of biologically based tests for psychiatric illnesses (Khandai & Aizenstein, 2013; Savitz, et al., 2013). The identification of neuroimaging biomarkers could enhance the differential diagnosis of mental disorders, improve the prediction of treatment response to pharmacotherapy and psychotherapy options, and potentially facilitate treating individuals at enhanced risk of disease prophylactically to prevent neurotoxicity and clinical deterioration (Savitz, et al., 2013).

Next to grey matter abnormalities, white matter pathology has also been linked to mood disorders. For example, increased white matter hyperintensities have been repeatedly found in BD and MDD patients (Kieseppa et al., 2014; Mahon, Burdick, & Szeszko, 2010; Tham, Woon, Sum, Lee, & Sim, 2011). Moreover, regional volumetric white matter abnormalities have been frequently observed in BD, particularly in (pre)frontal brain regions (Bruno, Barker, Cercignani, Symms, & Ron, 2004; McIntosh et al., 2005; Stanfield, et al., 2009). In addition to these findings, decreases in white matter integrity in mood disorders as reflected by lower fractional anisotropy have been documented in a broad range of white matter tracts. For BD and MDD, decreased fractional anisotropy has been most commonly observed along prefrontal-subcortical tracts (Adler et al., 2006; Mahon, et al., 2010; Tham, et al., 2011). Importantly, widespread white matter integrity reductions have also been observed in the familial high-risk subjects of this study cohort (Sprooten et al., 2011) and in individuals at high risk of BD because of subthreshold bipolar symptoms (Paillere Martinot et al., 2014).

White matter fibre bundles contain myelinated neuronal axons and glial cells. Myelinated axons connect grey matter regions and are integral to neuronal communication between grey matter regions (Cannon, 2010). Given the interdependency of grey and white matter, it appears important for our understanding of the neuropathology of mood disorders to focus both on grey and white matter.

1.6 Neurocognitive functioning in mood disorders

A broad range of cognitive deficits have been observed in mood disorders patients and their unaffected relatives. This paragraph focuses on neuropsychological deficits in BD and MDD patients as well as their unaffected relatives. The main focus of the neuropsychological findings in unaffected relatives of mood disorders patients is based on relatives of BD patients as this subgroup of high-risk participants is examined further in the analyses provided in this thesis.

1.6.1 Neurocognitive tasks commonly employed in mood disorder research

To assess sustained attention, variants of the Continuous Performance Task (CPT; Kurtz, Ragland, Bilker, Gur, & Gur, 2001) are commonly employed. During the task, visual stimuli are presented on a computer screen and participants are asked to respond as quickly as possible when a specific stimuli appears and to ignore all other stimuli. Relevant stimuli appear only at low frequencies on the computer screen and the task usually lasts at least 20 minutes to assess sustained attention during a monotonous condition. The most commonly studied performance measures include the mean or median of the time needed to respond to relevant stimuli (reaction time), the number of incorrect responses (i.e. responding to non-target stimuli) and the number of omissions (i.e. non-responding to a target stimuli). Since neuropsychological deficits in unaffected first-degree relatives of

BD patients have been shown to be relatively subtle (Arts, Jabben, Krabbendam, & van Os, 2008), it is important to analyse task parameters that vary sufficiently between subjects to be able to detect mild cognitive deficits. Therefore, it appears advantageous to study the reaction time rather than the number of incorrect responses or omissions.

To assess visual attention and processing speed, the Trail Making Test Part A (Reitan, 1958) is commonly used. During this part of the task, participants are asked to connect numbers between one and 25 in a consecutive order as fast and accurate as possible. The time (measured in seconds) that the subject needs to successfully complete the task is a good indicator of psychomotor speed. The Trail Making Test A is well-suited to study cognitive performance in individuals with mood disorders as well as their unaffected relatives since the assessment of time needed to complete the task is measured in seconds and thus allows to detect even subtle differences in processing speed. By contrast, an analysis of errors made during task performance appears not to be a sensible approach when aiming to detect cognitive dysfunction in individuals at high risk of mood disorders since neuropsychological deficits have been shown to be mild in this study population (Arts, et al., 2008).

To assess processing speed and cognitive control, variants of Digit Symbol/Coding tasks such as the Digit Symbol Substitution Test (DSST; Wechsler, 1955) or the Symbol Digit Modalities Test (SDMT; Smith, 1982) are commonly used. During the DSST, subjects are required to learn a list of digit-symbol associations (e.g., 1 = ^, 2 = #, 3 = +, etc.). Subsequently, a list of digits is shown and subjects are asked to write down the corresponding symbol for each digit as quickly as possible. The SDMT is identical to the DSST but reverses the presentation of the material so that a list of symbols is presented and subjects are asked to write down the corresponding number for each symbol as quickly as possible. The number of correct digit-symbol or symbol-digit associations during a time period of 90 or 120 seconds provides an indicator of processing speed and cognitive control. Both tasks

are highly correlated with each other but SDMT raw scores are consistently lower than those of the DSST (Bowler, Sudia, Mergler, Harrison, & Cone, 1992). Since neuropsychological deficits in unaffected first-degree relatives of BD patients have been shown to be relatively subtle (Arts, et al., 2008), it is important to analyse task parameters that vary sufficiently between subjects to be able to detect mild cognitive deficits. Therefore, it appears advantageous to assess processing speed and cognitive control in high-risk individuals using the SDMT as task performance is slightly more challenging.

Attentional processes and short-term auditory or visuospatial memory can be assessed using the Digit Span forwards (Wechsler, 1955), the Spatial Span forwards of the Cambridge Neuropsychological Test Automated Battery (CANTAB; De Luca et al., 2003) or the Letter-Number Span (Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997). During the Digit Span forwards, participants are asked to repeat a string of numbers that is read out to them. During the Letter-Number Span, participants are asked to repeat a string of numbers and digits that is read out to them. During both tasks, the length of strings presented progressively increases (described in detail in paragraph 2.3.3 of Chapter 2). The highest amount of digits (or digits and letters for the Letter-Number Span) that a participant is able to successfully repeat serves as a measure of attention and auditory short-term memory. The Spatial Span forwards can be considered as the non-verbal analogue of the Digit Span forwards. During the Spatial Span forwards, participants are asked to remember a sequence of squares lighting up on a computer screen. Similar to the Digit Span task, the length of sequences of locations presented progressively increases. The highest amount of sequences of boxes that a participant is able to successfully recall serves as a measure of attention and visuospatial short-term memory. Since the highest amount of digits, letters and digits or sequences that a participant is able to remember is the only available task performance parameter, no alternative performance measure is available that may

be superior when examining task performance in unaffected first-degree relatives of BD patients.

To assess attentional processes and short-term verbal (working) memory, the Digit Span backwards (Wechsler, 1955) is commonly administered. The task is identical to the Digit Span forwards. However, the participant is asked to repeat the strings of numbers in reverse order. The task is also fully described in paragraph 2.3.3 of Chapter 2. The highest amount of digits that a participant is able to successfully repeat in reverse order serves as a measure of attention and auditory short-term memory.

Visual memory can be assessed using subtests of the Wechsler Memory Scale (Wechsler, 1987) or the delayed recall of the Rey-Osterrieth Complex Figure Test (Rey, 1964b). The Wechsler Memory Scale consists of several subtests to assess various components of memory. Tasks assessing visual memory components can be grouped together and form the Visual Memory Index, a composite score of global visual memory performance. Subtests of the Visual Memory Index of the Wechsler Memory Scale include Visual Reproduction I and II. During these two tasks, immediate (Visual Reproduction I) and delayed (Visual Reproduction II) recall of a visual drawing task is assessed separately. The immediate and 30 minutes delayed recall of the Rey-Osterrieth Complex Figure Test requires participants to reproduce a complex line drawing figure from memory that had been previously copied. Since neuropsychological deficits in unaffected first-degree relatives of BD patients have been shown to be relatively subtle (Arts, et al., 2008), it is important to analyse task parameters that vary sufficiently between subjects to be able to detect mild cognitive deficits. Therefore, it appears advantageous to particularly study performance during the delayed stage of the task.

Verbal learning and memory can be assessed using the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 2000) or the Rey Verbal Learning

Task (RVLT; Rey, 1964a). During both tasks, a list of words is read out loud and the participant is asked to recall as many words as possible. For the CVLT, the list contains 16 words which can be semantically grouped into four categories. For the RVLT, the list contains 15 words which are not semantically related to each other. The word list of the CVLT or RVLT is then read out four more times to the subject and the words recalled are notified each time. Then, an interference list containing new words is read out once to the subject and the participant is asked to recall as many words as possible from this new list. Subsequently, the short-delay free recall of the task is taking place during which subjects are asked to recall again all words from the first list. For the CVLT only, this short-delay free recall is followed by the short-delay cued recall. During the short-delay cued recall, the subject is presented with four categories and asked to recall as many words as possible from the first list that belong to a specific category. Then, a 20 minutes delay is taking place when administering the CVLT or RVLT, followed by another free recall of the first list (long-delay free recall). For the CVLT, this long-delay free recall is followed by another cued recall of the first list (long-delay cued recall). Finally, a list of words is read out to the participants and they are asked to indicate whether or not they think that a given word had been part of the first list or not. For the CVLT and RVLT, various performance indices of verbal learning and memory can be analysed such as the amount of words correctly recalled during the initial list learning stage of the task, the short- and long-delay free recall (and cued recall for the CVLT only), the amount of perseveration errors and the amount of intrusions. The CVLT is described in more detail in section 2.3.2 of Chapter 2. When studying high-risk cohorts, it appears advantageous to administer the RVLT rather than the CVLT to assess verbal learning and memory. The RVLT does not rely as extensively as the CVLT on executive functions since words cannot be grouped into semantic categories and thus provides a purer measure of learning and memory. Since words cannot be grouped into categories during the RVLT, this verbal learning and memory test is more difficult than the CVLT. Since neuropsychological deficits in unaffected first-degree relatives of BD patients have

been shown to be relatively subtle (Arts, et al., 2008), it is important to analyse task parameters that vary sufficiently between subjects to be able to detect mild cognitive deficits. Therefore, it appears advantageous to study verbal learning and memory using the RVLT rather than the CVLT.

Semantic and phonematic verbal fluency can be assessed using the semantic and categorical subtests of the Verbal Fluency Test (Benton & Hamsher, 1978). During this task, subjects are asked to generate as many words as possible beginning with a specific letter (e.g., 'S') or belonging to a specific category (e.g., 'animals') during a period of 60 seconds. The amount of words correctly identified serves as a measure of task performance. Additionally, perseverations and errors can be assessed. Since neuropsychological deficits in unaffected first-degree relatives of BD patients have been shown to be relatively subtle (Arts, et al., 2008), it is important to analyse task parameters that vary sufficiently between subjects to be able to detect mild cognitive deficits. Therefore, it appears advantageous to study the amount of words generated rather than perseverations or errors as these are more likely to appear frequently in patients with severe cognitive disabilities.

To examine executive control and cognitive flexibility, the Trail Making Test part B (Reitan, 1958) can be administered. During part B of the Trail Making Test, participants are asked to connect numbers and digits in an alternating fashion (e.g., 1, A, 2, B, 3, C, etc.) as fast and accurate as possible. Task performance relies to a large extent on executive function since attention needs to be switched from one stimulus dimension to another. Accordingly, the time (measured in seconds) that the subject needs to successfully complete the task is a good indicator of cognitive control and flexibility. Version B of the Trail Making Test is well-suited to study cognitive performance in individuals with mood disorders as well as their unaffected relatives since the time needed to complete the task is measured in seconds which allows to detect even subtle differences in executive control. By contrast, analysing errors made during task performance appears not to be a sensible approach when aiming to detect cognitive dysfunction in high-risk

individuals since neuropsychological deficits have been shown to be mild in this study population (Arts, et al., 2008).

Apart from the Trail Making Test B, cognitive flexibility can be assessed using various card-sorting tests such as the Wisconsin Card Sorting Task (WCST; Heaton, 1981), the Modified Card Sorting Task (Nelson, 1976) or the Intra-/Extradimensional Set Shifting Task (IED) of the CANTAB (Roberts, Robbins, & Everitt, 1988). During the WCST, participants are asked to sort cards according to a changing rule that is unknown to the subject. The cards can be sorted according to shape, colour or number according to visual feedback. When the participant has correctly sorted ten cards according to one rule, the rule is changed and the participant has to switch to another rule in order to perform the task correctly. In total, the rule is changed six times during the task. If a subject fails to learn a rule, the task is terminated. By contrast, the Modified Card Sorting Task is a variant of the WCST that is less complex and therefore more suitable for cognitively impaired individuals. The IED is also a modified version of the WCST that assesses different components of cognitive flexibility in a purer form. A detailed description of the IED can be found in paragraph 2.3.1 of Chapter 2. Performance measures of the WCST include the amount of categories achieved and perseveration responses. Categories achieved refers to the number of correct runs of ten sorts and can range between zero to six. A score of zero indicates that a subject failed to learn the rule so that the task was automatically terminated. A score of six indicates that the subject successful learned to sort cards according to the changing rules during the task. Since only patients with severe cognitive impairments fail to switch responding according to changing rules, this performance measure is not suitable when examining high-risk of mood disorders cohorts. Perseveration responses occur when a participant continues to sort cards according to a previously reinforced rule. Perseveration errors are a good indicator of overall task performance and the ability to set shift. Accordingly, it is commonly used in high-risk of mood disorders participants.

To examine executive control, the Stroop Color-Word Inference Task (Stroop, 1935) is often administered. During the task, single words are presented in coloured ink to the subjects and the subject is asked to name the colour of the ink as quickly as possible. The presented words include names of colours that either match or conflict with the ink name. Performance indices include the accuracy and response time. Since neuropsychological deficits in unaffected first-degree relatives of BD patients have been shown to be relatively subtle (Arts, et al., 2008), it is important to analyse task parameters that vary sufficiently between subjects to be able to detect mild cognitive deficits. Therefore, it appears advantageous to study response time rather than the number of errors made.

To assess immediate and delayed declarative facial memory, the Penn Facial Memory Test (Gur et al., 2001) can be administered. During the task, 20 faces are presented to the subject. After the initial learning stage of the task, the 20 target faces are presented together with 20 new distractor faces. The participant's score reflects the number of correctly recognized target faces and correctly rejected non-target distractor faces. Moreover, the median response time for correct responses serves as a measure of processing speed. After 20 minutes, a long-delay recall is taking place. Since neuropsychological deficits in unaffected first-degree relatives of BD patients have been shown to be relatively subtle (Arts, et al., 2008), it is important to analyse task parameters that vary sufficiently between subjects to be able to detect mild cognitive deficits. Therefore, it appears advantageous to particularly examine performance during the delayed recall of the task.

1.6.2 Neurocognitive functioning in affected patients

A broad range of cognitive deficits have been observed in mood disorders patients. A recent meta-analysis of neuropsychological performance in first-episode MDD patients (Lee, Hermens, Porter, & Redoblado-Hodge, 2012) found cognitive impairments in psychomotor speed as measured with the Trail Making Test A

(Reitan, 1958), the DSST (Wechsler, 1955) or the SDMT (Smith, 1982). Moreover, the authors observed lower attention in the BD patients, assessed with the Digit Span forwards (Wechsler, 1955) or the Spatial Span forwards of the CANTAB (De Luca, et al., 2003), and visual learning and memory impairments as measured with the Visual Reproduction I and II of the Wechsler Memory Scale (Wechsler, 1987), the 30 minutes delayed recall of the Rey-Osterrieth Complex Figure Test (Rey, 1964b) or the Visual Memory Index of the Wechsler Memory Scale (Wechsler, 1987). Moreover, performance during several domains of executive function was significantly worse in first-episode patients than in healthy controls, including attentional switching as measured with the Trail Making Test B (Reitan, 1958), verbal fluency as assessed with the semantic and categorical subtests of the verbal fluency test (Benton & Hamsher, 1978) and cognitive flexibility using the WCST (Heaton, 1981), the Modified Card Sorting Task (Nelson, 1976) or the IED of the CANTAB (Roberts, et al., 1988). The effect sizes for the identified cognitive deficits were small to medium.

The advantage of studying first-episode MDD subjects is that the neuropsychological performance measures are considered to be less confounded by duration of illness or medication effects. Research has for example demonstrated that neuropsychological impairments are generally greater among MDD patients taking psychotropic medication (Snyder, 2013), having more major depressive episodes and longer durations of illness (Elgamal, Denburg, Marriott, & MacQueen, 2010).

In a meta-analysis of studies investigating euthymic BD patients, BD was associated with worse performance on a number of neuropsychological tests (Arts, et al., 2008). The largest effect sizes were observed for working memory as measured with the Digit Span backwards (Wechsler, 1955), executive control as assessed with the Trailmaking Test B, concept shifting as indicated by perseverative errors during the WCST, verbal fluency using words from a defined category as measured with the Verbal Fluency Test, delayed and immediate verbal

recall during the CVLT (Delis, et al., 2000) and mental speed as measured with the DSST. Of note, medium effect sizes were observed for executive control as measured with the Stroop Color-Word Inference Task (Stroop, 1935), mental speed as assessed with the Trailmaking Test A, delayed visual memory using the Rey-Osterrieth Complex Figure Test, verbal fluency using words beginning with a certain letter as part of the Verbal Fluency Test, sustained attention during the CPT (Kurtz, et al., 2001) and concept shifting as indexed by the categories achieved during the WCST.

Similarly, in a large mega-analysis of 1267 euthymic BD patients from 31 individual studies, Bourne et al. (2013) reported cognitive impairments for all studied neuropsychological tests, including various verbal learning and memory measures of the CVLT and RVLT (Rey, 1964a), processing speed as assessed with the Trail Making Test A and the Digit Span forwards, executive control during performance of the Trail Making Test B and cognitive flexibility as assessed with the WCST. Since the criteria for an endophenotype include that an endophenotype has to be state independent, it is advantageous to study euthymic patients as neurocognitive deficits have to be demonstrated in remitted patients to fulfil Gottesman and Gould's (Gottesman & Gould, 2003) criteria.

1.6.3 Neurocognitive functioning in unaffected relatives

Meta-analyses of neuropsychological performance measures in unaffected first-degree relatives of BD patients have yielded less widespread neuropsychological impairments than observed in mood disorders patients. In particular, executive control as measured with the Stroop task and the Trailmaking Task B as well as immediate verbal recall during performance of the CVLT have been found to be significantly worse in unaffected relatives as compared to control subjects (Arts, et al., 2008). By contrast, studying a large sample of multiplex multigenerational families, Glahn and colleagues (2010) reported worse performance of first-degree

BD relatives on measures of processing speed (Digit-Symbol Coding; Glahn et al., 2007), working memory as measured with the Object Delayed Reponse task (Glahn et al., 2006) and the Letter-Number Span (Gold, et al., 1997) as well as immediate and delayed declarative facial memory (Penn Facial Memory Test; Gur, et al., 2001). BD patients were also impaired on all of these subtests but only three of them showed a genetic correlation with affection status and may therefore serve as potential neurocognitive endophenotypes: Digit-Symbol Coding, Object Delayed Response and immediate facial memory as part of the Penn Facial Memory Test.

Glahn et al. (2012) also applied a novel approach to facilitate the identification of optimal endophenotypes for recurrent MDD. They developed the Endophenotype Ranking Value which examines the genetic utility of an endophenotype for a disorder by taking in to account the heritability of the disorder, the heritability of the candidate endophenotype and their genetic correlation. In a large study containing randomly-selected pedigrees, they found several neuropsychological measures that may serve as good candidate neurocognitive endophenotypes for MDD. These include verbal memory (recognition subtest of the CVLT), working memory (Digit Span forwards and Letter-Number Span), facial memory (immediate and delayed facial memory subtests of the Penn Facial Memory Test), attention (CPT hits and Trail Making Test A) and emotion recognition (Penn Facial Memory Test).

1.7 Summary and aims

The mood disorders BD and MDD are unified by a fundamental disturbance of mood. However, they also show a strong heterogeneity with respect to their clinical presentation and epidemiological characteristics. For example, MDD has higher prevalence estimates, is more common among women than men, is associated with an older age at onset and is treated with a different pharmacological approach as compared to BD. Despite these differences between the two conditions that suggest

a distinct underlying pathology of MDD and BD, research on their genetic aetiology also points towards an at least partly shared biological basis.

Several structural grey matter abnormalities and neurocognitive deficits have been detected in mood disorders patients but many findings have been inconsistent across studies. Similarly, research on unaffected close relatives has yielded inconsistent results. Accordingly, it remains largely unknown which of the commonly observed structural brain abnormalities and neuropsychological impairments in mood disorders patients are also present in close unaffected relatives and may therefore classify as a neuroanatomical or neurocognitive endophenotype for the condition. Similarly, it is unclear to what extent structural brain abnormalities and cognitive deficits in mood disorders patients and their close unaffected relatives change over time. Moreover, there is a lack of prospective longitudinal studies examining the nature of neuroanatomical and neurocognitive markers in mood disorders patients. For example, it remains controversial if they predate the onset of illness, only emerge as a function of illness onset or if they are related to medication effects, duration of illness or severity of symptoms.

Here, I present findings from a prospective longitudinal study of structural MRI measures and cognitive functioning in individuals at high familial risk of mood disorders who either remained well during a two-year time interval or developed MDD. I investigate whether volumes of subcortical brain structures (Chapter 3), thickness measures (Chapter 4) or surface area estimates of various cortical ROI (Chapter 5) may serve as an endophenotype for mood disorders and if they predate the onset of MDD and may therefore distinguish between individuals who develop MDD and those who remain well. Moreover, I compare the time-course of the structural measures over a two-year time period in high-risk individuals who developed MDD, high-risk subjects who remained well and healthy controls. Chapter 6 examines whether there are neuropsychological deficits in the high-risk group that may form a neurocognitive endophenotype for the condition and if there are cognitive deficits that distinguish between individuals who subsequently

develop MDD and those who remain well. Moreover, the time-course of cognitive deficits over the two-year time interval is investigated. Chapter 7 provides a summary of the main findings, lists the methodological limitations of this study and suggests implications for future research.

Chapter 2

Methods

2.1 The Bipolar Family Study

The Scottish Bipolar Family Study is a large prospective longitudinal study of individuals at high and low familial risk for mood disorders. It is well designed to examine the timing of structural brain abnormalities in mood disorders and their relationship to familial risk and onset of illness. The Bipolar Family Study examines several neuroscientific dimensions. The initial baseline assessment has been conducted between 2007 and 2012 and includes extensive clinical interviews and neuropsychological examinations. Moreover, blood samples were taken for DNA analysis and structural and functional MRI scans were acquired. The follow-up assessment has been carried out between 2009 and 2013 and took place approximately two years after each participant had completed the initial baseline assessment. The second assessment also included a clinical interview and neuropsychological testing as well as the acquisition of structural and functional MRI scans. The research project is ongoing and currently recruiting individuals for their third assessment.

2.1.1 Participants

Participants of the Bipolar Family Study were either at high familial risk for mood disorders or at low familial risk for mood disorders. Individuals are considered at high risk of mood disorders because of a close family history of BD. In particular, high-risk participants have at least one first-degree or two second-degree relatives with a clinical diagnosis of BD-I. As outlined in Chapter 1, mood disorders highly co-segregate within families and first-degree relatives of BD patients have a 10-fold excess risk of BD as compared to the general population, and a 3-fold

increased risk of MDD (Smoller & Finn, 2003). Hence, the overall risk of MDD in first-degree BD relatives is with 15% about twice as high as the risk of BD since MDD is more prevalent in the population than BD (Smoller & Gardner-Schuster, 2007). Individuals were considered at low familial risk of mood disorders if they had no personal or family history of mood disorders.

Participants at high familial risk of mood disorders were identified via their affected relatives as follows: Case loads of psychiatrists across Scotland were searched for patients diagnosed with BD-I. The diagnosis of affected subjects was confirmed with the Operational Criteria Symptom Checklist (OPCRIT; McGuffin, Farmer, & Harvey, 1991) using information from clinical case notes and the Structured Clinical Interview for DSM-IV Axis-I Disorders (SCID; First, Spitzer, Gibbon, & Williams, 1996). The BD-I patients were asked to identify close family members aged 16-25 years. Following informed consent, unaffected individuals with at least one first degree, or two second degree relatives with BD-I were invited to participate. The high risk of mood disorders participants were currently well and did not fulfil any exclusion criteria outlined below.

Unaffected, unrelated control subjects with no personal or family history of bipolar disorder were identified from the social networks of the high risk subjects and group-matched for age, sex and premorbid intelligence estimated with the National Adult Reading Test (NART; H. Nelson, 1982). Comparison subjects were screened for Axis-I disorders using the SCID.

At baseline assessment, exclusion criteria for all study groups included a personal history of major depression, mania or hypomania, psychosis, or any major neurological or psychiatric disorder, a history of substance dependence, a history of learning disability, any history of head injury that included loss of consciousness and any contraindications to MRI.

Approximately two years after the initial baseline examination, all participants were invited for a follow-up assessment. The diagnostic status of consenting

subjects who did not return for a second assessment was determined through written contact with the National Health Service (NHS). Written informed consent was acquired from all subjects and the study was approved by the Multicentre Regional Ethics Committee for Scotland.

2.1.2 Clinical assessment

Clinical assessments were conducted at the time of the first and second MRI scan by experienced psychiatrists (Andrew M. McIntosh and Jessica E. Sussmann) using the SCID. At both assessments, current manic and depressive symptoms were rated using the Young Mania Rating Scale (YMRS; Young, Biggs, Ziegler, & Meyer, 2000) and Hamilton Depression Rating Scale (HAM-D; M. Hamilton, 1960).

Whilst there have been a lot of research efforts to determine and quantify prodromal symptoms in psychosis (Fusar-Poli et al., 2013), prodromal criteria have not yet been applied to the study of mood disorders (Howes et al., 2011). Accordingly, no specialised clinical instruments capturing prodromal symptoms for mood disorders were available when this study was designed. Moreover, since depressed mood, irritability, racing thoughts and physical agitation have been associated with a subsequent onset of BD (Howes, et al., 2011), the YMRS, HAM-D and SCID appear to be well-suited to capture these putative prodromal symptoms of BD. Moreover, they are the most commonly applied clinical instruments in research on BD and administering them was considered to be advantageous given that it maximizes the ability to synthesize findings with other studies and increase chances of later replication.

Based on the follow-up clinical examination or the information provided by the case notes, high-risk subjects were grouped into those who remained well (HR-well), and those who subsequently developed MDD (HR-MDD). In total, 114 high-risk individuals provided suitable MRI data along with clinical information at baseline assessment. Of these, two individuals developed BD-I during the two-year

follow-up period and were subsequently excluded from all analyses due to the small sample size. Overall, 20 high-risk participants received a diagnosis of MDD within the two-year period, but one individual had to be excluded from baseline analysis due to unsatisfactory quality of the MRI scan. Accordingly, our analyses included 92 HR-well and 19 HR-MDD subjects at baseline. Of the healthy control subjects (HC), 96 provided suitable MRI data along with clinical information at baseline. Three of them developed MDD in the follow-up period and were therefore excluded from all analyses, leading to a sample size of 93 HC subjects. At follow-up, 63 HR-well, 20 HR-MDD, and 62 HC subjects provided suitable data. Four HR-MDD participants were prescribed antidepressant medication at follow-up. Three subjects were taking selective serotonin reuptake inhibitors (1 fluoxetine, 1 citalopram, 1 sertraline) and one participant was on a tricyclic antidepressant (lofepramine). The remaining 16 HR-MDD subjects were unmedicated.

2.1.3 Neuropsychological assessment

Several neuropsychological tasks were administered at baseline and follow-up assessment by trained research assistants. These include the IED to assess various components of cognitive flexibility, the CVLT to measure verbal learning and memory performance and the Digit-Span to assess attentional processes and (working) memory functions. A detailed description of these tasks and their supposed underlying brain regions are provided in paragraph 2.3 of this chapter.

2.1.4 Magnetic resonance imaging acquisition and pre-processing

Structural brain imaging at baseline and follow-up assessment was carried out at the Brain Imaging Research Centre (BIRC) for Scotland on a GE 1.5 T Signa Horizon HDX scanner (GE Medical, Milwaukee, USA). The T_1 sequence was a

coronal gradient echo sequence with magnetisation preparation (MPRAGE) and yielded 180 contiguous 1.2 mm coronal slices (TI = 500 ms; TE = 4 ms; matrix = 192 x 192; flip angle = 8°).

The acquired T_1 brain images from baseline and follow-up assessments were converted to mgz format and processed using the surface-based and volume-based streams of the FreeSurfer software package version 5.1.0 (http://surfer.nmr.mgh.harvard.edu/fswiki/recon-all/). Measures of volume, cortical thickness and surface area of several brain ROI were acquired (see paragraph 2.2 for more details about the processing with FreeSurfer).

2.2 Principles of structural magnetic resonance brain imaging

Structural MRI is a powerful in-vivo imaging technique that allows to obtain noninvasively information regarding the brain's anatomy. Further processing of the acquired MRI data enables the estimation of volume, thickness and surface area of several brain structures.

2.2.1 Magnetic resonance imaging

The development of MRI dates back to the 1940s when scientists at Harvard and Stanford University discovered a resonance phenomenon in samples that were placed in a magnetic field. Edward Purcell and Felix Bloch were jointly awarded the 1952 Nobel Prize for Physics for this ground-breaking discovery (Jezzard & Clare, 2008).

MRI makes use of the magnetic property of the atomic nuclei of hydrogen (protons) which are predominantly found in soft tissue of the human body such as the brain. A fundamental physical property of protons is that they are positively charged and possess an angular momentum which is called 'spin'. Accordingly, protons have a minute magnetic field and tend to align along the direction of an

applied magnetic field, similar to a compass needle. If there was no thermal agitation of the nuclei, all spins would align parallel to the external magnetic field and the nuclear moment would be in the lowest energy state. However, there is great amount of thermal agitation of the nuclei at physiological temperature which causes the spins to align parallel or anti-parallel (highest energy state) to the magnetic field. Only the small imbalance of slightly higher numbers of spins pointing parallel than anti-parallel to the field contributes to the net magnetic moment of the ensemble (Jezzard & Clare, 2008).

When a person is placed inside the MRI scanner, a strong magnetic field is generated by a large electromagnet which causes the average magnetic moment of protons to become aligned along the direction of the magnetic field. Subsequently, radiofrequency (RF) energy is briefly applied which induces energy transitions between the two energy states and causes the protons aligned with the magnetic field to absorb the energy and to flip their spin. The specific frequency of the radiofrequency energy is called Lamor or resonance frequency. Depending on the selected intensity and duration of the resonance frequeny, the degree to which the spin of the proton is affected varies. When the RF magnetic field is switched off, the protons release the absorbed energy and their spins return to their initial state of equilibrium at a rate determined by the T_1 and T_2 relaxation times. It is this energy which is detected with large coils as a RF signal that forms the final MR signal. Depending on the physical and chemical properties of the tissue, the T_1 and T_2 relaxation times vary. Spatial localization is obtained by superimposing a spatially varying gradient magnetic field on the uniform main magnetic field which causes nuclei at different positions to precess at different frequencies. Multiple slices can be acquired simultaneously and the distribution of protons can be calculated from the signal using inverse Forier transformation (Edelman & Warach, 1993).

The resulting magnetic resonance images consist of spatially localized signal intensities which are represented by greyscale colour codings. Depending on the tissue properties, strength of the magnetic field, pulse sequence, and other factors,

the signal intensities vary. Images can be T_1-weighted which causes fat-containing tissues to appear brighter and liquor darker and provides a good contrast of white and grey matter. Using T_2-weighed images by contrast, liquor appears bright and the soft tissue dark. This contrast is for example particularly well suited when investigating ischemic lesions (Edelman & Warach, 1993; Jezzard & Clare, 2008).

2.2.2 Processing structural brain images with FreeSurfer

FreeSurfer is a set of automated tools for subcortical segmentation and reconstruction of the brain's surface based on structural MRI data. This freely available software package has been developed by the Laboratory for Computational Neuroimaging of the Athinoula A. Martinos Center for Biomedical Imaging in Boston. It contains a fully automated structural imaging stream for processing cross-sectional and longitudinal data. Broadly, there are two different processing pipelines. The volume-based stream facilitates the segmentation of macroscopically visible subcortical brain structures and calculation of their volumes. The surface-based stream facilitates the reconstruction of cortical surfaces and provides several measures, including cortical thickness and cortical surface area (Fischl, 2012). According to FreeSurfer's terminology, 'cortical' essentially refers to brain regions that are considered to be part of the cerebral cortex. By contrast, 'subcortical' refers to brain structures of the diencephalon (e.g., thalamus) or telencephalon (e.g., basal ganglia, amygdala and hippocampus) that are located inferior to the cortices of the cerebral cortex. Although the hippocampus is, strictly speaking, anatomically part of the cerebral cortex, the term 'subcortical' will be used throughout the thesis to collectively refer to the diencephalic and telencephalic brain structures outlined.

There are several advantages of using (semi-)automated segmentation tools such as FreeSurfer when compared to manual segmentation approaches. First of all, the segmentation of regional brain volumes, cortical thickness and surface area using

FreeSurfer is less time-consuming and thus less labour-intensive than hand-tracing methods. Accordingly, processing large data sets may not be feasible using manual segmentation approaches and relies on optimised automated tools (Wenger et al., 2014). Second, automated morphometry approaches are to a greater extent user independent as they do not rely as heavily on expert knowledge in neuroanatomy as manual morphometry approaches (Eggert, Sommer, Jansen, Kircher, & Konrad, 2012). However, it needs to be highlighted that FreeSurfer includes various processing steps that have to be checked visually for accuracy and edited where required (see paragraphs 2.2.2.1 and 2.2.2.2 for details) so that sufficient neuroanatomical expertise is needed, too. Third, when combining data sets from different research groups for mega-analysis, segmentation results obtained from one specific automated segmentation tool such as FreeSurfer are not prone to interindividual differences in employing hand tracing guidelines and may thus result in higher interrater reliabilities than manual segmentation methods. However, as already pointed out earlier, FreeSurfer also requires visual inspection and manual intervention if needed which may also lower interrater reliability.

Several disadvantages exist when applying automated morphometry approaches. First of all, differences in brain structure between individuals may be very subtle and automated segmentation tools may not (yet) be able to detect them as accurately as neuroanatomical experts do (Eggert, et al., 2012). Second, automated tools such as FreeSurfer apply various processing algorithms to the MRI data which influence the accuracy of the final neuroanatomical volume, thickness or surface area measurements. In particular, it has been shown that the choice of algorithms for intensity correction, skull-stripping and tissue class segmentation impacts on the quality of the resulting segmentations (Acosta-Cabronero, Williams, Pereira, Pengas, & Nestor, 2008; Clark, Woods, Rottenberg, Toga, & Mazziotta, 2006; Eggert, et al., 2012; Fein et al., 2006). Importantly, research has shown that the version of FreeSurfer as well as the type of workstation and operating system used interact with these segmentation algorithms and influence the final

segmentation results (Gronenschild et al., 2012). Moreover, scanner-specific parameters such as type of scanner, field strength and pulse sequence have been shown to impact on regional brain volume, cortical thickness and surface area estimates (Han et al., 2006; Jovicich et al., 2009; Schnack et al., 2010). Accordingly, scientists are advised to segment and parcellate brain regions with FreeSurfer for their study cohort in exactly the same manner without changing any scanner-specific parameters, updating the FreeSurfer version or changing the type of workstation or operating system used for data processing. Although following these guidelines will help to achieve high reliability of the resulting brain volumes, cortical thickness and surface area estimates, this methodological issue results in difficulties when trying to compare and combine data from different study cohorts.

The FreeSurfer processing pipeline also includes the registration of the MRI scan to a brain atlas template which - depending on age, gender, handedness, disease status and other features of a study cohort - may not always have a good fit. For the volume-based stream of FreeSurfer, registration to the MNI-305 (Evans et al., 1993) atlas is conducted (see paragraph 2.2.2.1 for details). Given that this atlas template is made up of 239 male and 66 female participants (all right-handed) that were between 19 and 27 years old, the MNI-305 atlas template appears to fit the BFS data satisfactorily given that our sample consists of a similar age range, is predominantly right-handed and unaffected by neurological diseases. However, it should be considered that the brains of male and right-handed participants may fit the brain atlas template better than the brains of female and left-handed subjects. For the surface-based stream, the FreeSurfer pipeline aligns the MRI scans to the Desikan-Killiany atlas template (Desikan et al., 2006). This gyral-based template has been created by averaging MRI scans of 40 subjects who were between 19 and 86 years old (26 females and 14 males; handedness has not been reported), with 10 of them suffering from Alzheimer's disease. The rationale for the inclusion of large age ranges and Alzheimer's disease patients has been to create an atlas template that captures the range of brain atrophy typically found in studies on ageing

(Desikan, et al., 2006). Accordingly, it could be assumed that this atlas template may not ideally fit the data of the BFS given the young age of the BFS participants. However, since the geometry of the grey and white matter boundary drives the registration and segmentation process which is completely invariant to grey matter atrophy and would require huge amounts of white matter atrophy to result in unsatisfactory cortical labelling (personal communication with Dr. Bruce Fischl, Professor in radiology and software developer of FreeSurfer), it appears to be suitable to use the Desikan-Killiany atlas across all age ranges. Moreover, all cortical parcellations were visually inspected for accuracy and edited where necessary (see paragraph 2.2.2.2).

The limitations of FreeSurfer addressed above are particularly important when analysing longitudinal data. Research has shown that inter-scan variability due to limited accuracy of FreeSurfer may be in the same magnitude as the changes in brain morphology over time (Klauschen, Goldman, Barra, Meyer-Lindenberg, & Lundervold, 2009). It needs to be highlighted, however, that Klauschen et al. (2009) did not manually edit the resulting brain volume segmentations and that careful inspection and editing of FreeSurfer's segmentation and parcellation boundaries may enhance accuracy and thus allow for detecting subtle changes in brain architecture over time.

Despite these disadvantages of using automated morphometry tools such as FreeSurfer that impact on the accuracy and reliability of the resulting brain volume, cortical thickness and surface area estimates, both the volume-based and surface-based approach of FreeSurfer have already been shown to have a good validity when compared to manual segmentation (Kuperberg et al., 2003; Morey et al., 2009; Salat et al., 2004) and to histological analyses (Rosas et al., 2002). Moreover, good test-retest reliability has been demonstrated (Clarkson et al., 2011; Han, et al., 2006; Morey et al., 2010; Wonderlick et al., 2009). To allow for feasibility of this research project given its time limitations, it has been decided to process the BFS MRI scans with FreeSurfer rather than employing hand-tracing methods. To

improve accuracy, all regional brain segmentations and parcellations were visually checked and edited where needed as described in paragraphs 2.2.2.1 and 2.2.2.2.

2.2.2.1 The volume-based stream

The volume-based stream consists of several processing steps that facilitate the segmentation and labelling of various brain structures and provides measures of their volumes (in mm^3). The volumes of the following brain structures were assessed using the volume-based stream of FreeSurfer for each hemisphere separately: caudate, hippocampus, lateral ventricles, pallidum, putamen, thalamus and amygdala (see Figure 2.1). The volume-based stream of FreeSurfer has been shown to have a good validity when compared to manual segmentation (Dewey et al., 2010) and a generally good test-retest reliability except for the amygdala (Morey, et al., 2009).

The volume-based stream consists of several processing stages (Fischl et al., 2002; Fischl et al., 2004; https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferAnalysisPipelineOverview). First, the data is converted into mgz format which is readable by the FreeSurfer software. Next, non-parametric non-uniform intensity normalization is carried out to correct for intensity non-uniformity in the MRI data. Then, an affine registration to the MNI-305 atlas is calculated, designed to be insensitive to pathology and to maximise the accuracy of the segmentations. The MNI-305 atlas is a template of the average of 305 volumetric MRI scans of healthy control subjects which were each registered to the Talairach atlas brain (Talairach & Tournoux, 1988). The work has been conducted at the Montreal Neurological Institute in Canada, hence the name MNI (Evans, et al., 1993). Several of the other processes use Talairach coordinates as seed points. Next, an intensity normalisation is carried out to correct for fluctuations in intensity due to field inhomogeneity and all voxels are scaled so

that the mean intensity of white matter is 110. Subsequently, non-brain tissue is removed using a hybrid watershed deformation procedure (Segonne et al., 2004).

