QUANTIFICATION OF GREY MATTER CHANGES FOLLOWING FIRST EPISODE MANIA IN BIPOLAR I DISORDER: A SYSTEMATIC REVIEW AND PROSPECTIVE STUDY

by

ARIANA CAHN

B.Sc., The University of Alberta, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(NEUROSCIENCE)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2021

© Ariana Cahn, 2021
The following individuals certify that they have read, and recommend to the Faculty of Graduate
and Postdoctoral Studies for acceptance, a thesis entitled:

Quantification of Grey Matter Changes Following First Episode Mania in Bipolar I Disorder: A
Systematic Review and Prospective Study

submitted by Ariana Cahn in partial fulfillment of the requirements for

the degree of Master of Science

in Neuroscience

Examining Committee:

Dr. Lakshmi Yatham, Professor, Department of Psychiatry, UBC

Co-supervisor

Dr. Trisha Chakrabarty, Assistant Professor, Department of Psychiatry, UBC

Co-supervisor

Dr. Todd Woodward, Professor, Department of Psychiatry, UBC

Supervisory Committee Member

Dr. Kamyar Keramatian, Clinical Assistant Professor, Department of Psychiatry, UBC

Supervisory Committee Member

Dr. Liisa Galea, Professor, Department of Psychology, UBC

Additional Examiner

Additional Supervisory Committee Members:

Dr. Stan Floresco, Professor, Department of Psychology, UBC

Supervisory Committee Member
Abstract

**Background:** While there is evidence of widespread grey matter (GM) changes in bipolar I disorder (BD-I), it is unclear how early in the illness such changes emerge. However, to date there has been little systematic examination of longitudinal grey matter changes early during BD-I. We conducted a systematic review to examine the literature regarding GM changes in BD-I patients following the first episode of mania (FEM), in addition to a quantitative analysis of grey matter changes in a prospective cohort of BD-I patients in the first three years of disease.

**Methods:** Following PRISMA guidelines, we conducted a systematic review of studies examining longitudinal changes in GM volume (GMV), cortical thickness and/or surface area in patients with BD-I following FEM. We qualitatively synthesized results regarding baseline differences between BD-I patients and HCs, and longitudinal GM changes in BD-I patients.

For the longitudinal study, FEM patients and HCs were recruited and completed structural 3T MRI at pre-determined intervals. Images were analyzed using FreeSurfer’s longitudinal pipeline and linear mixed models used to examine longitudinal changes in cortical thickness, surface area and subcortical volume.

**Results:** Fifteen studies met inclusion criteria for the systematic review, all examining changes in GMV. While results were highly heterogeneous, one replicated finding was decreased anterior cingulate cortex (ACC) volumes in first episode BD-I patients versus HC at baseline, and greater longitudinal ACC volume decrease in patients in the months following FEM. The potential impact of episode recurrence, substance use, age of onset and prior depressive episodes was inconsistently examined.

In the longitudinal study, BD-I patients and HCs showed comparable reductions in cortical surface area and subcortical volumes over 3 years. No significant time*group effects were found; BD-I patients had smaller right thalamic volumes versus HCs at year 3 at trend-level significance.

**Limitations:** The literature regarding GM changes early in BD-I is inconsistent. The sample size in our longitudinal study may have been insufficient to detect statistically significant changes.

**Conclusions:** Evidence for longitudinal grey matter changes in the early phases of BD-I is inconsistent. Larger longitudinal studies are needed to fully understand neuroprogression in early BD-I.
Lay Summary

Bipolar I Disorder (BD-I) is a psychiatric condition that lasts the whole lifetime. Patients experience repetitive manic and depressive episodes. Between episodes, decreased mental function and quality of life persists. Prior research has shown changes in certain brain regions, but these findings present a snapshot in time and do not tell us at what point these changes develop. To answer this question, we performed two studies. First, we performed a systematic review of previous studies tracking brain changes in newly diagnosed BD-I patients over time. We summarized the findings so other researchers can better understand what gaps need to be filled. Second, we completed a long-term study looking at BD-I brain scans over the first three years of the disease. Results from this project can provide a better understanding of the long-term course of bipolar disorder and provide guidance for early treatments to prevent further structural and cognitive decline.
Preface

The work performed for this thesis was completed as part of my MSc studies at the University of British Columbia under the supervision of Drs. Lakshmi N. Yatham and Trisha Chakrabarty.