The resulting skull-stripped brain images were carefully visually inspected and edited where necessary according to standard FreeSurfer guidelines (http://surfer.nmr.mgh.harvard.edu/fswiki/Edits). An example of the intensity normalised brain image before and after the removal of non-brain tissue is shown in Figure 2.2

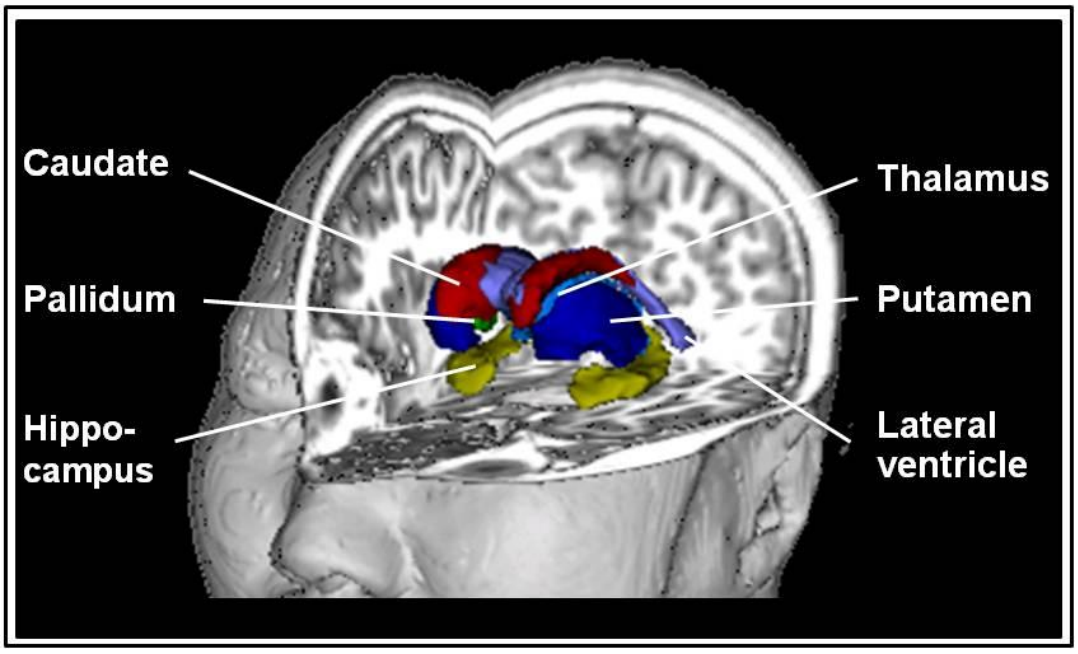

Figure 2.1 Three-dimensional representation of the subcortical ROI extracted with the volume-based stream of FreeSurfer. Not shown: amygdala.

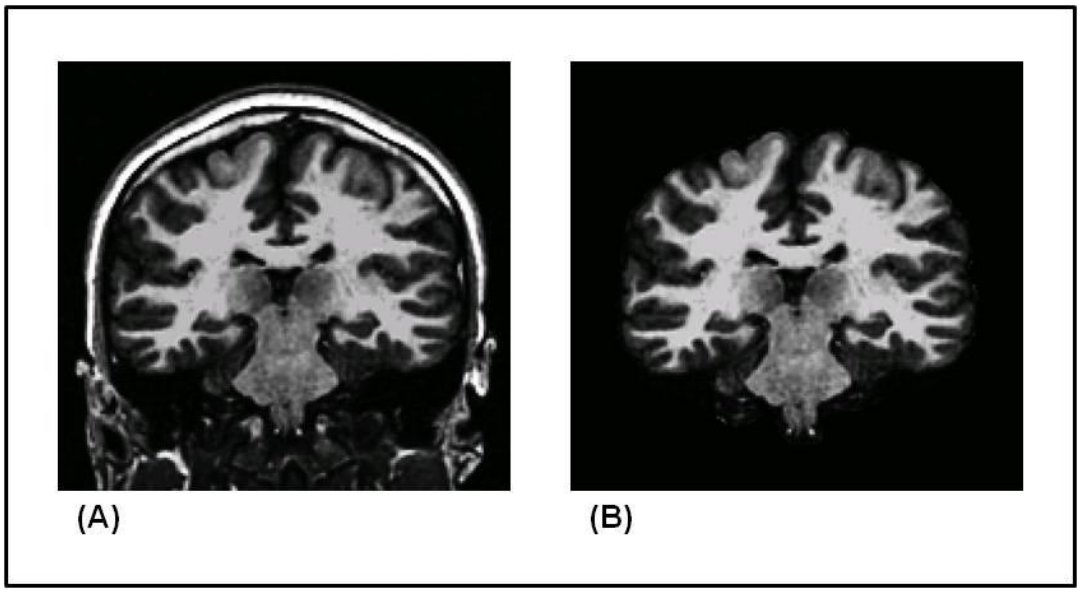

Figure 2.2 Visual illustration of a FreeSurfer processed brain scan before (A) and after (B) removal of non-brain tissue.

Finally, the procedure assigns a neuroanatomical label to each voxel in the resulting MRI volume based on a probabilistic atlas and segments the subcortical white matter and deep grey matter volumetric brain structures (Fischl, et al., 2002; Fischl, et al., 2004). This last step is based on a subject-independent probabilistic atlas as well as subject-specific measured values. The atlas is based on a data set of manually segmented images that were used to create statistics about the likelihood of a particular label being at any given location. The labels are then mapped into Talairach space to achieve voxel-wise correspondence for all subjects. Three types of probability are calculated at each voxel: (1) the probability that a given voxel belongs to each of the label classes (i.e. grey matter, white matter, cerebrospinal fluid); (2) the likelihood that a given voxel belongs to a label given the classification of its neighbouring voxels; (3) the probability distribution function of the measured intensity value. The classification of each voxel to a label is achieved by finding the segmentation that maximizes the probability of input given the prior probabilities from the probabilistic atlas (Fischl, et al., 2002; Fischl, et al., 2004). The final segmentations were carefully visually inspected for accuracy and edited

according to standard FreeSurfer guidelines (http://surfer.nmr.mgh.harvard.edu/fswiki/Edits) where necessary. A visual representation of the segmentations is provided in Figure 2.3.

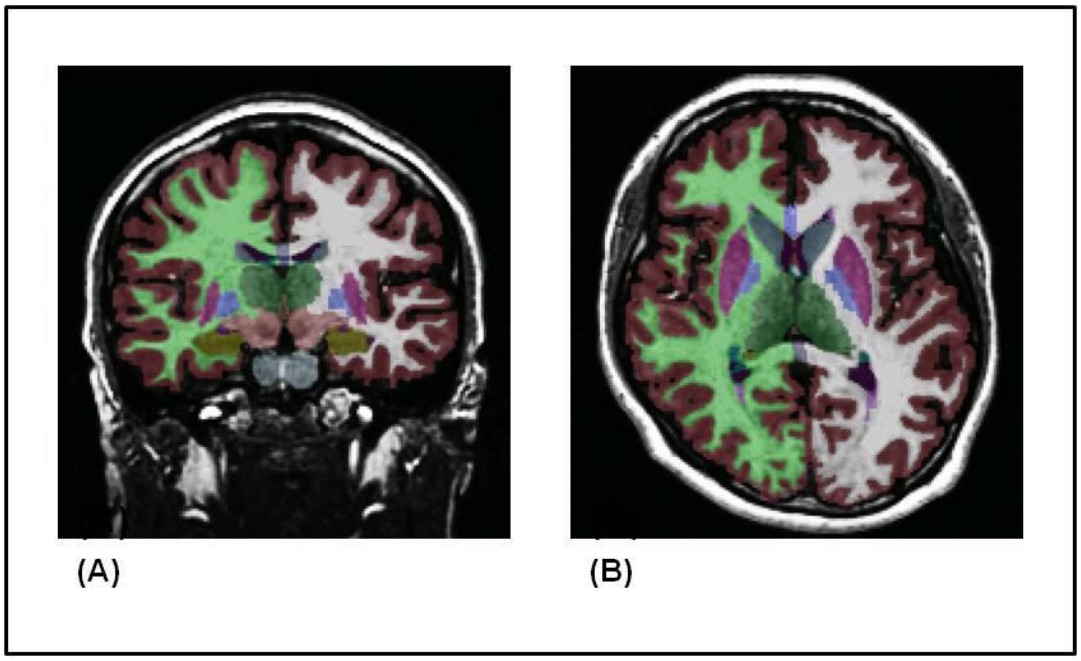

Figure 2.3 Coronal (A) and horizontal (B) visual illustration of volume-based segmented brain regions. Brown: grey matter; white & light green: white matter; yellow: hippocampus; dark green: thalamus; pink: putamen; blue: pallidum.

2.2.2.2 The surface-based stream

The surface-based stream consists of several stages that facilitate the reconstruction of cortical surfaces and provides measures of several cortical parameters (https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferAnalysisPipelineOverview). Amongst others, FreeSurfer provides information about volumes (in mm^3), cortical thickness in mm (defined as distance between white and pial surface) and surface area (mm^2) of many cortical brain regions. The surface-based stream of FreeSurfer has been shown to have a good validity when compared to manual measurement

(Kuperberg, et al., 2003; Salat, et al., 2004) and to histological analyses (Rosas, et al., 2002). Moreover, good test-retest reliability has been demonstrated (Clarkson, et al., 2011; Han, et al., 2006).

There are two types of cortical parcellations available: Cortical measures can either be based on the Desikan-Killiany atlas or on the Destrieux atlas. All brain measures of this thesis are based on the Desikan-Killiany atlas which subdivides the cerebral cortex into several gyral-based neuroanatomical regions (Desikan, et al., 2006). The following ROIs were assessed for subsequent analyses for each hemisphere (see Figure 2.4): frontal pole, inferior frontal gyrus, middle frontal gyrus, superior frontal gyrus, orbitofrontal gyrus (including lateral and medial orbitofrontal gyrus), anterior cingulate (including rostral and caudal anterior cingulate), precentral gyrus, superior temporal gyrus, parahippocampal gyrus and fusiform gyrus.

There are several advantages of selecting ROIs for analysis rather than performing vertex-wise comparisons. First, the selected ROIs encompass larger brain areas as compared to a single vertex and their grey matter estimates are thus supposed to vary to a larger extent between individuals and across time. Accordingly, comparing cortical surface area and thickness of ROIs cross-sectionally and longitudinally appears to be more powerful when aiming at detecting subtle differences between groups and over time. Second, findings deriving from the ROI approach chosen may provide more meaningful insights into the pathophysiology underlying mood disorders than vertex-wise approaches. For example, if cortical thickness of a single vertex in the inferior frontal gyrus is found to be reduced in high-risk individuals who go on to develop MDD as compared to subjects who remain unaffected by the disease, it appears challenging to further investigate and interpret the origin of this small-sized brain abnormality. By contrast, if cortical thickness of the inferior frontal gyrus as a whole is found to be reduced in high-risk individuals who go on to develop MDD as compared to subjects who remain unaffected by the disease, this finding can be more easily further investigated using fMRI paradigms, postmortem histological analyses or animal models and may lead

more easily to new hypotheses regarding the origin and effects of this widespread brain abnormality. Third, to avoid false positive findings, analyses are to be corrected for multiple testings. Given the relatively small sample size of the HR-MDD subjects however, it appears vital to select a small number of ROIs to have enough statistical power to be actually able to detect significant differences after correcting for multiple comparisons.

Nevertheless, the ROI approach also has several disadvantages. First, small circumscribed cortical thickness or surface area abnormalities may not be detected because only the mean value of a large brain structure is compared between groups and over time. Second, the precise location and extent of cortical thickness or surface area abnormalities remains unknown. For example, if cortical thickness in the inferior frontal gyrus as a whole is found to be reduced in high-risk individuals who go on to develop MDD as compared to subjects who remain unaffected by the disease, it remains unknown how the reduced mean cortical thickness emerged. It might have been caused by reduced thickness of all vertices of the inferior frontal gyrus or it might have been caused by some very extensive thickness reductions in just one part of the inferior frontal gyrus.

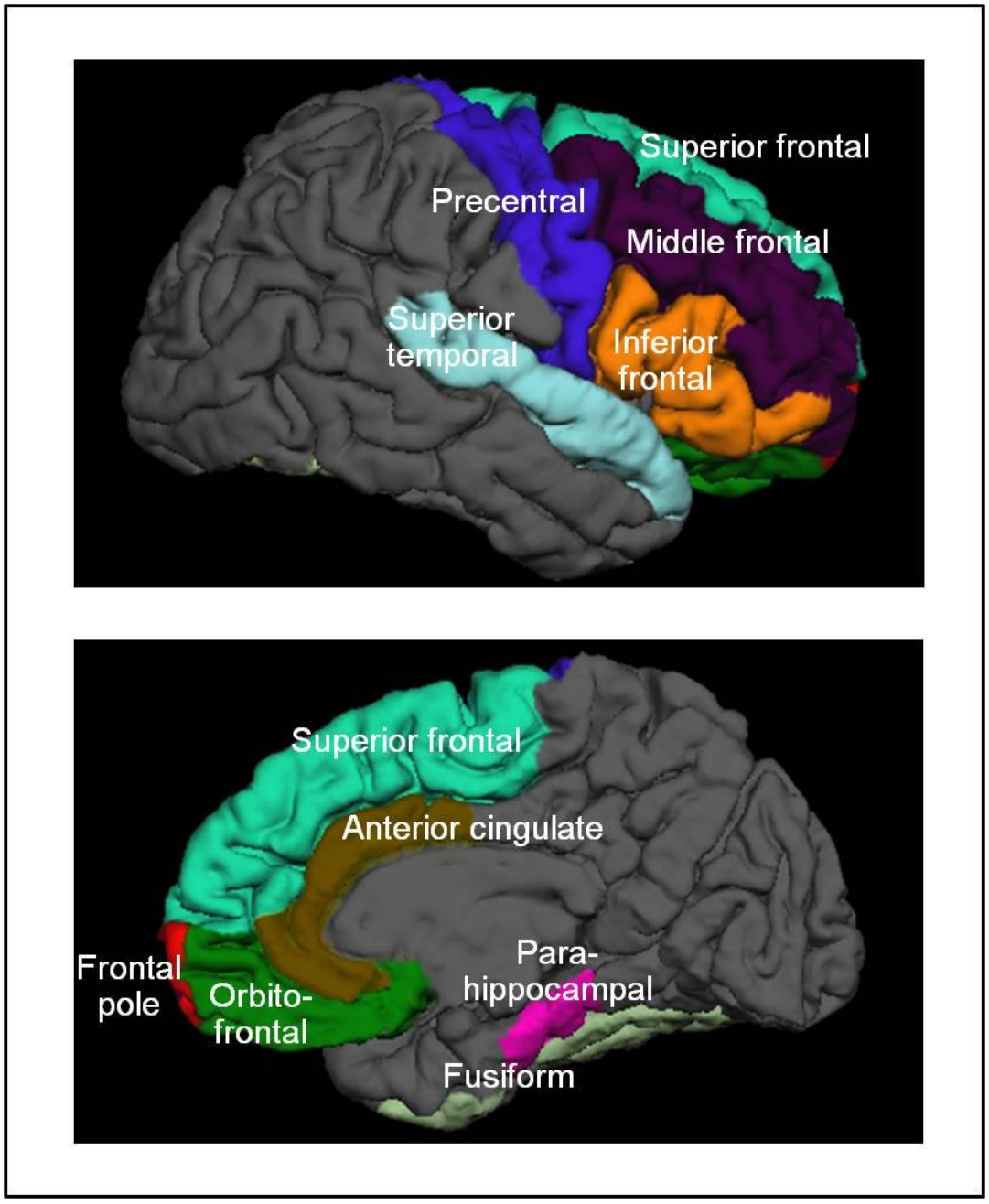

Figure 2.4 Visual representation of the cortical ROIs.

The surface-based analysis pipeline consists of several processing steps designed to create a three-dimensional model of the cortical surface to extract neuroanatomical measures such as cortical thickness and cortical surface area (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999). First, an affine registration to

Talairach space (Talairach & Tournoux, 1988) is conducted which allows FreeSurfer to compute seed points in later processing stages. This is a different algorithm than the one used for the volume-based stream. The B1 bias field is estimated automatically by measuring variation in white matter intensity. The main body of white matter is used to estimate the field across the entire volume. Based on their locations in Talairach space, their intensity and the intensity of their local neighbouring voxels, likely white matter points are selected. Then, a bias field correction is applied by dividing the intensity at each voxel by the estimated bias field at that location. Next, non-brain tissue is removed using a hybrid watershed deformation procedure (Segonne, et al., 2004), following segmentation of white matter by classifying voxels as either being white matter or non-white matter on the basis of intensity, neighbourhood and smoothness constraints (see Figure 2.5). The hemispheres are then separated from each other (see Figure 2.5) and the cerebellum and the brain stem are removed by cutting planes. The locations of the cutting planes are selected on the basis of the expected Talairach locations of the corpus callosum and pons as well as several rule-based algorithms that encode the expected shape of these structures. The software then generates an initial surface for each hemisphere by tiling the outside of the white matter mass for that hemisphere, followed by an automated topology correction that removes topological defects such as holes in a hemisphere (Fischl, Liu, & Dale, 2001; Segonne, Pacheco, & Fischl, 2007). Subsequently, the initial surface is refined following intensity gradients to optimally place the grey and white matter borders as well as the grey matter and cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to another tissue class (Dale, et al., 1999). Accordingly, the surface that follows the intensity gradients between the white and grey matter is referred to as white surface and the surface that follows the intensity gradients between grey matter and the cerebrospinal fluid is referred to as pial surface. The resulting tissue type segmented surfaces were visually inspected and edited where necessary according to standard FreeSurfer guidelines (http://surfer.nmr.mgh.harvard.edu/fswiki/Edits). An example, illustrating the white

and pial surfaces overlaid onto the skull stripped brain scan is provided in Figure 2.6.

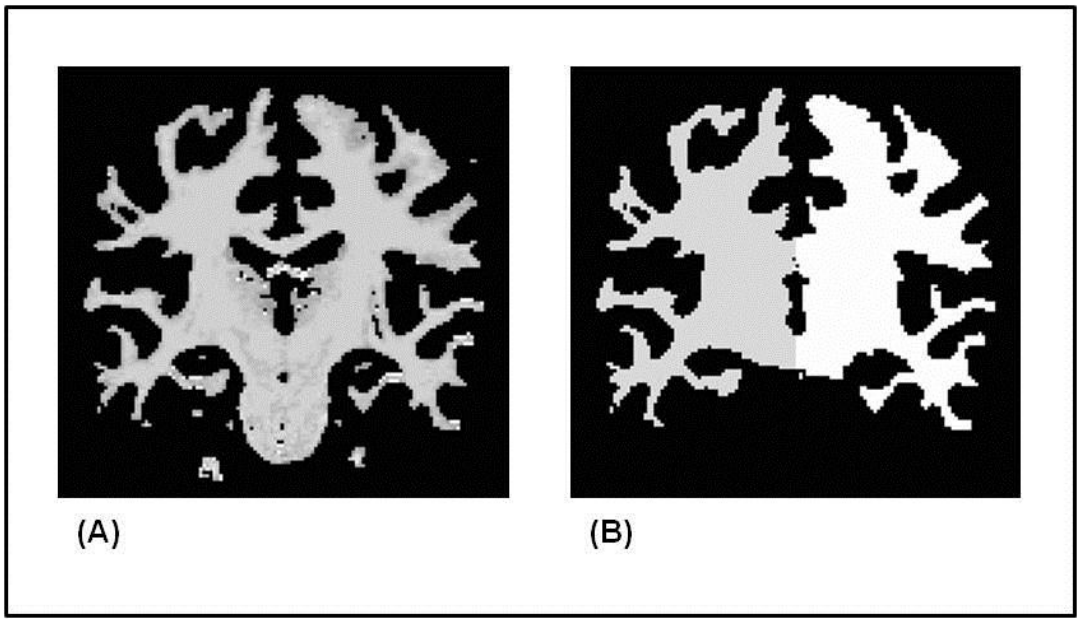

Figure 2.5 Illustration of the segmentation of white matter (A) and the subsequent separation of the two hemispheres (B), represented by distinct intensity values which are visualised as different shades of gray.

As a last processing step, the cortical surfaces are inflated, morphed and registered to an average spherical surface representation which utilises individual cortical folding patterns to improve the alignment (Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999). The cortex is subsequently parcellated into distinct brain regions based on the gyral and sulcal structure defined by the Desikan-Killiany atlas (Desikan, et al., 2006) (http://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation). Using this atlas, each vertex within each subject's inflated surface is probabilistically annotated with a neuroanatomical atlas label. The accuracy of all cortical parcellations was inspected visually blind to diagnostic status.

Measures of cortical thickness and surface area were then extracted for ROIs from the data. Cortical thickness is measured as the average shortest distance between the white and pial surfaces at each location (Fischl & Dale, 2000). Cortical surface area is computed as the average white surface area for each brain structure.

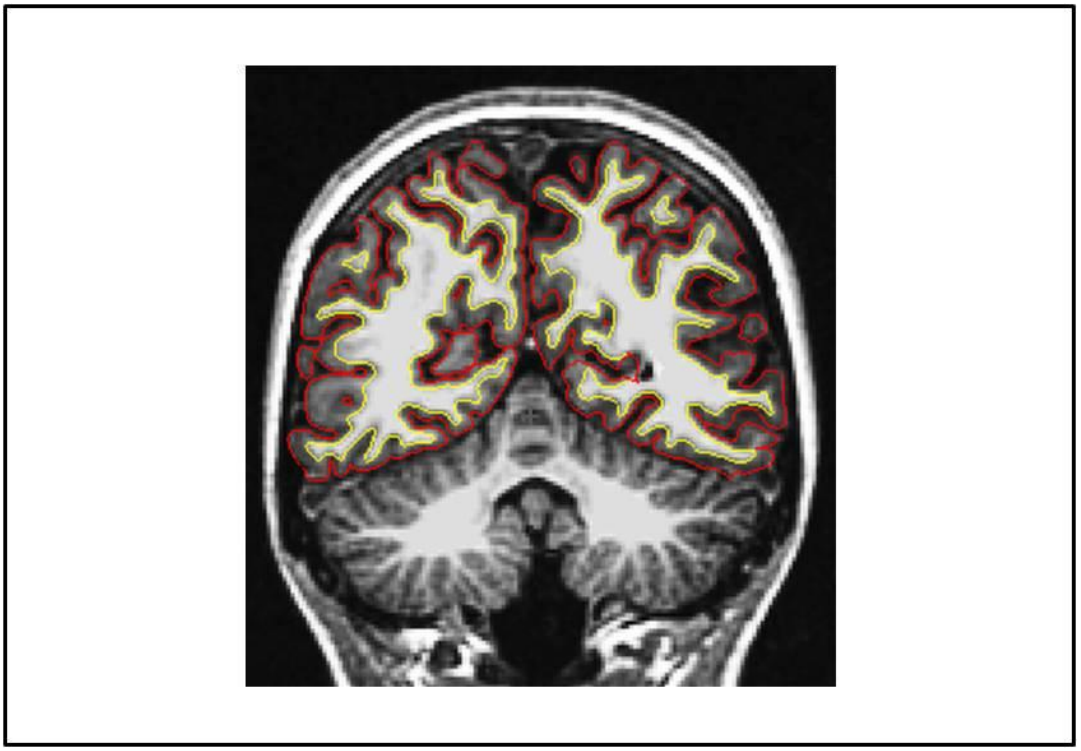

Figure 2.6 Visual illustration of the white (shown in yellow) and the pial (shown in red) surfaces, overlaid onto the original brain scan.

2.3 Principles of neuropsychological assessment

Major attempts have been undertaken to divide neuropsychological functions into small distinct cognitive components that are thought to rely at least partially on dissociable brain areas. In the BFS, the IED, CVLT and Digit-Span were used to cover the main domains of cognitive function. The IED was employed to assess

components of executive function (cognitive flexibility and attentional set-shifting), the CVLT was administered to measure verbal learning and memory performance and the Digit-Span was used to assess short term verbal memory and attentional processes. Moreover, it was initially decided to include these three neuropsychological tasks as studies had previously reported performance abnormalities in BD (Arts, et al., 2008). This paragraph introduces the neuropsychological concepts of the IED, CVLT and Digit-Span task and highlights their proposed underlying neural substrates as these tasks are the focus of Chapter 6.

2.3.1 The Intra-/Extradimensional Set Shifting Task

The IED is a computerized adaptation of the WCST and part of the CANTAB. The task has been developed to assess different components of cognitive flexibility, including attentional set-shifting. It is less complex than the original WCST and enables researchers "to study the formation, maintenance and shifting of cognitive set in a purer form" (Lawrence, Sahakian, & Robbins, 1998, p. 384).

The term attentional set-shifting refers to "the ability to switch attention from one aspect of a stimulus to another in an ongoing task, in accordance with changing reinforcement contingencies" (Chamberlain, Blackwell, Fineberg, Robbins, & Sahakian, 2005). This cognitive ability is crucial to flexibly alter behaviour according to environmental changes and therefore carries survival value (Kehagia, Murray, & Robbins, 2010). Two types of attentional set-shifting have been proposed (Downes et al., 1989), intradimensional and extradimensional set-shifting. An intradimensional shift requires switching attention from exemplars of one stimulus dimension (e.g., circles) to newly introduced stimuli of the same perceptual dimension or sensory modality (e.g., squares) on the basis of feedback. This cognitive ability reflects rule generalisation when novel stimuli are presented. In contrast, an extradimensional shift occurs when attention needs to be switched

between different perceptual dimensions or sensory modalities (e.g., from lines to shapes) on the basis of feedback, and is thought to reflect rule transfer from one stimulus dimension to another (Pantelis et al., 1999; Robbins & Arnsten, 2009).

The IED is made up of nine consecutive stages, with each stage consisting of at least six trials. On each trial, two visual stimuli are presented simultaneously on the computer screen and subjects are asked to touch (i.e. select) the one that subsequently leads to positive visual and auditory feedback. In the beginning of the task, the participants can only figure out which of the two stimuli is the relevant one by trial and error. However, once they learned which stimulus to select, the subjects are instructed to keep on touching the correct stimulus until they realise that the rule has changed, in which case they need to start choosing the newly relevant stimulus. During the task, the two visual stimuli randomly vary between four possible locations on the computer screen. If subjects fail to learn a rule at any stage of the task within 50 trials, the whole task terminates.

During the first stage of the IED, subjects are presented two visual stimuli of the same dimension (pink shapes), with one of them being reinforced by positive computer feedback. This block requires simple discrimination learning. After successfully completing this stage, the same two stimuli are presented during the second block of the task. However, the rule is now reversed so that the previously irrelevant stimulus becomes relevant. Therefore, stage two is a good indicator of reversal learning, "the ability to switch responding to a previously non-reinforced stimulus" (Kehagia, et al., 2010). During the third and fourth stage of the task, the rule remains unchanged. However, another stimulus dimension is introduced: in stage three, white stripes appear adjacent to the pink colour-filled shapes; and in stage four, the same white stripes are randomly superimposed on the relevant dimension (shapes). After successfully completing these stages, the fifth stage begins which contains identical stimuli as the fourth block, however the rule is reversed so that the previously irrelevant stimulus becomes relevant. Thus, this stage is another indicator of reversal learning. In the sixth stage of the task, novel

exemplars of stimuli for both dimensions are introduced and subjects are required to shift attention to the novel exemplar of the previously relevant perceptual dimension of shapes. This block of the IED is a good measure of intradimensional set-shifting, the ability to switch attention from exemplars of one stimulus dimension to novel exemplar of the same perceptual dimension. During stage seven, the rule is reversed so that the previously not reinforced exemplar of the same dimension becomes relevant. Thus, this stage requires reversal learning. The eighth stage introduces novel exemplars of stimuli for both dimensions and subjects are required to shift attention to the novel exemplar of the previously unrewarded perceptual dimension of lines. Thus, this stage assesses extradimensional set-shifting, the ability to shift attention between different perceptual dimensions. Finally, the rule is reversed in stage nine so that the previously irrelevant stimulus of the same perceptual dimension is now reinforced. Accordingly, this block is another indicator of reversal learning.

In summary, performance during stage one is a good indicator of simple discrimination learning; stages two, five, seven, and nine are a good measure of reversal learning; stage six is an indicator of intradimensional set-shifting; and stage eight is a measure of extradimensional set-shifting (see Figure 2.7).

Early research in marmosets has shown that lesions of the lateral prefrontal cortex (BA 9) impair extradimensional, but not intradimensional set-shifting (Dias, Robbins, & Roberts, 1996). Of interest, the same study reported that orbitofrontal lesions selectively impaired reversal learning, the ability to switch responding to a previously non-reinforced stimulus (Kehagia, et al., 2010). Further studies in humans have linked extradimensional set-shifting to the function of anterior and/or dorsolateral prefrontal regions (Nagahama et al., 2001; Rogers, Andrews, Grasby, Brooks, & Robbins, 2000).

Hampshire and Owen (2006) applied an event related functional MRI (fMRI) paradigm which led to the hypothesis that lateral prefrontal, orbital, and parietal brain areas may form a supervisory network that controls the focus of attention

during task performance but that there is also a potential functional specialisation of each brain region. In detail, extradimensional set-shifting has been associated with increased activation of the ventrolateral prefrontal cortex, while intradimensional set-shifting appears to be particularly influenced by the function of the dorsolateral prefrontal and posterior parietal cortex. By contrast, reversal-learning likely relies especially on the function of the lateral orbitofrontal cortex and changes in stimulus-response mapping appear to be mediated by the posterior parietal cortex. Finally, the authors showed that the dorsolateral prefrontal cortex appears to be generally involved in the search for a solution, but no other specific components of set-shifting.

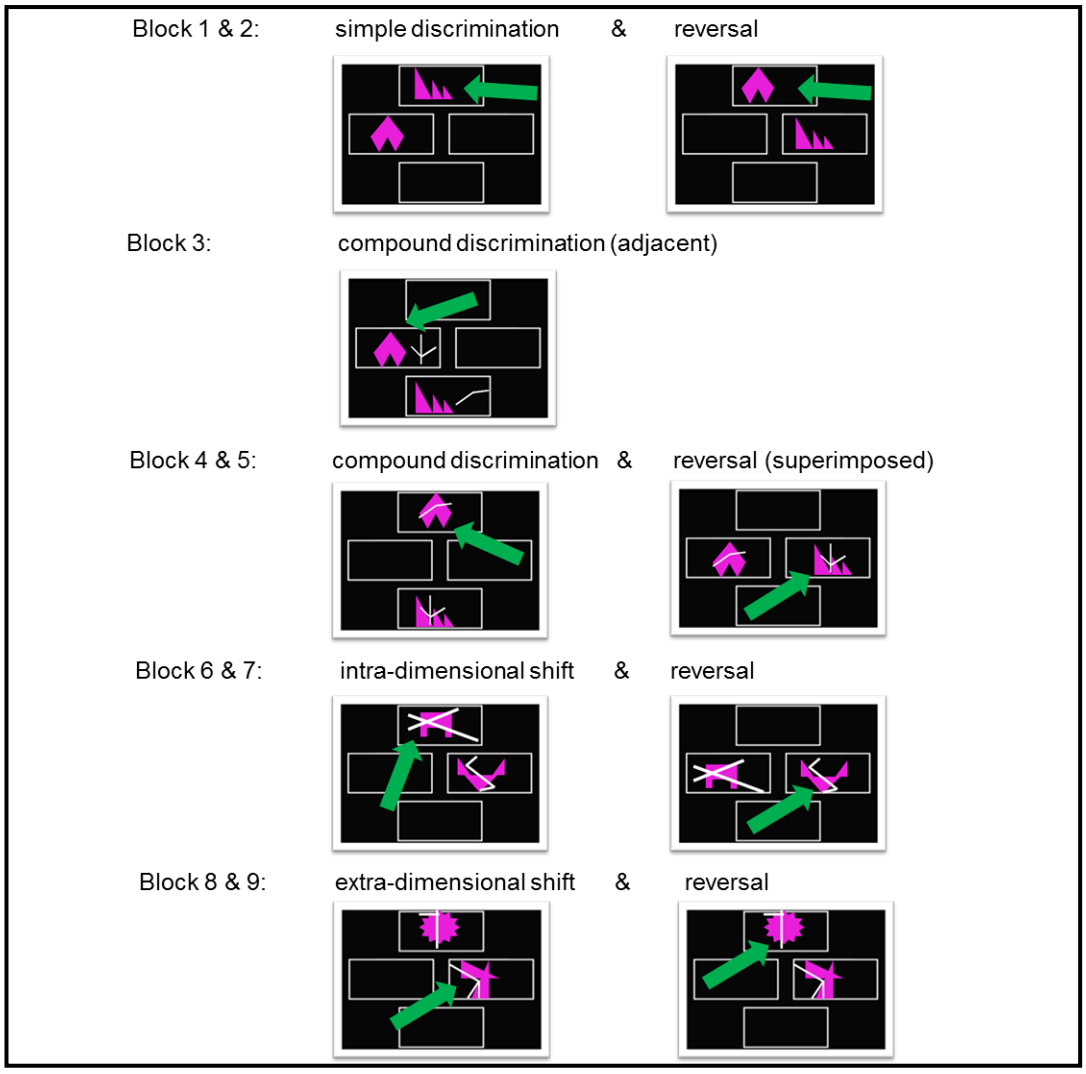

Figure 2.7 The IED task. Blocks 1 and 2 require discrimination and reversal learning, respectively. In blocks 3 and 4, a second dimension is introduced. In stage 5, the rule is reversed. Novel exemplars of stimuli are introduced in block 6 and require an intradimensional shift, followed by a reversal of the rule in block 7. In block 8, novel exemplars of stimuli are presented and an extradimensional shift is required, followed by a reversal of the rule in block 9. Green arrows indicate the correct stimulus during each stage of the task.

2.3.2 The California Verbal Learning Test

The CVLT assesses various measures of verbal learning and memory. It should be noted that a good performance on the CVLT presupposes organisational strategies as learning and memorising the word lists is easier when categorising them into different dimensions. Poor task performance therefore does not necessarily reflect verbal learning and memory deficits per se but may rather be caused by poor strategy formation.

The CVLT consists of two different lists of words (list A and list B), each consisting of 16 words. Every word can be subgrouped into one of four distinct categories, with each category being composed of four words. List A is composed of four fruits, four herbs and spices, four articles of clothing and four tools. List B also contains four names of fruits and four names of herbs and spices (different ones than included in List A) but also four types of fish and four types of kitchen utensils.

At first, the examiner reads the words of list A in a pseudo random order to the participant with the instruction to recall the words in any given order which enables one to assess the subject's spontaneous use of semantic associations. The list is presented for five consecutive times and the words remembered by the subject are recorded (free recall of list A). Subsequently, words of the interference list B are read to the participant for one single time and the subject again is asked to recall the words in any given order (free recall of list B). Next, the subject is asked to recall once more the words from list A (short delay recall), immediately followed by a cued recall. During the short delay cued recall, the participant is asked to recall the items from the different four semantic categories of list A. Subjects who have not used semantic clustering during learning trials typically benefit from cueing while the performance of participants who have already used this strategy usually does not improve. The task is then interrupted for 20 minutes in which the participant usually completes other neuropsychological tasks. After the 20 minutes

delay, the subject is again asked to recall all words from list A (long delay free recall), followed by a cued long delay recall of list A. Finally, the participant is presented with words that were either included in list A, items that were not included in list A but are from the same categories, items that phonetically resemble list A items and every-day items that were not part of list A (recognition).

Several measures of verbal learning and memory performance can be obtained. The number of items from list A that were correctly recalled during all five trials (recall of list A) provides a good indicator of verbal learning. The number of items remembered during the free recall of list B provides another measure of verbal learning performance. It assesses the effects of proactive interference on learning, i.e. the effect of prior learning (free recall trials of list A) on the retention of subsequently learned items (free recall of list B). The number of recalled words during the short delay free and cued recall provide a good estimate of short-term verbal memory. The number of recalled words during the long delay free and cued recall is a good indicator of long-term verbal memory. The recognition trial enables the assessment of discriminality and response biases. Moreover, perseverations (repetitions of words that have already been named), false positives (i.e. words from list A that were recalled during short or long delay recall trials or words from list B that were incorrectly named during other trials) and intrusions (words that were recalled during free or cued recall trials that were not part of list A or B) can be assessed. Intrusions can be divided into semantic intrusions (words that are semantically related to the target word) and non-semantic intrusions (words that are not semantically related to any target word).

fMRI studies indicate that complex functions such as verbal memory are subserved by a wide range of brain regions, with each of them being more or less specifically engaged during the distinct verbal memory components encoding, consolidation and retrieval. Encoding of verbal information has been linked to increased brain activation in the left anterior and posterior medial temporal lobe, prefrontal cortex, supplementary motor area and the inferior parietal cortex (Jansen et al., 2009). The

short-term storage and retrieval of verbal information has been linked to a bilateral frontal and parietal network of brain regions, including the posterior inferior frontal, anterior middle frontal, anterior cingulate and supramarginal gyrus (Henson, Burgess, & Frith, 2000). Long-term memory storage and retrieval of verbal information has been linked to a similar network of brain regions as short-term storage (Andreasen et al., 1995; S. Dupont, Samson, Le Bihan, & Baulac, 2002). Interestingly, a positive relationship between brain activation in the right hippocampus and right frontal lobe and CVLT performance has been established, indicating that better verbal memory processing results is higher engagement of these brain areas, potentially reflecting particularly efficient encoding and retrieval strategies (Johnson, Saykin, Flashman, McAllister, & Sparling, 2001). The recognition of learned words has been specifically associated with the function of the right dorsolateral prefrontal cortex (Johnson, et al., 2001).

2.3.3 The Digit Span

The Digit Span is a subtest of the Wechsler Adult Intelligence Scale (Wechsler, 1997) and assesses attentional processes and short-term verbal (working) memory. During the forward condition of the Digit Span, participants are asked to repeat a string of numbers. At first, a string of three numbers is read to the subject. During the task, the length of strings presented progressively increases. If a participant successfully passes the first two trials (e.g., 3 digits), the next two trials are read to the subject which contain one more digit than the previous trials (e.g., 4 digits). The task is terminated when a participant fails to repeat the string of digits correctly for a given length of digits for two consecutive times. The numbers should be read by the examiner at a rate one per second in a monotonous way to avoid any clustering that aids repetition performance. During the backward condition of the Digit Span, exactly the same rules apply. However, the participant is asked to repeat the strings of numbers in reverse order.

During serial digit learning and recall (forwards), brain activation of the hippocampus, superior frontal gyrus and cingulate have been observed for both conditions (Karakas & Karakas, 2006). It has been suggested this brain activation may reflect various executive function components that are required for a successful task performance. Others have reported that a broad bilateral frontoparietal network is implicated during task performance, encompassing mainly the dorsal prefrontal cortex, premotor cortex, anterior portion of the supplementary motor area and inferior parietal lobes (Jantzen, Anderson, Steinberg, & Kelso, 2004). Research on the brain regions involved in the backward version of the Digit Span which loads more heavily on working memory, has shown a positive relationship between task performance and grey matter volumes of the right anterior and posterior superior temporal gyrus, left inferior frontal gyrus and the left Rolandic operculum (Li, Qin, Zhang, Jiang, & Yu, 2012). Lesion and fMRI studies have suggested an involvement of the left inferior parietal cortex, left inferior frontal gyrus, dorsolateral prefrontal cortex, anterior cingulate, premotor and supplementary motor area in working memory tasks (Li, et al., 2012).

2.4 Statistical analyses

All statistical analyses were conducted in SPSS version 19 (http://www.spss.com), except for False-Discovery Rate (FDR) corrections which were computed in R version 2.13.0 (http://www.r-project.org/). Statistical analyses of demographic and clinical data were conducted using one way analyses of variance (ANOVA), chi-squares tests or Kruskal-Wallis tests where appropriate.

Both cross-sectional and longitudinal analyses in Chapter 3, 4, 5 and 6 were conducted to evaluate structural brain differences or neuropsychological performance differences between the HC, HR-well and HR-MDD groups at baseline and over time. First, a cross-sectional analysis of cortical and subcortical brain measures for each ROI and neuropsychological measures that were acquired

at baseline assessment was performed. Analyses of covariance (ANCOVA) were conducted for each ROI to compare the cortical and subcortical brain measures and neuropsychological data between the three groups at baseline, covarying for age and sex. For the subcortical volumes, intracranial volume was also included as a covariate in the analysis.

Due to the longitudinal nature and the fact that the data consists of nonuniform numbers of repeated measurements, linear mixed-effects models were applied to investigate each ROI and neuropsychological performance over time. The linear mixed effects model approach has several advantages over the often used repeated-measures ANOVA design. It takes into account that the baseline and follow-up data within each subject are correlated, thereby circumventing the need to make adjustments for heteroscedascity and sphericity assumption violations. Moreover, it permits the inclusion of multiple measurements per person as well as incomplete data sets so that casewise deletion of missing data is not needed, leading to increased statistical power (Diggle, Heagerty, Liang, & Zeger, 2013). In the linear mixed-effects model used, the intercept term is treated as a random effect that varies by individual so that intraindividual correlation among the structural brain measures of a particular individual is taken into account. The following independent variables were used as predictors of volume for the subcortical and cortical ROI and the neuropsychological measures: group, time (baseline versus follow-up assessment), group-by-time interaction. Age and sex served as covariates in all longitudinal analyses, with intracranial volume being used as an additional covariate in the longitudinal analyses of subcortical brain volumes.

A statistical significance level of $p \leq 0.05$ was chosen for both cross-sectional and longitudinal analyses, fully corrected for multiple comparisons using the Benjamini & Hochberg FDR procedure (Benjamini & Hochberg, 1995). To allow for comparison with previous studies, effect sizes for nominal significant group differences were additionally calculated using Cohen's *d* (Cohen, 1988). Wherever significant between-group differences were found, pairwise comparisons were

performed between the three groups, for which p-values were corrected according to Tukey's "Honest significance difference" method ($p_{HSD} \leq 0.05$). For significant interaction effects in the longitudinal analysis, subsequent pairwise comparisons were performed with p-values being adjusted according to Bonferroni procedure ($p_{Bonf} \leq 0.05$).

To assess the relationship between depression symptoms and each neuroanatomical and neurocognitive measures at baseline, Spearman's rank correlation coefficient between the HAM-D sum scores and each subcortical and cortical ROI and neuropsychological parameter for each group was calculated. All p-values were corrected according to Benjamini & Hochberg FDR procedure and considered significant when $p \leq 0.05$. To examine whether changes in structural brain measures and neuropsychological performance over time are related to changes in depression symptom severity, Spearman's rank correlation coefficients were calculated between the change in each neuroanatomical and neurocognitive measure and the change of the HAM-D sum score for each group.

To examine the potentially confounding effects of exposure to medication and relatedness of subjects on regional brain volumes, the following additional analyses were performed: The analyses were repeated excluding medicated HR-MDD subjects (n=4), followed by randomly excluding related subjects (n=2 HC; n=17 HR-well; n=2 HR-MDD).

Chapter 3

Subcortical volumes in individuals at high familial risk of mood disorders

3.1 Introduction

As outlined in chapter 1, mood disorders such as BD and MDD often aggregate within families and there is evidence pointing towards an overlap in the pathogenesis of the two conditions (Craddock & Forty, 2006). Close relatives of BD patients are at enhanced risk for developing mood disorders during the course of their lives themselves. First-degree relatives of BD patients have a 10-fold excess risk of BD as compared to the general population and a 3-fold increased risk of MDD (Smoller & Finn, 2003). Given that MDD is more prevalent in the population than BD, the overall risk of MDD in first-degree BD relatives is with 15% about twice as high as the risk of BD (Smoller & Gardner-Schuster, 2007).

A large body of structural MRI studies in mood disorders has been published to identify structural brain abnormalities that may help to elucidate the pathophysiology of the condition. Most consistently, volumetric abnormalities of the thalamus, limbic and basal ganglia subcortical regions in affected patients have been reported (Arnone, et al., 2009; Beyer & Krishnan, 2002; Bora, et al., 2012; Hallahan, et al., 2011; Kempton, et al., 2008; Kempton, et al., 2011; Konarski, et al., 2008; Koolschijn, et al., 2009; Savitz & Drevets, 2009).

In brief, reduced basal ganglia volumes of the caudate, putamen and pallidum have been repeatedly documented in MDD (Bora, et al., 2012; Kempton, et al., 2011; Koolschijn, et al., 2009), with inconsistent findings being observed for BD (Beyer & Krishnan, 2002; Savitz & Drevets, 2009). In MDD but not BD patients, the thalamus and hippocampus have been found to be reduced (Arnone, et al., 2009; Kempton, et al., 2008; Kempton, et al., 2011; Koolschijn, et al., 2009). Moreover,

there is some evidence for volume reductions of the amygdala in both conditions (Bora, et al., 2012; Hallahan, et al., 2011). Finally, a significant enlargement of the lateral ventricles has been repeatedly documented in both MDD and BD patients (Arnone, et al., 2009; Kempton, et al., 2008; Kempton, et al., 2011; McDonald, Zanelli, et al., 2004), potentially reflecting medial temporal lobe, lateral prefrontal cortex or basal ganglia volume reductions (Savitz & Drevets, 2009). A more detailed description of subcortical brain findings in mood disorders is provided in Chapter 1.

Despite strong evidence for subcortical brain abnormalities in mood disorders, their aetiology remains largely unknown as most imaging studies have assessed brain structure in already affected patients only. These studies cannot discern whether structural brain abnormalities represent neuropathological events linked to the illness onset, compensatory effects that help to attenuate the severity of the condition, nonspecific effects of chronic illness, or effects of medication. They also do not provide information to what extent structural brain abnormalities may represent neuroanatomical risk markers for mood disorders that are already present before illness onset. Only a few studies have examined whether subcortical grey matter abnormalities are also evident in unaffected close relatives so that they may serve as neuroanatomic endophenotypes for the condition. Imaging studies of unaffected relatives of BD patients have yielded inconsistent results, with a recent meta-analysis indicating no significant subcortical abnormalities (Fusar-Poli, et al., 2012).

The prospective longitudinal Bipolar Family Study is well suited to investigate the nature of structural brain changes in mood disorders. The volumes of subcortical ROIs were extracted from the baseline and follow-up MRI scans of Bipolar Family Study participants and compared cross-sectionally for baseline data and longitudinally between the HR-MDD, HR-well and HC groups. The rationale for the baseline analysis was to examine whether regional subcortical volume abnormalities exist in the HR-MDD group prior to illness onset that may serve as

neuroanatomical markers for a subsequent development of MDD and to assess their relationship to familial risk. The longitudinal analysis was intended to explore the time course of subcortical brain volumes during a period of two years to detect whether abnormal subcortical brain development in the high-risk participants is evidence that can be linked to an onset of MDD or familial risk. Similar to a repeated measures ANOVA, the longitudinal analysis also allowed for testing with an increased power as compared to baseline analysis whether significant volumetric differences between the study groups exist across both time points.

3.2 Methods

3.2.1 Participants and clinical assessment

Participants were recruited as part of the Scottish Bipolar Family Study as outlined in more detailed in Chapter 2. At baseline assessment, 93 HC, 92 HR-well and 19 HR-MDD participants provided suitable data. At follow-up assessment, full data-sets were available for 62 HC, 63 HR-well and 20 HR-MDD subjects. All groups were matched with respect to gender, age and verbal intelligence. The mean age at baseline assessment for all groups was 21 years. At follow-up assessment, the mean age was 23 years for the HC and HR-MDD groups and 24 years for the HR-well group. The mean interscan interval (i.e. the time period between the two assessments) for all groups was two years (HC: 2.13 years; HR-well: 2.15 years; HR-well: 2.10 years). There were no significant differences between the groups at baseline or follow-up assessment regarding the YMRS sum score. However, there were significant group differences both at baseline and follow-up for depressive symptoms assessed with the HAM-D. At baseline, the HR-well and the HR-MDD group had significantly higher depression scores ($p \leq 0.047$ and $p \leq 0.003$, respectively) than the HC subjects, with no significant differences between the high-risk groups. At follow-up assessment, the HR-MDD had higher depression

scores than the HC and the HR-well group ($p \leq 0.013$ and $p \leq 0.010$, respectively) as expected, with no significant differences between the HC and HR-well individuals. All demographic and clinical details are provided in Table 3.1.

3.2.2 Magnetic resonance imaging and processing

Structural MRI scans were acquired on a 1.5 Tesla scanner as outlined in Chapter 2. The T_1 weighted images were processed using the volume-based stream of FreeSurfer (fully described in Chapter 2). Based on the literature review provided in Chapter 1 and hence the scientific relevance for mood disorders, the volumes of the following subcortical brain structures were assessed for each hemisphere separately: amygdala, caudate, hippocampus, lateral ventricles, thalamus, pallidum and putamen.

Table 3.1 Demographic and clinical characteristics

	Baseline					Follow-up				
	HC (n=93)	**HR-well** (n=92)	**HR-MDD** (n=19)	**Statistics**		**HC** (n=62)	**HR-well** (n=63)	**HR-MDD** (n=20)	**Statistics**	
	Mean (SD)	Mean (SD)	Mean (SD)	F/ χ2	p	Mean (SD)	Mean (SD)	Mean (SD)	F/ χ2	p
Age (years)	21.01 (2.45)	21.20 (2.88)	21.10 (2.82)	0.13	0.88	22.82 (2.73)	23.71 (2.84)	23.33 (2.98)	1.77	0.17
Gender (M:F)	40:53	44:48	9:10	0.46	0.80	21:41	29:34	7:13	2.12	0.35
Handedness (Right:other)	88:5	81:11	19:0	5.54	0.24	61:1	57:6	20:0	5.43	0.07
NART IQ	110.31 (8.00)	108.39 (9.37)	107.26 (6.80)	1.64	0.20	-	-	-	-	-
ISI (years)	2.13 (0.22)	2.15 (0.22)	2.10 (0.13)	0.20	0.82	-	-	-	-	-
HAM-D*	0 (1)	0 (2)	1 (5)	9.79	**0.01**	1 (3)	1 (2)	5 (12)	7.59	**0.02**
YMRS*	0 (0)	0 (0)	0 (0)	3.48	0.18	0 (0)	0 (1)	0 (0)	0.79	0.68

Abbreviations: NART = National Adult Reading Test; ISI = interscan interval; HAM-D = Hamilton Depression Rating Scale; YMRS = Young Mania Rating Scale.

* Kruskal-Wallis test, median and interquartile presented for skewed variables.

3.2.3 Statistical analysis

To compare cross-sectionally regional subcortical volume differences between the HC, HR-well and HR-MDD groups at baseline, ANCOVA's were conducted. Age, sex and intracranial volume served as covariates in this analysis. Next, linear mixed-effects models were applied to investigate subcortical brain volumes for each region of interest over time (see Chapter 2 for a more detailed methodological description). In the linear mixed-effects model used, the intercept term is treated as a random effect that varies by individual so that intraindividual correlation among the structural brain measures of a particular individual is taken into account. The following independent variables were used as predictors of volume for the subcortical regions outlined: group, time (baseline versus follow-up assessment), group-by-time interaction. Age, sex and intracranial volume served as covariates in the analyses.

A statistical significance level of $p \leq 0.05$ was chosen, fully corrected for multiple comparisons using the Benjamini & Hochberg FDR procedure (Benjamini & Hochberg, 1995). To allow for comparison with previous studies, effect sizes for nominal significant group differences were additionally calculated using Cohen's *d* (Cohen, 1988). Wherever significant between-group differences were found, pairwise comparisons were performed between the three groups, for which p-values were corrected according to Tukey's HSD procedure ($p_{HSD} \leq 0.05$).

To assess the relationship between depression symptoms and the volumetric measures of each ROI, Spearman's rank correlation coefficients between the HAM-D scores and the subcortical regions for each group were calculated. To assess the relationship between volumetric changes and changes in depressive symptoms over time, Spearman's rank correlation coefficients were calculated between the volumetric differences of the ROIs and the differences of the HAM-D scores between the two assessments. For ease of interpretation, data derived from the follow-up assessment was subtracted from the data acquired at baseline so that

positive values reflect increases in volume or depression symptoms over time, while negative values represent decreases of these measures over time. All p-values were corrected according to Benjamini & Hochberg FDR procedure and considered significant when $p \leq 0.05$.

To examine the potentially confounding effects of exposure to medication and relatedness of subjects on regional brain volumes, the following additional analyses were performed: The analyses were repeated excluding medicated HR-MDD subjects, followed by randomly excluding related subjects.

3.3 Results

3.3.1 Cross-sectional analysis

3.3.1.1 Group differences at baseline

At baseline assessment, no significant FDR-corrected differences in regional brain volume were found between the groups (see Table 3.2). There was a nominal significant group difference for the volume of the right caudate ($p \leq 0.014$). Post-hoc tests indicated that the HR-MDD group had smaller volumes of this brain structure than the HC ($p_{HSD} \leq 0.033$, $d = 0.456$) and the HR-well group ($p_{HSD} \leq 0.011$, $d = 0.553$), with no differences between the HC and HR-well group ($p_{HSD} \leq 0.999$, $d = 0.109$).

3.3.1.2 Correlation analysis

There were no significant correlations between the severity of depressive symptoms as measured with the HAM-D and subcortical volumes that remained significant after FDR correction. However, one nominal significant result was detected that just failed the FDR procedure. For the HR-MDD group, the volume of

the left lateral ventricle was positively correlated with depressive symptom severity ($p \leq 0.005$), indicating that higher depression symptoms are associated with larger left lateral ventricles.

3.3.1.3 Analysis of potential confounders

To eliminate the potential confounding effects of familial relatedness of some subjects, the analyses were repeated, excluding randomly individuals from the same pedigree (Table 3.4). This analysis similarly yielded no significant FDR-corrected group differences. The nominal significant finding for the right caudate in the whole study sample remained nominally significant ($p \leq 0.023$).

Table 3.2 Cross-sectional analysis of regional subcortical volumes

	HC	HR-well	HR-MDD	Statistics	
Region	Mean (SD)	Mean (SD)	Mean (SD)	F	p
L lat ventricle	6.69 (2.05)	7.17 (2.50)	5.71 (2.45)	1.70	0.19
R lat ventricle	6.04 (2.44)	6.48 (2.20)	5.40 (2.83)	1.25	0.29
L caudate	3.68 (0.43)	3.71 (0.44)	3.53 (0.46)	2.41	0.09
R caudate	3.83 (0.45)	3.88 (0.47)	3.62 (0.47)	4.34	0.01
L putamen	5.91 (0.78)	6.03 (0.72)	6.09 (0.77)	0.43	0.65
R putamen	5.71 (0.67)	5.72 (0.61)	5.82 (0.65)	0.11	0.89
L pallidum	1.94 (0.29)	1.87 (0.29)	1.94 (0.29)	2.26	0.11
R pallidum	1.86 (0.28)	1.83 (0.29)	1.94 (0.29)	1.19	0.31
L thalamus	6.58 (0.75)	6.56 (0.73)	6.66 (0.75)	0.58	0.56
R thalamus	6.62 (0.67)	6.56 (0.73)	6.75 (0.76)	1.43	0.24
L hippocampus	3.50 (0.45)	3.49 (0.46)	3.61 (0.54)	0.44	0.64
R hippocampus	3.58 (0.46)	3.56 (0.47)	3.65 (0.59)	0.31	0.74
L amygdala	1.77 (0.32)	1.79 (0.35)	1.79 (0.34)	0.04	0.96
R amygdala	1.96 (0.30)	1.92 (0.33)	2.06 (0.33)	2.65	0.07

Volumes are measured in cm^3. Abbreviations: L = left; R = right; lat = lateral.

Table 3.3 Correlation between regional subcortical volumes and HAM-D sum scores

	HC		HR-well		HR-MDD	
Region	R	p	R	p	R	p
L lat ventricle	0.02	0.88	0.11	0.29	0.61	0.01
R lat ventricle	-0.01	0.93	0.09	0.39	0.20	0.42
L caudate	-0.04	0.71	-0.10	0.33	0.14	0.57
R caudate	0.01	0.96	-0.08	0.48	0.18	0.45
L putamen	0.11	0.29	-0.02	0.82	0.22	0.37
R putamen	-0.01	0.92	-0.18	0.08	0.16	0.52
L pallidum	-0.01	0.90	0.02	0.86	0.05	0.85
R pallidum	--0.02	0.87	-0.11	0.30	-0.06	0.81
L thalamus	-0.10	0.34	-0.01	0.94	0.16	0.50
R thalamus	-0.03	0.79	-0.13	0.22	0.15	0.55
L hippocampus	0.03	0.79	-0.07	0.52	0.35	0.14
R hippocampus	-0.05	0.64	0.06	0.58	0.19	0.45
L amygdala	-0.06	0.55	0.05	0.61	-0.37	0.12
R amygdala	0.12	0.27	-0.06	0.56	0.05	0.85

Abbreviations: L = left; R = right; lat = lateral.

Table 3.4 Cross-sectional analysis of regional subcortical volumes in unrelated subjects

	HC	HR-well	HR-MDD	Statistics	
Region	Mean (SD)	Mean (SD)	Mean (SD)	F	p
L lat ventricle	6.68 (2.09)	7.07 (2.78)	5.69 (2.79)	1.40	0.25
R lat ventricle	6.08 (2.46)	6.52 (2.43)	5.40 (1.89)	1.10	0.34
L caudate	3.68 (0.43)	3.69 (0.47)	3.54 (0.47)	2.15	0.12
R caudate	3.83 (0.45)	3.86 (0.49)	3.62 (0.48)	3.84	0.02
L putamen	5.92 (0.79)	6.04 (0.74)	6.07 (0.79)	0.19	0.83
R putamen	5.71 (0.67)	5.73 (0.64)	5.80 (0.66)	0.15	0.86
L pallidum	1.94 (0.29)	1.87 (0.29)	1.95 (0.29)	2.82	0.06
R pallidum	1.86 (0.28)	1.82 (0.30)	1.95 (0.29)	1.99	0.14
L thalamus	6.58 (0.76)	6.58 (0.73)	6.69 (0.76)	0.76	0.47
R thalamus	6.62 (0.68)	6.56 (0.71)	6.77 (0.78)	1.97	0.14
L hippocampus	3.50 (0.45)	3.48 (0.46)	3.59 (0.55)	0.64	0.53
R hippocampus	3.59 (0.46)	3.56 (0.48)	3.65 (0.61)	0.65	0.52
L amygdala	1.76 (0.32)	1.79 (0.34)	1.82 (0.31)	0.09	0.91
R amygdala	1.96 (0.30)	1.91 (0.33)	2.04 (0.33)	2.65	0.07

Volumes are measured in cm^3. Abbreviations: L = left; R = right; lat = lateral.

3.3.2 Longitudinal analysis

3.3.2.1 Group differences over time

The linear mixed-effects model analyses revealed no significant effects of group, time or group-by-time interaction (see Table 3.5). A trend towards a significant group effect was observed for the right caudate ($p \leq 0.024$), with the HR-MDD group having smaller volumes than the HC ($p_{HSD} \leq 0.024$, $d = 0.639$) and the HR-well ($p_{HSD} \leq 0.016$, $d = 0.687$) group across time. This effect was only significant at nominal level and did not survive FDR correction. No nominal significant effect of time or group-by-time interaction for the right caudate was detected. There was a nominally significant group-by-time interaction for the volume of the left amygdala ($p \leq 0.036$), with no (nominal) significant effect of group or time. Post-hoc analyses revealed that the volume of this structure decreased over time in the HR-well as compared to the HC group that in turn displayed volumetric increases ($p_{HSD} \leq 0.024$, $d = 0.376$). No significant interaction effects for the HR-MDD as compared to the HC group ($p_{HSD} \leq 0.504$, $d = 0.291$) and the HR-well as compared to HR-MDD participants ($p_{HSD} \leq 0.158$, $d = 0.134$) were found.

3.3.2.2 Correlation analysis

There were no significant FDR-corrected correlations between changes in the severity of depressive symptoms as measured with the HAM-D and changes in subcortical volumetric measures over time (see Table 3.6). However, several correlations that were significant at a nominal significance level were observed for the HR-well group. Volumetric changes of the left thalamus were positively correlated with changes in depressive symptoms ($p \leq 0.009$), indicating that volume increases of this subcortical structure over time are associated with increases in depressive symptoms. Moreover, changes of the volume of the right hippocampus ($p \leq 0.030$) and right putamen ($p \leq 0.036$) were positively correlated

with depressive symptom changes, suggesting that volumetric increases in these brain regions over time are associated with increasing depressive symptoms. No nominal significant correlations were observed for the HC and HR-MDD group.

3.3.2.3 Analysis of potential confounders

To eliminate the potential confounding effects of familial relatedness of some subjects, the longitudinal analyses were repeated, excluding randomly individuals from the same pedigree (Tables 3.7). This analysis similarly yielded no significant FDR-corrected effect of group, time or group-by-time interaction. Also, when excluding medicated HR-MDD subjects from the analysis, no significant effects were observed (Table 3.8).

Table 3.5 Longitudinal analysis of regional subcortical volumes

	HC		HR-well		HR-MDD		Statistics					
	Baseline	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Group effect**		**Time effect**		**GroupXTime**	
Region	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L lat ventricle	6.69 (2.05)	6.98 (2.74)	7.17 (2.50)	7.28 (2.69)	5.71 (2.45)	5.57 (2.65)	1.25	0.29	1.80	0.18	0.62	0.54
R lat ventricle	6.04 (2.44)	6.27 (2.06)	6.48 (2.20)	6.56 (2.38)	5.40 (2.83)	5.46 (2.01)	0.79	0.46	0.88	0.35	0.13	0.88
L caudate	3.68 (0.43)	3.70 (0.53)	3.71 (0.44)	3.60 (0.45)	3.53 (0.46)	3.44 (0.51)	1.74	0.18	2.03	0.16	1.27	0.28
R caudate	3.83 (0.45)	3.89 (0.59)	3.88 (0.47)	3.83 (0.49)	3.62 (0.47)	3.51 (0.50)	3.81	0.02	0.48	0.49	0.92	0.40
L putamen	5.91 (0.78)	6.02 (0.83)	6.03 (0.72)	6.01 (0.82)	6.09 (0.77)	5.77 (0.83)	0.11	0.90	1.22	0.27	2.42	0.09
R putamen	5.71 (0.67)	5.81 (0.71)	5.72 (0.61)	5.69 (0.72)	5.82 (0.65)	5.40 (0.79)	0.45	0.64	2.34	0.13	2.52	0.08
L pallidum	1.94 (0.29)	1.93 (0.29)	1.87 (0.29)	1.88 (0.33)	1.94 (0.29)	1.90 (0.27)	1.27	0.28	0.19	0.66	0.26	0.78
R pallidum	1.86 (0.28)	1.85 (0.32)	1.83 (0.29)	1.79 (0.33)	1.94 (0.29)	1.83 (0.25)	1.04	0.36	2.93	0.09	0.58	0.56
L thalamus	6.58 (0.75)	6.60 (0.66)	6.56 (0.73)	6.38 (0.76)	6.66 (0.75)	6.39 (0.74)	0.88	0.42	2.98	0.09	1.42	0.24
R thalamus	6.62 (0.67)	6.68 (0.70)	6.56 (0.73)	6.48 (0.79)	6.75 (0.76)	6.55 (0.79)	1.10	0.33	0.85	0.36	1.04	0.36
L hippocampus	3.50 (0.45)	3.47 (0.49)	3.49 (0.46)	3.41 (0.40)	3.61 (0.54)	3.44 (0.53)	0.37	0.69	2.64	0.11	0.28	0.76
R hippocampus	3.58 (0.46)	3.65 (0.52)	3.56 (0.47)	3.47 (0.38)	3.65 (0.59)	3.38 (0.54)	1.58	0.21	3.33	0.07	2.54	0.08
L amygdala	1.77 (0.32)	1.83 (0.33)	1.79 (0.35)	1.67 (0.27)	1.79 (0.34)	1.72 (0.23)	1.55	0.22	1.14	0.29	3.40	0.04
R amygdala	1.96 (0.30)	1.97 (0.30)	1.92 (0.33)	1.92 (0.27)	2.06 (0.33)	1.91 (0.26)	1.31	0.27	1.20	0.28	1.12	0.33

Volumes are measured in cm^3. Abbreviations: L = left; R = right; lat = lateral.

Table 3.6 Correlation between changes in regional subcortical volumes and changes in HAM-D sum scores

	HC		HR-well		HR-MDD	
Region	R	p	R	p	R	p
L lat ventricle	-0.06	0.71	0.10	0.51	0.22	0.44
R lat ventricle	-0.19	0.43	0.13	0.37	0.20	0.48
L caudate	-0.11	0.47	0.27	0.06	0.04	0.88
R caudate	-0.13	0.40	0.27	0.06	0.02	0.95
L putamen	-0.16	0.29	0.05	0.74	0.02	0.94
R putamen	-0.15	0.31	0.30	0.04	-0.10	0.72
L pallidum	-0.05	0.75	0.28	0.05	0.19	0.51
R pallidum	-0.20	0.18	0.13	0.36	0.31	0.26
L thalamus	-0.11	0.47	0.37	0.01	0.11	0.69
R thalamus	-0.18	0.23	0.07	0.64	0.22	0.43
L hippocampus	-0.06	0.71	0.19	0.19	0.15	0.59
R hippocampus	-0.17	0.26	0.31	0.03	0.28	0.32
L amygdala	0.02	0.91	-0.05	0.74	-0.07	0.80
R amygdala	-0.12	0.48	0.03	0.86	0.37	0.17

Abbreviations: L = left; R = right; lat = lateral.

Table 3.7 Longitudinal analysis of regional subcortical volumes in unrelated subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Group effect		Time effect		GroupXTime	
Region	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L lat ventricle	6.68 (2.09)	6.99 (2.07)	7.07 (2.78)	7.08 (2.38)	5.69 (2.79)	5.59 (2.33)	0.91	0.40	1.20	0.28	0.73	0.48
R lat ventricle	6.08 (2.46)	6.27 (2.31)	6.52 (2.43)	6.47 (2.35)	5.40 (1.89)	5.46 (2.19)	0.62	0.54	0.52	0.47	0.28	0.75
L caudate	3.68 (0.43)	3.70 (0.53)	3.69 (0.47)	3.57 (0.43)	3.54 (0.47)	3.43 (0.52)	1.53	0.22	2.75	0.10	1.65	0.20
R caudate	3.83 (0.45)	3.89 (0.59)	3.86 (0.49)	3.78 (0.48)	3.62 (0.48)	3.51 (0.51)	2.89	0.06	0.92	0.34	1.25	0.29
L putamen	5.92 (0.79)	6.02 (0.82)	6.04 (0.74)	5.91 (0.74)	6.07 (0.79)	5.75 (0.84)	0.01	0.99	2.50	0.12	3.22	0.04
R putamen	5.71 (0.67)	5.81 (0.70)	5.73 (0.64)	5.62 (0.70)	5.80 (0.66)	5.36 (0.79)	0.84	0.45	3.67	0.06	2.89	0.06
L pallidum	1.94 (0.29)	1.93 (0.29)	1.87 (0.29)	1.84 (0.34)	1.95 (0.29)	1.89 (0.27)	2.08	0.13	0.98	0.33	0.32	0.72
R pallidum	1.86 (0.28)	1.85 (0.32)	1.82 (0.30)	1.73 (0.30)	1.95 (0.29)	1.83 (0.26)	2.64	0.07	5.01	0.03	1.05	0.35
L thalamus	6.58 (0.76)	6.60 (0.66)	6.58 (0.73)	6.34 (0.70)	6.69 (0.76)	6.32 (0.69)	1.08	0.34	4.81	0.03	2.16	0.12
R thalamus	6.62 (0.68)	6.68 (0.70)	6.56 (0.71)	6.38 (0.72)	6.77 (0.78)	6.51 (0.78)	2.13	0.12	2.21	0.14	1.92	0.15
L hippocampus	3.50 (0.45)	3.46 (0.49)	3.48 (0.46)	3.38 (0.39)	3.59 (0.55)	3.41 (0.53)	0.50	0.61	2.95	0.09	0.34	0.71
R hippocampus	3.59 (0.46)	3.65 (0.52)	3.56 (0.48)	3.44 (0.37)	3.65 (0.61)	3.33 (0.52)	2.19	0.12	4.52	0.04	2.88	0.06
L amygdala	1.76 (0.32)	1.83 (0.33)	1.79 (0.34)	1.67 (0.27)	1.82 (0.31)	1.73 (0.23)	1.34	0.26	1.46	0.23	3.40	0.04
R amygdala	1.96 (0.30)	1.97 (0.30)	1.91 (0.33)	1.90 (0.27)	2.04 (0.33)	1.90 (0.27)	1.55	0.22	1.15	0.26	0.82	0.44

Volumes are measured in cm^3. Abbreviations: L = left; R = right; lat = lateral.

Table 3.8 Longitudinal analysis of regional subcortical volumes in unmedicated HR-MDD subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Group effect		Time effect		GroupXTime	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L lat ventricle	6.68 (2.09)	6.99 (2.07)	7.07 (2.78)	7.08 (2.38)	5.87 (2.60)	5.83 (2.23)	0.86	0.43	2.01	0.16	0.37	0.69
R lat ventricle	6.08 (2.46)	6.27 (2.31)	6.52 (2.43)	6.47 (2.35)	5.44 (1.75)	5.63 (2.28)	0.66	0.52	0.91	0.34	0.09	0.91
L caudate	3.68 (0.43)	3.70 (0.53)	3.69 (0.47)	3.57 (0.43)	3.59 (0.44)	3.45 (0.45)	1.36	0.26	2.21	0.14	1.35	0.26
R caudate	3.83 (0.45)	3.89 (0.59)	3.86 (0.49)	3.78 (0.48)	3.70 (0.44)	3.50 (0.42)	2.78	0.06	1.01	0.32	1.28	0.28
L putamen	5.92 (0.79)	6.02 (0.82)	6.04 (0.74)	5.91 (0.74)	6.21 (0.75)	5.73 (0.67)	0.12	0.89	1.92	0.17	2.92	0.06
R putamen	5.71 (0.67)	5.81 (0.70)	5.73 (0.64)	5.62 (0.70)	5.96 (0.60)	5.57 (0.58)	0.25	0.78	1.81	0.18	2.08	0.13
L pallidum	1.94 (0.29)	1.93 (0.29)	1.87 (0.29)	1.84 (0.34)	1.96 (0.29)	1.88 (0.28)	1.25	0.29	0.55	0.46	0.54	0.58
R pallidum	1.86 (0.28)	1.85 (0.32)	1.82 (0.30)	1.73 (0.30)	1.97 (0.30)	1.81 (0.23)	1.10	0.34	4.40	0.04	1.16	0.32
L thalamus	6.58 (0.76)	6.60 (0.66)	6.58 (0.73)	6.34 (0.70)	6.75 (0.67)	6.38 (0.67)	0.94	0.39	3.97	0.05	1.79	0.17
R thalamus	6.62 (0.68)	6.68 (0.70)	6.56 (0.71)	6.38 (0.72)	6.87 (0.67)	6.50 (0.63)	1.22	0.30	1.98	0.16	1.72	0.18
L hippocampus	3.50 (0.45)	3.46 (0.49)	3.48 (0.46)	3.38 (0.39)	3.60 (0.55)	3.45 (0.56)	0.34	0.72	1.81	0.18	0.14	0.87
R hippocampus	3.59 (0.46)	3.65 (0.52)	3.56 (0.48)	3.44 (0.37)	3.67 (0.60)	3.34 (0.56)	1.65	0.20	4.49	0.04	3.14	0.05
L amygdala	1.76 (0.32)	1.83 (0.33)	1.79 (0.34)	1.67 (0.27)	1.76 (0.37)	1.72 (0.19)	1.61	0.20	0.72	0.40	3.38	0.04
R amygdala	1.96 (0.30)	1.97 (0.30)	1.91 (0.33)	1.90 (0.27)	2.02 (0.34)	1.90 (0.25)	1.07	0.35	0.68	0.41	0.67	0.51

Volumes are measured in cm^3. Abbreviations: L = left; R = right; lat = lateral.

3.4 Discussion

The baseline and longitudinal analyses yielded no significant volumetric differences of subcortical brain structures between the groups. However, a nominal significant group effect for the right caudate was found both cross-sectionally and longitudinally as well as a nominal significant group-by-time interaction for the left amygdala. Since this is, to the best of my knowledge, the first prospective longitudinal study examining regional subcortical volumes in high-risk of mood disorders individuals who were unaffected at initial assessment and subsequently either developed MDD or remained well, the discussion focuses on cross-sectional analyses.

The baseline analysis yielded no significant volumetric differences of subcortical brain structures between the groups. This finding suggests that young adults at high risk of mood disorders because of a close family history of BD do not exhibit subcortical volumetric brain abnormalities, regardless of a subsequent onset of MDD or not. These results are in line with a recent meta-analysis that did not find evidence towards subcortical volumetric brain abnormalities in unaffected relatives of BD patients (Fusar-Poli, et al., 2012) and indicate that subcortical brain abnormalities do not emerge as a consequence of being at high familial risk for mood disorders. Moreover, they appear not to predate an onset of MDD and are thus unlikely to be neurodevelopmental in nature.

The longitudinal analysis also did not reveal any significant effect of group, time or group-by-time interaction. These non-significant findings are in contrast to the often observed subcortical brain abnormalities in mood disorder patients (Konarski, et al., 2008; Koolschijn, et al., 2009). Accordingly, volumetric subcortical brain abnormalities in MDD do not appear to emerge prior to illness onset or as a consequence of illness-specific mechanisms that are directly linked to the onset of the disorder. Rather, it appears likely that volumetric differences only emerge during the course of the disorder or in conjunction with psychopharmacological

treatment effects. Subcortical brain volumes may also be influenced by the severity or length of illness, the age at illness onset or the length and type of psychopharmacological treatment.

Although not significant when FDR-corrected, a nominal significant group effect for the volume of the right caudate was found during the baseline and longitudinal analyses, indicating smaller volumes in the HR-MDD as compared to the other two study groups across both time points. Volume reductions of the caudate in MDD patients have been repeatedly documented in volume-based and VBM-based meta-analyses (Arnone, et al., 2009; Bora, et al., 2012; Kempton, et al., 2011; Koolschijn, et al., 2009). Given that our HR-MDD group consisted of 20 participants only, it may well be that our study sample was underpowered to detect this subtle volumetric abnormality. For example, a study by Bremner et al. (2000) which consisted of 16 MDD patients did not find volumetric differences of the caudate, while Dupont and colleagues (1995) observed a trend towards significant volume reductions in a study sample of 30 MDD participants, and in a large study sample of 50 MDD patients significant caudate reductions were evident (Krishnan et al., 1992). Furthermore, our analysis might have lacked power given that we investigated young adults with an early onset MDD as it has been shown that volume reductions of the caudate are more pronounced in older individuals with a late onset of depression as compared to young MDD patients (Greenwald et al., 1997). It has also been established that the caudate is particularly decreased in severe and treatment-resistant subtypes of depression (Shah, Glabus, Goodwin, & Ebmeier, 2002). Given that the majority of the HR-MDD participants of our study were not receiving pharmacological treatment for depression and that the average HAM-D depression sum score at follow-up assessment was only 5 suggests that the symptoms of several individuals were already in remission or mild rather than severe. This difference in clinical presentation of the disorder may consequently have led to smaller caudate atrophy that is more difficult to detect.

Clearly, further longitudinal research with larger sample sizes is needed to clarify if the observed nominal significant caudate abnormalities are deemed correct. Nevertheless, the findings provide a first hint to suggest that subtle volume reductions of the right caudate exist in individuals at familial risk of mood disorders who go on to develop MDD and that these remain relatively stable across a two-year period of time. Given that there was no evidence towards an effect of time or group-by-time interaction, the volumetric decreases cannot be associated with the onset of illness but are rather likely to either represent illness-associated processes that are already evident two years before the illness onset or are neurodevelopmental in nature which may enhance vulnerability to MDD.

From a functional perspective, the caudate forms together with the putamen the striatum which has been traditionally associated with the regulation of motion (Savitz & Drevets, 2009). More recently, the striatum has been shown to play an important role in mood regulation, cognitive processes and motivation through its involvement in several parallel organised cortical-subcortical circuits (Tekin & Cummings, 2002). Accordingly, disturbed function of these processes holds the potential to mimic the symptoms of depression and mania. In concert with the just said, a tryptophan depletion study found depression severity to be linked to reduced caudate activity (K. A. Smith, Morris, Friston, Cowen, & Dolan, 1999), while manic episodes have been associated with striatal overactivity during reward-related tasks (Savitz & Drevets, 2009).

A nominal significant group-by-time interaction was observed for the volume of the left amygdala that did not survive FDR correction. The interaction was driven by the HR-well group displaying decreasing left amygdala volume over time as compared to the HC group which in turn showed a volumetric increase. Although not being significant when corrected for FDR, this result may give a first hint of evidence to suggest that young unaffected relatives of BD patients display an abnormal volumetric decrease of the left amygdala in their early 20s that may represent neurodevelopmental processes associated with enhanced vulnerability for

mood disorders. Previous cross-sectional studies have not detected volume reductions of the amygdala in close relatives of BD patients (Fusar-Poli, et al., 2012), with studies on BD patients providing inconsistent findings of volumetric increases, decreases or no changes (Konarski, et al., 2008). In particular, amygdala volume reductions are often found in young BD patients, with enlarged volumes being present in older BD patients and it has been shown that the volume is inversely influenced by age in BD patients versus control subjects (Konarski, et al., 2008). Our findings accordingly provide a first line of evidence to suggest that the same also holds true for high-risk subjects.

The amygdala is part of a network of brain areas that are charged with the identification of the emotional significance of environmental stimuli, the generation of a subsequent affective state, and the production of an autonomic response (Phillips, et al., 2003a). Disturbed function of the amygdala therefore likely influences a wide range of other brain regions and holds the potential to mimic the symptoms of mood disorders (Phillips, et al., 2003b).

The strengths of this study are its longitudinal nature, the assessment of subjects prior to illness onset, the relatively young age of the participants and the comparatively large sample size of high-risk subjects and controls. In addition, all subjects underwent careful clinical assessment at both time points and the effects of medication and relatedness of subjects were ruled out. All brain scans were obtained at the same scanner using the identical protocol at both visits and the MRI data were processed in an identical way using thoroughly validated methods.

Several limitations of this study cohort need to be addressed. First, it remains unknown whether currently unaffected HR-well subjects may develop MDD in the future. Second, previous longitudinal studies have reported that the majority of the high-risk subjects who developed BD themselves experienced depressive episodes years before conversion (Duffy, 2010; Hillegers et al., 2005) so that it appears likely that some of our HR-MDD subjects may develop BD in the future. The follow-up assessments of our study cohort will clarify if some of the HR-MDD

participants will convert to BD and if some of our HR-well subjects go on to develop a mood disorder. Third, our study groups differed with respect to depression symptom severity at baseline. However, the median of the HAM-D total score was only 1 in the HR-MDD group, suggesting only sub-syndromal depressive symptoms. Therefore, it appears unlikely that general mood differences at baseline between the groups have influenced our findings.

In summary, no significant volumetric subcortical brain abnormalities were detected in individuals at high familial risk of mood disorders, regardless of an onset of MDD or not. Accordingly, it can be concluded that subcortical volumetric brain abnormalities that are often observed in mood disorders only emerge during the course of illness and are not related to the illness onset or enhanced familial risk for the disorder. However, nominal significant findings provide a first line of evidence to suggest that the volume of the right caudate is decreased in individuals at high familial risk of mood disorders who go on to develop MDD and that this volume reduction is not related to the illness onset but rather appears to be relatively stable across time. Moreover, individuals at high familial risk of mood disorders who did not develop MDD had a nominal significant abnormal decrease of the left amygdala over time as compared to control subjects. Future longitudinal studies are needed to clarify if the nominal significant findings are deemed correct.

Chapter 4

Cortical thickness in individuals at high familial risk of mood disorders

4.1 Introduction

As outlined in chapter 1, close relatives of BD patients have an up to 10-fold excess risk of BD as compared to the general population, and an up to 3-fold increased risk of MDD (Smoller & Finn, 2003). The aggregation of mood disorders within families together with moderate to high heritability estimates and the finding of a substantially shared genetic architecture in genome-wide association analyses provide strong support for overlapping causal pathways in BD and MDD (Barnett & Smoller, 2009; Craddock, 2006; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; McGuffin et al., 2003; Schulze et al., 2012; Sullivan, et al., 2000).