Chapter 1. The introduction to this thesis was written by me with input from my supervisors. Images in Figures 1-2A, and 1-3 were adapted with permission from R.S. Desikan et al. Neuroimage 31, 968-980 (2006). The images in Figures 1-4 were provided by volBrain, produced by Manjón & Coupé (https://www.volbrain.upv.es/index.php).

Chapter 2. The systematic review in Chapter 2 was conducted primarily by me with assistance from T.C. and C.F. I wrote the manuscript and created the tables and figures. L.Y., T.C., and K.K. provided feedback and aided in the submission process. A version of Chapter 2 has been accepted for publication and is published as:


Chapter 3. Imaging data was collected as part of the Systematic Treatment Optimization Program for Early Mania (STOP-EM, UBC Clinical Ethics Research Board Certificate Number H04-701679). STOP-EM is a longitudinal research project naturalistically following bipolar I disorder (BD-I) patients over the first 15 years of their disorder, acquiring structural, cognitive, and clinical information. Dr. Yatham is responsible for the design, implementation, and continuation of the STOP-EM program. Drs. Frangou and Ge performed the morphometric analysis on the brain data collected from the STOP-EM project and provided me with brain measures. I conducted statistical analyses on these brain measures, giving the results for the study. I wrote the entirety of Chapter 3, except for section 3.2.3 which was written by Drs. Frangou and Ge. I created all the tables in Chapter 3. The graphics in Figure 3-1A were adapted with permission from R.S. Desikan et al. Neuroimage 31, 968-980 (2006), and the image in Figure 3-1B was adapted from: Xueyi Shen, et al., CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons. Figures 3-2 – 3-4 were created by me. Chapter 4 was also written by me with input from Drs. Yatham and Chakrabarty.
Table of Contents

Abstract ........................................................................................................................................... iii
Lay Summary ..................................................................................................................................... iv
Preface ............................................................................................................................................... v
Table of Contents .............................................................................................................................. vi
List of Tables ...................................................................................................................................... x
List of Figures ..................................................................................................................................... xi
List of Abbreviations ........................................................................................................................ xii
Acknowledgements .......................................................................................................................... xiv
Dedication .......................................................................................................................................... xvi

Chapter 1: Introduction .................................................................................................................... 1

1.1 Overview of Bipolar I Disorder ................................................................................................. 1
   1.1.1 Diagnostic Criteria of BD-I ................................................................................................. 1
   1.1.2 Historical Perspective of Bipolar Disorder ......................................................................... 2
   1.1.3 Etiology of Bipolar Disorder .............................................................................................. 3

1.2 Clinical Progression in BD-I .................................................................................................... 4

1.3 Neuroimaging in Bipolar Disorder ............................................................................................ 5
   1.3.1 Neuroimaging during the First Episode of Mania ............................................................... 6
   1.3.2 Cross-sectional Neuroimaging Studies ............................................................................... 7
       1.3.2.1 Frontal Regions .............................................................................................................. 7
       1.3.2.2 Temporal Lobes ............................................................................................................ 11
       1.3.2.3 Subcortical Regions .................................................................................................... 14
   1.3.3 Longitudinal Neuroimaging............................................................................................... 16

1.4 The Systematic Treatment Optimization Program for Early Mania ....................................... 18
Chapter 2: Longitudinal Grey Matter Changes Following First Episode Mania in Bipolar I Disorder: A Systematic Review