Accumulating evidence from structural MRI studies indicates that neuroanatomical changes in the frontal and temporal lobes are associated with mood disorders (Arnone, et al., 2009; Beyer & Krishnan, 2002; Bora, et al., 2012; Hallahan, et al., 2011; Kempton, et al., 2011; Konarski, et al., 2008; Koolschijn, et al., 2009; Savitz & Drevets, 2009). Volumetric grey matter reductions of the prefrontal lobe have been consistently found in both MDD and BD, predominantly in the orbitofrontal gyrus (Kempton, et al., 2011; Konarski, et al., 2008; Koolschijn, et al., 2009; Stanfield, et al., 2009; Wilke, et al., 2004) and the anterior cingulate cortex (Bora, et al., 2012; Bora, et al., 2010; Caetano et al., 2006; Koolschijn, et al., 2009; Lochhead, Parsey, Oquendo, & Mann, 2004; Sassi et al., 2004), but also repeatedly in the inferior frontal (Bora, et al., 2012; Bora, et al., 2010; Du, et al., 2012; Matsuo, et al., 2012; Selvaraj, et al., 2012), middle frontal (Bora, et al., 2012; Du, et al., 2012; Lopez-Larson, et al., 2002; Lyoo, et al., 2004; Peng et al., 2011) and superior frontal gyrus (Bora, et al., 2012; Lopez-Larson, et al., 2002). Furthermore,

grey matter decreases have been observed in the precentral gyrus (Bora, et al., 2012; Lyoo, et al., 2004), superior temporal gyrus (Abe et al., 2010; Peng, et al., 2011; Selvaraj, et al., 2012) and in the medial temporal lobe, particularly in the parahippocampal gyrus (Abe, et al., 2010; Bora, et al., 2012; Peng, et al., 2011). A detailed review on frontal and temporal cortical findings for MDD and BD is provided in Chapter 1.

The majority of research has examined grey matter volume in mood disorders but not cortical thickness. Grey matter volume is a composite of cortical thickness and surface area, and research suggests that grey matter volume is more closely linked to surface area than to thickness (Winkler, et al., 2009). Accordingly, the contribution of cortical thickness towards structural brain abnormalities in mood disorders remains largely unknown. A small number of existing studies have however reported cortical thickness abnormalities in both MDD and BD patients in the frontal lobe (Elvsashagen et al., 2013; Foland-Ross et al., 2011; Grieve, Korgaonkar, Koslow, Gordon, & Williams, 2013; Hartberg et al., 2011; Lan et al., 2014; Lyoo et al., 2006; Qiu et al., 2014; Reynolds et al., 2014; Rimol et al., 2010; Truong et al., 2013; Tu et al., 2012; van Eijndhoven et al., 2013; van Tol et al., 2013) as well as in the superior temporal gyrus (Elvsashagen, et al., 2013; Hartberg, et al., 2011; Rimol, et al., 2010; van Tol, et al., 2013). Moreover, thickness reductions in the inferior and middle temporal gyrus (Elvsashagen, et al., 2013), parahippocampal (Rimol, et al., 2010) and fusiform gyri (Hartberg, et al., 2011; Lyoo, et al., 2006) has been found in BD patients.

Prefrontal brain regions are closely involved in emotion processing and affect regulation – functions that are clearly disturbed in mood disorders (Phillips, et al., 2003a, 2003b). Distinct frontal lobe structures maintain reciprocal connections to temporal brain areas and are intensively interconnected with limbic regions. On the basis of these findings, it has been postulated that a medial prefrontal network, highly connected to superior and medial temporal lobe, is centrally involved in mood disorders (Drevets, Price, & Furey, 2008; Price & Drevets, 2010).

All imaging studies of mood disorders so far have assessed cortical thickness in already affected individuals only. These studies can't discern whether structural brain abnormalities reflect early neurodevelopmental disruptions predisposing to illness, events linked to the illness onset, or whether they are adaptive or secondary to the effects of chronic illness or its treatment. Neuroimaging studies of individuals at high risk of mood disorders because of a close family history of BD hold the potential to identify the structural brain abnormalities related to enhanced familial vulnerability, unconfounded by the presence of illness. However, to the best of my knowledge, no study has yet examined the effects of familial risk on cortical thickness in a prospective study.

The longitudinal Scottish Bipolar Family Study is well-designed to examine the timing of structural brain abnormalities in mood disorders and their relationship to familial risk and onset of illness. In the analyses presented here, cortical thickness of frontal and temporal ROIs were compared cross-sectionally for data acquired at baseline assessment and longitudinally over time between the HR-MDD, HR-well and HC groups. The rationale for the baseline analysis was to examine whether regional cortical thickness abnormalities exist in the HR-MDD group prior to illness onset that may serve as neuroanatomical markers for a subsequent development of MDD and to assess their relationship to familial risk. The longitudinal analysis was intended to explore the time course of cortical thickness during a period of two years to detect whether abnormal brain development in the high-risk participants is evident that can be linked to an onset of MDD or familial risk. Similar to a repeated measures ANOVA, the longitudinal analysis also allowed for testing with an increased power as compared to baseline analysis whether significant thickness differences between the study groups exist across both time points.

4.2 Methods

4.2.1 Participants and clinical assessment

Participants were recruited as part of the Scottish Bipolar Family Study as outlined more detailed in Chapter 2. The HC, HR-well and HR-MDD groups were matched for age, gender, handedness and intelligence and detailed demographic and clinical descriptions of the study sample are provided in Table 3.1 in Chapter 3.

4.2.2 Magnetic resonance imaging and processing

Structural MRI scans were acquired on a 1.5 Tesla scanner as outlined in Chapter 2. The T_1 weighted images were processed using the surface-based stream of FreeSurfer (fully described in Chapter 2). Based on the literature review provided in Chapter 1 and hence the scientific relevance for mood disorders, cortical thickness of the following brain structures was assessed for each hemisphere separately: frontal pole, inferior frontal gyrus, middle frontal gyrus, superior frontal gyrus, orbitofrontal gyrus (including lateral and medial orbitofrontal gyrus), anterior cingulate (including rostral and caudal anterior cingulate), precentral gyrus, superior temporal gyrus, parahippocampal gyrus and fusiform gyrus.

4.2.3 Statistical analysis

An ANCOVA was conducted to compare cross-sectionally regional thickness differences between the HC, HR-well and HR-MDD groups at baseline. Age and sex served as covariates in this analysis. Next, linear mixed-effects models were applied to investigate cortical thickness for each ROI over time (see Chapter 2 for a more detailed methodological description). In the linear mixed-effects model used, the intercept term is treated as a random effect that varies by individual so that intraindividual correlation among the structural brain measures of a particular

individual is taken into account. The following independent variables were used as predictors of cortical thickness for the regions outlined: group, time (baseline versus follow-up assessment), group-by-time interaction. Age and sex served as covariates in the analyses.

A statistical significance level of $p \leq 0.05$ was chosen, fully corrected for multiple comparisons using the Benjamini & Hochberg FDR procedure (Benjamini & Hochberg, 1995). To allow for comparison with previous studies, effect sizes for nominal significant group differences were additionally calculated using Cohen's *d* (Cohen, 1988). Wherever significant between-group differences were found, pairwise comparisons were performed between the three groups, for which p-values were corrected according to Tukey's HSD procedure ($p_{HSD} \leq 0.05$).

To assess the relationship between depression symptoms and cortical thickness of each ROI, Spearman's rank correlation coefficients between the HAM-D scores and thickness measures for each group were calculated. To assess the relationship between changes in thickness and changes in depressive symptoms over time, Spearman's rank correlation coefficients were calculated between the thickness differences of the ROIs and the differences of the HAM-D scores between the two assessments. For ease of interpretation, data derived from the follow-up assessment was subtracted from the data acquired at baseline so that positive values reflect increases in thickness or depression symptoms over time, while negative values represent decreases of these measures over time. All p-values were corrected according to Benjamini & Hochberg FDR procedure and considered significant when $p \leq 0.05$.

To examine the potentially confounding effects of exposure to medication and relatedness of subjects on regional brain volumes, the following additional analyses were performed: The analyses were repeated excluding medicated HR-MDD subjects, followed by randomly excluding related subjects.

4.3 Results

4.3.1 Cross-sectional analysis

4.3.1.1 Group differences at baseline

At baseline assessment, no significant FDR-corrected differences in regional cortical thickness were found between the groups (see Table 4.1). There was a nominal significant group difference for right parahippocampal ($p \leq 0.011$), right fusiform ($p \leq 0.025$) and right frontal pole thickness ($p \leq 0.033$). Post-hoc tests indicated that the HR-MDD group had reduced right parahippocampal thickness as compared to HC participants ($p_{HSD} \leq 0.010$, $d = 0.757$), with no thickness differences between the HC and HR-well group ($p_{HSD} \leq 0.251$, $d = 0.232$) and the two high-risk groups ($p_{HSD} \leq 0.114$, $d = 0.501$). Post-hoc tests for the right fusiform gyrus revealed no significant pairwise group differences ($p_{HSD} \leq 0.086$, $d = 0.299$ for HC versus HR-well; $p_{HSD} \leq 0.062$, $d = 0.648$ for HC versus HR-MDD; $p_{HSD} \leq 0.559$, $d = 0.292$ for HR-well versus HR-MDD). For the right frontal pole, post-hoc analyses showed cortical thickness differences between the two high-risk groups, with the HR-well group having decreased thickness as compared to the HR-MDD participants ($p_{HSD} \leq 0.027$, $d = 0.685$). No differences between the HC and the HR-well subjects ($p_{HSD} \leq 0.445$, $d = 0.156$) and the HC as compared to the HR-MDD group were observed ($p_{HSD} \leq 0.146$, $d = 0.478$).

4.3.1.2 Correlation analysis

For the cortical thickness of the ROIs, there were no significant correlations with depression symptom severity that remained significant after FDR correction. There were however three nominal significant correlations. The left fusiform gyrus thickness was negatively correlated with the HAM-D sum score in the HC ($p \leq 0.036$) and HR-well group ($p \leq 0.035$), indicating that thinning of this brain region

is associated with higher depression symptoms. For the HR-MDD group, there was a nominal significant positive correlation between the left anterior cingulate thickness and the severity of depression ($p \leq 0.050$), indicating that thickening of this brain region is associated with higher depression symptoms (see Table 4.2).

4.3.1.3 Analysis of potential confounders

To eliminate the potential confounding effects of familial relatedness of some subjects, the analyses were repeated, excluding randomly individuals from the same pedigree (Table 4.3). This analysis similarly yielded no significant FDR-corrected group differences. The nominal significant finding for the right parahippocampal gyrus ($p \leq 0.006$) and right fusiform gyrus ($p \leq 0.017$) remained nominally significant, but not the observed nominal frontal pole thickness differences ($p \leq 0.080$).

Table 4.1 Cross-sectional analysis of cortical thickness

	HC	HR-well	HR-MDD	Statistics	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	F	p
L inf frontal G	2.47 (0.22)	2.49 (0.15)	2.52 (0.20)	0.69	0.50
R inf frontal G	2.46 (0.17)	2.42 (0.15)	2.50 (0.21)	2.48	0.09
L mid frontal G	2.38 (0.18)	2.42 (0.16)	2.43 (0.19)	1.89	0.15
R mid frontal G	2.35 (0.16)	2.34 (0.14)	2.39 (0.20)	1.02	0.36
L sup frontal G	2.70 (0.18)	2.70 (0.18)	2.72 (0.18)	0.18	0.83
R sup frontal G	2.65 (0.19)	2.64 (0.19)	2.70 (0.19)	0.89	0.41
L orbitofrontal G	2.44 (0.23)	2.45 (0.20)	2.42 (0.28)	0.23	0.80
R orbitofrontal G	2.41 (0.23)	2.38 (0.19)	2.40 (0.21)	0.44	0.64
L frontal pole	2.95 (0.40)	3.03 (0.43)	3.06 (0.38)	1.30	0.28
R frontal pole	2.90 (0.41)	2.84 (0.36)	3.08 (0.34)	3.48	0.03
L ant cingulate	2.48 (0.33)	2.47 (0.32)	2.48 (0.33)	0.04	0.96
R ant cingulate	2.39 (0.31)	2.43 (0.32)	2.43 (0.36)	0.36	0.70
L precentral G	2.49 (0.14)	2.49 (0.13)	2.48 (0.18)	0.07	0.94
R precentral G	2.45 (0.12)	2.44 (0.13)	2.44 (0.14)	0.19	0.82
L fusiform G	2.33 (0.21)	2.32 (0.18)	2.29 (0.21)	0.44	0.65
R fusiform G	2.37 (0.18)	2.31 (0.22)	2.25 (0.19)	3.77	0.03
L sup temporal G	2.42 (0.22)	2.45 (0.19)	2.45 (0.18)	0.73	0.49
R sup temporal G	2.44 (0.19)	2.42 (0.21)	2.40 (0.18)	0.73	0.48
L parahippoc G	2.24 (0.34)	2.19 (0.34)	2.14 (0.37)	1.19	0.31
R parahippoc G	2.23 (0.34)	2.15 (0.35)	1.98 (0.32)	4.61	0.01

Cortical thickness measures are provided in mm. Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

Table 4.2 Correlation between cortical thickness and HAM-D sum score

	HC		HR-well		HR-MDD	
Region (Gyrus)	R	p	R	p	R	p
L inf frontal G	-0.16	0.14	0.03	0.81	0.01	0.99
R inf frontal G	-0.20	0.06	-0.02	0.86	0.29	0.23
L mid frontal G	-0.09	0.42	-0.08	0.45	0.11	0.67
R mid frontal G	-0.17	0.11	-0.12	0.25	0.18	0.46
L sup frontal G	-0.15	0.16	-0.06	0.54	0.16	0.51
R sup frontal G	-0.18	0.10	-0.10	0.36	0.03	0.91
L orbitofrontal G	-0.07	0.51	-0.05	0.66	0.21	0.39
R orbitofrontal G	-0.01	0.92	-0.07	0.53	0.19	0.44
L frontal pole	-0.13	0.23	-0.03	0.79	0.03	0.91
R frontal pole	-0.11	0.29	-0.02	0.83	0.17	0.49
L ant cingulate	-0.08	0.48	0.04	0.69	0.46	0.05
R ant cingulate	-0.03	0.81	0.05	0.61	0.36	0.13
L precentral G	-0.20	0.06	-0.15	0.15	-0.40	0.09
R precentral G	-0.14	0.19	-0.07	0.50	-0.31	0.20
L fusiform G	-0.22	0.04	-0.22	0.04	0.06	0.79
R fusiform G	-0.04	0.70	-0.11	0.29	0.22	0.37
L sup temporal G	-0.19	0.07	0.04	0.73	0.12	0.63
R sup temporal G	-0.12	0.27	-0.06	0.54	0.24	0.32
L parahippoc G	-0.16	0.14	0.03	0.78	-0.10	0.70
R parahippoc G	-0.02	0.83	0.12	0.28	0.11	0.65

Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

Table 4.3 Cross-sectional analysis of cortical thickness in unrelated subjects

	HC	HR-well	HR-MDD	Statistics	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	F	p
L inf frontal G	2.47 (0.22)	2.48 (0.16)	2.53 (0.27)	0.70	0.50
R inf frontal G	2.47 (0.17)	2.42 (0.15)	2.50 (0.21)	2.83	0.06
L mid frontal G	2.38 (0.18)	2.42 (0.16)	2.43 (0.19)	1.42	0.25
R mid frontal G	2.35 (0.16)	2.34 (0.14)	2.40 (0.20)	1.36	0.26
L sup frontal G	2.70 (0.18)	2.70 (0.18)	2.73 (0.23)	0.32	0.73
R sup frontal G	2.65 (0.19)	2.63 (0.18)	2.71 (0.20)	1.38	0.25
L orbitofrontal G	2.44 (0.23)	2.44 (0.20)	2.41 (0.29)	0.18	0.83
R orbitofrontal G	2.42 (0.23)	2.38 (0.19)	2.39 (0.25)	0.58	0.56
L frontal pole	2.94 (0.40)	3.06 (0.43)	3.06 (0.39)	1.83	0.16
R frontal pole	2.90 (0.41)	2.84 (0.37)	3.06 (0.33)	2.56	0.08
L ant cingulate	2.48 (0.33)	2.47 (0.31)	2.49 (0.39)	0.06	0.95
R ant cingulate	2.39 (0.31)	2.42 (0.32)	2.40 (0.35)	0.18	0.84
L precentral G	2.49 (0.14)	2.48 (0.14)	2.48 (0.18)	0.04	0.96
R precentral G	2.45 (0.12)	2.43 (0.14)	2.44 (0.14)	0.45	0.64
L fusiform G	2.33 (0.21)	2.31 (0.18)	2.29 (0.21)	0.47	0.63
R fusiform G	2.37 (0.19)	2.29 (0.22)	2.25 (0.20)	4.15	0.02
L sup temporal G	2.42 (0.22)	2.45 (0.19)	2.44 (0.17)	0.48	0.62
R sup temporal G	2.45 (0.19)	2.41 (0.21)	2.41 (0.18)	0.83	0.44
L parahippoc G	2.25 (0.35)	2.20 (0.34)	2.11 (0.36)	1.42	0.25
R parahippoc G	2.23 (0.35)	2.14 (0.34)	1.95 (0.30)	5.34	≤0.01

Cortical thickness measures are provided in mm. Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

4.3.2 Longitudinal analysis

4.3.2.1 Group differences over time

Tables 4.4 and Figure 4.1 depict the results of the linear mixed-effects model analyses. A significant group effect was found for the right parahippocampal gyrus ($p \leq 0.002$) and right fusiform gyrus ($p \leq 0.005$) that passed the FDR procedure. Post-hoc analyses revealed that the HR-well ($p_{HSD} \leq 0.049$, $d = 0.309$) and the HR-MDD ($p_{HSD} \leq 0.011$, $d = 0.851$) group have reduced parahippocampal thickness across time as compared to HC subjects. This brain region was on average 4.00%

thinner in the HR-well group and 10.89% thinner in the HR-MDD group relative to controls. Moreover, the HR-MDD group displayed more pronounced thinning in this region as compared to the HR-well ($p_{HSD} \leq 0.041$, $d = 0.506$) group. For the right fusiform gyrus, the HR-well ($p_{HSD} \leq 0.028$, $d = 0.379$) and the HR-MDD ($p_{HSD} \leq 0.014$, $d = 0.700$) subjects have reduced cortical thickness as compared to the HC group, with no significant difference between the high-risk groups ($p_{HSD} \leq 0.307$, $d = 0.259$). The fusiform gyrus was on average 2.11% thinner in the HR-well and 4.22% thinner in the HR-MDD group as compared to controls.

A significant group-by-time interaction ($p \leq 0.05$) was detected for the left inferior frontal gyrus ($p \leq 0.002$) and the left precentral gyrus ($p \leq 0.001$). For the inferior frontal region, HR-well subjects had a greater cortical thickness decline (3.61% thickness decline) relative to the HC group ($p_{HSD} \leq 0.002$, $d = 0.377$; 1.22% thickness decline) over time and exhibited a distinct pattern of cortical thickness development as compared to the HR-MDD group ($p_{HSD} \leq 0.002$, $d = 0.503$) that showed an increasing thickening over time (1.19% thickness increase). For the left precentral gyrus, the HR-well group exhibited greater cortical thickness decline (2.44% thickness decline) relative to the HC group ($p_{HSD} \leq 0.032$, $d = 0.280$; 1.61% thickness decline) over time, while the HR-MDD subjects showed cortical thickness expansions (2.02% thickness increase) over time which was in contrast to the regional thickness decline observed in the HC ($p_{HSD} \leq 0.001$, $d = 0.616$) and the HR-well ($p_{HSD} \leq 0.001$, $d = 0.786$) group.

4.3.2.2 Correlation analysis

There was no FDR-corrected or nominal significant correlation between regional cortical thickness changes over time and changes of depressive symptoms severity (see Table 4.5).

4.3.2.3 Analysis of potential confounders

All results remained significant after FDR correction when randomly excluding related subjects, except for the group effect of right fusiform gyrus thickness ($p \leq 0.009$) which however remained highly significant at a nominal level (Table 4.3). For the right parahippocampus, there was a significant group effect ($p \leq 0.002$) and significant group-by-time interactions were observed for the left inferior frontal and precentral gyrus ($p \leq 0.002$ and $p \leq 0.001$, respectively).

Similarly, all results remained significant when excluding medicated subjects, except for the group effect of right fusiform gyrus thickness ($p \leq 0.009$) which was only highly significant at a nominal level (see Table 4.4). A significant group effect for the right parahippocampus ($p \leq 0.001$) and significant group-by-time interactions for the left inferior frontal and precentral gyrus were observed ($p \leq 0.005$ and $p \leq 0.001$, respectively).

Table 4.4 Longitudinal analysis of cortical thickness

	HC		HR-well		HR-MDD		Statistics					
	Baseline	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Group effect**		**Time effect**		**GroupXTime**	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L inf frontal G	2.47 (0.22)	2.44 (0.17)	2.49 (0.15)	2.40 (0.17)	2.52 (0.20)	2.55 (0.22)	3.14	0.05	3.22	0.08	5.35	**<0.01**
R inf frontal G	2.46 (0.17)	2.44 (0.19)	2.42 (0.15)	2.42 (0.17)	2.50 (0.21)	2.45 (0.18)	1.69	0.19	3.51	0.06	1.30	0.28
L mid frontal G	2.38 (0.18)	2.39 (0.16)	2.42 (0.16)	2.39 (0.17)	2.43 (0.19)	2.51 (0.16)	3.41	0.04	0.93	0.34	1.80	0.15
R mid frontal G	2.35 (0.16)	2.33 (0.17)	2.34 (0.14)	2.34 (0.16)	2.39 (0.20)	2.39 (0.16)	1.54	0.22	0.57	0.45	0.47	0.71
L sup frontal G	2.70 (0.18)	2.69 (0.19)	2.70 (0.18)	2.68 (0.18)	2.72 (0.18)	2.81 (0.20)	2.40	0.10	0.10	0.76	1.04	0.38
R sup frontal G	2.65 (0.19)	2.61 (0.17)	2.64 (0.19)	2.61 (0.18)	2.70 (0.19)	2.74 (0.17)	3.11	0.05	0.34	0.56	1.65	0.18
L orbitofrontal G	2.44 (0.23)	2.39 (0.20)	2.45 (0.20)	2.41 (0.17)	2.42 (0.28)	2.48 (0.20)	0.73	0.48	0.39	0.54	2.66	0.05
R orbitofrontal G	2.41 (0.23)	2.39 (0.20)	2.38 (0.19)	2.38 (0.19)	2.40 (0.21)	2.42 (0.19)	0.24	0.79	0.04	0.85	0.79	0.50
L frontal pole	2.95 (0.40)	2.95 (0.41)	3.03 (0.43)	3.03 (0.38)	3.06 (0.38)	3.20 (0.43)	3.00	0.05	0.73	0.40	0.44	0.73
R frontal pole	2.90 (0.41)	2.87 (0.40)	2.84 (0.36)	2.85 (0.31)	3.08 (0.34)	3.06 (0.44)	2.94	0.02	0.20	0.66	0.22	0.88
L ant cingulate	2.48 (0.33)	2.46 (0.29)	2.47 (0.32)	2.44 (0.28)	2.48 (0.33)	2.54 (0.29)	0.22	0.80	0.00	1.00	1.57	0.20
R ant cingulate	2.39 (0.31)	2.37 (0.26)	2.43 (0.32)	2.35 (0.29)	2.43 (0.36)	2.42 (0.23)	0.25	0.78	1.88	0.17	1.90	0.13
L precentral G	2.49 (0.14)	2.45 (0.13)	2.49 (0.13)	2.43 (0.15)	2.48 (0.18)	2.53 (0.12)	2.10	0.34	2.83	0.10	8.67	**<0.01**
R precentral G	2.45 (0.12)	2.43 (0.14)	2.44 (0.13)	2.42 (0.15)	2.44 (0.14)	2.46 (0.16)	0.51	0.60	2.02	0.16	2.19	0.09
L fusiform G	2.33 (0.21)	2.29 (0.18)	2.32 (0.18)	2.27 (0.18)	2.29 (0.21)	2.24 (0.30)	1.04	0.35	3.35	0.07	1.72	0.17
R fusiform G	2.37 (0.18)	2.38 (0.17)	2.31 (0.22)	2.33 (0.19)	2.25 (0.19)	2.29 (0.23)	5.20	**<0.01**	0.66	0.42	0.23	0.88
L sup temporal G	2.42 (0.22)	2.43 (0.21)	2.45 (0.19)	2.41 (0.20)	2.45 (0.18)	2.41 (0.25)	0.04	0.96	0.98	0.32	1.03	0.38
R sup temporal G	2.44 (0.19)	2.50 (0.18)	2.42 (0.21)	2.43 (0.21)	2.40 (0.18)	2.40 (0.18)	3.60	0.03	0.50	0.48	1.55	0.21
L parahippoc G	2.24 (0.34)	2.25 (0.32)	2.19 (0.34)	2.12 (0.37)	2.14 (0.37)	2.13 (0.45)	1.76	0.17	1.73	0.19	1.43	0.24
R parahippoc G	2.23 (0.34)	2.27 (0.29)	2.15 (0.35)	2.17 (0.31)	1.98 (0.32)	2.03 (0.37)	6.27	**<0.01**	0.44	0.51	0.20	0.90

Cortical thickness measures are provided in mm. Abbreviations: L = left; R = right; inf = inferior; mid = middle; sup = superior; G = gyrus; parahippoc = parahippocampal; **bold** indicates significant effect after FDR correction.

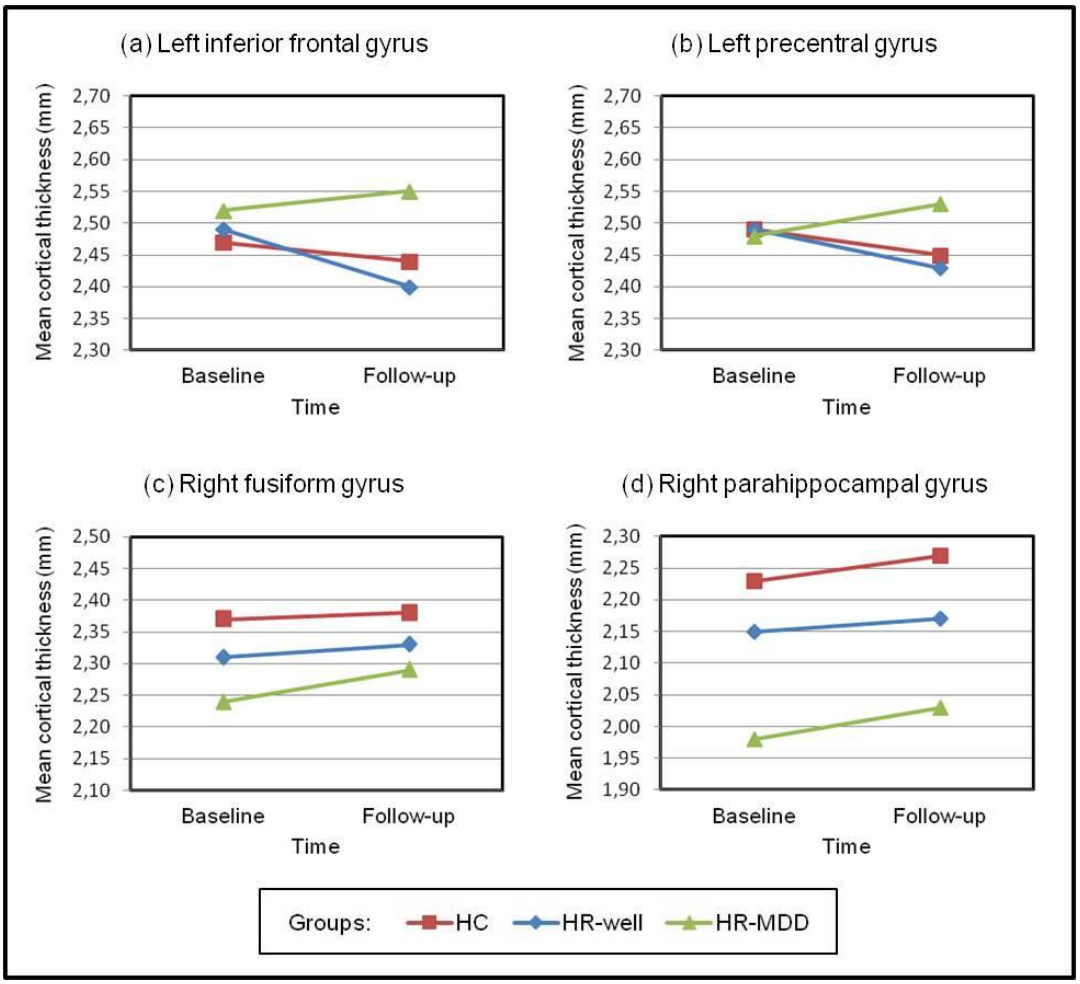

Figure 4.1 Significant group-by-time interactions (a, b) and group effects (c, d) of longitudinal cortical thickness analyses.

Table 4.5 Correlation between changes in cortical thickness and changes in HAM-D sum score

	HC		HR-well		HR-MDD	
Region (Gyrus)	R	p	R	p	R	p
L inf frontal G	0.01	0.96	0.05	0.76	-0.04	0.89
R inf frontal G	0.09	0.55	0.08	0.61	0.12	0.67
L mid frontal G	0.07	0.63	0.09	0.56	-0.06	0.83
R mid frontal G	-0.01	0.97	-0.05	0.72	0.15	0.59
L sup frontal G	0.04	0.81	0.06	0.68	0.14	0.61
R sup frontal G	-0.01	0.99	0.03	0.86	-0.04	0.90
L orbitofrontal G	-0.03	0.84	0.21	0.15	-0.05	0.85
R orbitofrontal G	0.11	0.48	0.02	0.89	-0.18	0.53
L frontal pole	-0.08	0.61	0.01	0.99	-0.21	0.46
R frontal pole	-0.10	0.49	0.12	0.42	-0.17	0.55
L ant cingulate	0.09	0.56	0.14	0.34	0.14	0.61
R ant cingulate	-0.01	0.97	0.09	0.52	0.23	0.42
L precentral G	-0.04	0.80	0.03	0.83	0.02	0.95
R precentral G	-0.04	0.77	0.17	0.23	0.08	0.77
L fusiform G	0.01	0.94	-0.05	0.75	-0.02	0.95
R fusiform G	0.06	0.71	0.11	0.46	0.13	0.65
L sup temporal G	-0.04	0.80	0.07	0.63	-0.05	0.85
R sup temporal G	-0.04	0.79	0.05	0.75	0.08	0.77
L parahippoc G	-0.27	0.07	0.13	0.35	-0.22	0.43
R parahippoc G	-0.16	0.29	0.26	0.07	-0.10	0.72

Abbreviations: L = left; R = right; inf = inferior; mid = middle; sup = superior; G = gyrus; parahippoc = parahippocampal.

Table 4.6 Longitudinal analysis of cortical thickness in unrelated subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Group effect**		**Time effect**		**GroupXTime**	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L inf frontal G	2.47 (0.22)	2.44 (0.16)	2.48 (0.16)	2.39 (0.18)	2.53 (0.27)	2.56 (0.23)	3.82	0.02	2.97	0.09	4.51	**<0.01**
R inf frontal G	2.48 (0.17)	2.44 (0.19)	2.42 (0.15)	2.40 (0.18)	2.50 (0.21)	2.43 (0.18)	2.25	0.11	4.86	0.03	0.43	0.65
L mid frontal G	2.38 (0.18)	2.39 (0.16)	2.42 (0.19)	2.38 (0.16)	2.43 (0.19)	2.52 (0.16)	3.58	0.03	0.73	0.39	2.50	0.09
R mid frontal G	2.35 (0.16)	2.33 (0.17)	2.33 (0.14)	2.32 (0.16)	2.40 (0.20)	2.39 (0.16)	2.13	0.12	1.07	0.30	0.09	0.92
L sup frontal G	2.70 (0.18)	2.68 (0.19)	2.69 (0.18)	2.67 (0.18)	2.73 (0.18)	2.80 (0.20)	2.46	0.09	0.00	1.16	0.97	0.32
R sup frontal G	2.65 (0.19)	2.61 (0.17)	2.63 (0.18)	2.59 (0.18)	2.71 (0.20)	2.74 (0.18)	3.70	0.03	0.61	0.44	1.16	0.32
L orbitofrontal G	2.44 (0.23)	2.39 (0.20)	2.44 (0.20)	2.40 (0.17)	2.44 (0.22)	2.46 (0.19)	0.21	0.81	0.56	0.45	1.79	0.18
R orbitofrontal G	2.42 (0.23)	2.39 (0.20)	2.38 (0.19)	2.36 (0.17)	2.39 (0.25)	2.41 (0.19)	0.62	0.54	0.18	0.67	0.72	0.49
L frontal pole	2.94 (0.40)	2.95 (0.41)	3.06 (0.43)	3.06 (0.36)	3.06 (0.39)	3.16 (0.41)	3.17	0.04	0.43	0.52	0.33	0.72
R frontal pole	2.90 (0.41)	2.87 (0.40)	2.84 (0.37)	2.85 (0.31)	3.06 (0.33)	3.06 (0.45)	2.71	0.07	0.02	0.90	0.36	0.70
L ant cingulate	2.48 (0.33)	2.46 (0.29)	2.47 (0.31)	2.44 (0.26)	2.49 (0.40)	2.53 (0.30)	0.20	0.82	0.01	0.92	1.45	0.24
R ant cingulate	2.39 (0.31)	2.37 (0.26)	2.42 (0.32)	2.34 (0.29)	2.40 (0.35)	2.39 (0.21)	0.02	0.98	1.44	0.23	0.54	0.59
L precentral G	2.49 (0.14)	2.45 (0.13)	2.48 (0.14)	2.42 (0.15)	2.48 (0.18)	2.53 (0.12)	1.78	0.17	2.93	0.09	7.68	**<0.01**
R precentral G	2.45 (0.12)	2.43 (0.14)	2.43 (0.14)	2.41 (0.16)	2.45 (0.14)	2.46 (0.16)	0.87	0.42	2.53	0.11	0.37	0.69
L fusiform G	2.33 (0.21)	2.29 (0.18)	2.31 (0.18)	2.28 (0.18)	2.29 (0.21)	2.26 (0.30)	0.74	0.48	1.90	0.17	0.06	0.94
R fusiform G	2.37 (0.19)	2.38 (0.17)	2.29 (0.22)	2.33 (0.19)	2.25 (0.20)	2.30 (0.23)	4.87	<0.01	1.22	0.27	0.19	0.82
L sup temporal G	2.42 (0.22)	2.43 (0.21)	2.45 (0.19)	2.39 (0.21)	2.44 (0.20)	2.42 (0.26)	0.00	1.00	0.61	0.44	0.82	0.44
R sup temporal G	2.45 (0.19)	2.50 (0.18)	2.41 (0.21)	2.42 (0.21)	2.41 (0.18)	2.40)0.18)	3.74	0.03	0.38	0.54	1.41	0.25
L parahippoc G	2.25 (0.35)	2.25 (0.32)	2.20 (0.34)	2.13 (0.36)	2.11 (0.36)	2.13 (0.36)	1.72	0.18	1.20	0.28	0.48	0.62
R parahippoc G	2.23 (0.35)	2.27 (0.30)	2.14 (0.34)	2.16 (0.32)	1.95 (0.30)	2.03 (0.38)	6.94	**<0.01**	0.67	0.42	0.23	0.80

Cortical thickness measures are provided in mm. Abbreviations: L = left; R = right; inf = inferior; mid = middle; sup = superior; G = gyrus; parahippoc = parahippocampal; **bold** indicates significant effect after FDR correction.

Table 4.7 Longitudinal analysis of cortical thickness in unmedicated subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Group effect**		**Time effect**		**GroupXTime**	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L inf frontal G	2.47 (0.22)	2.44 (0.17)	2.49 (0.15)	2.40 (0.17)	2.52 (0.28)	2.58 (0.22)	3.08	0.05	2.07	0.15	5.33	**<0.01**
R inf frontal G	2.46 (0.17)	2.44 (0.19)	2.42 (0.15)	2.42 (0.17)	2.50 (0.21)	2.46 (0.15)	1.85	0.16	1.64	0.11	0.41	0.66
L mid frontal G	2.38 (0.18)	2.39 (0.16)	2.42 (0.16)	2.39 (0.17)	2.42 (0.21)	2.52 (0.15)	2.81	0.06	1.26	0.26	2.91	0.06
R mid frontal G	2.35 (0.16)	2.33 (0.17)	2.34 (0.14)	2.34 (0.16)	2.40 (0.20)	2.40 (0.14)	1.52	0.22	0.89	0.35	0.27	0.76
L sup frontal G	2.70 (0.18)	2.69 (0.19)	2.70 (0.18)	2.68 (0.18)	2.72 (0.23)	2.77 (0.17)	0.84	0.43	0.10	0.75	0.40	0.67
R sup frontal G	2.65 (0.19)	2.61 (0.17)	2.64 (0.19)	2.61 (0.18)	2.69 (0.21)	2.74 (0.16)	2.22	0.11	0.38	0.54	1.06	0.35
L orbitofrontal G	2.44 (0.23)	2.39 (0.20)	2.45 (0.20)	2.41 (0.17)	2.41 (0.27)	2.50 (0.19)	0.92	0.40	0.07	0.79	2.44	0.09
R orbitofrontal G	2.41 (0.23)	2.39 (0.20)	2.38 (0.19)	2.38 (0.19)	2.40 (0.22)	2.44 (0.20)	0.35	0.70	0.00	0.95	0.88	0.42
L frontal pole	2.95 (0.40)	2.95 (0.41)	3.03 (0.43)	3.03 (0.38)	3.08 (0.40)	3.19 (0.47)	2.93	0.06	0.38	0.54	0.34	0.71
R frontal pole	2.90 (0.41)	2.87 (0.40)	2.84 (0.36)	2.85 (0.31)	3.05 (0.34)	3.04 (0.44)	2.01	0.05	0.24	0.62	0.19	0.83
L ant cingulate	2.48 (0.33)	2.46 (0.29)	2.47 (0.32)	2.44 (0.28)	2.49 (0.39)	2.58 (0.32)	0.48	0.62	0.00	0.96	1.54	0.22
R ant cingulate	2.39 (0.31)	2.37 (0.26)	2.43 (0.32)	2.35 (0.29)	2.43 (0.36)	2.45 (0.24)	0.47	0.63	1.54	0.22	0.69	0.50
L precentral G	2.49 (0.14)	2.45 (0.13)	2.49 (0.13)	2.43 (0.15)	2.47 (0.19)	2.51 (0.11)	0.35	0.71	3.00	0.09	8.35	**<0.01**
R precentral G	2.45 (0.12)	2.43 (0.14)	2.44 (0.13)	2.42 (0.15)	2.44 (0.14)	2.33 (0.14)	0.26	0.77	3.17	0.08	0.16	0.85
L fusiform G	2.33 (0.21)	2.29 (0.18)	2.32 (0.18)	2.27 (0.18)	2.31 (0.22)	2.26 (0.32)	0.47	0.63	2.65	0.11	0.03	0.97
R fusiform G	2.37 (0.18)	2.38 (0.17)	2.31 (0.22)	2.33 (0.19)	2.27 (0.20)	2.30 (0.21)	4.88	<0.01	0.67	0.42	0.12	0.89
L sup temporal G	2.42 (0.22)	2.43 (0.21)	2.45 (0.19)	2.41 (0.20)	2.44 (0.20)	2.43 (0.22)	0.05	0.95	0.56	0.46	0.73	0.48
R sup temporal G	2.44 (0.19)	2.50 (0.18)	2.42 (0.21)	2.43 (0.21)	2.38 (0.19)	2.41 (0.19)	3.79	0.02	0.94	0.33	1.41	0.25
L parahippoc G	2.24 (0.34)	2.25 (0.32)	2.19 (0.34)	2.12 (0.37)	2.16 (0.37)	2.17 (0.44)	1.45	0.24	0.69	0.41	0.81	0.45
R parahippoc G	2.23 (0.34)	2.27 (0.29)	2.15 (0.35)	2.17 (0.31)	2.00 (0.29)	2.06 (0.27)	5.42	**<0.01**	0.54	0.46	0.22	0.81

Cortical thickness measures are provided in mm. Abbreviations: L = left; R = right; inf = inferior; mid = middle; sup = superior; G = gyrus; parahippoc = parahippocampal; **bold** indicates significant effect after FDR correction.

4.4 Discussion

This is, to the best of my knowledge, the first prospective longitudinal study examining cortical thickness in high-risk of mood disorders individuals who were unaffected at initial assessment and either developed MDD or remained well during the follow-up period. Reduced cortical thickness in the right parahippocampal and fusiform gyrus across the two time points were found in both high-risk groups relative to controls, with the HR-MDD group displaying a thinner parahippocampus gyrus than the HR-well group. Over time, HR-well subjects had progressive thickness reductions in the left inferior frontal and precentral gyrus relative to controls, while the HR-MDD group showed cortical thickening of these areas.

The finding of a thinner parahippocampal and fusiform gyrus in high-risk individuals, irrespective of time or the onset of MDD, suggests that thinning in these temporal brain regions constitutes a familial trait marker for vulnerability to mood disorders. Whether these structural brain abnormalities are a consequence of shared genetic and/or environmental effects cannot be determined from the data. Given that they are already present in early adulthood, they are unlikely to be of degenerative origin but likely represent disturbances of normal brain development predisposing to illness. Since the HR-MDD subjects displayed a thinner parahippocampal gyrus than the HR-well group, these reductions may be related to risk of developing MDD. Parahippocampal and fusiform gyrus reductions are thus a potential neuroanatomic endophenotype for mood disorders, as thinning is evident in both unaffected relatives and affected patients and independent of state.

Previous studies support the possibility that right parahippocampal and fusiform thickness reductions are linked to increased vulnerability for mood disorders. Right parahippocampal thinning has been associated with higher genetic liability to BD (Hulshoff Pol et al., 2012), and research focussing on candidate genes for mood disorders has detected associations between risk allele carriers of the *DISC1* or

BDNF gene and reductions in parahippocampal volume/thickness (Carless et al., 2011; Montag, Weber, Fliessbach, Elger, & Reuter, 2009; Takahashi et al., 2008) and fusiform volume (Montag, et al., 2009). Right parahippocampal thinning has been observed in BD-I patients (Rimol, et al., 2010) and right fusiform thinning in a BD cohort (Hartberg, et al., 2011; Lyoo, et al., 2006). The few studies investigating cortical thickness in MDD have not detected similar findings (Ajilore et al., 2011; Grieve, et al., 2013; Jarnum et al., 2011; Koolschijn et al., 2010; Lan, et al., 2014; Qiu, et al., 2014; Reynolds, et al., 2014; Truong, et al., 2013; Tu, et al., 2012; van Tol, et al., 2013), but since they included predominantly medicated and/or older adults, it is likely that age, medication or duration of illness effects account for this discrepancy. In keeping with this, a recent voxel-based morphometry meta-analysis indeed showed that only first-episode, mainly medication naïve, MDD patients have decreased grey matter in a cluster encompassing the right parahippocampal gyrus (Bora, et al., 2012). Moreover, fusiform thinning in high-risk of depression individuals because of a close family history of MDD has been found to be associated with higher depression severity (Peterson et al., 2009).

The parahippocampal gyrus is of particular interest for the aetiology of mood disorders because of its potential role in emotional regulation. Functional MRI studies applying facial affect processing paradigms found that BD and MDD patients have increased activation in the right parahippocampus as compared to controls (Delvecchio et al., 2012). Our research group has recently shown that high-risk of mood disorders individuals who are homozygous for risk haplotype of the DGKH gene show relatively greater brain activation of the right parahippocampus during a verbal fluency task as compared to low-risk haplotype subjects, with the reverse pattern being observed for healthy control subjects (Whalley et al., 2012). Furthermore, it has been shown that remitted depressed patients maintain an increased connectivity of the posterior cingulate cortex with the parahippocampal gyrus and that greater connectivity appears to represent a

prognostic factor for future depressive episodes (Zamoscik, Huffziger, Ebner-Priemer, Kuehner, & Kirsch, 2014).

Our analysis yielded significant group-by-time interactions for the left inferior frontal and precentral gyrus. The finding of abnormal thinning in these brain areas in the HR-well group over time relative to controls suggests that thinning in regionally specific left frontal lobe areas forms a familial trait marker for vulnerability to mood disorders and that abnormal thinning already takes place in early adulthood, potentially reflecting early neurodegenerative processes.

Our results are in line with a twin study that found liability for BD to be associated with inferior frontal and precentral grey matter density reductions (van der Schot, et al., 2010), with the precentral grey matter reductions being limited to the right hemisphere however. Despite potential differences in underlying environmental and genetic risk factors, cortical thinning in both of these frontal areas has also been observed in cohort of unaffected relatives of MDD patients, with the inferior frontal thinning being restricted to the right hemisphere though (Peterson, et al., 2009). Also, reduced grey matter volumes of the left precentral gyrus have been detected in individuals at high risk of MDD because of negative cognitive styles (X. Zhang, Yao, Zhu, Wang, & Zhong, 2012). Moreover, our findings are in concert with several neuroimaging studies reporting cortical thinning or grey matter volume reductions in the circumscribed brain regions in BD (Bora, et al., 2010; Doris, Belton, Ebmeier, Glabus, & Marshall, 2004; Foland-Ross, et al., 2011; Hartberg, et al., 2011; Rimol, et al., 2010) and MDD patients (Tu, et al., 2012; J. Zhang, Xiao, Zhu, Wang, & Yao, 2011; X. Zhang, et al., 2012).

Importantly, we observed a distinct pattern of increasing relative cortical thickness over time in the HR-MDD as compared to the HR-well group due to an absence of regional thinning of the left inferior frontal and precentral gyrus in the HR-MDD cohort. For the precentral gyrus, the cortical thickness development in the HR-MDD group was also significantly different from the HC subjects. Since our results remained significant after excluding medicated individuals, the observed findings

in the HR-MDD group cannot be attributed to medication effects but rather appear to be linked to the onset of illness and underlying disease-associated processes. Given that human brain maturation involves frontal grey matter loss beyond adolescence (Gogtay, et al., 2004; Sowell et al., 2003; Westlye et al., 2010), the absence of cortical thinning in the HR-MDD group may reflect a lack or delay of normal synaptic pruning processes.

Although these findings are in contrast to the frequently observed thinning or grey matter decrease in MDD patients, they are in line with two of the three existing longitudinal studies of MDD patients that found cortical thickening of frontal or temporal brain structures over time (Ahdidan et al., 2011; Jarnum, et al., 2011) and with recent findings of cortical thickening of various brain areas in MDD patients (Grieve, et al., 2013; Qiu, et al., 2014; Reynolds, et al., 2014; van Eijndhoven, et al., 2013). Interestingly, our findings overlap with a longitudinal study of paediatric prodromal BD subjects, displaying grey matter increases in the left ventrolateral prefrontal cortex (including the inferior frontal gyrus) over time as they convert to BD (Gogtay et al., 2007).

The strengths of this study are its longitudinal nature, the assessment of subjects prior to illness onset, the relatively young age of the participants and the comparatively large sample size of high-risk subjects and controls. In addition, all subjects underwent careful clinical assessment at both time points and the effects of medication and relatedness of subjects were ruled out. All brain scans were obtained at the same scanner using the identical protocol at both visits and the MRI data were processed in an identical way using thoroughly validated methods.

Nevertheless, some limitations need to be addressed. First, it remains unknown whether currently unaffected HR-well subjects may develop MDD in the future. Second, previous longitudinal studies have reported that the majority of the high-risk subjects who developed BD themselves experienced depressive episodes years before conversion (Duffy, 2010; Hillegers, et al., 2005) so that it appears likely that some of our HR-MDD subjects may develop BD in the future. The follow-up

assessments of our study cohort will clarify if some of the HR-MDD participants will convert to BD and if some of our HR-well subjects go on to develop a mood disorder. Third, our study groups differed with respect to depression symptom severity at baseline. However, the median of the HAM-D total score was only 1 in the HR-MDD group, suggesting only sub-syndromal depression symptoms. Moreover, our correlation analysis revealed no relationship between depression symptom severity and our structural brain measures. Therefore, it appears unlikely that general mood differences at baseline between the groups have influenced our findings.

In summary, our findings suggest that thinning in the right parahippocampal and right fusiform gyrus across time constitutes a familial trait marker for vulnerability to mood disorders. Moreover, enhanced liability to mood disorders is associated with abnormal left inferior frontal and precentral gyrus thinning in early adulthood, potentially reflecting early neurodegenerative processes. By contrast, the onset of MDD is linked to initially thickening of these brain areas, possibly linked to disease-associated processes through a lack of synaptic pruning. These findings advance our understanding of the neuropathological processes underlying mood disorders and future longitudinal studies should examine their validity.

Chapter 5

Cortical surface area in individuals at high familial risk of mood disorders

5.1 Introduction

Close relatives of BD patients are at enhanced risk of developing BD or MDD during their lifetime in comparison to the general population as outlined in more detail in Chapter 1 (Smoller & Finn, 2003). Several neuroanatomical abnormalities in the frontal and temporal lobes have been associated with mood disorders (Beyer & Krishnan, 2002; Bora, et al., 2012; Kempton, et al., 2011; Konarski, et al., 2008; Savitz & Drevets, 2009). In brief, grey matter volume reductions of the prefrontal cortex (Arnone, et al., 2009; Koolschijn, et al., 2009), particularly in the orbitofrontal gyrus (Bora, et al., 2012; Kempton, et al., 2011; Koolschijn, et al., 2009; Stanfield, et al., 2009; Wilke, et al., 2004), inferior, middle and superior frontal gyrus (Bora, et al., 2012; Bora, et al., 2010; Du, et al., 2012; Lopez-Larson, et al., 2002; Lyoo, et al., 2004; Selvaraj, et al., 2012) have been observed in both MDD and BD patients. Moreover, grey matter decreases in the anterior cingulate (Bora, et al., 2012; Bora, et al., 2010; Koolschijn, et al., 2009) and precentral gyrus (Bora, et al., 2012; Lyoo, et al., 2004) have been detected. For the temporal lobe, there is evidence towards volumetric abnormalities of the superior temporal and medial temporal lobe (Bora, et al., 2012; Selvaraj, et al., 2012). A detailed review of frontal and temporal brain abnormalities in MDD as compared to BD is provided in Chapter 1.

The majority of research has examined grey matter volume but not cortical surface area in mood disorders. Since grey matter volume is a product of cortical thickness and surface area (Winkler, et al., 2009), it remains therefore largely unknown if the frequently observed regional volume reductions in mood disorder patients reflect reduced cortical surface, reduced cortical thickness, or both. Only very few studies

have examined cortical surface area in mood disorders patients as yet. Pubmed was searched on 19th of August 2013 with the terms ("mood disorder" or "bipolar" or "major depressive disorder" or "depression" or "depressive") and ("surface area") to identify previous studies that have examined cortical surface area abnormalities in mood disorders patients or their unaffected close relatives. None of the papers matching the search criteria were found to have investigated surface area in MDD patients or unaffected close relatives of mood disorders patients. Only a few studies were found that examined cortical surface area in BD.

In a large study sample of 139 BD patients and 207 control subjects, Hartberg and colleagues (2011) found larger surface areas of the left superior frontal gyrus and right temporal pole in a combined cohort of BD-I and BD-II patients as compared to controls. However, a recent analysis of this study cohort revealed no significant differences for surface area between the groups when comparing healthy controls with BD-I patients only (Rimol et al., 2012). Research focussing on the surface area of the anterior cingulate gyrus similarly did not find evidence for surface area abnormalities in BD-I patients or BD patients who were experiencing a first psychotic episode (Fornito et al., 2008; Fornito et al., 2009).

To examine whether cortical surface area abnormalities exist in individuals at high familial risk of mood disorders and to study their timing and relationship to an onset of MDD, cortical surface area for frontal and temporal ROIs were extracted from baseline and follow-up scans of Scottish Bipolar Family Study participants. In the analyses presented here, cortical surface area of frontal and temporal ROIs were compared cross-sectionally for data acquired at baseline assessment and longitudinally over time between the HR-MDD, HR-well and HC groups. The rationale for the baseline analysis was to examine whether regional cortical surface area abnormalities exist in the HR-MDD group prior to illness onset that may serve as neuroanatomical markers for a subsequent development of MDD and to assess their relationship to familial risk. The longitudinal analysis was intended to explore the time course of surface area during a period of two years to detect whether

abnormal brain development in the high-risk participants is evident that can be linked to an onset of MDD or familial risk. Similar to a repeated-measures ANOVA, the longitudinal analysis also allowed for testing with an increased power as compared to baseline analysis whether significant thickness differences between the study groups exist across both time points.

5.2 Methods

5.2.1 Participants and clinical assessment

Participants were recruited as part of the Scottish Bipolar Family Study as outlined more detailed in Chapter 2. The HC, HR-well and HR-MDD groups were matched for age, gender, handedness and intelligence and detailed demographic and clinical descriptions of the study sample are provided in Table 3.1 in Chapter 3.

5.2.2 Magnetic resonance imaging and processing

Structural MRI scans were acquired on a 1.5 Tesla scanner as outlined in Chapter 2. The T_1 weighted images were processed using the surface-based stream of FreeSurfer (fully described in Chapter 2). Based on the literature review provided in Chapter 1 and hence the scientific relevance for mood disorders, cortical surface area of the following brain structures was assessed for each hemisphere separately: frontal pole, inferior frontal gyrus, middle frontal gyrus, superior frontal gyrus, orbitofrontal gyrus (including lateral and medial orbitofrontal gyrus), anterior cingulate (including rostral and caudal anterior cingulate), precentral gyrus, superior temporal gyrus, parahippocampal gyrus and fusiform gyrus.

5.2.3 Statistical analysis

An ANCOVA was conducted to compare cross-sectionally regional surface area differences between the HC, HR-well and HR-MDD groups at baseline. Age and sex served as covariates in this analysis. Next, linear mixed-effects models were applied to investigate surface area for each ROI over time (see Chapter 2 for a more detailed methodological description). In the linear mixed-effects model used, the intercept term is treated as a random effect that varies by individual so that intraindividual correlation among the structural brain measures of a particular individual is taken into account. The following independent variables were used as predictors of surface area for the cortical regions outlined: group, time (baseline versus follow-up assessment), group-by-time interaction. Age and sex served as covariates in the analyses.

A statistical significance level of $p \leq 0.05$ was chosen, fully corrected for multiple comparisons using the Benjamini & Hochberg FDR procedure (Benjamini & Hochberg, 1995). To allow for comparison with previous studies, effect sizes for nominal significant group differences were additionally calculated using Cohen's *d* (Cohen, 1988). Wherever significant between-group differences were found, pairwise comparisons were performed between the three groups, for which p-values were corrected according to Tukey's HSD procedure ($p_{HSD} \leq 0.05$).

To assess the relationship between depression symptoms and surface area of each ROI, Spearman's rank correlation coefficients between the HAM-D scores and surface area for each group were calculated. To assess the relationship between changes in surface area and changes in depressive symptoms over time, Spearman's rank correlation coefficients were calculated between the surface area differences of the ROIs and the differences of the HAM-D scores between the two assessments. For ease of interpretation, data derived from the follow-up assessment was subtracted from the data acquired at baseline so that positive values reflect increases in surface area or depression symptoms over time, while negative values

represent decreases of these measures over time. All p-values were corrected according to Benjamini & Hochberg FDR procedure and considered significant when $p \leq 0.05$.

To examine the potentially confounding effects of exposure to medication and relatedness of subjects on regional brain volumes, the following additional analyses were performed: The analyses were repeated excluding medicated HR-MDD subjects, followed by randomly excluding related subjects.

5.3 Results

5.3.1 Cross-sectional analysis

5.3.1.1 Group differences at baseline

At baseline assessment, no significant differences in regional cortical surface area were found between the groups (see Table 5.1). Also, there were no nominal significant findings.

5.3.1.2 Correlation analysis

For the ROIs, there was a significant positive correlation between the HAM-D sum score and the left frontal pole surface area in the HR-MDD group that remained significant after FDR correction ($p \leq 0.002$), indicating that surface area increases in this brain region are associated with higher depression symptoms in this study group. There were no other correlations that remained significant after FDR procedure. However, there was a nominal significant negative correlation between the HAM-D sum score and the right orbitofrontal surface area in the HC group ($p \leq 0.022$), suggesting that reducing surface area is associated with increasing

depressive symptom severity. No other nominal significant correlations were found. All results are shown in Table 5.2.

5.3.1.3 Analysis of potential confounders

To eliminate the potential confounding effects of familial relatedness of some subjects, the analyses were repeated, excluding randomly individuals from the same pedigree (Table 5.3). This analysis similarly yielded no significant group differences.

Table 5.1 Cross-sectional analysis of cortical surface area

	HC	HR-well	HR-MDD	Statistics	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	F	P
L inf frontal G	35.24 (4.37)	35.76 (5.27)	35.51 (2.47)	0.28	0.76
R inf frontal G	35.71 (4.73)	36.32 (5.38)	36.44 (4.45)	0.41	0.67
L mid frontal G	76.04 (12.12)	77.75 (10.95)	76.38 (9.26)	0.54	0.59
R mid frontal G	75.54 (11.58)	77.72 (11.59)	75.83 (10.35)	0.87	0.42
L sup frontal G	69.41 (8.49)	71.12 (8.66)	69.07 (7.07)	1.12	0.33
R sup frontal G	67.37 (8.40)	68.62 (9.22)	67.17 (8.97)	0.54	0.59
L orbitofrontal G	41.60 (5.55)	41.74 (6.13)	42.45 (4.72)	0.17	0.84
R orbitofrontal G	40.86 (6.12)	41.20 (6.00)	42.09 (4.70)	0.35	0.71
L frontal pole	2.38 (0.40)	2.40 (0.38)	2.45 (0.51)	0.26	0.77
R frontal pole	3.20 (0.55)	3.29 (0.51)	3.14 (0.43)	1.03	0.36
L ant cingulate	13.40 (2.76)	14.31 (3.19)	13.79 (2.02)	2.24	0.11
R ant cingulate	13.54 (2.76)	13.24 (2.64)	12.64 (2.00)	0.99	0.37
L precentral G	45.87 (4.65)	47.12 (5.82)	46.33 (4.50)	1.34	0.27
R precentral G	46.73 (5.18)	46.62 (5.94)	46.35 (5.54)	0.04	0.96
L fusiform G	31.95 (4.71)	32.46 (4.31)	32.22 (3.54)	0.30	0.74
R fusiform G	30.19 (4.24)	30.94 (4.65)	30.81 (3.77)	0.68	0.51
L sup temporal G	36.05 (4.59)	36.79 (4.31)	36.00 (4.55)	0.70	0.50
R sup temporal G	34.11 (3.68)	34.60 (4.12)	33.06 (4.54)	1.26	0.29
L parahippoc G	6.88 (0.99)	7.00 (0.97)	6.79 (1.00)	0.51	0.60
R parahippoc G	6.41 (0.81)	6.47 (0.93)	6.58 (0.88)	0.32	0.73

Cortical surface area measures are provided in cm^2. Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

Table 5.2 Correlation of cortical surface area and HAM-D sum score

	HC		HR-well		HR-MDD	
Region (Gyrus)	R	p	R	p	R	p
L inf frontal G	-0.06	0.60	-0.04	0.68	-0.02	0.94
R inf frontal G	-0.19	0.07	-0.05	0.61	0.16	0.52
L mid frontal G	-0.18	0.09	-0.04	0.71	-0.04	0.89
R mid frontal G	-0.17	0.12	0.03	0.78	-0.12	0.64
L sup frontal G	-0.04	0.71	0.00	0.99	0.30	0.23
R sup frontal G	-0.16	0.13	-0.10	0.39	0.14	0.58
L orbitofrontal G	-0.16	0.14	-0.09	0.37	-0.05	0.84
R orbitofrontal G	-0.24	0.02	0.03	0.77	-0.04	0.88
L frontal pole	-0.13	0.24	-0.11	0.31	0.68	**<0.01**
R frontal pole	-0.01	0.90	-0.05	0.63	-0.05	0.84
L ant cingulate	-0.10	0.34	-0.08	0.46	-0.02	0.95
R ant cingulate	-0.13	0.22	-0.05	0.67	0.19	0.46
L precentral G	0.06	0.61	0.08	0.45	0.25	0.33
R precentral G	0.02	0.86	0.00	0.98	0.30	0.23
L fusiform G	-0.02	0.82	0.04	0.74	0.21	0.41
R fusiform G	-0.02	0.88	0.02	0.83	0.14	0.57
L sup temporal G	-0.07	0.50	-0.06	0.56	0.20	0.42
R sup temporal G	-0.06	0.56	0.03	0.76	0.09	0.74
L parahippoc G	0.10	0.38	0.06	0.55	0.09	0.71
R parahippoc G	-0.11	0.31	0.03	0.81	0.37	0.13

Abbreviations: L = left; R = right; inf = inferior; mid = middle; sup = superior; G = gyrus; parahippoc = parahippocampal; **bold** indicates significant effect after FDR correction.

Table 5.3 Cross-sectional analysis of cortical surface area in unrelated subjects

	HC	HR-well	HR-MDD	Statistics	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	F	P
L inf frontal G	35.22 (4.39)	35.94 (5.12)	35.56 (2.53)	0.52	0.60
R inf frontal G	35.59 (4.68)	36.58 (5.24)	36.62 (4.52)	0.96	0.39
L mid frontal G	75.89 (12.22)	78.22 (10.93)	76.94 (9.19)	0.86	0.42
R mid frontal G	75.51 (11.66)	78.04 (11.43)	76.21 (10.52)	1.02	0.36
L sup frontal G	69.37 (8.56)	71.60 (8.23)	69.49 (7.03)	1.60	0.21
R sup frontal G	67.24 (8.45)	68.99 (8.91)	67.71 (8.90)	0.85	0.43
L orbitofrontal G	41.49 (5.55)	41.81 (6.19)	42.65 (4.77)	0.32	0.73
R orbitofrontal G	40.76 (6.13)	41.35 (6.08)	42.35 (4.69)	0.60	0.55
L frontal pole	2.38 (0.41)	2.39 (0.40)	2.46 (0.53)	0.28	0.76
R frontal pole	3.20 (0.54)	3.35 (0.47)	3.13 (0.44)	2.57	0.08
L ant cingulate	13.31 (2.72)	14.32 (3.14)	13.91 (2.01)	2.63	0.08
R ant cingulate	13.47 (2.75)	13.17 (2.57)	12.76 (1.98)	0.66	0.52
L precentral G	45.86 (4.70)	47.51 (5.68)	46.65 (4.41)	2.14	0.12
R precentral G	46.69 (5.21)	46.91 (5.86)	46.77 (5.37)	0.03	0.97
L fusiform G	31.87 (4.73)	32.53 (4.33)	32.42 (3.53)	0.47	0.62
R fusiform G	30.11 (4.22)	31.00 (4.79)	30.88 (3.86)	0.88	0.42
L sup temporal G	35.99 (4.60)	36.85 (4.42)	35.98 (4.68)	0.82	0.44
R sup temporal G	34.10 (3.72)	34.47 (4.15)	33.19 (4.63)	0.78	0.46
L parahippoc G	6.86 (0.99)	7.01 (1.00)	6.76 (1.00)	0.67	0.51
R parahippoc G	6.40 (0.81)	6.44 (0.92)	6.58 (0.90)	0.33	0.72

Cortical surface area measures are provided in cm^2. Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

5.3.2 Longitudinal analysis

5.3.2.1 Group differences over time

Table 5.4 provides the results of the linear mixed-effects model analyses. The analysis yielded no significant FDR-adjusted effects of group, time or group-by-time interactions. There were, however, four nominal significant findings. First, a nominal significant group effect ($p \leq 0.039$) was found for the right superior temporal gyrus and the right anterior cingulate ($p \leq 0.049$). Post-hoc analyses revealed that the HR-MDD group has smaller right superior temporal surface area

across both time points as compared to HC ($p_{HSD} \leq 0.023$, $d = 0.704$) and HR-well subjects ($p_{HSD} \leq 0.036$, $d = 0.578$), with no significant differences between the other two groups ($p_{HSD} \leq 0.999$, $d = 0.036$). For the right anterior cingulate, post-hoc analyses revealed no significant pair-wise group differences ($p_{HSD} \leq 0.170$ for HC versus HR-well, $d = 0.278$; $p_{HSD} \leq 0.101$, $d = 0.539$ for HC versus HR-MDD; $p_{HSD} \leq 0.988$, $d = 0.243$ for HR-well versus HR-MDD). Moreover, nominal significant group-by-time interactions for the left fusiform ($p \leq 0.028$) and left precentral gyrus ($p \leq 0.034$) were found. Post-hoc tests showed that the HR-MDD group had a pronounced decrease in fusiform surface area over time as opposed to the HC group ($p_{HSD} \leq 0.022$, $d = 0.590$), with no other significant interaction effects among the groups ($p_{HSD} \leq 0.099$, $d = 0.244$ for HC versus HR-well; $p_{HSD} \leq 0.128$, $d = 0.388$ for HR-well versus HR-MDD). Similar results were obtained for the left precentral gyrus, with the HR-MDD group displaying an abnormal decrease in surface area as compared to the HC group ($p_{HSD} \leq 0.016$, $d = 0.622$) and no interaction effects between the other groups ($p_{HSD} \leq 0.165$, $d = 0.210$ for HC versus HR-well; $p_{HSD} \leq 0.128$, $d = 0.377$ for HR-well versus HR-MDD).

5.3.2.2 Correlation analysis

There were no significant correlations between changes in surface area of each ROI and changes in depressive symptom severity. There was a nominal significant positive correlation between changes in right fusiform gyrus surface area and changes in the HAM-D sum score in the HR-well group ($p \leq 0.036$), indicating that increasing surface area of this brain region is associated with increasing depression symptom severity (see Table 5.5).

5.3.2.3 Analysis of potential confounders

To eliminate the potential confounding effects of familial relatedness of some subjects, the analyses were repeated, excluding randomly individuals from the same pedigree (Table 5.6). This analysis similarly yielded no significant effect of group, time or group-by-time interaction. The nominal significant group effects for the right superior temporal gyrus ($p \leq 0.046$) and the right anterior cingulate surface area ($p \leq 0.044$) remained nominal significant. Similarly, the nominal significant group-by-time interactions for the left precentral gyrus ($p \leq 0.033$) and left fusiform gyrus ($p \leq 0.037$) remained nominal significant.

When excluding medicated HR-MDD subjects from the analysis, there were no FDR-adjusted significant findings. By contrast, three out of the four nominal significant results of the original analysis did not reach nominal significance anymore. These were the group effects for the right superior temporal gyrus ($p \leq 0.210$) and the right anterior cingulate ($p \leq 0.125$) as well as the group-by-time interaction effect for the left fusiform gyrus ($p \leq 0.055$). Only the group-by-time interaction for the surface area of the left precentral gyrus remained significant at a nominal level ($p \leq 0.035$).

Table 5.4 Longitudinal analysis of cortical surface area

	HC		HR-well		HR-MDD		Statistics					
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Group effect		Time effect		GroupXTime	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L inf frontal G	35.24 (4.37)	35.43 (4.43)	35.76 (5.27)	35.34 (5.02)	35.51 (2.47)	34.97 (3.82)	0.07	0.93	1.18	0.28	0.25	0.78
R inf frontal G	35.71 (4.73)	36.31 (4.63)	36.32 (5.38)	35.92 (5.17)	36.44 (4.45)	35.16 (4.50)	0.09	0.91	2.48	0.12	1.56	0.21
L mid frontal G	76.04 (12.12)	77.38 (10.86)	77.75 (10.95)	77.00 (9.55)	76.38 (9.26)	72.70 (9.30)	0.95	0.39	2.80	0.10	1.19	0.31
R mid frontal G	75.54 (11.58)	76.77 (10.51)	77.72 (11.59)	77.23 (11.29)	75.83 (10.35)	72.27 (9.96)	1.17	0.31	1.30	0.26	1.23	0.30
L sup frontal G	69.41 (8.49)	70.77 (7.60)	71.12 (8.66)	70.49 (7.06)	69.07 (7.07)	65.62 (7.76)	2.39	0.09	2.70	0.10	2.33	0.10
R sup frontal G	67.37 (8.40)	68.28 (7.30)	68.62 (9.22)	68.26 (8.41)	67.17 (8.97)	65.01 (9.88)	0.96	0.39	1.53	0.22	1.35	0.26
L orbitofrontal G	41.60 (5.55)	41.48 (4.78)	41.74 (6.13)	42.06 (5.07)	42.45 (4.72)	40.61 (5.47)	0.13	0.88	1.07	0.30	2.32	0.73
R orbitofrontal G	40.86 (6.12)	41.10 (5.19)	41.20 (6.00)	41.25 (5.02)	42.09 (4.70)	40.29 (5.11)	0.03	0.97	0.85	0.36	0.73	0.48
L frontal pole	2.38 (0.40)	2.38 (0.40)	2.40 (0.38)	2.42 (0.41)	2.45 (0.51)	2.22 (0.35)	0.21	0.81	0.89	0.35	2.77	0.07
R frontal pole	3.20 (0.55)	3.23 (0.52)	3.29 (0.51)	3.24 (0.51)	3.14 (0.43)	3.22 (0.63)	0.52	0.59	0.28	0.60	0.28	0.76
L ant cingulate	13.40 (2.76)	13.59 (2.74)	14.31 (3.19)	13.88 (2.91)	13.79 (2.02)	13.52 (2.94)	0.93	0.40	0.05	0.83	1.05	0.35
R ant cingulate	13.54 (2.76)	13.86 (2.46)	13.24 (2.64)	12.86 (2.66)	12.64 (2.00)	12.76 (2.35)	3.06	0.05	0.28	0.60	1.25	0.29
L precentral G	45.87 (4.65)	46.67 (3.99)	47.12 (5.82)	46.77 (4.09)	46.33 (4.50)	44.54 (5.02)	1.08	0.34	1.31	0.26	3.47	0.03
R precentral G	46.73 (5.18)	47.53 (4.51)	46.62 (5.94)	46.23 (4.48)	46.35 (5.54)	44.54 (4.86)	1.70	0.19	0.81	0.37	2.24	0.11
L fusiform G	31.95 (4.71)	32.91 (4.87)	32.46 (4.31)	32.02 (3.68)	32.22 (3.54)	30.10 (5.56)	1.18	0.31	1.45	0.23	3.66	0.03
R fusiform G	30.19 (4.24)	30.55 (4.44)	30.94 (4.65)	30.72 (4.60)	30.81 (3.77)	29.84 (4.72)	0.42	0.66	0.01	0.94	0.18	0.83
L sup temporal G	36.05 (4.59)	36.42 (4.30)	36.79 (4.31)	36.70 (3.89)	36.00 (4.55)	35.00 (5.29)	0.89	0.41	1.23	0.27	0.19	0.83
R sup temporal G	34.11 (3.68)	34.80 (2.89)	34.60 (4.12)	34.19 (3.56)	33.06 (4.54)	32.15 (4.70)	3.30	0.04	0.63	0.43	1.63	0.20
L parahippoc G	6.88 (0.99)	6.71 (0.80)	7.00 (0.97)	7.04 (1.40)	6.79 (1.00)	6.55 (0.90)	1.72	0.18	1.40	0.24	0.85	0.43
R parahippoc G	6.41 (0.81)	6.50 (0.79)	6.47 (0.93)	6.33 (0.86)	6.58 (0.88)	6.45 (0.93)	0.23	0.80	0.40	0.53	1.22	0.30

Cortical surface area measures are provided in cm^2. Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

Table 5.5 Correlation between changes in surface area and changes in HAM-D sum score

	HC		HR-well		HR-MDD	
Region (Gyrus)	R	p	R	p	R	p
L inf frontal G	0.08	0.60	0.21	0.15	0.07	0.80
R inf frontal G	-0.04	0.82	0.19	0.19	0.30	0.27
L mid frontal G	-0.01	0.95	0.13	0.38	0.30	0.28
R mid frontal G	0.04	0.80	0.17	0.24	0.14	0.61
L sup frontal G	0.08	0.58	0.23	0.11	0.39	0.15
R sup frontal G	-0.08	0.60	0.19	0.19	0.11	0.70
L orbitofrontal G	0.05	0.76	0.01	0.94	-0.14	0.62
R orbitofrontal G	0.05	0.73	0.09	0.55	0.23	0.41
L frontal pole	-0.12	0.44	0.18	0.21	-0.12	0.66
R frontal pole	-0.01	0.93	-0.03	0.84	-0.07	0.80
L ant cingulate	-0.10	0.52	0.20	0.16	0.26	0.34
R ant cingulate	0.05	0.75	0.07	0.64	-0.02	0.93
L precentral G	0.04	0.77	0.26	0.07	0.30	0.28
R precentral G	0.01	0.98	0.19	0.19	0.17	0.54
L fusiform G	0.05	0.73	0.20	0.16	0.20	0.47
R fusiform G	0.09	0.54	0.30	0.04	0.05	0.85
L sup temporal G	-0.01	0.97	0.18	0.21	0.25	0.38
R sup temporal G	-0.07	0.65	0.05	0.72	0.21	0.45
L parahippoc G	-0.09	0.55	-0.03	0.82	0.08	0.77
R parahippoc G	0.05	0.76	0.06	0.70	0.15	0.58

Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

Table 5.6 Longitudinal analysis of cortical surface area in unrelated subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Group effect		Time effect		GroupXTime	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
L inf frontal G	35.22 (4.39)	35.43 (4.43)	35.22 (5.12)	35.54 (4.88)	35.63 (2.59)	35.26 (3.70)	0.15	0.87	1.01	0.32	0.23	0.79
R inf frontal G	35.59 (4.68)	36.31 (4.63)	36.58 (5.24)	35.99 (4.84)	36.58 (4.65)	35.28 (4.59)	0.15	0.87	2.65	0.11	1.90	0.16
L mid frontal G	75.89 (12.22)	77.38 (10.86)	78.22 (10.93)	77.11 (9.60)	76.69 (9.41)	73.41 (8.97)	0.54	0.58	2.74	0.10	1.27	0.28
R mid frontal G	75.51 (11.66)	76.77 (10.51)	78.04 (11.43)	77.51 (11.54)	75.94 (10.77)	73.17 (9.36)	0.74	0.48	1.06	0.31	1.09	0.34
L sup frontal G	69.37 (8.56)	70.77 (7.60)	71.60 (8.23)	70.86 (6.48)	69.08 (7.03)	66.05 (7.72)	2.22	0.11	2.58	0.11	2.33	0.10
R sup frontal G	67.24 (8.44)	68.28 (7.30)	68.99 (8.90)	68.83 (8.55)	67.33 (9.01)	65.46 (9.94)	0.78	0.46	1.40	0.24	1.38	0.26
L orbitofrontal G	41.89 (5.55)	41.48 (4.78)	41.81 (6.19)	42.22 (4.50)	42.44 (4.83)	40.77 (5.57)	0.21	0.81	0.56	0.45	0.37	0.69
R orbitofrontal G	40.76 (6.13)	41.10 (5.19)	41.35 (6.08)	41.22 (5.16)	42.13 (4.73)	40.50 (5.17)	0.01	0.99	0.76	0.39	0.79	0.46
L frontal pole	2.38 (0.41)	2.38 (0.40)	2.39 (0.40)	2.43 (0.43)	2.41 (0.50)	2.21 (0.36)	0.43	0.65	0.38	0.54	2.34	0.10
R frontal pole	3.20 (0.54)	3.23 (0.52)	3.35 (0.47)	3.33 (0.48)	3.11 (0.45)	3.22 (0.64)	2.28	0.11	0.51	0.48	0.29	0.75
L ant cingulate	13.31 (2.72)	13.59 (2.74)	14.32 (3.14)	13.98 (3.01)	13.90 (2.07)	13.55 (3.02)	1.05	0.35	0.06	0.81	1.09	0.34
R ant cingulate	13.47 (2.75)	13.86 (2.46)	13.17 (2.57)	12.81 (2.79)	12.72 (2.04)	12.57 (2.26)	3.17	0.04	0.14	0.71	1.49	0.23
L precentral G	45.86 (4.70)	46.67 (3.99)	47.51 (5.68)	47.04 (3.94)	46.32 (4.31)	44.52 (5.16)	1.55	0.22	1.46	0.23	3.51	0.03
R precentral G	46.69 (5.21)	47.53 (4.51)	46.91 (5.86)	46.49 (4.32)	46.55 (5.44)	44.67 (4.96)	1.13	0.33	0.81	0.37	2.33	0.10
L fusiform G	31.87 (4.73)	32.91 (4.87)	32.53 (4.33)	32.28 (3.66)	32.25 (3.56)	30.26 (5.67)	0.87	0.42	1.01	0.32	3.38	0.04
R fusiform G	30.11 (4.21)	30.55 (4.44)	31.00 (4.79)	31.12 (4.68)	30.76 (3.94)	30.30 (4.38)	0.44	0.64	0.12	0.73	0.09	0.92
L sup temporal G	35.99 (4.60)	36.42 (4.30)	36.85 (4.42)	36.67 (3.77)	35.84 (4.79)	35.02 (5.43)	0.99	0.37	1.35	0.25	0.21	0.81
R sup temporal G	34.10 (3.72)	34.80 (2.89)	34.47 (4.15)	34.11 (3.69)	32.91 (4.63)	32.14 (4.83)	3.13	0.05	0.52	0.47	1.52	0.22
L parahippoc G	6.86 (0.99)	6.71 (0.80)	7.01 (0.99)	6.99 (1.46)	6.68 (0.99)	6.54 (0.92)	1.79	0.17	1.32	0.25	0.42	0.66
R parahippoc G	6.40 (0.81)	6.50 (0.79)	6.44 (0.92)	6.28 (0.89)	6.58 (0.93)	6.49 (0.94)	0.41	0.67	0.30	0.59	1.27	0.29

Cortical surface area measures are provided in cm^2. Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

Table 5.7 Longitudinal analysis of cortical surface area in unmedicated HR-MDD subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Group effect		Time effect		GroupXTime	
Region (Gyrus)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	P
L inf frontal G	35.24 (4.37)	35.43 (4.43)	35.76 (5.27)	35.34 (5.02)	35.51 (2.47)	35.29 (3.83)	0.21	0.81	1.47	0.23	0.32	0.72
R inf frontal G	35.71 (4.73)	36.31 (4.63)	36.32 (5.38)	35.92 (5.17)	36.44 (4.45)	35.08 (4.60)	0.02	0.99	3.37	0.07	1.94	0.15
L mid frontal G	76.04 (12.12)	77.38 (10.86)	77.75 (10.95)	77.00 (9.55)	76.38 (9.26)	73.35 (9.29)	0.52	0.60	2.61	0.11	1.15	0.32
R mid frontal G	75.54 (11.58)	76.77 (10.51)	77.72 (11.59)	77.23 (11.29)	75.83 (10.35)	72.25 (10.83)	1.00	0.37	1.05	0.31	1.08	0.34
L sup frontal G	69.41 (8.49)	70.77 (7.60)	71.12 (8.66)	70.49 (7.06)	69.07 (7.07)	66.41 (7.99)	0.90	0.41	3.07	0.08	2.50	0.09
R sup frontal G	67.37 (8.40)	68.28 (7.30)	68.62 (9.22)	68.26 (8.41)	67.17 (8.97)	65.80 (10.09)	0.34	0.71	1.29	0.26	1.24	0.29
L orbitofrontal G	41.60 (5.55)	41.48 (4.78)	41.74 (6.13)	42.06 (5.07)	42.45 (4.72)	41.75 (4.53)	0.17	0.84	0.48	0.49	0.14	0.87
R orbitofrontal G	40.86 (6.12)	41.10 (5.19)	41.20 (6.00)	41.25 (5.02)	42.09 (4.70)	40.88 (5.03)	0.05	0.95	0.75	0.39	0.68	0.51
L frontal pole	2.38 (0.40)	2.38 (0.40)	2.40 (0.38)	2.42 (0.41)	2.45 (0.51)	2.27 (0.33)	0.09	0.92	0.44	0.51	1.82	0.17
R frontal pole	3.20 (0.55)	3.23 (0.52)	3.29 (0.51)	3.24 (0.51)	3.14 (0.43)	3.29 (0.68)	0.36	0.70	0.14	0.71	0.15	0.86
L ant cingulate	13.40 (2.76)	13.59 (2.74)	14.31 (3.19)	13.88 (2.91)	13.79 (2.02)	13.61 (3.09)	0.81	0.45	0.17	0.68	0.99	0.38
R ant cingulate	13.54 (2.76)	13.86 (2.46)	13.24 (2.64)	12.86 (2.66)	12.64 (2.00)	13.07 (2.30)	2.10	0.13	0.60	0.44	1.38	0.26
L precentral G	45.87 (4.65)	46.67 (3.99)	47.12 (5.82)	46.77 (4.09)	46.33 (4.50)	44.36 (5.19)	1.22	0.30	1.46	0.23	3.45	0.04
R precentral G	46.73 (5.18)	47.53 (4.51)	46.62 (5.94)	46.23 (4.48)	46.35 (5.54)	45.04 (5.03)	1.14	0.32	0.53	0.47	1.88	0.16
L fusiform G	31.95 (4.71)	32.91 (4.87)	32.46 (4.31)	32.02 (3.68)	32.22 (3.54)	30.53 (5.15)	0.66	0.52	0.93	0.34	2.96	0.06
R fusiform G	30.19 (4.24)	30.55 (4.44)	30.94 (4.65)	30.72 (4.60)	30.81 (3.77)	30.59 (4.85)	0.17	0.84	0.04	0.85	0.11	0.89
L sup temporal G	36.05 (4.59)	36.42 (4.30)	36.79 (4.31)	36.70 (3.89)	36.00 (4.55)	35.44 (4.53)	0.28	0.76	1.74	0.19	0.40	0.67
R sup temporal G	34.11 (3.68)	34.80 (2.89)	34.60 (4.12)	34.19 (3.56)	33.06 (4.54)	32.54 (4.71)	1.57	0.21	0.79	0.38	1.73	0.18
L parahippoc G	6.88 (0.99)	6.71 (0.80)	7.00 (0.97)	7.04 (1.40)	6.79 (1.00)	6.58 (0.97)	1.46	0.23	1.23	0.27	0.83	0.44
R parahippoc G	6.41 (0.81)	6.50 (0.79)	6.47 (0.93)	6.33 (0.86)	6.58 (0.88)	6.51 (0.98)	0.72	0.49	0.78	0.38	1.35	0.26

Cortical surface area measures are provided in cm^2. Abbreviations: inf = inferior; mid = middle; sup = superior; ant = anterior; G = gyrus; parahippoc = parahippocampal.