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Introduction</td>
<td>20</td>
</tr>
<tr>
<td>2.2 Methods</td>
<td>20</td>
</tr>
<tr>
<td>2.2.1 Search Strategy</td>
<td>20</td>
</tr>
<tr>
<td>2.2.2 Eligibility Criteria</td>
<td>21</td>
</tr>
<tr>
<td>2.2.3 Study Selection, Data Collection, and Data Extraction</td>
<td>21</td>
</tr>
<tr>
<td>2.2.4 Subgroup Analysis</td>
<td>22</td>
</tr>
<tr>
<td>2.2.5 Risk of Bias</td>
<td>22</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>22</td>
</tr>
<tr>
<td>2.3.1 Study Selection</td>
<td>22</td>
</tr>
<tr>
<td>2.3.2 Study Characteristics</td>
<td>23</td>
</tr>
<tr>
<td>2.3.3 Risk of Bias Within Studies</td>
<td>29</td>
</tr>
<tr>
<td>2.3.4 Baseline GM Comparison Between FEM Patients and Controls</td>
<td>29</td>
</tr>
<tr>
<td>2.3.5 Longitudinal GM Changes in BD-I Patients Following FEM</td>
<td>30</td>
</tr>
<tr>
<td>2.3.5.1 Longitudinal Frontal GM Changes</td>
<td>30</td>
</tr>
<tr>
<td>2.3.5.2 Longitudinal Parietal-Occipital GM Changes</td>
<td>31</td>
</tr>
<tr>
<td>2.3.5.3 Longitudinal Temporal GM Changes</td>
<td>31</td>
</tr>
<tr>
<td>2.3.5.4 Longitudinal Subcortical GM Changes</td>
<td>31</td>
</tr>
<tr>
<td>2.3.6 Subgroup Analyses</td>
<td>36</td>
</tr>
<tr>
<td>2.3.6.1 VBM vs ROI</td>
<td>36</td>
</tr>
<tr>
<td>2.3.6.2 Episode Recurrence and Relationship to Longitudinal GM Change</td>
<td>36</td>
</tr>
<tr>
<td>2.3.6.3 Substance Use</td>
<td>36</td>
</tr>
<tr>
<td>2.3.6.4 Psychotic Features in FEM</td>
<td>37</td>
</tr>
<tr>
<td>2.3.6.5 Pediatric vs Adult-Onset BD-I</td>
<td>37</td>
</tr>
</tbody>
</table>
2.3.6.6 Depressive Episodes Prior to FEM ......................................................... 37

2.4 Discussion ............................................................................................................ 37

2.4.1 Conclusions and Implications ......................................................................... 42

Chapter 3: Comparison of Grey Matter Brain Measures Over Time in Bipolar I Disorder

Patients versus Healthy Controls ........................................................................... 44

3.1 Brief Introduction .................................................................................................. 44

3.2 Methods ................................................................................................................ 44

3.2.1 Participants ......................................................................................................... 44

3.2.2 Neuroimaging ..................................................................................................... 48

3.2.3 Image Analysis .................................................................................................. 48

3.2.4 Statistical Analysis ............................................................................................ 48

3.2.4.1 Missing Data .................................................................................................. 49

3.3 Results .................................................................................................................. 52

3.3.1 Demographic Variables at Baseline ................................................................. 52

3.3.2 Cortical Thickness Measures ............................................................................. 52

3.3.3 Surface Area Measures ...................................................................................... 57

3.3.4 Subcortical Volume Measures .......................................................................... 61

3.4 Discussion ............................................................................................................. 65

Chapter 4: Discussion .................................................................................................. 69

4.1 Overview of Findings ........................................................................................... 69

4.2 Longitudinal Gray Matter Changes Following First Episode Mania in Bipolar I Disorder: A Systematic Review ............................................................................................................. 69

4.3 Comparison of GM Brain Measures Over Time in BD-I Patients vs Healthy Controls ................................................................. 71

4.4 Strength and Limitations ..................................................................................... 72

4.5 Conclusions ......................................................................................................... 74
References ................................................................................................................................. 76
List of Tables

Table 2-1. Characteristics of the 15 studies included in the systematic review .......................... 26
Table 2-2 Results for differences at baseline and during longitudinal follow up from the 15 studies included in the systematic review .................................................................................................................. 32
Table 3-1 Demographic and clinical features of STOP-EM study participants .......................... 47
Table 3-2 Cortical Thickness Estimated Marginal Means measures and Mixed Linear Model results ........................................................................................................................................................................... 52
Table 3-3 Pairwise comparison of Cortical Thickness measures at Baseline, Year 1, and Year 3 between first episode BD-I patients and healthy controls ............................................................. 55
Table 3-4 Surface Area Estimated Marginal Means measures and Mixed Linear Model results 58
Table 3-5 Pairwise comparison of Baseline, Year 1, and Year 3 separated by group for Surface Area measures ......................................................................................................................................................................................... 59
Table 3-6 Subcortical Volume Estimated Marginal Means measures and Mixed Linear Model results ............................................................................................................................................................................................ 62
Table 3-7 Pairwise comparison of Baseline, Year 1, and Year 3 separated by group for Subcortical Volume measures ........................................................................................................................................................................................................ 63
List of Figures

Figure 1-1 A representative 3 Tesla (3T) T1-weighted magnetic resonance imaging (MRI) brain scan image.......................................................................................................................................................................................... 6
Figure 1-2 Frontal brain regions as parcellated by the Desikan-Killiany Atlas and subregions on the anterior cingulate cortex (ACC)........................................................................................................................................................................ 10
Figure 1-3 Temporal brain regions as parcellated by the Desikan-Killiany Atlas .................. 13
Figure 1-4 A representative figure of subcortical regions as segmented by volBrain.......... 16
Figure 2-1 Flowchart for the systematic review and study selection process following the Preferred Reporting Items for Systematic review and Meta Analyses (PRISMA) statement .................... 23
Figure 3-1 Visual representation of regions of interest (ROIs) analyzed ................................ 51
Figure 3-2 Change in cortical thickness (CT) over time in five bilateral cortical regions ....... 56
Figure 3-3 Change in Surface Area (SA) over time in five bilateral cortical regions ........... 60
Figure 3-4 Change in Subcortical Volume (SCV) over time in four bilateral subcortical regions.65
List of Abbreviations

3D: 3 Dimensions
A.D.: Anno Domini
ACC: Anterior Cingulate Cortex
ADHD: Attention Deficit Hyperactive Disorder
BD: Bipolar Disorder
BD-I: Bipolar I Disorder
BD-II: Bipolar II Disorder
BL: Baseline
BMI: Body Mass Index
CANMAT: Canadian Network for Mood and Anxiety Treatments
CANTAB: Cambridge Neuropsychological Test Automated Battery
CSF: Cerebrospinal Fluid
CT: Cortical Thickness
DLPFC: Dorsolateral Prefrontal Cortex
DSM: Diagnostic and Statistical Manual of Mental Disorders
EM: Estimated Marginal
ENIGMA: ENIGMA Consortium
FDR: False Discovery Rate
FEAF: First Episode Affective Psychosis
FEM: First Episode of Mania
FOV: Field of View
FWE: Family-Wise Error
GAF: Global Assessment of Functioning Scale
GM: Grey Matter
GMV: Grey Matter Volume
HAMD: Hamilton Depression Rating Scale
HC: Healthy Controls
ICV: Intracranial Volume
IQ: Intelligence Quotient
MADRS: Montgomery-Åsberg Depression Rating Scale
MDD: Major Depressive Disorder
MINI: Mini International Neuropsychiatric Interview
MPRAGE: Magnetization Prepared - Rapid Gradient Echo
MRI: Magnetic Resonance Imaging
N.B.: Nota Bene
OFC: Orbitofrontal Cortex
PFC: Prefrontal Cortex
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
ROI: Region of Interest
SA: Surface Area
SCV: Subcortical Volume
STOP-EM: Systematic Treatment Optimization Program for Early Mania
SUD: Substance Use Disorder
T: Tesla
TE: Echo Time
TR: Repetition Time
UBC: University of British Columbia
UBCH: University of British Columbia Hospital
VBM: Voxel-Based Morphometry
WM: White Matter
Y1: Year 1
Y3: Year 3
YMRS: Young Mania Rating Scale
Acknowledgements