5.4 Discussion

The baseline and longitudinal analyses of surface area yielded no significant results. Nevertheless, several nominal significant findings were observed that will be discussed here. Given that this is, to the best of my knowledge, the first study examining cortical surface area in high-risk of mood disorders subjects and MDD patients, the discussion hampers comparison with previous research findings.

The baseline analysis yielded no significant surface area differences in the ROIs between the groups. This finding suggests that young adults at high risk of mood disorders because of a close family history of BD do not exhibit cortical surface area abnormalities, regardless of a subsequent onset of MDD or not. These results are in line with several studies which did not find any frontal or temporal grey matter abnormalities in unaffected BD relatives (Kempton, et al., 2009; Kieseppa, et al., 2003; McIntosh, et al., 2004; McIntosh, et al., 2006; van der Schot, et al., 2009). They are in contrast, however, to two studies that did detect an association between genetic liability to BD and decreased grey matter volume in the medial/dorsolateral frontal gyrus and precentral gyrus (van der Schot, et al., 2010) and the medial frontal gyrus and anterior cingulate cortex (McDonald, Bullmore, et al., 2004). Furthermore, they are opposed to findings in unaffected young relatives of BD patients of increased right inferior frontal gyri (Hajek, et al., 2013) and increased parahippocampal volumes (Ladouceur, et al., 2008). All in all, our findings suggest that vulnerability to BD in young adolescents is not associated with surface area abnormalities and that observed grey matter or volumetric differences in previous studies may be related to cortical thickness rather than cortical surface abnormalities.

A significant correlation at baseline between the left frontal pole surface area and the HAM-D sum score that passed FDR-correction was found in the HR-MDD group, indicating that the surface area of this brain region increases with higher depression symptom severity in this study group. Given that this relationship was not detected in the HR-well group and that the two high-risk groups did not differ

with respect to their baseline depression symptoms as measured with the HAM-D, it appears unlikely that the left frontal pole increases in response to higher depression symptoms or vice versa in individuals at high familial risk per se. Similarly, no significant correlation for this structure was observed when examining if changes in surface area are linked to changes in depression symptom severity over time. Accordingly, the nature of this finding remains debatable.

The longitudinal analysis also did not reveal any significant effect of group, time or group-by-time interaction. These non-significant findings suggest that the often observed frontal and temporal grey matter reductions in MDD patients do not emerge as a function of abnormal cortical surface area development that is related the onset of MDD or familial risk. Rather, it appears likely that the frequently observed cortical brain abnormalities in mood disorders patients only emerge as a consequence of cortical thickness pathology or that they only develop with increasing length of illness, age, severity of symptoms or as a consequence of medication effects.

Several nominal significant results were found for the longitudinal analysis that did not survive FDR-correction. These were nominal group effects for the right superior temporal gyrus and the right anterior cingulate, with post-hoc tests suggesting smaller surface area across time in the superior temporal gyrus in the HR-MDD group as compared to the other study groups and no significant pair-wise differences in the anterior cingulate. However, these findings did not remain nominally significant when excluding medicated HR-MDD subjects from the analysis, indicating that psychopharmacological treatment effects may have influenced surface area decreases in these brain regions in the HR-MDD group. Furthermore, nominal significant group-by-time interactions were observed for the left fusiform and left precentral gyrus, indicating abnormal surface area decreases over time in the HR-MDD group as compared to HC subjects. However, only the group-by-time interaction for the left precentral gyrus remained significant after excluding medicated HR-MDD subjects from the analysis, indicating that the

abnormal left fusiform surface development over time in the HR-MDD may have been related to the effects of psychopharmacological treatment. Future studies are required to examine whether surface area in this region is influenced by medication. The group-by-time interaction for the left precentral gyrus that remained nominally significant after excluding medicated HR-MDD subjects gives a first line of evidence to suggest that abnormal surface area decreases in this brain region may be linked to an onset of MDD and reflect illness-associated effects. Given that there was no significant correlation between changes in cortical surface area in this brain region and changes in the HAM-D sum score, it is unlikely that left precentral surface area decreases are linked to depression symptom severity. More longitudinal studies with larger sample sizes are clearly needed to investigate if this nominal significant finding is deemed correct.

The strengths of this study are its longitudinal nature, the assessment of subjects prior to illness onset, the relatively young age of the participants and the comparatively large sample size of high-risk subjects and controls. In addition, all subjects underwent careful clinical assessment at both time points and the effects of medication and relatedness of subjects were ruled out. All brain scans were obtained at the same scanner using the identical protocol at both visits and the MRI data were processed in an identical way using thoroughly validated methods.

Nevertheless, some limitations need to be addressed. First, it remains unknown whether currently unaffected HR-well subjects may develop MDD in the future. Second, previous longitudinal studies have reported that the majority of the high-risk subjects who developed BD themselves experienced depressive episodes years before conversion (Duffy, 2010; Hillegers, et al., 2005) so that it appears likely that some of our HR-MDD subjects may develop BD in the future. The follow-up assessments of our study cohort will clarify if some of the HR-MDD participants will convert to BD and if some of our HR-well subjects go on to develop a mood disorder. Third, our study groups differed with respect to depression symptom

severity at baseline. However, the median of the HAM-D total score was only 1 in the HR-MDD group, suggesting only sub-syndromal depression symptoms.

In summary, our findings suggest that individuals at high familial risk of mood disorders do not exhibit surface area abnormalities, regardless of an onset of MDD or not. Moreover, the frequently observed frontal and temporal grey matter decreases in MDD patients do not appear to be driven by abnormal cortical surface area development in relation to illness onset or familial high risk in young adults. However, a nominal significant interaction effect for the left precentral gyrus provides partial support for abnormal surface area decreases over time as high-risk subjects get ill, but this findings needs to be further examined in future studies.

Chapter 6

Neurocognitive performance in individuals at high familial risk of mood disorders

6.1 Introduction

Close relatives of BD patients are at enhanced risk of developing BD or MDD during their lifetime as outlined in more detail in Chapter 1. There is strong support for an overlap in the causal pathways of both disorders coming from familial aggregation and genetic association studies (Barnett & Smoller, 2009; Craddock, 2006; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; McGuffin, et al., 2003; Schulze, et al., 2012; Sullivan, et al., 2000).

A broad range of cognitive deficits have been observed in mood disorders patients, with neuropsychological performance impairments being less widespread in unaffected BD or MDD relatives (see paragraph 1.6). Amongst others, first-episode MDD patients have shown performance deficits during tasks involving attentional processing speed such as the Digit Span forwards and cognitive flexibility as measured with the WCST or IED (Lee, et al., 2012). Similarly, meta-analyses of first-episode or euthymic BD patients have found impairments during attentional processing speed and cognitive flexibility (Arts, et al., 2008; Bourne, et al., 2013). Moreover, various verbal learning and memory performance measures of the CVLT have been found to be impaired (Arts, et al., 2008; Bourne, et al., 2013) in both of these patient populations, with working memory impairments as measured with the Digit Span backwards being only evident in the euthymic BD subgroup (Arts, et al., 2008).

First-degree relatives of BD patients have also been shown to have deficits during tasks involving attentional processing (Glahn, et al., 2010), verbal learning and memory (Arts, et al., 2008; Balanza-Martinez et al., 2008; Bora, Yucel, & Pantelis,

2009), cognitive flexibility (Bora, et al., 2009) and working memory (Balanza-Martinez, et al., 2008; Glahn, et al., 2010). It should however be noted that the magnitude and consistency of these findings is generally reduced in comparison to affected individuals. Further evidence for the importance of attention, verbal memory and working memory deficits in the aetiology of mood disorders comes from a study by Glahn et al. (2012). Using a large sample of randomly-selected pedigrees, the authors showed that neuropsychological impairments in these domains may potentially serve as candidate neurocognitive endophenotypes for MDD.

To the best of my knowledge, all neuropsychological studies on BD and MDD thus far have assessed neurocognition in affected individuals or their unaffected relatives only. These studies can't discern whether cognitive deficits reflect early neurodevelopmental disruptions predisposing to illness, events linked to the illness onset, or whether they are adaptive or secondary to the effects of chronic illness or its treatment. Neuropsychological examinations of individuals at high risk of mood disorders because of a close family history of BD hold the potential to identify neurocognitive deficits related to enhanced familial vulnerability, unconfounded by the presence of illness. However, to the best of my knowledge, no study has yet examined the effects of familial risk on neurocognition in a prospective study.

The longitudinal Scottish Bipolar Family Study is well-designed to examine the timing of neuropsychological impairment in mood disorders and their relationship to familial risk and onset of illness. In the analyses presented here, neuropsychological measures of attentional processing speed, verbal learning and memory, cognitive flexibility and working memory were compared cross-sectionally for data acquired at baseline assessment and longitudinally over time between the HR-MDD, HR-well and HC groups. The rationale for the baseline analysis was to examine whether specific neuropsychological deficits exists in the HR-MDD group prior to illness onset that may serve as neurocognitive marker for a subsequent development of MDD and to assess their relationship to familial risk.

The longitudinal analysis was intended to explore the time course of neuropsychological performance during a period of two years to detect whether a decline in neuropsychological function in the high-risk participants is evident that can be linked to an onset of MDD or familial risk. Similar to a repeated measures ANOVA, the longitudinal analysis also allowed for testing with an increased power as compared to baseline analysis whether significant performance differences between the study groups exist across both time points.

6.2 Methods

6.2.1 Participants and clinical assessment

Participants were recruited as part of the Scottish Bipolar Family Study as outlined more detailed in Chapter 2. The HC, HR-well and HR-MDD groups were matched for age, gender, handedness and intelligence and detailed demographic and clinical descriptions of the study sample are provided in Table 3.1 in Chapter 3.

6.2.2 Neuropsychological assessment

At baseline and follow-up assessment, the Digit Span forwards and backwards, CVLT and IED were administered to assess attentional processing speed, working memory, verbal learning, verbal memory and cognitive flexibility. Detailed information about the neuropsychological tests is provided in Chapter 2.3. The following performance measures were extracted for analysis (see Table 6.1): For the Digit Span, the number of correctly recalled strings of numbers during the forwards condition (Digit Span forwards) and the number of correctly recalled strings of numbers in reversed order during the backwards condition (Digit Span backwards) were calculated as measures of attentional processing speed and working memory, respectively. For the CVLT, the number of words correctly recalled during trials 1-5 (CVLT learning) as an indice of verbal learning ability

was computed as well as the number of words recalled during the free short delay recall (CVLT short delay) and free long delay recall (CVLT long delay) as estimates of verbal memory. For the IED, the number of trials needed to complete task stage 1 (IED SDL) and the number of trials needed to complete task stages 2, 5 and 7 (IED RL) were calculated to assess simple discrimination learning and reversal learning performance, respectively. Furthermore, the number of trials needed to complete task stage 6 (IED IDS) and the number of trials needed to complete task stage 8 (IED EDS) were computed to extract intradimensional set-shifting and extradimensional set-shifting ability, respectively.

Table 6.1 Neuropsychological performance parameters

Task parameter	Measure	Neuropsychological domain
Digit Span forwards	Number of correctly recalled strings of numbers	Attentional processing speed
Digit Span backwards	the number of correctly recalled strings of numbers in reversed order	Working memory
CVLT learning	number of words correctly recalled during trials 1-5	Verbal learning
CVLT short delay	number of words correctly recalled during free short delay recall	Verbal memory
CVLT long delay	number of words correctly recalled during free long delay recall	Verbal memory
IED SDL	Number of trials needed to complete stage 1	Simple discrimination learning
IED RL	Number of trials needed to complete stages 2, 5, 7	Reversal learning
IED IDS	Number of trials needed to complete stage 6	Intradimensional set-shifting
IED EDS	Number of trials needed to complete stage 8	Extradimensional set-shifting

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting.

6.2.3 Statistical analysis

An ANCOVA was conducted to compare cross-sectionally neuropsychological performance measures between the HC, HR-well and HR-MDD groups at baseline. Neurocognitive performance measures that did not conform to assumptions of normality and/or homogeneity of variance were transformed using the Box-Cox procedure (Box & Cox, 1964) which automatically selects transformations that maximise the approximation of the transformed data to a normal distribution using a likelihood function. Age and sex served as covariates in this analysis. Next, linear mixed-effects models were applied to investigate neuropsychological performance over time (see Chapter 2 for a more detailed methodological description). In the linear mixed-effects model used, the intercept term is treated as a random effect that varies by individual so that intraindividual correlation among the neurocognitive measures of a particular individual is taken into account. The following independent variables were used as predictors of neurocognitive parameters outlined: group, time (baseline versus follow-up assessment), group-by-time interaction. Age and sex served as covariates in the analyses.

A statistical significance level of $p \leq 0.05$ was chosen, fully corrected for multiple comparisons using the Benjamini & Hochberg FDR procedure (Benjamini & Hochberg, 1995). To allow for comparison with previous studies, effect sizes for nominal significant group differences were additionally calculated using Cohen's *d* (Cohen, 1988). Wherever significant between-group differences were found, pairwise comparisons were performed between the three groups, for which p-values were corrected according to Tukey's HSD procedure ($p_{HSD} \leq 0.05$).

To assess the relationship between depression symptoms and neuropsychological performance of each measure, Spearman's rank correlation coefficients between the HAM-D scores and performance measures for each group were calculated. To assess the relationship between changes in performance and changes in depressive symptoms over time, Spearman's rank correlation coefficients were calculated between the performance differences of the neurocognitive measures and the

differences of the HAM-D scores between the two assessments. For ease of interpretation, data derived from the follow-up assessment was subtracted from the data acquired at baseline so that positive values reflect increases in performance or depression symptoms over time, while negative values represent decreases of these measures over time. All p-values were corrected according to Benjamini & Hochberg FDR procedure and considered significant when $p \leq 0.05$.

To examine the potentially confounding effects of exposure to medication and relatedness of subjects on neuropsychological performance, the following additional analyses were performed: The analyses were repeated excluding medicated HR-MDD subjects, followed by randomly excluding related subjects.

6.3 Results

6.3.1 Cross-sectional analysis

6.3.1.1 Group differences at baseline

At baseline assessment, no significant FDR-corrected differences in neuropsychological performance were found between the groups (see Table 6.2). There was a nominal significant group difference for extradimensional set shifting ($p \leq 0.021$). Post-hoc tests indicated that the HR-MDD group ($p_{HSD} \leq 0.019$, d = 0.655) and the HR-well group ($p_{HSD} \leq 0.038$, d = 0.201) needed more trials to successfully complete the extradimensional set shifting stage of the IED as compared to HC participants, with no performance differences between the high-risk groups ($p_{HSD} \leq 0.232$, d = 0.459).

6.3.1.2 Correlation analysis

For the neurocognitive performance measures, there were no significant correlations with depression symptom severity that remained significant after FDR

correction. There was one nominal significant correlation. The free long delay recall of the CVLT was negatively correlated with the HAM-D sum score in the HC sample ($p \leq 0.028$), indicating that poorer long delay memory performance is associated with higher depression symptoms (see Table 6.3).

6.3.1.3 Analysis of potential confounders

To eliminate the potential confounding effects of familial relatedness of some subjects, the analyses were repeated, excluding randomly individuals from the same pedigree (Table 6.4). This analysis similarly yielded no significant FDR-corrected group differences. The nominal significant group effect for extradimensional set-shifting ($p \leq 0.028$) remained nominally significant.

Table 6.2 Cross-sectional analysis of neurocognitive performance

	HC	HR-well	HR-MDD	Statistics	
Task parameter	Mean (SD)	Mean (SD)	Mean (SD)	F	p
Digit Span forwards	6.77 (1.28)	6.87 (1.21)	6.68 (1.53)	0.31	0.74
Digit Span backwards	5.18 (1.24)	4.80 (1.14)	5.22 (1.22)	2.47	0.09
CVLT learning	54.55 (10.87)	54.35 (9.45)	50.47 (17.72)	0.02	0.98
CVLT short delay	11.94 (2.74)	11.35 (2.78)	10.18 (4.76)	0.97	0.38
CVLT long delay	12.73 (2.67)	11.90 (2.88)	11.18 (4.75)	1.50	0.23
IED SDL	6.61 (0.80)	7.00 (1.87)	6.75 (0.86)	1.75	0.18
IED RL	22.29 (4.47)	21.97 (2.71)	24.63 (8.78)	0.46	0.63
IED IDS	6.50 (1.33)	6.50 (1.03)	6.50 (0.89)	0.23	0.79
IED EDS	24.53 (18.03)	28.07 (17.74)	37.50 (17.49)	3.93	0.02

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting.

Table 6.3 Correlation between neurocognitive performance measures and HAM-D sum scores

	HC		HR-well		HR-MDD	
Task parameter	R	p	R	p	R	p
Digit Span forwards	0.14	0.23	0.07	0.55	-0.46	0.06
Digit Span backwards	-0.04	0.71	-0.04	0.73	-0.42	0.11
CVLT learning	-0.21	0.06	-0.10	0.39	0.11	0.71
CVLT short delay	-0.16	0.14	-0.06	0.58	0.13	0.67
CVLT long delay	-0.24	0.03	-0.10	0.39	0.10	0.74
IED SDL	-0.13	0.27	0.10	0.37	0.13	0.65
IED RL	-0.01	0.95	-0.15	0.18	-0.19	0.52
IED IDS	-0.01	0.99	-0.11	0.34	-0.39	0.16
IED EDS	0.10	0.39	-0.01	0.93	0.03	0.93

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting.

Table 6.4 Cross-sectional analysis of neurocognitive performance in unrelated subjects

	HC	HR-well	HR-MDD	Statistics	
Task parameter	Mean (SD)	Mean (SD)	Mean (SD)	F	p
Digit Span forwards	6.75 (1.27)	6.81 (1.17)	6.95 (1.55)	0.01	0.99
Digit Span backwards	5.16 (1.24)	4.83 (1.16)	5.33 (1.19)	1.98	0.14
CVLT learning	54.44 (10.89)	54.14 (9.93)	48.65 (17.62)	0.08	0.93
CVLT short delay	11.97 (2.77)	11.34 (2.79)	9.82 (4.52)	1.44	0.24
CVLT long delay	12.73 (2.70)	11.81 (2.90)	10.76 (4.56)	1.84	0.16
IED SDL	6.60 (0.81)	7.06 (2.01)	6.63 (0.89)	1.88	0.16
IED RL	22.27 (4.51)	22.06 (2.86)	24.63 (8.78)	0.54	0.59
IED IDS	6.51 (1.35)	6.48 (1.05)	6.63 (0.96)	0.10	0.91
IED EDS	24.37 (17.97)	27.98 (17.90)	36.38 (18.68)	3.45	0.03

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting.

6.3.2 Longitudinal analysis

6.3.2.1 Group differences over time

Table 6.5 provides the results of the linear mixed-effects model analyses. A significant group effect was found for the long delay free recall ($p \leq 0.003$) and extradimensional set-shifting ($p \leq 0.004$) that passed the FDR procedure (see Figure 6.1). Post-hoc analyses revealed that HC subjects recall significantly more words over both assessment time points during the long delay free recall than HR-well participants ($p_{HSD} \leq 0.002$, $d = 0.482$), with no significant differences between the HC and HR-MDD group ($p_{HSD} \leq 0.136$, $d = 0.461$) or the HR-well and HR-MDD group ($p_{HSD} \leq 0.485$, $d = 0.041$). For the extradimensional set-shifting performance, post-hoc analyses showed that HC subjects have a significantly superior task performance across both time points as compared to HR-well ($p_{HSD} \leq 0.031$, $d = 0.231$) and HR-MDD subjects ($p_{HSD} \leq 0.004$, $d = 0.849$), with no significant difference between the two high-risk groups ($p_{HSD} \leq 0.077$, $d = 0.571$).

Moreover, a significant effect of time was observed for the Digit Span forwards ($p \leq 0.009$), verbal learning ($p \leq 0.003$), free short delay recall ($p \leq 0.006$), simple discrimination learning ($p \leq 0.003$) and extradimensional set-shifting ($p \leq 0.001$). For all neurocognitive measures except for simple discrimination learning, the effect was driven by all participant groups displaying enhanced performance during the follow-up assessment as compared to the baseline assessment. By contrast, the time effect for simple discrimination learning ability was in the opposite direction, with all participant groups showing worse performance at follow-up appointment as compared to baseline appointment.

The analyses revealed no significant group-by-time interactions. There was one nominally significant group effect for the free short delay recall ($p \leq 0.018$) that did not pass FDR correction. Post-hoc analyses showed that the HC group performed better during this task as compared to the HR-well group ($p_{HSD} \leq 0.010$, $d = 0.412$),

with no differences between HC and HR-MDD subjects ($p_{HSD} \leq 0.195$, $d = 0.364$) or between the two high-risk groups ($p_{HSD} \leq 0.622$, $d = 0.062$).

6.3.2.2 Correlation analysis

There were no significant correlations between changes in neurocognitive performance measures and changes in depressive symptom severity (see Table 6.6). There was a nominal significant negative correlation between changes in the Digit Span forwards performance and changes in the HAM-D sum score in the HC group ($p \leq 0.045$), indicating that decreasing task performance in HC subjects is associated with increasing depression symptom severity. By contrast, there was a nominal significant positive correlation between changes in the Digit Span forwards performance and changes in the HAM-D sum score in the HR-well group ($p \leq 0.031$), indicating that increasing task performance is associated with increasing depression symptom severity. As highlighted before, these results were only nominally significant and did not hold when applying FDR corrections.

6.3.2.3 Analysis of potential confounders

All results remained significant after FDR correction when randomly excluding related subjects. A significant group effect was found for the long delay free recall ($p \leq 0.003$) and extradimensional set-shifting ($p \leq 0.003$) that passed the FDR procedure. Of note, the nominally significant group effect for the free short delay recall that did not pass FDR correction in the original analysis reached FDR-corrected significance when analysing unrelated subjects only ($p \leq 0.015$). Moreover, a significant effect of time was observed as in the original analysis for the Digit Span forwards ($p \leq 0.012$), verbal learning ($p \leq 0.005$), free short delay recall ($p \leq 0.012$), simple discrimination learning ($p \leq 0.004$) and extradimensional set-shifting ($p \leq 0.001$).

When excluding medicated HR-MDD subjects from the analysis, the significant group effect for the long delay free recall ($p \leq 0.003$) remained significant after FDR correction. The group effect for extradimensional set-shifting ($p \leq 0.035$) remained significant at nominal level only. The significant effects of time for simple discrimination learning ($p \leq 0.003$) and extradimensional set-shifting ($p \leq 0.001$) remained significant after the FDR procedure. The observed time effects for the Digit Span forwards ($p \leq 0.019$), verbal learning ($p \leq 0.029$) and free short delay recall ($p \leq 0.045$) remained significant at a nominal level only.

Table 6.5 Longitudinal analysis of neurocognitive performance

	HC		HR-well		HR-MDD		Statistics					
	Baseline	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Group effect**		**Time effect**		**GroupXTime**	
Region	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
Digit Span forwards	6.77 (1.28)	7.50 (1.48)	6.87 (1.21)	6.98 (1.34)	6.68 (1.53)	7.10 (1.41)	0.36	0.70	7.02	**0.01**	1.35	0.26
Digit Span backwards	5.18 (1.24)	5.46 (1.36)	4.80 (1.14)	5.00 (1.11)	5.22 (1.22)	5.29 (1.71)	2.57	0.08	0.78	0.38	0.20	0.82
CVLT learning	54.55 (10.87)	60.48 (9.47)	54.35 (9.45)	56.57 (10.34)	50.47 (17.72)	56.23 (10.79)	1.63	0.20	9.32	**<0.01**	2.05	0.13
CVLT short delay	11.94 (2.74)	13.18 (2.70)	11.35 (2.78)	11.75 (3.00)	10.18 (4.76)	12.00 (3.18)	4.09	0.02	7.77	**0.01**	1.38	0.25
CVLT long delay	12.73 (2.67)	13.73 (2.59)	11.90 (2.88)	12.05 (2.77)	11.18 (4.75)	12.27 (3.04)	6.11	**<0.01**	1.74	0.19	1.47	0.23
IED SDL	6.61 (0.80)	6.98 (1.60)	7.00 (1.87)	7.31 (2.37)	6.75 (0.86)	7.30 (1.53)	0.54	0.58	8.89	**<0.01**	1.28	0.28
IED RL	22.29 (4.47)	22.30 (2.15)	21.97 (2.71)	24.36 (11.95)	24.63 (8.78)	22.70 (2.58)	0.61	0.54	3.59	0.06	0.21	0.81
IED IDS	6.50 (1.33)	6.83 (1.67)	6.50 (1.03)	7.10 (2.79)	6.50 (0.89)	6.25 (0.55)	0.76	0.47	0.80	0.37	1.64	0.20
IED EDS	24.53 (18.03)	17.25 (15.42)	28.07 (17.74)	22.07 (17.47)	37.50 (17.49)	29.20 (19.88)	5.80	**<0.01**	19.23	**<0.01**	0.22	0.80

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting; **bold** indicates significant effect after FDR correction.

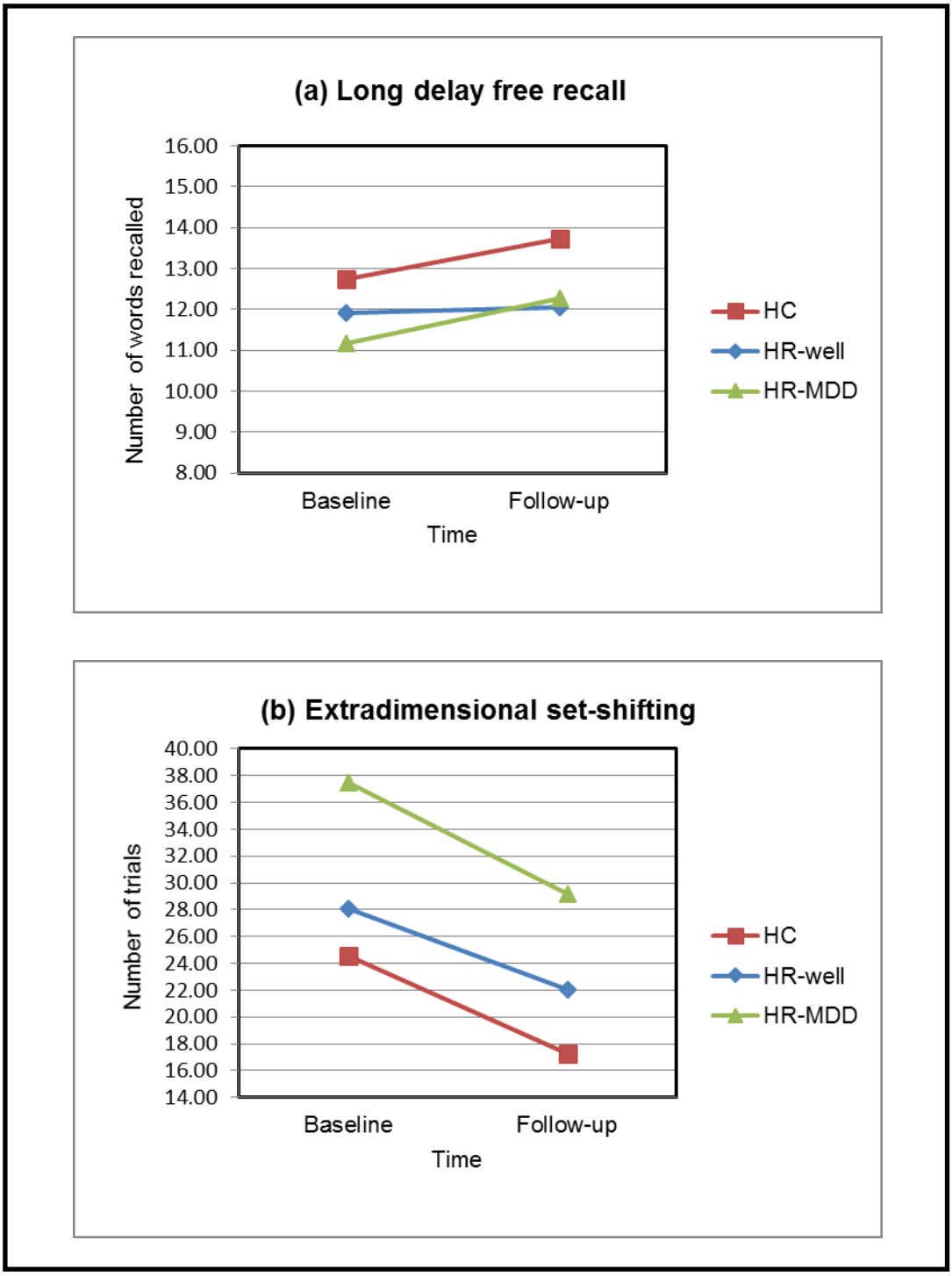

Figure 6.1 Significant group effects (a, b) of longitudinal neurocognitive performance analyses.

Table 6.6 Correlation between changes in neurocognitive performance and changes in HAM-D sum scores

	HC		HR-well		HR-MDD	
Region	R	p	R	p	R	p
Digit Span forwards	-0.32	0.05	0.34	0.03	0.03	0.94
Digit Span backwards	-0.01	0.98	-0.23	0.16	-0.03	0.94
CVLT learning	-0.09	0.54	-0.21	0.16	0.35	0.32
CVLT short delay	-0.15	0.31	-0.05	0.76	0.15	0.67
CVLT long delay	-0.06	0.70	0.03	0.84	0.02	0.96
IED SDL	0.29	0.08	-0.23	0.14	0.11	0.77
IED RL	0.04	0.83	0.01	0.95	-0.39	0.27
IED IDS	0.22	0.19	-0.12	0.45	-0.22	0.55
IED EDS	0.06	0.72	-0.29	0.07	0.49	0.15

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting.

Table 6.7 Longitudinal analysis of neurocognitive performance in unrelated subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	**Follow-up**	**Baseline**	**Follow-up**	**Baseline**	**Follow-up**	**Group effect**		**Time effect**		**GroupXTime**	
Region	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
Digit Span forwards	6.75 (1.27)	7.51 (1.46)	6.81 (1.17)	6.87 (1.36)	6.95 (1.55)	7.15 (1.42)	0.54	0.58	6.52	**0.01**	1.41	0.25
Digit Span backwards	5.16 (1.24)	5.49 (1.36)	4.83 (1.16)	4.98 (1.13)	5.33 (1.19)	5.30 (1.75)	1.94	0.15	0.55	0.46	0.16	0.85
CVLT learning	54.44 (10.89)	60.35 (9.44)	54.14 (9.93)	56.31 (10.79)	48.65 (17.62)	55.62 (10.67)	1.71	0.18	8.09	**<0.01**	1.95	0.15
CVLT short delay	11.97 (2.77)	13.14 (2.70)	11.34 (2.79)	11.73 (3.06)	9.82 (4.52)	11.81 (3.12)	4.31	**0.02**	6.54	**0.01**	1.25	0.29
CVLT long delay	12.73 (2.70)	13.66 (2.62)	11.81 (2.90)	12.04 (2.89)	10.76 (4.56)	12.10 (3.00)	5.99	**<0.01**	1.87	0.17	0.99	0.37
IED SDL	6.60 (0.81)	6.96 (1.59)	7.06 (2.01)	7.40 (2.57)	6.63 (0.89)	7.32 (1.57)	0.59	0.55	8.39	**<0.01**	1.13	0.33
IED RL	22.27 (4.51)	22.28 (2.14)	22.06 (2.86)	24.77 (13.07)	24.63 (8.78)	22.32 (1.97)	0.41	0.66	2.36	0.13	0.34	0.71
IED IDS	6.51 (1.35)	6.82 (1.66)	6.48 (1.05)	6.75 (1.54)	6.63 (0.96)	6.21 (0.54)	0.81	0.45	0.30	0.58	2.08	0.13
IED EDS	24.37 (17.97)	17.85 (15.91)	27.98 (17.90)	21.06 (16.90)	36.38 (18.68)	30.21 (19.89)	6.05	**<0.01**	16.47	**<0.01**	0.08	0.92

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting; **bold** indicates significant effect after FDR correction.

Table 6.8 Longitudinal analysis of neurocognitive performance in unmedicated HR-MDD subjects

	HC		HR-well		HR-MDD		Statistics					
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Group effect		Time effect		GroupXTime	
Region	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F	p	F	p	F	p
Digit Span forwards	6.77 (1.28)	7.50 (1.48)	6.87 (1.21)	6.98 (1.34)	6.68 (1.53)	7.12 (1.29)	0.38	0.69	5.64	0.02	1.45	0.24
Digit Span backwards	5.18 (1.24)	5.46 (1.36)	4.80 (1.14)	5.00 (1.11)	5.22 (1.22)	5.47 (1.66)	3.03	0.05	1.49	0.23	0.05	0.95
CVLT learning	54.55 (10.87)	60.48 (9.47)	54.35 (9.45)	56.57 (10.34)	50.47 (17.72)	55.11 (11.03)	1.55	0.21	4.88	0.03	2.85	0.06
CVLT short delay	11.94 (2.74)	13.18 (2.70)	11.35 (2.78)	11.75 (3.00)	10.18 (4.76)	11.67 (3.25)	4.01	0.02	4.10	0.05	1.33	0.27
CVLT long delay	12.73 (2.67)	13.73 (2.59)	11.90 (2.88)	12.05 (2.77)	11.18 (4.75)	12.11 (3.27)	5.96	**<0.01**	0.50	0.48	1.79	0.17
IED SDL	6.61 (0.80)	6.98 (1.60)	7.00 (1.87)	7.31 (2.37)	6.75 (0.86)	7.44 (1.67)	0.68	0.51	9.28	**<0.01**	1.43	0.24
IED RL	22.29 (4.47)	22.30 (2.15)	21.97 (2.71)	24.36 (11.95)	24.63 (8.78)	22.69 (2.70)	0.59	0.56	2.59	0.11	0.34	0.72
IED IDS	6.50 (1.33)	6.83 (1.67)	6.50 (1.03)	7.10 (2.79)	6.50 (0.89)	6.25 (0.58)	0.35	0.70	0.20	0.65	2.36	0.10
IED EDS	24.53 (18.03)	17.25 (15.42)	28.07 (17.74)	22.07 (17.47)	37.50 (17.49)	26.06 (19.84)	3.41	0.04	19.05	**<0.01**	0.32	0.73

Abbreviations: CVLT = California Verbal Learning Test; IED = Intra-/Extradimensional Set Shifting Task; SDL = simple discrimination learning; RL = reversal learning; IDS = intradimensional set-shifting; EDS = extradimensional set-shifting; **bold** indicates significant effect after FDR correction.

6.4 Discussion

This is, to the best of my knowledge, the first prospective longitudinal study examining neurocognitive performance in high-risk of mood disorders individuals who were unaffected at initial assessment and either developed MDD or remained well during the follow-up period.

Reduced long delay verbal memory and extradimensional set-shifting performance across the two time points were found in the HR-well group relative to controls, with the HR-MDD group displaying decreased extradimensional set-shifting abilities as compared to the HC group only. Moreover, significant effects of time were observed for attentional processing speed, verbal learning, short delay verbal memory and extradimensional set-shifting, indicating superior performance during these tasks at follow-up assessment as compared to baseline assessment across all participant groups. There was also a significant effect of time for simple discrimination learning due to decreased task performance at follow-up as compared to baseline across all groups.

The finding of reduced long delay verbal memory and extradimensional set-shifting performance in the HR-well group across time suggests that neurocognitive deficits in these domains constitute a familial trait marker for vulnerability to mood disorders. It cannot be determined from the data if the decreased task performance is a consequence of shared genetic and/or environmental effects. Given that they are already present in early adulthood, they are unlikely to be of degenerative origin but likely represent disturbances of normal brain development predisposing to illness. Since no significant differences between the two high-risk groups and no group-by-time interaction emerged, the results do not speak towards reduced verbal memory and extradimensional set-shifting abilities to be directly linked to an onset of MDD. It should be highlighted, however, that visual inspection of extradimensional set-shifting performance (Figure 6.1) shows that the HR-MDD group performed worse than the HR-well

group across time. Accordingly, it cannot be ruled out that the small sample size of the HR-MDD group did not allow for detection of significant effects due to a lack of power. Indeed, a meta-analyses by Bora and colleagues (2009) showed that effect sizes for cognitive flexibility in healthy relatives of BD patients were small. It appears therefore important to investigate neurocognition further in a larger sample size of high-risk subjects who go on to develop MDD.

The results of decreased verbal memory and extradimensional set-shifting in relatives of BD patients are in line with findings from recent meta-analyses (Arts, et al., 2008; Balanza-Martinez, et al., 2008; Bora, et al., 2009). However, it should be noted that the analyses presented here do not confirm previous reports of attentional processing, verbal learning and working memory to be reduced in high-risk BD subjects (Arts, et al., 2008; Balanza-Martinez, et al., 2008; Bora, et al., 2009; Glahn, et al., 2010). As outlined earlier, one potential reason for the absence of significant findings may well derive from the fact that neurocognitive deficits in relatives of BD patients are better detectable in particularly large sample sizes as effect sizes generally tend to be small (Arts, et al., 2008).

From a neuroanatomical point of view, both short-term and long-term storage and retrieval of verbal information have been linked to a bilateral frontal and parietal network of brain regions including the posterior inferior frontal, anterior middle frontal, anterior cingulate and supramarginal gyrus (Andreasen, et al., 1995; S. Dupont, et al., 2002; Henson, et al., 2000). Moreover, it has been shown that enhanced performance during the CVLT is associated with higher engagement of the right hippocampus and right frontal lobe (Johnson, et al., 2001). Dysfunction of this network of brain regions may well be in line with studies suggesting that there may be diminished prefrontal modulation of various brain regions including the anterior cingulate that results in dysregulation of mood as evident in BD (Strakowski, Delbello, & Adler, 2005).

While complex tasks involving extradimensional set-shifting undoubtedly rely on the interplay of various brain regions including lateral prefrontal, orbital, and

parietal brain areas that may serve as a supervisory attentional network, the ventrolateral prefrontal cortex in particular has been hypothesized to be functionally specialised for extradimensional set-shifting (Hampshire & Owen, 2006). Reduced extradimensional set-shifting performance may thus be in line with hypothesis of malfunction of the ventral brain system to be underlying the pathogenesis of BD which is thought to be essential for affective processing and modulation, with the ventrolateral prefrontal cortex playing a central role (Phillips, et al., 2003a, 2003b).