To my supervisors: Dr. Lakshmi Yatham and Dr. Trisha Chakrabarty. Thank you for supervising me during my graduate studies. During my time with you, I have learned many new skills and have been fully immersed in field of clinical research. This has been an interesting experience – one I am sure I will never forget.

To my Supervisory Committee, Dr. Kamyar Keramatian, Dr. Todd Woodward, and Dr. Stan Floresco – I thank you for agreeing to serve on my committee. Thank you for being there to answer my questions, help with statistics and look over my writing.

Next, I would like to profoundly thank the UBC Institutes of Mental Health for funding my research. By awarding me the Marshalls Scholar’s Award two years running, you allowed this important research to take place. In addition to receiving the Faculty of Medicine Entrance award, which covered the costs of my first year’s tuition, I received the inaugural Djavad Mowafagian Center for Brain Research’s Benjamin Feldman and Family Endowment Fund for Transformational Activity in Mental Health, allowing for the continuation of this research. I would like to thank all the funding bodies that aided me financially during my time at UBC.

To all our First Episode Mania study participants, and to the many patients with Bipolar Disorder around the world participating in other studies, thank you for agreeing to a plethora of clinic visits and brain scans. Without you and your dedication to these projects, these massive undertakings would not have been possible. Thank you for your selfless acts to ensure the continuation of such important research, that I truly hope will improve your lives as well as all those who are afflicted by Bipolar Disorder.

To the team at the Mood Disorders Clinic – Nazlin Walji, Jayasree Basivireddy, Adam White, Jing Liu, Minnie Teng, Sharon Ahn, Gayatri Saraf, Christian Frysh, Dr. Raymond Lam, Amna Khan, and Vanessa Evans – COVID may have kept us apart, but it is truly thanks to all of you that I felt welcome in my new city. You all made the MDC my family and for that I am eternally grateful. I would like to extend a special thanks to Nazlin Walji for always being a shoulder to cry on. Calling me during the pandemic to make sure I was ok made me feel like a valued member of the team, as well as an important person to you. To Amna Khan and Vanessa Evans: our friendship extends
beyond the short walls of our cubicle office space, and for that I am overjoyed. I look forward to seeing how our friendship blossoms in the future.

To the friends I made along the way: Tristan, Hong, Ellen, Kelly, Peter, Hannah, and the countless others, thank you for always listening to me rant about school. Coffee dates, forest walks, and ski days were all an essential part of this journey. Thank you for taking me away from school sometimes and allowing us to just have fun instead. To the people who acted as family to me: Mike, Alana, Lorene and Don, thank you for watching over me as I’ve been separated from my own family here. And to my best friend back home, Harman Khinda, thank you for always promptly answering my texts and picking up the phone when I call you at weird hours. Life is more fun with you in it.

To my partner, Bradley Meyer, you are my rock. You keep me sane and balanced. Through the insanity of life as a graduate student, and especially during a pandemic, no less, your never-ending support is something I admire and thank you endlessly for. I am a better person thanks to you. I love you.