The finding of significant time effects for attentional processing speed, verbal learning, short delay verbal memory and extradimensional set-shifting in the direction of superior performance during follow-up as compared to baseline assessment across groups most likely reflects the effects of repeated task presentation. These practice effects during neuropsychological examination have been well documented in the literature (Bartels, Wegrzyn, Wiedl, Ackermann, & Ehrenreich, 2010). There was also a significant effect of time for simple discrimination learning due to decreased task performance at follow-up as compared to baseline assessment across all groups. One plausible explanation for this conflicting result may be that participants remembered the task from the baseline assessment and already shifted their attention to the currently irrelevant stimulus as they were expecting the reversal of the rule to occur at follow-up examination.

The strengths of this study are its longitudinal nature, the assessment of subjects prior to illness onset, the relatively young age of the participants and the comparatively large sample size of high-risk subjects and controls. In addition, all subjects underwent careful clinical assessment at both time points and the effects of medication and relatedness of subjects were ruled out.

Nevertheless, some limitations need to be addressed. First, it remains unknown whether currently unaffected HR-well subjects may develop MDD in the future. Second, previous longitudinal studies have reported that the majority of the high-

risk subjects who developed BD themselves experienced depressive episodes years before conversion (Duffy, 2010; Hillegers, et al., 2005) so that it appears likely that some of our HR-MDD subjects may develop BD in the future. The follow-up assessments of our study cohort will clarify if some of the HR-MDD participants will convert to BD and if some of our HR-well subjects go on to develop a mood disorder. Third, our study groups differed with respect to depression symptom severity at baseline. However, the median of the HAM-D total score was only 1 in the HR-MDD group, suggesting only sub-syndromal depression symptoms. Moreover, our correlation analysis revealed no relationship between depression symptom severity and our neurocognitive performance measures. Therefore, it appears unlikely that general mood differences at baseline between the groups have influenced our findings. Last but not least and as already highlighted previously, the relatively small HR-MDD sample size might have resulted in a lack of power to detect significant effects.

In summary, our findings suggest that reduced long delay verbal memory and extradimensional set-shifting performance across time constitute a familial trait marker for vulnerability to mood disorders. Both neurocognitive performance deficits appear to be relatively stable over a two-year time period and do not appear to be linked to the onset of MDD. These findings advance our understanding of the neuropathological processes underlying mood disorders and future longitudinal studies should examine their validity.

Chapter 7

General conclusions and implications for the future

In this thesis, I examined regional subcortical brain volumes, cortical thickness and cortical surface area of various brain regions and neuropsychological function in 111 initially unaffected relatives of BD patients of whom 20 developed MDD within a period of two years and matched healthy control subjects. Cross-sectional analyses at baseline were conducted to assess whether regional structural brain abnormalities and specific neuropsychological dysfunctions exist in the HR-MDD group prior to illness onset that may serve as neuroanatomic or neurocognitive markers for a subsequent onset of MDD and to assess their relationship to familial risk. The longitudinal analysis was intended to explore the time course of regional structural brain parameters and neuropsychological function during a period of two years to detect whether abnormal brain development or neurocognitive deterioration in the high-risk participants is evident that can be linked to an onset of MDD or familial risk. In this last chapter, I will summarise the key findings along its limitations and suggest implications for future research.

7.1 Summary of main findings

The cross-sectional and longitudinal analyses of the volumes of subcortical brain regions yielded no significant results. This indicates that subcortical brain abnormalities that have been reported in mood disorder patients do not appear to emerge as a consequence of being at high familial risk for mood disorders and are therefore unlikely to form a neuroanatomic endophenotype for the disorder. They also do not appear to predate an onset of MDD and are thus unlikely to be neurodevelopment in nature. Moreover, they do not appear to emerge as a consequence of illness-specific mechanisms that are directly linked to the onset of

the disorder. Rather, it appears likely that volumetric differences only emerge during the course of the disorder or in conjunction with psychopharmacological treatment effects.

For regional cortical thickness, the cross-sectional analyses at baseline yielded no significant results. The longitudinal analyses however revealed several significant findings. First, reduced cortical thickness in the right parahippocampal and fusiform gyrus across the two time points were found in both high-risk groups relative to controls, with the HR-MDD group displaying a thinner parahippocampus gyrus than the HR-well group. The fact that thinner parahippocampal and fusiform gyrus are evident in high-risk individuals, irrespective of time or the onset of MDD, suggests that thinning in these temporal brain regions constitutes a familial trait marker for vulnerability to mood disorders. Given that they are already present in early adulthood, they are unlikely to be of degenerative origin but likely represent disturbances of normal brain development predisposing to illness. Since the HR-MDD subjects displayed a thinner parahippocampal gyrus than the HR-well group, particular strong thickness reductions of this brain area may be related to risk of developing MDD. Parahippocampal and fusiform gyrus reductions are thus potential neuroanatomic endophenotypes for mood disorders, as thinning is evident in both unaffected relatives and affected patients and independent of state.

Second, the longitudinal analyses of cortical thickness revealed that HR-well subjects had progressive thickness reductions in the left inferior frontal and precentral gyrus relative to controls, while the HR-MDD group showed cortical thickening of these areas over time. The abnormal thinning of these brain areas in the HR-well group over time relative to controls suggests that thinning in regionally specific left frontal lobe areas forms a familial trait marker for vulnerability to mood disorders and that abnormal thinning already takes place in early adulthood, potentially reflecting early neurodegenerative processes. Importantly, a distinct pattern of increasing relative cortical thickness of the left

inferior frontal and precentral gyrus over time in the HR-MDD as compared to the HR-well group due to an absence of regional thinning of these brain areas in the HR-MDD cohort was found. For the precentral gyrus, the cortical thickness development in the HR-MDD group was also significantly different from the HC subjects. Absence of cortical thinning of these brain regions thus appears to be linked to the onset of illness and underlying disease-associated processes. Given that human brain maturation involves frontal grey matter loss beyond adolescence (Gogtay, et al., 2004; Sowell, et al., 2003; Westlye, et al., 2010), the absence of cortical thinning in the HR-MDD group may reflect a lack or delay of normal synaptic pruning processes.

For cortical surface area, the cross-sectional and longitudinal analyses yielded no significant results. This indicates that young unaffected adults at enhanced risk of mood disorders because of a close family history of BD do not exhibit cortical surface area abnormalities so that it appears unlikely that these may form a neuroanatomic endophenotype for the disorder. They also do not appear to predate an onset of MDD or to emerge as a consequence of illness-specific mechanisms that are directly linked to the onset of the disorder. Nevertheless, it cannot be ruled out that regional cortical surface area abnormalities may emerge during the course of the disorder or in conjunction with psychopharmacological treatment effects.

For neuropsychological performance measures, the cross-sectional analyses at baseline yielded no significant results. The longitudinal analyses however revealed several significant findings. First, reduced long delay verbal memory and extradimensional set-shifting performance across the two time points were found in the HR-well group relative to controls, with the HR-MDD group displaying decreased extradimensional set-shifting abilities as compared to the HC group only. The finding of reduced long delay verbal memory and extradimensional set-shifting performance in the HR-well group across time suggests that neurocognitive deficits in these domains constitute a familial trait marker for vulnerability to mood disorders. Given that they are already present in early

adulthood, they are unlikely to be of degenerative origin but likely represent disturbances of normal brain development predisposing to illness. Since no significant differences between the two high-risk groups and no group-by-time interaction emerged, the results do not speak towards reduced verbal memory and extradimensional set-shifting abilities to be directly linked to an onset of MDD.

Second, significant effects of time were observed for attentional processing speed, verbal learning, short delay verbal memory and extradimensional set-shifting, indicating superior performance during these tasks at follow-up assessment as compared to baseline assessment across all participant groups. These most likely reflect practice effects during neuropsychological examination which are well documented in the literature (Bartels, et al., 2010). There was also a significant effect of time for simple discrimination learning due to decreased task performance at follow-up as compared to baseline across all groups. One plausible explanation for this conflicting result may be that participants remembered the task from the baseline assessment and already shifted their attention to the currently irrelevant stimulus as they were expecting the reversal of the rule to occur at follow-up examination.

Taken together, these results suggest that the pathophysiology underlying mood disorders in early adulthood is driven by (at least) two factors. First of all, early structural brain and neurocognitive abnormalities that are potentially neurodevelopmental in nature and enhance disease vulnerability are evident. In detail, structural brain abnormalities in the right parahippocampus and fusiform gyrus as well as verbal memory and extra-dimensional set-shifting performance deficits that are relatively stable over time appear to enhance disease vulnerability. This suggests that early abnormal brain developmental processes appear to be taking place in mood disorders, leading to decreased cortical thickness in the temporal lobe and decreased verbal memory and set-shifting ability. Of note, particularly strong thickness reductions of the parahippocampal gyrus appear be related to the onset of MDD within a time period of two years. Second,

pathophysiological disease-associated processes are taking place. In detail, the onset of MDD is associated with increases in cortical thickness in the left inferior frontal and precentral gyrus, suggesting that disease-associated processes are taking place in these regions. Future studies should investigate the nature of these structural brain and neurocognitive abnormalities associated with enhanced disease vulnerability and illness onset.

Our finding of reduced cortical thickness in the right parahippocampal and fusiform gyrus associated with enhanced disease vulnerability is partly in line with previous findings of reduced white matter integrity in HR subjects of this study cohort (Sprooten, et al., 2011). Using diffusion tensor imaging, Sprooten et al. (2011) found reduced fractional anisotropy in the HR as compared to the HC group in a large cluster encompassing white matter tracts connecting temporal lobe regions with each other or other brain regions. Given that brain maturation in late adolescence and early adulthood involves increases in white matter volume and fractional anisotropy (Faria et al., 2010; Giorgio et al., 2010), these results similarly to the grey matter findings point towards disturbed brain maturation processes in the temporal lobe to be associated with an increased risk for mood disorder. A similar finding of reduced fractional anisotropy in numerous white matter tracks has also been observed in adolescents at high risk of BD because of subthreshold bipolar symptoms (Paillere Martinot, et al., 2014). In line with our findings, research on white matter integrity in first-episode MDD young treatment naïve adults reported decreased fractional anisotropy in the right parahippocampal gyrus of the MDD cohort (Zhu et al., 2011). Using diffusion tensor imaging in combination with fMRI, de Kwaasteniet et al. (2013) recently showed that white matter integrity of the uncinate fasciculus which connects medial temporal lobe structures with the anterior cingulate cortex is reduced in MDD and that there is a negative correlation between uncinate fasciculus integrity and functional connectivity between these brain regions. Taken together, these findings highlight the importance of studying structural (grey and white matter) abnormalities in

temporal lobe structures of mood disorders as well as their underlying pathophysiological mechanisms and their relationship to brain activation.

Our finding of decreases in cortical thickness in the left inferior frontal and precentral gyrus over time in the HR-well subjects may be well in line with previous findings of reduced white matter integrity in HR subjects of this study cohort, encompassing frontal brain areas (Sprooten, et al., 2011). However, no longitudinal analysis of fractional anisotropy of this study cohort has been conducted as yet. Similarly, decreased fractional anisotropy in BD and MDD has been commonly observed along frontal-subcortical tracts (Adler, et al., 2006; Mahon, et al., 2010; Tham, et al., 2011), but there is a lack of longitudinal studies. Moreover, there is evidence towards volumetric frontal white matter abnormalities in BD patients (Bruno, et al., 2004; McIntosh, et al., 2005; Stanfield, et al., 2009) which may be related to our findings. Future longitudinal studies of high-risk samples investigating white matter development over time will help to clarify if our results are in line with other neuroimaging findings.

7.2 Methodological remarks and limitations

The methodological strengths of this study are its longitudinal nature, the assessment of subjects prior to illness onset, the relatively young age of the participants and the comparatively large sample size of high-risk subjects and controls. In addition, all subjects underwent careful clinical assessment at both time points and the effects of medication and relatedness of subjects were ruled out. All brain scans were obtained at the same scanner using the identical protocol at both visits and the MRI data were processed in an identical way using thoroughly validated methods.

Several limitations of this study cohort need to be addressed. First, the sample size of the HR-MDD subjects was relatively small so that it cannot be ruled out that some analyses were under-powered to detect significant differences between the

HR-MDD and the other two study groups. Second, it remains unknown whether currently unaffected HR-well subjects may develop MDD in the future. Third, previous longitudinal studies have reported that the majority of the high-risk subjects who developed BD themselves experienced depressive episodes years before conversion (Duffy, 2010; Hillegers, et al., 2005) so that it appears likely that some of our HR-MDD subjects may develop BD in the future. The follow-up assessments of our study cohort will clarify if some of the HR-MDD participants will convert to BD and if some of our HR-well subjects go on to develop a mood disorder.

Fourth, the DSM-IV criteria employed to diagnose MDD (and BD) are a subject of debate. For example, the DSM-IV criteria for MDD have raised concerns about increasing numbers of false positive diagnoses since the criteria for MDD appear to be fulfilled by a large percentage of the population at one point in their lives and may be considered a normal response to stress (Wakefield, Schmitz, & Baer, 2010). Moreover, there is large heterogeneity in symptoms that ultimately lead to a diagnosis of MDD so that the concept of MDD as being a homogeneous disorder has been challenged (Goldberg, 2011). It would have thus been better to account for this issue by conducting analyses with respect to separate symptom dimensions of MDD. However, given the small sample size of the HR-MDD group, this approach was not feasible.

Fifth, our study groups differed with respect to depression symptom severity at baseline. However, the median of the HAM-D total score was only 1 in the HR-MDD group, suggesting only sub-syndromal depressive symptoms. Therefore, it appears unlikely that general mood differences at baseline between the groups have influenced our findings. Nevertheless, it needs to be highlighted that the clinical instruments employed at baseline and follow-up assessments might have not been ideal to detect sub-syndromal depressive and manic symptoms. However, no prodromal criteria for BD or MDD had been developed when this study was originally designed and putative prodromal symptoms such as depressed mood,

irritability, racing thoughts and physical agitation (Howes, et al., 2011) are covered by the YMRS, HAM-D and SCID. It would have still been desirable to include specialised clinical interviews capturing a broader range of putative prodromal symptoms and their varying degrees of severity.

Sixth, the (semi-)automated brain imaging software FreeSurfer was employed to obtain subcortical volumes, cortical thickness and cortical surface area estimates. As described in detail in paragraph 2.2.2, this approach has been shown to have various disadvantages in comparison to manual hand-tracing methods with respect to accuracy and reliability of findings. To minimize this limitation, the performance of FreeSurfer was visually checked at various processing stages and regional brain segmentations and parcellations were edited where necessary. Nevertheless, it cannot be ruled out that these methodological factors might have influenced our findings.

Seventh, given the relatively young age of our study cohort, it appears likely that brain development processes known to occur during adolescence and early adulthood (see also paragraph 1.5.3) might have influenced the results. It has been shown that during this time period, cortical grey matter volume decreases in a non-linear fashion across different brain regions; reductions first appear in sensorimotor areas, followed by decreases in the frontal cortex, parietal cortex and finally in the temporal cortex (Gogtay, et al., 2004). These grey matter reductions have been linked to increased synaptic and/or neuronal pruning processes taking place at late adolescence and early adulthood (Gogtay, et al., 2004). Our findings of abnormal cortical thickness reductions in the inferior frontal and precentral gyrus over the two-year time period in the HR-well subjects may thus be in line with the fact that grey matter reduces in the frontal cortex first. By contrast, reductions in temporal brain structures take place at later stages so that they might have been too subtle to detect them over the two-year time period studied and may only be evident at older age. It would be advantageous to study cortical and subcortical brain development

in mood disorders and high-risk populations during a larger time frame. The follow-up examinations of the BFS will help to clarify this aspect.

Eighth, it has been shown that the brain development trajectories differ substantially between the sexes. For nearly all cortical and subcortical brain areas, grey matter volume increases peak about 1-2 years earlier in women than men (Lenroot et al., 2007). For example, the volume of the whole cerebral cortex peaks at around the age of 10.5 years in females and around the age of 14.5 years in men (Lenroot, et al., 2007). These initial increases in grey and white matter are followed by decreases taking place until early adulthood. Accordingly, differences between the sexes in brain development trajectories might have influenced the results. It would have been advantageous to conduct separate statistical analyses for male and female participants. Given the small sample size of the HR-MDD subjects, this approach would have lowered the statistical power to be able to detect significant differences between the groups or over time so that this analysis was unfortunately not feasible. To minimize this limitation, we covaried for this factor in the analyses and checked that there were no significant differences in gender distribution between the study groups.

Ninth, the neuropsychological tasks employed appear not to be ideal for detecting longitudinal changes in cognitive performance since they have been shown to be prone to practice effects (Bartels, et al., 2010). It would have been advantageous to assess cognitive function at the second assessment point with parallel versions of the CVLT, IED and Digit Span. However, to the best of my knowledge, no parallel version for the IED does exist as yet. Moreover, task performance of the CVLT, IED and Digit Span relies at least to some extend on executive functions such as planning abilities and executive control mechanisms which might have influenced task performance. In particular, it can be expected that having performed the CVLT previously leads to enhanced performance at the second assessment as the subject is already familiar with grouping words according to semantic categories. Accordingly, it would have been advantageous to select a verbal learning and

memory task that does not include grouping words into semantic categories. Similarly, if subjects completed the IED previously, it appears likely that they remembered the rules at the second visit and accordingly performed better on intra- and extradimensional set-shifting than would have been expected if they were asked to complete an unknown task. Last but not least, general differences in IQ might have influenced the neuropsychological performance results.

7.3 Implications for the future

While structural brain abnormalities and neuropsychological performance deficits have been repeatedly demonstrated in patients with mood disorders and their unaffected close relatives, findings have been largely inconsistent. The observed discrepancies likely stem from the effects of age, gender, differences in symptom presentation, symptom severity, duration of illness, number of depressive and/or manic episodes, age of onset, comorbidity, medication as well as the effects of genetic and environmental susceptibility factors and their complex interplay which act upon human brain development and functioning. Further, sample sizes may often be not large enough to detect significant small structural brain and neuropsychological performance abnormalities. Moreover, the use of different methodological designs may add to the heterogeneity in findings.

Future studies should therefore address the complex interactions of various socio-demographic, clinical and susceptibility factors and take sample size and methodological issues into account to investigate the aetiology of neuroanatomic and neurocognitive abnormalities in more detail. Several approaches should be considered. On the one hand, very large study samples of young individuals at familial risk for BD or MDD who are followed longitudinally for years or even decades are required to disentangle the time course of structural brain and neuropsychological abnormalities in mood disorders. Such large-scale projects may however only be feasible if many research centres around the world are

collaboratively working together and share their data for mega-analysis. Given the large sample size, it should be feasible to analyse particular subgroups of the study cohort without losing power to detect significant differences. In detail, subgroups covering the particular symptom presentation of affected subjects, age of onset or medication status would add important information to the aetiology of structural brain and cognitive abnormalities. Moreover, analyses should disentangle the effects of age, gender and various other important factors on brain structure and function.

On the other hand, the effects of genetic and environmental susceptibility factors and their complex interplay on brain structure and function need to be addressed in large sample sizes of healthy control subjects and mood disorder patients. One example of such an approach is for example the ENIGMA consortium (Thompson et al., 2014), a large network of researchers working together on a range of large-scale studies that integrate data from 70 institutions worldwide. The consortium has been able to detect common variants in the genome that affect subcortical brain structure that no individual site could detect on its own.

Furthermore, new research projects should benefit from improvements made in technology and methodology of brain imaging. For example, using 7 tesla MRI scanners would provide a higher resolution of the brain and may thus help to detect even subtle changes in brain structure. As it has been recently possible to assess cortical thickness and surface area separately from each other, future research should make use of this advancement in brain imaging software and avoid standard VBM analyses which represent the combined effects of thickness and surface area only.

Finally, future studies should try to develop tools to predict the risk of an onset of mood disorders as accurately as possible. As this study suggests, these algorithms should include cortical thickness measures for the right parahippocampal and fusiform gyrus as well as the left inferior frontal and precentral gyrus. Moreover, verbal memory and extra-dimensional set-shifting performance indices may help to

calculate the relative risks of developing a mood disorder. It needs to be noted, however, that the results of this study were obtained in individuals with a family history of BD so that future studies need to address if the findings of this study also hold in subjects who have no family history of mood disorders. The development of such tools should also include measures of clinical (prodromal) symptoms and potentially (resting state) fMRI findings, white matter integrity indices and genetic measures.

7.4 General conclusion

Reduced cortical thickness in the right parahippocampal and fusiform gyrus constitute a familial trait marker for vulnerability to mood disorders and may thus form potential neuroanatomic endophenotypes. Particularly strong thickness reductions of the parahippocampal gyrus appear be related to the risk of MDD. Moreover, progressive thickness reductions in the left inferior frontal and precentral gyrus in early adulthood form a familial trait marker for vulnerability to mood disorders, potentially reflecting early neurodegenerative processes. By contrast, an absence of cortical thinning of these brain regions in early adulthood appears to be linked to the onset of MDD and underlying disease-associated processes, potentially reflecting a lack or delay of normal synaptic pruning processes. From a neuropsychological perspective, reduced long delay verbal memory and extradimensional set-shifting performance across time constitute a familial trait marker for vulnerability to mood disorders, likely representing disturbances of normal brain development predisposing to illness.

References

Abe, O., Yamasue, H., Kasai, K., Yamada, H., Aoki, S., Inoue, H., et al. (2010). Voxel-based analyses of gray/white matter volume and diffusion tensor data in major depression. *Psychiatry Res, 181*(1), 64-70.

Abkevich, V., Camp, N. J., Hensel, C. H., Neff, C. D., Russell, D. L., Hughes, D. C., et al. (2003). Predisposition locus for major depression at chromosome 12q22-12q23.2. *Am J Hum Genet, 73*(6), 1271-1281.

Acosta-Cabronero, J., Williams, G. B., Pereira, J. M., Pengas, G., & Nestor, P. J. (2008). The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. *Neuroimage, 39*(4), 1654-1665.

Adler, C. M., Adams, J., DelBello, M. P., Holland, S. K., Schmithorst, V., Levine, A., et al. (2006). Evidence of white matter pathology in bipolar disorder adolescents experiencing their first episode of mania: a diffusion tensor imaging study. *Am J Psychiatry, 163*(2), 322-324.

Ahdidan, J., Hviid, L. B., Chakravarty, M. M., Ravnkilde, B., Rosenberg, R., Rodell, A., et al. (2011). Longitudinal MR study of brain structure and hippocampus volume in major depressive disorder. *Acta Psychiatr Scand, 123*(3), 211-219.

Ajilore, O., Narr, K., Rosenthal, J., Pham, D., Hamilton, L., Watari, K., et al. (2011). Regional cortical gray matter thickness differences associated with type 2 diabetes and major depression. *Psychiatry Res, 184*(2), 63-70.

Alloy, L. B., Abramson, L. Y., Urosevic, S., Walshaw, P. D., Nusslock, R., & Neeren, A. M. (2005). The psychosocial context of bipolar disorder: environmental, cognitive, and developmental risk factors. *Clin Psychol Rev, 25*(8), 1043-1075.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4 ed.). Washington, DC: American Psychiatric Press.

Andreasen, N. C., O'Leary, D. S., Arndt, S., Cizadlo, T., Hurtig, R., Rezai, K., et al. (1995). Short-term and long-term verbal memory: a positron emission tomography study. *Proc Natl Acad Sci U S A, 92*(11), 5111-5115.

Angst, J., & Merikangas, K. (1997). The depressive spectrum: diagnostic classification and course. *J Affect Disord, 45*(1-2), 31-39; discussion 39-40.

Angst, J., & Preisig, M. (1995). Course of a clinical cohort of unipolar, bipolar and schizoaffective patients. Results of a prospective study from 1959 to 1985. *Schweiz Arch Neurol Psychiatr, 146*(1), 5-16.

Anguelova, M., Benkelfat, C., & Turecki, G. (2003). A systematic review of association studies investigating genes coding for serotonin receptors and the serotonin transporter: I. Affective disorders. *Mol Psychiatry, 8*(6), 574-591.

Aragam, N., Wang, K. S., & Pan, Y. (2011). Genome-wide association analysis of gender differences in major depressive disorder in the Netherlands NESDA and NTR population-based samples. *J Affect Disord, 133*(3), 516-521.

Arnone, D., Cavanagh, J., Gerber, D., Lawrie, S. M., Ebmeier, K. P., & McIntosh, A. M. (2009). Magnetic resonance imaging studies in bipolar disorder and schizophrenia: meta-analysis. *Br J Psychiatry, 195*(3), 194-201.

Arnone, D., McIntosh, A. M., Ebmeier, K. P., Munafo, M. R., & Anderson, I. M. (2012). Magnetic resonance imaging studies in unipolar depression: systematic review and meta-regression analyses. *Eur Neuropsychopharmacol, 22*(1), 1-16.

Arts, B., Jabben, N., Krabbendam, L., & van Os, J. (2008). Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. *Psychol Med, 38*(6), 771-785.

Badner, J. A., & Gershon, E. S. (2002). Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry, 7*(4), 405-411.

Balanza-Martinez, V., Rubio, C., Selva-Vera, G., Martinez-Aran, A., Sanchez-Moreno, J., Salazar-Fraile, J., et al. (2008). Neurocognitive endophenotypes (endophenocognitypes) from studies of relatives of bipolar disorder subjects: a systematic review. *Neurosci Biobehav Rev, 32*(8), 1426-1438.

Barnett, J. H., & Smoller, J. W. (2009). The genetics of bipolar disorder. *Neuroscience, 164*(1), 331-343.

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci, 11*, 118.

Baum, A. E., Akula, N., Cabanero, M., Cardona, I., Corona, W., Klemens, B., et al. (2008). A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry, 13*(2), 197-207.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological, 57*(1), 289-300.

Benton, A., & Hamsher, K. (1978). *Multilingual Aphasia Examination Manual, revised.* Iowa: University of Iowa.

Beyer, J. L., & Krishnan, K. R. (2002). Volumetric brain imaging findings in mood disorders. *Bipolar Disord, 4*(2), 89-104.

Beyer, J. L., Kuchibhatla, M., Payne, M., Moo-Young, M., Cassidy, F., MacFall, J., et al. (2004). Caudate volume measurement in older adults with bipolar disorder. *Int J Geriatr Psychiatry, 19*(2), 109-114.

Blanco, C., Laje, G., Olfson, M., Marcus, S. C., & Pincus, H. A. (2002). Trends in the treatment of bipolar disorder by outpatient psychiatrists. *Am J Psychiatry, 159*(6), 1005-1010.

Boccardi, M., Almici, M., Bresciani, L., Caroli, A., Bonetti, M., Monchieri, S., et al. (2010). Clinical and medial temporal features in a family with mood disorders. *Neurosci Lett, 468*(2), 93-97.

Bora, E., Fornito, A., Pantelis, C., & Yucel, M. (2012). Gray matter abnormalities in Major Depressive Disorder: a meta-analysis of voxel based morphometry studies. *J Affect Disord, 138*(1-2), 9-18.

Bora, E., Fornito, A., Yucel, M., & Pantelis, C. (2010). Voxelwise meta-analysis of gray matter abnormalities in bipolar disorder. *Biol Psychiatry, 67*(11), 1097-1105.

Bora, E., Yucel, M., & Pantelis, C. (2009). Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. *J Affect Disord, 113*(1-2), 1-20.

Bosker, F. J., Hartman, C. A., Nolte, I. M., Prins, B. P., Terpstra, P., Posthuma, D., et al. (2011). Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry, 16*(5), 516-532.

Bourne, C., Aydemir, O., Balanza-Martinez, V., Bora, E., Brissos, S., Cavanagh, J. T., et al. (2013). Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. *Acta Psychiatr Scand*.

Bowler, R. M., Sudia, S., Mergler, D., Harrison, R., & Cone, J. (1992). Comparison of Digit Symbol and Symbol Digit Modalities tests for assessing neurotoxic exposure. *The Clinical Neuropsychologist, 6*, 103-104.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B-Methodological, 26*(2), 211-252.

Breen, G., Webb, B. T., Butler, A. W., van den Oord, E. J., Tozzi, F., Craddock, N., et al. (2011). A genome-wide significant linkage for severe depression on chromosome 3: the depression network study. *Am J Psychiatry, 168*(8), 840-847.

Bremner, J. D., Narayan, M., Anderson, E. R., Staib, L. H., Miller, H. L., & Charney, D. S. (2000). Hippocampal volume reduction in major depression. *Am J Psychiatry, 157*(1), 115-118.

Bruno, S. D., Barker, G. J., Cercignani, M., Symms, M., & Ron, M. A. (2004). A study of bipolar disorder using magnetization transfer imaging and voxel-based morphometry. *Brain, 127*(Pt 11), 2433-2440.

Caetano, S. C., Kaur, S., Brambilla, P., Nicoletti, M., Hatch, J. P., Sassi, R. B., et al. (2006). Smaller cingulate volumes in unipolar depressed patients. *Biol Psychiatry, 59*(8), 702-706.

Camp, N. J., Lowry, M. R., Richards, R. L., Plenk, A. M., Carter, C., Hensel, C. H., et al. (2005). Genome-wide linkage analyses of extended Utah pedigrees identifies loci that influence recurrent, early-onset major depression and anxiety disorders. *Am J Med Genet B Neuropsychiatr Genet, 135B*(1), 85-93.

Cannon, D. M. (2010). Neuroimaging and the pathophysiology and treatment of depression: recent advances and future needs. In J. F. Cryan & B. E. Leonard (Eds.), *Modern Trends in Pharmacopsychiatry* (Vol. 27, pp. 101-123). Basel: Karger.

Carless, M. A., Glahn, D. C., Johnson, M. P., Curran, J. E., Bozaoglu, K., Dyer, T. D., et al. (2011). Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. *Mol Psychiatry, 16*(11), 1096-1104, 1063.

Carter, T. D., Mundo, E., Parikh, S. V., & Kennedy, J. L. (2003). Early age at onset as a risk factor for poor outcome of bipolar disorder. *J Psychiatr Res, 37*(4), 297-303.

Chamberlain, S. R., Blackwell, A. D., Fineberg, N. A., Robbins, T. W., & Sahakian, B. J. (2005). The neuropsychology of obsessive compulsive disorder: the importance of failures in cognitive and behavioural inhibition as candidate endophenotypic markers. *Neurosci Biobehav Rev, 29*(3), 399-419.

Cichon, S., Muhleisen, T. W., Degenhardt, F. A., Mattheisen, M., Miro, X., Strohmaier, J., et al. (2011). Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet, 88*(3), 372-381.

Clark, K. A., Woods, R. P., Rottenberg, D. A., Toga, A. W., & Mazziotta, J. C. (2006). Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *Neuroimage, 29*(1), 185-202.

Clarkson, M. J., Cardoso, M. J., Ridgway, G. R., Modat, M., Leung, K. K., Rohrer, J. D., et al. (2011). A comparison of voxel and surface based cortical thickness estimation methods. *Neuroimage, 57*(3), 856-865.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates.

Connor, S. E., Ng, V., McDonald, C., Schulze, K., Morgan, K., Dazzan, P., et al. (2004). A study of hippocampal shape anomaly in schizophrenia and in families multiply affected by schizophrenia or bipolar disorder. *Neuroradiology, 46*(7), 523-534.

Cotter, D., Hudson, L., & Landau, S. (2005). Evidence for orbitofrontal pathology in bipolar disorder and major depression, but not in schizophrenia. *Bipolar Disord, 7*(4), 358-369.

Craddock, N. (2006). Genetics of mood disorders. *Psychiatry, 5*(5), 170-174.

Craddock, N., & Forty, L. (2006). Genetics of affective (mood) disorders. *Eur J Hum Genet, 14*(6), 660-668.

Craddock, N., & Sklar, P. (2009). Genetics of bipolar disorder: successful start to a long journey. *Trends Genet, 25*(2), 99-105.

Craddock, N., & Sklar, P. (2013). Genetics of bipolar disorder. *Lancet, 381*(9878), 1654-1662.

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*.

Dahabra, S., Ashton, C. H., Bahrainian, M., Britton, P. G., Ferrier, I. N., McAllister, V. A., et al. (1998). Structural and functional abnormalities in elderly patients clinically recovered from early- and late-onset depression. *Biol Psychiatry, 44*(1), 34-46.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage, 9*(2), 179-194.

de Kwaasteniet, B., Ruhe, E., Caan, M., Rive, M., Olabarriaga, S., Groefsema, M., et al. (2013). Relation between structural and functional connectivity in major depressive disorder. *Biol Psychiatry, 74*(1), 40-47.

De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J. A., Proffitt, T. M., Mahony, K., et al. (2003). Normative data from the CANTAB. I: development of executive function over the lifespan. *J Clin Exp Neuropsychol, 25*(2), 242-254.

DelBello, M. P., Zimmerman, M. E., Mills, N. P., Getz, G. E., & Strakowski, S. M. (2004). Magnetic resonance imaging analysis of amygdala and other subcortical brain regions in adolescents with bipolar disorder. *Bipolar Disord, 6*(1), 43-52.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test: Second Edition*. San Antonio, TX: Psychological Corporation.

Delvecchio, G., Fossati, P., Boyer, P., Brambilla, P., Falkai, P., Gruber, O., et al. (2012). Common and distinct neural correlates of emotional processing in Bipolar Disorder and Major Depressive Disorder: a voxel-based meta-analysis of functional magnetic resonance imaging studies. *Eur Neuropsychopharmacol, 22*(2), 100-113.

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980.

Dewey, J., Hana, G., Russell, T., Price, J., McCaffrey, D., Harezlak, J., et al. (2010). Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. *Neuroimage, 51*(4), 1334-1344.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature, 380*(6569), 69-72.

Dick, D. M., Foroud, T., Flury, L., Bowman, E. S., Miller, M. J., Rau, N. L., et al. (2003). Genomewide linkage analyses of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. *Am J Hum Genet, 73*(1), 107-114.

Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. C. (2013). *Analysis of longitudinal data* (2 ed.). Oxford: Oxford University Press.

Doris, A., Belton, E., Ebmeier, K. P., Glabus, M. F., & Marshall, I. (2004). Reduction of cingulate gray matter density in poor outcome bipolar illness. *Psychiatry Res, 130*(2), 153-159.

Downes, J. J., Roberts, A. C., Sahakian, B. J., Evenden, J. L., Morris, R. G., & Robbins, T. W. (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific attentional dysfunction. *Neuropsychologia, 27*(11-12), 1329-1343.

Drevets, W. C., Price, J. L., & Furey, M. L. (2008). Brain structural and functional abnormalities in mood disorders: implications for neurocircuitry models of depression. *Brain Struct Funct, 213*(1-2), 93-118.

Du, M. Y., Wu, Q. Z., Yue, Q., Li, J., Liao, Y., Kuang, W. H., et al. (2012). Voxelwise meta-analysis of gray matter reduction in major depressive disorder. *Prog Neuropsychopharmacol Biol Psychiatry, 36*(1), 11-16.

Duffy, A. (2010). From predisposition to illness: genetically sensitive intermediate pathways to mood disorders. *Br J Psychiatry, 197*(5), 341-342.

Dupont, R. M., Jernigan, T. L., Heindel, W., Butters, N., Shafer, K., Wilson, T., et al. (1995). Magnetic resonance imaging and mood disorders. Localization of white matter and other subcortical abnormalities. *Arch Gen Psychiatry, 52*(9), 747-755.

Dupont, S., Samson, Y., Le Bihan, D., & Baulac, M. (2002). Anatomy of verbal memory: a functional MRI study. *Surg Radiol Anat, 24*(1), 57-63.

Edelman, R. R., & Warach, S. (1993). Magnetic resonance imaging (1). *N Engl J Med, 328*(10), 708-716.

Eggert, L. D., Sommer, J., Jansen, A., Kircher, T., & Konrad, C. (2012). Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. *PLoS One, 7*(9), e45081.

Elgamal, S., Denburg, S., Marriott, M., & MacQueen, G. (2010). Clinical factors that predict cognitive function in patients with major depression. *Can J Psychiatry, 55*(10), 653-661.

Elvsashagen, T., Westlye, L. T., Boen, E., Hol, P. K., Andreassen, O. A., Boye, B., et al. (2013). Bipolar II disorder is associated with thinning of prefrontal and temporal cortices involved in affect regulation. *Bipolar Disord, 15*(8), 855-864.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). *3D statistical neuroanatomical models from 305 MRI volumes*. Paper presented at the Nuclear Science Symposium and Medical Imaging Conference.

Faria, A. V., Zhang, J., Oishi, K., Li, X., Jiang, H., Akhter, K., et al. (2010). Atlas-based analysis of neurodevelopment from infancy to adulthood using diffusion tensor imaging and applications for automated abnormality detection. *Neuroimage, 52*(2), 415-428.

Fein, G., Landman, B., Tran, H., Barakos, J., Moon, K., Di Sclafani, V., et al. (2006). Statistical parametric mapping of brain morphology: sensitivity is dramatically increased by using brain-extracted images as inputs. *Neuroimage, 30*(4), 1187-1195.

Ferreira, M. A., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., et al. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet, 40*(9), 1056-1058.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinical Version (SCID-CV)*. Washington, D.C.: American Psychiatric Press.

Fischl, B. (2012). FreeSurfer. *Neuroimage, 62*(2), 774-781.

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A, 97*(20), 11050-11055.

Fischl, B., Liu, A., & Dale, A. M. (2001). Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging, 20*(1), 70-80.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron, 33*(3), 341-355.

Fischl, B., Salat, D. H., van der Kouwe, A. J., Makris, N., Segonne, F., Quinn, B. T., et al. (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage, 23 Suppl 1*, S69-84.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9*(2), 195-207.

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp, 8*(4), 272-284.

Foland-Ross, L. C., Thompson, P. M., Sugar, C. A., Madsen, S. K., Shen, J. K., Penfold, C., et al. (2011). Investigation of cortical thickness abnormalities in lithium-free adults with bipolar I disorder using cortical pattern matching. *Am J Psychiatry, 168*(5), 530-539.

Foley, D. L., Neale, M. C., & Kendler, K. S. (1998). Reliability of a lifetime history of major depression: implications for heritability and co-morbidity. *Psychol Med, 28*(4), 857-870.

Fornito, A., Malhi, G. S., Lagopoulos, J., Ivanovski, B., Wood, S. J., Saling, M. M., et al. (2008). Anatomical abnormalities of the anterior cingulate and paracingulate cortex in patients with bipolar I disorder. *Psychiatry Res, 162*(2), 123-132.

Fornito, A., Yucel, M., Wood, S. J., Bechdolf, A., Carter, S., Adamson, C., et al. (2009). Anterior cingulate cortex abnormalities associated with a first psychotic episode in bipolar disorder. *Br J Psychiatry, 194*(5), 426-433.

Fusar-Poli, P., Borgwardt, S., Bechdolf, A., Addington, J., Riecher-Rossler, A., Schultze-Lutter, F., et al. (2013). The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry, 70*(1), 107-120.

Fusar-Poli, P., Howes, O., Bechdolf, A., & Borgwardt, S. (2012). Mapping vulnerability to bipolar disorder: a systematic review and meta-analysis of neuroimaging studies. *J Psychiatry Neurosci, 37*(3), 170-184.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci, 2*(10), 861-863.

Giorgio, A., Watkins, K. E., Chadwick, M., James, S., Winmill, L., Douaud, G., et al. (2010). Longitudinal changes in grey and white matter during adolescence. *Neuroimage, 49*(1), 94-103.

Glahn, D. C., Almasy, L., Barguil, M., Hare, E., Peralta, J. M., Kent, J. W., Jr., et al. (2010). Neurocognitive endophenotypes for bipolar disorder identified in multiplex multigenerational families. *Arch Gen Psychiatry, 67*(2), 168-177.

Glahn, D. C., Almasy, L., Blangero, J., Burk, G. M., Estrada, J., Peralta, J. M., et al. (2007). Adjudicating neurocognitive endophenotypes for schizophrenia. *Am J Med Genet B Neuropsychiatr Genet, 144B*(2), 242-249.

Glahn, D. C., Barrett, J., Bearden, C. E., Mintz, J., Green, M. F., Serap Monkul, E., et al. (2006). Dissociable mechanisms for memory impairment in bipolar disorder and schizophrenia. *Psychol Med, 36*(8), 1085-1095.

Glahn, D. C., Curran, J. E., Winkler, A. M., Carless, M. A., Kent, J. W., Jr., Charlesworth, J. C., et al. (2012). High dimensional endophenotype ranking in the search for major depression risk genes. *Biol Psychiatry, 71*(1), 6-14.