Last but not least, to my parents, Margo and Dana, I would not be here without you. Whether through emotional, or financial support, every decision you have made in your lives has culminated in my being here. You are the best parents a girl could ever have. Thank you for supporting me, believing in me, and not allowing me to give up. I miss you terribly as I am away from you here, but our bonds persist. I love you more than I have the ability to express in words, so I hope this acknowledgment serves as such.
To my Uncle Robert,

who passed away suddenly before I finished this thesis.

I was looking forward to you reading it.
Chapter 1: Introduction

1.1 Overview of Bipolar I Disorder

Bipolar I Disorder (BD-I) is a leading cause of disability worldwide in the 10-24 year old age group (Gore et al., 2011), with an estimated lifetime prevalence of 4.4% in the United States (Merikangas et al., 2011). The hallmark of BD-I is recurrent manic episodes. These patients also commonly experience depressive episodes lasting longer than two weeks, although their occurrence is not required for a diagnosis of BD-I. These episodes can be accompanied by psychotic symptoms. In addition, patients can experience episodes with mixed features, wherein depressive and manic symptoms are experienced concurrently (American Psychiatric Association, 2013). It was originally believed that patients experienced full functional recovery in the time between mood episodes (also known as ‘euthymic’ periods), however it is now known that patients often experience sub-syndromal mood symptoms, psychosocial and cognitive impairments, and an overall reduced quality of life during these inter-episode periods (Thompson et al., 2005; Van Gorp et al., 1998; Zarate et al., 2000). Patients with BD are also at a 10-30 fold higher rate of suicide than the general population (Dome et al., 2019). When polled, patients with BD stated that their condition moderately to severely impaired their function and wellbeing (Kessler et al., 1994).

1.1.1 Diagnostic Criteria of BD-I

BD-I is one diagnosis within a broader category of mood disorders. To be diagnosed with BD-I, the patient must have experienced at least one lifetime manic episode (American Psychiatric Association, 2013). To meet the criteria for a manic episode, patients must have experienced: 1) abnormally and persistently elevated or irritable mood and abnormally and persistently increased goal directed activity or energy present for at least one week, and 2) at least three of the following additional symptoms during this time: grandiosity, decreased need for sleep, increased talkativeness, racing thoughts, distractibility, increase in goal-directed activity or psychomotor agitation, and involvement in activities which have high risk factors. The symptoms during this time must have been sufficiently severe to cause marked impairment in functioning or require hospitalization, and the episode could have not been caused by the effects of a substance (American Psychiatric Association, 2013). Manic episodes are frequently accompanied by
psychotic symptoms, defined as delusions or hallucinations (American Psychiatric Association, 2013).

A hypomanic episode is a period of distinctly elevated or irritable mood and abnormally and persistently increased goal directed activity or energy lasting at least four consecutive days, is associated with an uncharacteristic change in the individual’s functioning and behavior that is observable to others, and is also not due to an exogenous substance (American Psychiatric Association, 2013). Symptoms of hypomania are qualitatively similar to mania; however, they do not require hospitalization, are not associated with significant functional impairment and are not associated with psychotic symptoms. Finally, patients can also experience major depressive episodes, which must include at least five of the following symptoms for at least two weeks: 1) depressed mood; 2) diminished pleasure in previously enjoyed activities; 3) significant weight loss or weight gain; 4) persistent insomnia or hypersomnia; 5) psychomotor agitation or retardation; 6) fatigue; 7) feelings of worthlessness or excessive guilt; 8) diminished concentration; or 9) recurrent thoughts of death or suicidal ideation. At least one of symptoms 1) or 2) must be present. These symptoms must cause significant distress or impairment, and must not be due to the physiological effects of another substance or medical condition (American Psychiatric Association, 2013).

BD-I differs from BD-II, where a manic episode is not required for the latter, but rather the patient meets criteria for a past or current hypomanic episode and a past or current major depressive episode (American Psychiatric Association, 2013). While a major depressive episode is not required for the diagnosis of BD-I, the majority of patients with BD-I will experience depression; indeed, depressive symptoms tend to predominate the course of BD-I and are a major contributor to functional impairment (Judd et al., 2003).