Glahn, D. C., Thompson, P. M., & Blangero, J. (2007). Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. *Hum Brain Mapp, 28*(6), 488-501.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A, 101*(21), 8174-8179.

Gogtay, N., Ordonez, A., Herman, D. H., Hayashi, K. M., Greenstein, D., Vaituzis, C., et al. (2007). Dynamic mapping of cortical development before and after the onset of pediatric bipolar illness. *J Child Psychol Psychiatry, 48*(9), 852-862.

Gold, J. M., Carpenter, C., Randolph, C., Goldberg, T. E., & Weinberger, D. R. (1997). Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. *Arch Gen Psychiatry, 54*(2), 159-165.

Goldberg, D. (2011). The heterogeneity of "major depression". *World Psychiatry, 10*(3), 226-228.

Gottesman, II, & Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry, 160*(4), 636-645.

Greenwald, B. S., Kramer-Ginsberg, E., Bogerts, B., Ashtari, M., Aupperle, P., Wu, H., et al. (1997). Qualitative magnetic resonance imaging findings in geriatric depression. Possible link between later-onset depression and Alzheimer's disease? *Psychol Med, 27*(2), 421-431.

Grieve, S. M., Korgaonkar, M. S., Koslow, S. H., Gordon, E., & Williams, L. M. (2013). Widespread reductions in gray matter volume in depression. *Neuroimage Clin, 3*, 332-339.

Gronenschild, E. H., Habets, P., Jacobs, H. I., Mengelers, R., Rozendaal, N., van Os, J., et al. (2012). The effects of FreeSurfer version, workstation type, and

Macintosh operating system version on anatomical volume and cortical thickness measurements. *PLoS One, 7*(6), e38234.

Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., et al. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology, 25*(5), 766-776.

Hajek, T., Cullis, J., Novak, T., Kopecek, M., Blagdon, R., Propper, L., et al. (2013). Brain structural signature of familial predisposition for bipolar disorder: replicable evidence for involvement of the right inferior frontal gyrus. *Biol Psychiatry, 73*(2), 144-152.

Hajek, T., Gunde, E., Slaney, C., Propper, L., MacQueen, G., Duffy, A., et al. (2009a). Amygdala and hippocampal volumes in relatives of patients with bipolar disorder: a high-risk study. *Can J Psychiatry, 54*(11), 726-733.

Hajek, T., Gunde, E., Slaney, C., Propper, L., MacQueen, G., Duffy, A., et al. (2009b). Striatal volumes in affected and unaffected relatives of bipolar patients--high-risk study. *J Psychiatr Res, 43*(7), 724-729.

Hallahan, B., Newell, J., Soares, J. C., Brambilla, P., Strakowski, S. M., Fleck, D. E., et al. (2011). Structural magnetic resonance imaging in bipolar disorder: an international collaborative mega-analysis of individual adult patient data. *Biol Psychiatry, 69*(4), 326-335.

Hamilton, M. (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry, 23*, 56-62.

Hamilton, S. P. (2011). A new lead from genetic studies in depressed siblings: assessing studies of chromosome 3. *Am J Psychiatry, 168*(8), 783-789.

Hampshire, A., & Owen, A. M. (2006). Fractionating attentional control using event-related fMRI. *Cereb Cortex, 16*(12), 1679-1689.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., et al. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage, 32*(1), 180-194.

Hartberg, C. B., Sundet, K., Rimol, L. M., Haukvik, U. K., Lange, E. H., Nesvag, R., et al. (2011). Brain cortical thickness and surface area correlates of neurocognitive performance in patients with schizophrenia, bipolar disorder, and healthy adults. *J Int Neuropsychol Soc, 17*(6), 1080-1093.

Hauser, P., Matochik, J., Altshuler, L. L., Denicoff, K. D., Conrad, A., Li, X., et al. (2000). MRI-based measurements of temporal lobe and ventricular structures in patients with bipolar I and bipolar II disorders. *J Affect Disord, 60*(1), 25-32.

Haznedar, M. M., Roversi, F., Pallanti, S., Baldini-Rossi, N., Schnur, D. B., Licalzi, E. M., et al. (2005). Fronto-thalamo-striatal gray and white matter volumes and anisotropy of their connections in bipolar spectrum illnesses. *Biol Psychiatry, 57*(7), 733-742.

Heaton, R. K. (1981). *A manual for the Wisconsin Card Sorting Test.* Odessa, FL: Psychological Assessment Resources.

Hek, K., Demirkan, A., Lahti, J., Terracciano, A., Teumer, A., Cornelis, M. C., et al. (2013). A genome-wide association study of depressive symptoms. *Biol Psychiatry, 73*(7), 667-678.

Henson, R. N., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia, 38*(4), 426-440.

Hillegers, M. H., Reichart, C. G., Wals, M., Verhulst, F. C., Ormel, J., & Nolen, W. A. (2005). Five-year prospective outcome of psychopathology in the adolescent offspring of bipolar parents. *Bipolar Disord, 7*(4), 344-350.

Holma, K. M., Melartin, T. K., Haukka, J., Holma, I. A., Sokero, T. P., & Isometsa, E. T. (2010). Incidence and predictors of suicide attempts in DSM-IV major depressive disorder: a five-year prospective study. *Am J Psychiatry, 167*(7), 801-808.

Holmans, P., Weissman, M. M., Zubenko, G. S., Scheftner, W. A., Crowe, R. R., Depaulo, J. R., Jr., et al. (2007). Genetics of recurrent early-onset major

depression (GenRED): final genome scan report. *Am J Psychiatry, 164*(2), 248-258.

Howes, O. D., Lim, S., Theologos, G., Yung, A. R., Goodwin, G. M., & McGuire, P. (2011). A comprehensive review and model of putative prodromal features of bipolar affective disorder. *Psychol Med, 41*(8), 1567-1577.

Hulshoff Pol, H. E., van Baal, G. C., Schnack, H. G., Brans, R. G., van der Schot, A. C., Brouwer, R. M., et al. (2012). Overlapping and segregating structural brain abnormalities in twins with schizophrenia or bipolar disorder. *Arch Gen Psychiatry, 69*(4), 349-359.

Jansen, A., Sehlmeyer, C., Pfleiderer, B., Sommer, J., Konrad, C., Zwitserlood, P., et al. (2009). Assessment of verbal memory by fMRI: lateralization and functional neuroanatomy. *Clin Neurol Neurosurg, 111*(1), 57-62.

Jantzen, K. J., Anderson, B., Steinberg, F. L., & Kelso, J. A. (2004). A prospective functional MR imaging study of mild traumatic brain injury in college football players. *AJNR Am J Neuroradiol, 25*(5), 738-745.

Jarnum, H., Eskildsen, S. F., Steffensen, E. G., Lundbye-Christensen, S., Simonsen, C. W., Thomsen, I. S., et al. (2011). Longitudinal MRI study of cortical thickness, perfusion, and metabolite levels in major depressive disorder. *Acta Psychiatr Scand, 124*(6), 435-446.

Jezzard, P., & Clare, S. (2008). Principles of nuclear magnetic resonance and MRI. In P. Jezzard, P. M. Matthews & S. M. Smith (Eds.), *Functional MRI. An introduction to methods.* (pp. 67-92). New York: Oxford University Press.

Johnson, S. C., Saykin, A. J., Flashman, L. A., McAllister, T. W., & Sparling, M. B. (2001). Brain activation on fMRI and verbal memory ability: functional neuroanatomic correlates of CVLT performance. *J Int Neuropsychol Soc, 7*(1), 55-62.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., et al. (2009). MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition

sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage, 46*(1), 177-192.

Judd, L. L., Akiskal, H. S., Schettler, P. J., Coryell, W., Maser, J., Rice, J. A., et al. (2003). The comparative clinical phenotype and long term longitudinal episode course of bipolar I and II: a clinical spectrum or distinct disorders? *J Affect Disord, 73*(1-2), 19-32.

Karakas, H. M., & Karakas, S. (2006). Learning of serial digits leads to frontal activation in functional MR imaging. *Diagn Interv Radiol, 12*(1), 9-13.

Kehagia, A. A., Murray, G. K., & Robbins, T. W. (2010). Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. *Curr Opin Neurobiol, 20*(2), 199-204.

Kempton, M. J., Geddes, J. R., Ettinger, U., Williams, S. C., & Grasby, P. M. (2008). Meta-analysis, database, and meta-regression of 98 structural imaging studies in bipolar disorder. *Arch Gen Psychiatry, 65*(9), 1017-1032.

Kempton, M. J., Ruberto, G., Vassos, E., Tatarelli, R., Girardi, P., Collier, D., et al. (2009). Effects of the CACNA1C risk allele for bipolar disorder on cerebral gray matter volume in healthy individuals. *Am J Psychiatry, 166*(12), 1413-1414.

Kempton, M. J., Salvador, Z., Munafo, M. R., Geddes, J. R., Simmons, A., Frangou, S., et al. (2011). Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. *Arch Gen Psychiatry, 68*(7), 675-690.

Kendler, K. S. (2001). Twin studies of psychiatric illness: an update. *Arch Gen Psychiatry, 58*(11), 1005-1014.

Kendler, K. S., & Gardner, C. O. (2001). Monozygotic twins discordant for major depression: a preliminary exploration of the role of environmental experiences in the aetiology and course of illness. *Psychol Med, 31*(3), 411-423.

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2002). Toward a comprehensive developmental model for major depression in women. *Am J Psychiatry, 159*(7), 1133-1145.

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2006). Toward a comprehensive developmental model for major depression in men. *Am J Psychiatry, 163*(1), 115-124.

Kendler, K. S., Pedersen, N. L., Neale, M. C., & Mathe, A. A. (1995). A pilot Swedish twin study of affective illness including hospital- and population-ascertained subsamples: results of model fitting. *Behav Genet, 25*(3), 217-232.

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry, 60*(9), 929-937.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., et al. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA, 289*(23), 3095-3105.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry, 62*(6), 593-602.

Kessler, R. C., Merikangas, K. R., & Wang, P. S. (2007). Prevalence, comorbidity, and service utilization for mood disorders in the United States at the beginning of the twenty-first century. *Annu Rev Clin Psychol, 3*, 137-158.

Khandai, A. C., & Aizenstein, H. J. (2013). Recent advances in neuroimaging biomarkers in geriatric psychiatry. *Curr Psychiatry Rep, 15*(6), 360.

Kieseppa, T., Mantyla, R., Tuulio-Henriksson, A., Luoma, K., Mantere, O., Ketokivi, M., et al. (2014). White matter hyperintensities and cognitive

performance in adult patients with bipolar I, bipolar II, and major depressive disorders. *Eur Psychiatry, 29*(4), 226-232.

Kieseppa, T., van Erp, T. G., Haukka, J., Partonen, T., Cannon, T. D., Poutanen, V. P., et al. (2003). Reduced left hemispheric white matter volume in twins with bipolar I disorder. *Biol Psychiatry, 54*(9), 896-905.

Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., & Lundervold, A. (2009). Evaluation of automated brain MR image segmentation and volumetry methods. *Hum Brain Mapp, 30*(4), 1310-1327.

Kohli, M. A., Lucae, S., Saemann, P. G., Schmidt, M. V., Demirkan, A., Hek, K., et al. (2011). The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron, 70*(2), 252-265.

Konarski, J. Z., McIntyre, R. S., Kennedy, S. H., Rafi-Tari, S., Soczynska, J. K., & Ketter, T. A. (2008). Volumetric neuroimaging investigations in mood disorders: bipolar disorder versus major depressive disorder. *Bipolar Disord, 10*(1), 1-37.

Koolschijn, P. C., van Haren, N. E., Lensvelt-Mulders, G. J., Hulshoff Pol, H. E., & Kahn, R. S. (2009). Brain volume abnormalities in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Hum Brain Mapp, 30*(11), 3719-3735.

Koolschijn, P. C., van Haren, N. E., Schnack, H. G., Janssen, J., Hulshoff Pol, H. E., & Kahn, R. S. (2010). Cortical thickness and voxel-based morphometry in depressed elderly. *Eur Neuropsychopharmacol, 20*(6), 398-404.

Krishnan, K. R., McDonald, W. M., Escalona, P. R., Doraiswamy, P. M., Na, C., Husain, M. M., et al. (1992). Magnetic resonance imaging of the caudate nuclei in depression. Preliminary observations. *Arch Gen Psychiatry, 49*(7), 553-557.

Kuperberg, G. R., Broome, M. R., McGuire, P. K., David, A. S., Eddy, M., Ozawa, F., et al. (2003). Regionally localized thinning of the cerebral cortex in schizophrenia. *Arch Gen Psychiatry, 60*(9), 878-888.

Kupfer, D. J. (2005). The increasing medical burden in bipolar disorder. *JAMA, 293*(20), 2528-2530.

Kurtz, M. M., Ragland, J. D., Bilker, W., Gur, R. C., & Gur, R. E. (2001). Comparison of the continuous performance test with and without working memory demands in healthy controls and patients with schizophrenia. *Schizophr Res, 48*(2-3), 307-316.

Ladouceur, C. D., Almeida, J. R., Birmaher, B., Axelson, D. A., Nau, S., Kalas, C., et al. (2008). Subcortical gray matter volume abnormalities in healthy bipolar offspring: potential neuroanatomical risk marker for bipolar disorder? *J Am Acad Child Adolesc Psychiatry, 47*(5), 532-539.

Lan, M. J., Chhetry, B. T., Oquendo, M. A., Sublette, M. E., Sullivan, G., Mann, J. J., et al. (2014). Cortical thickness differences between bipolar depression and major depressive disorder. *Bipolar Disord*.

Lau, J. Y., & Eley, T. C. (2010). The genetics of mood disorders. *Annu Rev Clin Psychol, 6*, 313-337.

Lawrence, A. D., Sahakian, B. J., & Robbins, T. W. (1998). Cognitive functions and corticostriatal circuits: insights from Huntington's disease. *Trends Cogn Sci, 2*(10), 379-388.

Lee, R. S., Hermens, D. F., Porter, M. A., & Redoblado-Hodge, M. A. (2012). A meta-analysis of cognitive deficits in first-episode Major Depressive Disorder. *J Affect Disord, 140*(2), 113-124.

Lenroot, R. K., & Giedd, J. N. (2008). The changing impact of genes and environment on brain development during childhood and adolescence: initial findings from a neuroimaging study of pediatric twins. *Dev Psychopathol, 20*(4), 1161-1175.

Lenroot, R. K., Gogtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., et al. (2007). Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *Neuroimage, 36*(4), 1065-1073.

Levinson, D. F. (2006). The genetics of depression: a review. *Biol Psychiatry, 60*(2), 84-92.

Lewis, C. M., Ng, M. Y., Butler, A. W., Cohen-Woods, S., Uher, R., Pirlo, K., et al. (2010). Genome-wide association study of major recurrent depression in the U.K. population. *Am J Psychiatry, 167*(8), 949-957.

Li, R., Qin, W., Zhang, Y., Jiang, T., & Yu, C. (2012). The neuronal correlates of digits backward are revealed by voxel-based morphometry and resting-state functional connectivity analyses. *PLoS One, 7*(2), e31877.

Lichtenstein, P., Yip, B. H., Bjork, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., et al. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet, 373*(9659), 234-239.

Lochhead, R. A., Parsey, R. V., Oquendo, M. A., & Mann, J. J. (2004). Regional brain gray matter volume differences in patients with bipolar disorder as assessed by optimized voxel-based morphometry. *Biol Psychiatry, 55*(12), 1154-1162.

Lopez-Larson, M. P., DelBello, M. P., Zimmerman, M. E., Schwiers, M. L., & Strakowski, S. M. (2002). Regional prefrontal gray and white matter abnormalities in bipolar disorder. *Biol Psychiatry, 52*(2), 93-100.

Lorenzetti, V., Allen, N. B., Fornito, A., & Yucel, M. (2009). Structural brain abnormalities in major depressive disorder: a selective review of recent MRI studies. *J Affect Disord, 117*(1-2), 1-17.

Lyoo, I. K., Kim, M. J., Stoll, A. L., Demopulos, C. M., Parow, A. M., Dager, S. R., et al. (2004). Frontal lobe gray matter density decreases in bipolar I disorder. *Biol Psychiatry, 55*(6), 648-651.

Lyoo, I. K., Sung, Y. H., Dager, S. R., Friedman, S. D., Lee, J. Y., Kim, S. J., et al. (2006). Regional cerebral cortical thinning in bipolar disorder. *Bipolar Disord, 8*(1), 65-74.

Mahon, K., Burdick, K. E., & Szeszko, P. R. (2010). A role for white matter abnormalities in the pathophysiology of bipolar disorder. *Neurosci Biobehav Rev, 34*(4), 533-554.

Matsuo, K., Kopecek, M., Nicoletti, M. A., Hatch, J. P., Watanabe, Y., Nery, F. G., et al. (2012). New structural brain imaging endophenotype in bipolar disorder. *Mol Psychiatry, 17*(4), 412-420.

McDonald, C., Bullmore, E. T., Sham, P. C., Chitnis, X., Wickham, H., Bramon, E., et al. (2004). Association of genetic risks for schizophrenia and bipolar disorder with specific and generic brain structural endophenotypes. *Arch Gen Psychiatry, 61*(10), 974-984.

McDonald, C., Marshall, N., Sham, P. C., Bullmore, E. T., Schulze, K., Chapple, B., et al. (2006). Regional brain morphometry in patients with schizophrenia or bipolar disorder and their unaffected relatives. *Am J Psychiatry, 163*(3), 478-487.

McDonald, C., Zanelli, J., Rabe-Hesketh, S., Ellison-Wright, I., Sham, P., Kalidindi, S., et al. (2004). Meta-analysis of magnetic resonance imaging brain morphometry studies in bipolar disorder. *Biol Psychiatry, 56*(6), 411-417.

McGuffin, P., Farmer, A., & Harvey, I. (1991). A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry, 48*(8), 764-770.

McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., & Cardno, A. (2003). The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry, 60*(5), 497-502.

McIntosh, A. M., Job, D. E., Moorhead, T. W., Harrison, L. K., Forrester, K., Lawrie, S. M., et al. (2004). Voxel-based morphometry of patients with schizophrenia or bipolar disorder and their unaffected relatives. *Biol Psychiatry, 56*(8), 544-552.

McIntosh, A. M., Job, D. E., Moorhead, T. W., Harrison, L. K., Lawrie, S. M., & Johnstone, E. C. (2005). White matter density in patients with

schizophrenia, bipolar disorder and their unaffected relatives. *Biol Psychiatry, 58*(3), 254-257.

McIntosh, A. M., Job, D. E., Moorhead, W. J., Harrison, L. K., Whalley, H. C., Johnstone, E. C., et al. (2006). Genetic liability to schizophrenia or bipolar disorder and its relationship to brain structure. *Am J Med Genet B Neuropsychiatr Genet, 141B*(1), 76-83.

McQueen, M. B., Devlin, B., Faraone, S. V., Nimgaonkar, V. L., Sklar, P., Smoller, J. W., et al. (2005). Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet, 77*(4), 582-595.

Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M., Petukhova, M., et al. (2007). Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry, 64*(5), 543-552.

Montag, C., Weber, B., Fliessbach, K., Elger, C., & Reuter, M. (2009). The BDNF Val66Met polymorphism impacts parahippocampal and amygdala volume in healthy humans: incremental support for a genetic risk factor for depression. *Psychol Med, 39*(11), 1831-1839.

Morey, R. A., Petty, C. M., Xu, Y., Hayes, J. P., Wagner, H. R., 2nd, Lewis, D. V., et al. (2009). A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage, 45*(3), 855-866.

Morey, R. A., Selgrade, E. S., Wagner, H. R., 2nd, Huettel, S. A., Wang, L., & McCarthy, G. (2010). Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Hum Brain Mapp, 31*(11), 1751-1762.

Morys, J. M., Bobek-Billewicz, B., Dziewiatkowski, J., Ratajczak, I., Pankiewicz, P., Narkiewicz, O., et al. (2003). A magnetic resonance volumetric study of the temporal lobe structures in depression. *Folia Morphol (Warsz), 62*(4), 347-352.

Mueller, T. I., Leon, A. C., Keller, M. B., Solomon, D. A., Endicott, J., Coryell, W., et al. (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *Am J Psychiatry, 156*(7), 1000-1006.

Muglia, P., Tozzi, F., Galwey, N. W., Francks, C., Upmanyu, R., Kong, X. Q., et al. (2010). Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry, 15*(6), 589-601.

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Oyanagi, C., et al. (2001). Dissociable mechanisms of attentional control within the human prefrontal cortex. *Cereb Cortex, 11*(1), 85-92.

Nelson, H. (1982). *The National Adult Reading Test Manual*. Windsor: NFER-Nelson.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12*(4), 313-324.

Nemeroff, C. B., & Owens, M. J. (2002). Treatment of mood disorders. *Nat Neurosci, 5 Suppl*, 1068-1070.

Nery, F. G., Monkul, E. S., & Lafer, B. (2013). Gray matter abnormalities as brain structural vulnerability factors for bipolar disorder: A review of neuroimaging studies of individuals at high genetic risk for bipolar disorder. *Aust N Z J Psychiatry*.

Noga, J. T., Vladar, K., & Torrey, E. F. (2001). A volumetric magnetic resonance imaging study of monozygotic twins discordant for bipolar disorder. *Psychiatry Res, 106*(1), 25-34.

Nugent, N. R., Tyrka, A. R., Carpenter, L. L., & Price, L. H. (2011). Gene-environment interactions: early life stress and risk for depressive and anxiety disorders. *Psychopharmacology (Berl), 214*(1), 175-196.

Nurnberger, J. I., Jr. (2012). Genetics of bipolar disorder: where we are and where we are going. *Depress Anxiety, 29*(12), 991-993.

Osby, U., Brandt, L., Correia, N., Ekbom, A., & Sparen, P. (2001). Excess mortality in bipolar and unipolar disorder in Sweden. *Arch Gen Psychiatry, 58*(9), 844-850.

Paillere Martinot, M. L., Lemaitre, H., Artiges, E., Miranda, R., Goodman, R., Penttila, J., et al. (2014). White-matter microstructure and gray-matter volumes in adolescents with subthreshold bipolar symptoms. *Mol Psychiatry, 19*(4), 462-470.

Pantelis, C., Barber, F. Z., Barnes, T. R., Nelson, H. E., Owen, A. M., & Robbins, T. W. (1999). Comparison of set-shifting ability in patients with chronic schizophrenia and frontal lobe damage. *Schizophr Res, 37*(3), 251-270.

Papakostas, G. I., Petersen, T., Mahal, Y., Mischoulon, D., Nierenberg, A. A., & Fava, M. (2004). Quality of life assessments in major depressive disorder: a review of the literature. *Gen Hosp Psychiatry, 26*(1), 13-17.

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends Cogn Sci, 9*(2), 60-68.

Peng, J., Liu, J., Nie, B., Li, Y., Shan, B., Wang, G., et al. (2011). Cerebral and cerebellar gray matter reduction in first-episode patients with major depressive disorder: a voxel-based morphometry study. *Eur J Radiol, 80*(2), 395-399.

Pergadia, M. L., Glowinski, A. L., Wray, N. R., Agrawal, A., Saccone, S. F., Loukola, A., et al. (2011). A 3p26-3p25 genetic linkage finding for DSM-IV major depression in heavy smoking families. *Am J Psychiatry, 168*(8), 848-852.

Perlis, R. H., Brown, E., Baker, R. W., & Nierenberg, A. A. (2006). Clinical features of bipolar depression versus major depressive disorder in large multicenter trials. *Am J Psychiatry, 163*(2), 225-231.

Peterson, B. S., Warner, V., Bansal, R., Zhu, H., Hao, X., Liu, J., et al. (2009). Cortical thinning in persons at increased familial risk for major depression. *Proc Natl Acad Sci U S A, 106*(15), 6273-6278.

Peterson, B. S., & Weissman, M. M. (2011). A brain-based endophenotype for major depressive disorder. *Annu Rev Med, 62*, 461-474.

Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003a). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biol Psychiatry, 54*(5), 504-514.

Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003b). Neurobiology of emotion perception II: Implications for major psychiatric disorders. *Biol Psychiatry, 54*(5), 515-528.

Piccinelli, M., & Wilkinson, G. (2000). Gender differences in depression. Critical review. *Br J Psychiatry, 177*, 486-492.

Price, J. L., & Drevets, W. C. (2010). Neurocircuitry of mood disorders. *Neuropsychopharmacology, 35*(1), 192-216.

Qiu, L., Lui, S., Kuang, W., Huang, X., Li, J., Zhang, J., et al. (2014). Regional increases of cortical thickness in untreated, first-episode major depressive disorder. *Transl Psychiatry, 4*, e378.

Rajkowska, G., Miguel-Hidalgo, J. J., Wei, J., Dilley, G., Pittman, S. D., Meltzer, H. Y., et al. (1999). Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. *Biol Psychiatry, 45*(9), 1085-1098.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills, 8*, 271-276.

Rey, A. (1964a). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rey, A. (1964b). *L'examen psychologique dans les cas d'encephalopathie traumatique.* Paris: Presses Universitaires de France.

Reynolds, S., Carrey, N., Jaworska, N., Langevin, L. M., Yang, X. R., & Macmaster, F. P. (2014). Cortical thickness in youth with major depressive disorder. *BMC Psychiatry, 14*, 83.

Richards, D. (2011). Prevalence and clinical course of depression: a review. *Clin Psychol Rev, 31*(7), 1117-1125.

Rietschel, M., Mattheisen, M., Frank, J., Treutlein, J., Degenhardt, F., Breuer, R., et al. (2010). Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biol Psychiatry, 68*(6), 578-585.

Rimol, L. M., Hartberg, C. B., Nesvag, R., Fennema-Notestine, C., Hagler, D. J., Jr., Pung, C. J., et al. (2010). Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. *Biol Psychiatry, 68*(1), 41-50.

Rimol, L. M., Nesvag, R., Hagler, D. J., Jr., Bergmann, O., Fennema-Notestine, C., Hartberg, C. B., et al. (2012). Cortical volume, surface area, and thickness in schizophrenia and bipolar disorder. *Biol Psychiatry, 71*(6), 552-560.

Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., et al. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry, 18*(4), 497-511.

Robbins, T. W., & Arnsten, A. F. (2009). The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annu Rev Neurosci, 32*, 267-287.

Roberts, A. C., Robbins, T. W., & Everitt, B. J. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *Q J Exp Psychol B, 40*(4), 321-341.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *J Cogn Neurosci, 12*(1), 142-162.

Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., et al. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology, 58*(5), 695-701.

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S., Busa, E., et al. (2004). Thinning of the cerebral cortex in aging. *Cereb Cortex, 14*(7), 721-730.

Sassi, R. B., Brambilla, P., Hatch, J. P., Nicoletti, M. A., Mallinger, A. G., Frank, E., et al. (2004). Reduced left anterior cingulate volumes in untreated bipolar patients. *Biol Psychiatry, 56*(7), 467-475.

Savitz, J., & Drevets, W. C. (2009). Bipolar and major depressive disorder: neuroimaging the developmental-degenerative divide. *Neurosci Biobehav Rev, 33*(5), 699-771.

Savitz, J., Rauch, S. L., & Drevets, W. C. (2013). Clinical application of brain imaging for the diagnosis of mood disorders: the current state of play. *Mol Psychiatry, 18*(5), 528-539.

Schnack, H. G., van Haren, N. E., Brouwer, R. M., van Baal, G. C., Picchioni, M., Weisbrod, M., et al. (2010). Mapping reliability in multicenter MRI: voxel-based morphometry and cortical thickness. *Hum Brain Mapp, 31*(12), 1967-1982.

Schulze, T. G., Akula, N., Breuer, R., Steele, J., Nalls, M. A., Singleton, A. B., et al. (2012). Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *World J Biol Psychiatry*.

Schurhoff, F., Bellivier, F., Jouvent, R., Mouren-Simeoni, M. C., Bouvard, M., Allilaire, J. F., et al. (2000). Early and late onset bipolar disorders: two different forms of manic-depressive illness? *J Affect Disord, 58*(3), 215-221.

Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., et al. (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage, 22*(3), 1060-1075.

Segonne, F., Pacheco, J., & Fischl, B. (2007). Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans Med Imaging, 26*(4), 518-529.

Segurado, R., Detera-Wadleigh, S. D., Levinson, D. F., Lewis, C. M., Gill, M., Nurnberger, J. I., Jr., et al. (2003). Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *Am J Hum Genet, 73*(1), 49-62.

Selvaraj, S., Arnone, D., Job, D., Stanfield, A., Farrow, T. F., Nugent, A. C., et al. (2012). Grey matter differences in bipolar disorder: a meta-analysis of voxel-based morphometry studies. *Bipolar Disord, 14*(2), 135-145.

Shah, P. J., Glabus, M. F., Goodwin, G. M., & Ebmeier, K. P. (2002). Chronic, treatment-resistant depression and right fronto-striatal atrophy. *Br J Psychiatry, 180*, 434-440.

Shi, J., Potash, J. B., Knowles, J. A., Weissman, M. M., Coryell, W., Scheftner, W. A., et al. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. *Mol Psychiatry, 16*(2), 193-201.

Shyn, S. I., Shi, J., Kraft, J. B., Potash, J. B., Knowles, J. A., Weissman, M. M., et al. (2011). Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry, 16*(2), 202-215.

Simon, G. E. (2003). Social and economic burden of mood disorders. *Biol Psychiatry, 54*(3), 208-215.

Simon, G. E., Bauer, M. S., Ludman, E. J., Operskalski, B. H., & Unutzer, J. (2007). Mood symptoms, functional impairment, and disability in people with bipolar disorder: specific effects of mania and depression. *J Clin Psychiatry, 68*(8), 1237-1245.

Simon, N. M., Otto, M. W., Wisniewski, S. R., Fossey, M., Sagduyu, K., Frank, E., et al. (2004). Anxiety disorder comorbidity in bipolar disorder patients: data from the first 500 participants in the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Am J Psychiatry, 161*(12), 2222-2229.

Sklar, P., Ripke, S., Scott, L. J., Andreassen, O. A., Cichon, S., Craddock, N., et al. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet, 43*(10), 977-983.

Sklar, P., Smoller, J. W., Fan, J., Ferreira, M. A., Perlis, R. H., Chambert, K., et al. (2008). Whole-genome association study of bipolar disorder. *Mol Psychiatry, 13*(6), 558-569.

Smith, A. (1982). *Symbol Digit Modalities Test*. Los Angeles: Western Psychological Services.

Smith, K. A., Morris, J. S., Friston, K. J., Cowen, P. J., & Dolan, R. J. (1999). Brain mechanisms associated with depressive relapse and associated cognitive impairment following acute tryptophan depletion. *Br J Psychiatry, 174*, 525-529.

Smoller, J. W., & Finn, C. T. (2003). Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet, 123C*(1), 48-58.

Smoller, J. W., & Gardner-Schuster, E. (2007). Genetics of bipolar disorder. *Curr Psychiatry Rep, 9*(6), 504-511.

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychol Bull, 139*(1), 81-132.

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nat Neurosci, 6*(3), 309-315.

Sprooten, E., Sussmann, J. E., Clugston, A., Peel, A., McKirdy, J., Moorhead, T. W., et al. (2011). White matter integrity in individuals at high genetic risk of bipolar disorder. *Biol Psychiatry, 70*(4), 350-356.

St Clair, D., Blackwood, D., Muir, W., Carothers, A., Walker, M., Spowart, G., et al. (1990). Association within a family of a balanced autosomal translocation with major mental illness. *Lancet, 336*(8706), 13-16.

Stanfield, A. C., Moorhead, T. W., Job, D. E., McKirdy, J., Sussmann, J. E., Hall, J., et al. (2009). Structural abnormalities of ventrolateral and orbitofrontal cortex in patients with familial bipolar disorder. *Bipolar Disord, 11*(2), 135-144.

Strakowski, S. M., Delbello, M. P., & Adler, C. M. (2005). The functional neuroanatomy of bipolar disorder: a review of neuroimaging findings. *Mol Psychiatry, 10*(1), 105-116.

Strakowski, S. M., DelBello, M. P., Zimmerman, M. E., Getz, G. E., Mills, N. P., Ret, J., et al. (2002). Ventricular and periventricular structural volumes in first- versus multiple-episode bipolar disorder. *Am J Psychiatry, 159*(11), 1841-1847.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Sullivan, P. F., Daly, M. J., & O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet, 13*(8), 537-551.

Sullivan, P. F., de Geus, E. J., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T., et al. (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry, 14*(4), 359-375.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry, 157*(10), 1552-1562.

Suominen, K., Mantere, O., Valtonen, H., Arvilommi, P., Leppamaki, S., Paunio, T., et al. (2007). Early age at onset of bipolar disorder is associated with more severe clinical features but delayed treatment seeking. *Bipolar Disord, 9*(7), 698-705.

Takahashi, T., Suzuki, M., Tsunoda, M., Kawamura, Y., Takahashi, N., Tsuneki, H., et al. (2008). Association between the brain-derived neurotrophic factor Val66Met polymorphism and brain morphology in a Japanese sample of schizophrenia and healthy comparisons. *Neurosci Lett, 435*(1), 34-39.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: an approach to cerebral imaging*. New York: Thieme.

Tekin, S., & Cummings, J. L. (2002). Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. *J Psychosom Res, 53*(2), 647-654.

Tham, M. W., Woon, P. S., Sum, M. Y., Lee, T. S., & Sim, K. (2011). White matter abnormalities in major depression: evidence from post-mortem, neuroimaging and genetic studies. *J Affect Disord, 132*(1-2), 26-36.

The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*(7145), 661-678.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav*.

Todtenkopf, M. S., Vincent, S. L., & Benes, F. M. (2005). A cross-study meta-analysis and three-dimensional comparison of cell counting in the anterior cingulate cortex of schizophrenic and bipolar brain. *Schizophr Res, 73*(1), 79-89.

Truong, W., Minuzzi, L., Soares, C. N., Frey, B. N., Evans, A. C., MacQueen, G. M., et al. (2013). Changes in cortical thickness across the lifespan in major depressive disorder. *Psychiatry Res, 214*(3), 204-211.

Tsuchiya, K. J., Byrne, M., & Mortensen, P. B. (2003). Risk factors in relation to an emergence of bipolar disorder: a systematic review. *Bipolar Disord, 5*(4), 231-242.

Tu, P. C., Chen, L. F., Hsieh, J. C., Bai, Y. M., Li, C. T., & Su, T. P. (2012). Regional cortical thinning in patients with major depressive disorder: a surface-based morphometry study. *Psychiatry Res, 202*(3), 206-213.

van der Schot, A. C., Vonk, R., Brans, R. G., van Haren, N. E., Koolschijn, P. C., Nuboer, V., et al. (2009). Influence of genes and environment on brain volumes in twin pairs concordant and discordant for bipolar disorder. *Arch Gen Psychiatry, 66*(2), 142-151.

van der Schot, A. C., Vonk, R., Brouwer, R. M., van Baal, G. C., Brans, R. G., van Haren, N. E., et al. (2010). Genetic and environmental influences on focal brain density in bipolar disorder. *Brain, 133*(10), 3080-3092.

van Eijndhoven, P., van Wingen, G., Katzenbauer, M., Groen, W., Tepest, R., Fernandez, G., et al. (2013). Paralimbic cortical thickness in first-episode depression: evidence for trait-related differences in mood regulation. *Am J Psychiatry, 170*(12), 1477-1486.

van Tol, M. J., Li, M., Metzger, C. D., Hailla, N., Horn, D. I., Li, W., et al. (2013). Local cortical thinning links to resting-state disconnectivity in major depressive disorder. *Psychol Med*, 1-13.

van Weel-Baumgarten, E., van den Bosch, W., van den Hoogen, H., & Zitman, F. G. (1998). Ten year follow-up of depression after diagnosis in general practice. *Br J Gen Pract, 48*(435), 1643-1646.

Viana, M. C., & Andrade, L. H. (2012). Lifetime Prevalence, age and gender distribution and age-of-onset of psychiatric disorders in the Sao Paulo Metropolitan Area, Brazil: results from the Sao Paulo Megacity Mental Health Survey. *Rev Bras Psiquiatr, 34*(3), 249-260.

Wakefield, J. C., Schmitz, M. F., & Baer, J. C. (2010). Does the DSM-IV clinical significance criterion for major depression reduce false positives? Evidence from the National Comorbidity Survey Replication. *Am J Psychiatry, 167*(3), 298-304.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale (Manual).* New York: Psychological Corporation.

Wechsler, D. (1987). *Wechsler memory scale - revised.* San Antonio, Texas: Psychological Corporation.

Wechsler, D. (1997). *WAIS-III administration and scoring manual.* San Antonio, Texas: Psychological Corporation.

Wenger, E., Martensson, J., Noack, H., Bodammer, N. C., Kuhn, S., Schaefer, S., et al. (2014). Comparing manual and automatic segmentation of

hippocampal volumes: reliability and validity issues in younger and older brains. *Hum Brain Mapp, 35*(8), 4236-4248.

Westlye, L. T., Walhovd, K. B., Dale, A. M., Bjornerud, A., Due-Tonnessen, P., Engvig, A., et al. (2010). Differentiating maturational and aging-related changes of the cerebral cortex by use of thickness and signal intensity. *Neuroimage, 52*(1), 172-185.

Whalley, H. C., Papmeyer, M., Romaniuk, L., Johnstone, E. C., Hall, J., Lawrie, S. M., et al. (2012). Effect of variation in diacylglycerol kinase eta (DGKH) gene on brain function in a cohort at familial risk of bipolar disorder. *Neuropsychopharmacology, 37*(4), 919-928.

Wilke, M., Kowatch, R. A., DelBello, M. P., Mills, N. P., & Holland, S. K. (2004). Voxel-based morphometry in adolescents with bipolar disorder: first results. *Psychiatry Res, 131*(1), 57-69.

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., et al. (2009). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage, 53*(3), 1135-1146.

Wonderlick, J. S., Ziegler, D. A., Hosseini-Varnamkhasti, P., Locascio, J. J., Bakkour, A., van der Kouwe, A., et al. (2009). Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. *Neuroimage, 44*(4), 1324-1333.

World Health Organization. (2004). *ICD-10 International Statistical Classification of Diseases and Health Related Problems* (2 ed.). Geneva: World Health Organization.

Wray, N. R., Pergadia, M. L., Blackwood, D. H., Penninx, B. W., Gordon, S. D., Nyholt, D. R., et al. (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry, 17*(1), 36-48.

Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (2000). Young Mania Rating Scale. In A. P. Association (Ed.), *Handbook of Psychiatric*

Measures (pp. 540-542). Washington, D.C.: American Psychiatric Association.

Zamoscik, V., Huffziger, S., Ebner-Priemer, U., Kuehner, C., & Kirsch, P. (2014). Increased involvement of the parahippocampal gyri in a sad mood predicts future depressive symptoms. *Soc Cogn Affect Neurosci*.

Zhang, J., Xiao, J., Zhu, X., Wang, X., & Yao, S. (2011). [Voxel-based morphometry on grey matter concentration of the brain in first-episode, antipsychotic-naive major depressive disorder]. *Zhong Nan Da Xue Xue Bao Yi Xue Ban, 36*(4), 307-311.

Zhang, X., Yao, S., Zhu, X., Wang, X., & Zhong, M. (2012). Gray matter volume abnormalities in individuals with cognitive vulnerability to depression: a voxel-based morphometry study. *J Affect Disord, 136*(3), 443-452.

Zhu, X., Wang, X., Xiao, J., Zhong, M., Liao, J., & Yao, S. (2011). Altered white matter integrity in first-episode, treatment-naive young adults with major depressive disorder: a tract-based spatial statistics study. *Brain Res, 1369*, 223-229.

Zipursky, R. B., Seeman, M. V., Bury, A., Langevin, R., Wortzman, G., & Katz, R. (1997). Deficits in gray matter volume are present in schizophrenia but not bipolar disorder. *Schizophr Res, 26*(2-3), 85-92.

Zisook, S., Lesser, I., Stewart, J. W., Wisniewski, S. R., Balasubramani, G. K., Fava, M., et al. (2007). Effect of age at onset on the course of major depressive disorder. *Am J Psychiatry, 164*(10), 1539-1546.

Zubenko, G. S., Maher, B., Hughes, H. B., 3rd, Zubenko, W. N., Stiffler, J. S., Kaplan, B. B., et al. (2003). Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. *Am J Med Genet B Neuropsychiatr Genet, 123B*(1), 1-18.