1.1.2 Historical Perspective of Bipolar Disorder

The first to categorize bipolar disorder as an illness was Jean Pierre Falret, in 1853, dubbing the disorder “circular insanity” or “la folie circulaire” (J. P. Falret, 1853). While he may have been the first to pathologize the disorder, it has been rudimentarily studied for centuries. In fact, such an observation was made by Areteus in the second century A.D., describing a group of euphoric patients who would “laugh, play, dance night and day, and sometimes go openly to the market crowned, as if victors in some contest of skill, only later to appear torpid, dull, and sorrowful” (Altschule, 1976; Sedler, 1983).
The concept of mixed states was concomitantly being studied in the 19th century (Angst & Sellaro, 2019), called “mixtures” (Heinroth, 1818) and “middle forms” (Griesinger, 1845), but it was not until 1861 when Falret’s son, Jules, was the first to use the modern term (J. Falret, 1861). These mixed states were described as: “manic stupor, agitated melancholia (depression with flight of ideas and agitation), and unproductive mania (elated mood, increased motor activity, and inhibition of thinking)” (Weygandt, 1899).

Kraepelin, who is most famous for his work in schizophrenia, took a unifying approach to the classification of mood disorders (1899). He included bipolar disorders under the umbrella term of “manic-depressive insanity”, which included single-episode and recurrent depression (Angst & Sellaro, 2019). However, his contemporaries disagreed with this conflation of what is now known as unipolar or major depressive disorder and bipolar disorders. Eventually, major depressive (i.e. unipolar) and bipolar disorders were segregated, largely on the basis of their differential response to antidepressant versus mood stabilizing medications (Angst & Sellaro, 2019).

It was finally in 1980 when the Diagnostic and Statistical Manual of Mental Disorders (DSM)-III introduced the term “bipolar disorder”, foregoing the more disparaging title of “manic-depressive” as this led patients to be met with stigma and judgement (Mason et al., 2016), followed by separation of BD-I and BD-II as separate entities in the DSM-IV, based on diagnostic validity of psychiatric disorders as proposed by Robins and Guze (Benazzi, 2007; Robins & Guze, 1970). According to DSM-5 terminology, specifiers describing the severity and type of the most recent episode experienced, as well as whether there were accompanying psychotic features, are added to a BD-I diagnosis. Additional specifiers are used to describe unique longitudinal characteristics experienced by some individuals with BD-I, such as the presence of rapid cycling (at least four mood episodes in a year) or a seasonal pattern (e.g., experiencing manic episodes in the summer and depressive episodes in the winter).

1.1.3 Etiology of Bipolar Disorder

While the cause of bipolar disorder is largely unknown, converging factors such as genetics, changes in brain structure and function, and psychosocial factors are thought to be involved.

Bipolar disorder is highly heritable, with an estimated 10-15% risk of developing mood episodes when a first-degree relative has BD (Faraone et al., 2003; Hillegers et al., 2005; Smoller and Finn, 2003). Heritability estimates of 70-90% have been shown in familial and twin studies.
(Craddock & Jones, 1999; Smoller & Finn, 2003). However, it remains difficult to distinguish between effects of shared environments from genetic factors (Smoller & Finn, 2003). Factors such as the age of onset (Strober, 1992), psychotic symptoms (Goes et al., 2007; Potash et al., 2003), frequency of episodes (Fisfalen et al., 2005), and polarity of initial episode (manic or depressive) (Kassem et al., 2006) are highly familial whilst comorbid anxiety and substance use are less so (Schulze et al., 2006).

Severe trauma, maltreatment, or abuse during childhood is a common risk factor, with approximately 50% of patients with BD reporting this experience (Agnew-Blais et al., 2016; Brown et al., 2005; Post et al., 2006). In fact, studies looking at childhood trauma found grey matter volume (GMV) alterations, especially in the thalamus, in patient with BD reporting trauma versus those without (Duarte et al., 2016; Poletti et al., 2016; Song et al., 2020). Additionally, patients with BD with psychosis have shown increased levels of inflammation which were associated with childhood trauma (Quidé et al., 2021). Finally, prior exposure to antidepressants or psychostimulants has been linked to emergence of manic episodes (Delbello et al., 2001; Goldsmith et al., 2011; Soutullo et al., 2002).

1.2 Clinical Progression in BD-I

While it was classically believed that periods of euthymia presented a complete remission, we now know that many patients with BD-I will experience persistent functional and cognitive impairment during this time (Thompson et al., 2005; Van Gorp et al., 1998; Zarate et al., 2000). Furthermore, the course of BD-I for some patients appears to be characterized by progressive worsening of clinical and functional outcomes (Solé et al., 2017). These observations have formed the basis for clinical staging models (Berk et al., 2017; Kapczinski et al., 2009; Yatham et al., 2018) which broadly describe early, intermediate, and late stages of BD-I. These staging models start in the pre-symptomatic at-risk phase, moving into non-specific prodromal symptoms, then the onset of mood symptoms, followed by a relapse-remitting course, and finally moving into unrelenting, treatment-resistant episodes (Kapczinski et al., 2009; McGorry et al., 2006; Post, 2010). Genotype and environmental factors have been hypothesized to interact in the early stage, resulting in neurobiological changes that increase propensity for mood episodes (Craddock and Forty, 2006). This progresses through further cycling of mood episodes alongside less favourable responses to treatments, finally resulting in impaired cellular resilience. Macroscopic structural
brain changes are posited to evolve with the above described cellular and clinical changes (Berk et al., 2017; Kapczinski et al., 2009; Yatham et al., 2018).

Such staging models re-conceptualize BD-I as a neuroprogressive disorder and suggest that effective intervention at an early stage may help arrest further neurodegeneration. Thus, understanding structural brain changes early in the disease course can enhance understanding of the expected impacts of early intervention in BD-I. Moreover, looking at morphological changes using neuroimaging techniques can help us better understand clinical patterns and the relationships between symptomology and pathophysiology.

1.3 Neuroimaging in Bipolar Disorder

Magnetic resonance imaging (MRI) is a powerful tool which allows for the visualization of high-resolution images of internal structures, including the brain. High-strength magnets in the scanner create a strong and constant magnetic field, causing hydrogen atoms (protons) in the subject to align with the field (Plewes & Kucharczyk, 2012). A radiofrequency pulse applied to the subject in the scanner excites the protons, causing them to spin out of equilibrium before returning gradually in phase with the magnetic field (relaxation). This return to equilibrium is recorded by a receiver coil, and tissue classification occurs based on the time it took (McRobbie et al., 2006). Protons in white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) of the brain have unique relaxation times, allowing the scanner to yield high-contrast images. In T1-weighted images, CSF appears darkest, followed by GM, and then WM with the brightest signal (Bojorquez et al., 2017). MRI can be used to quantify the total volumes of WM, GM, and CSF. In addition, volumes of specific brain structures can be measured. Studies examining regional brain volumes typically follow one of two approaches: a region of interest (ROI) analysis, where potentially relevant brain regions are a priori identified by the researcher, or voxel-based morphometry (VBM), where changes across the brain are examined voxel-by-voxel between groups. Grey matter volume is the product of cortical thickness and surface area; more recently, MRI has been used to measure thickness and surface area separately across the cortex.
MRI has been widely used to probe morphological changes in BD, allowing for better understanding of the neurobiology of the disorder. As studies largely focus on GM imaging in BD, the below literature review and the studies conducted for this thesis will reflect this.

1.3.1 Neuroimaging during the First Episode of Mania

The first episode of mania (FEM) is a crucial time. In addition to being the key diagnostic criterium for BD-I, FEM presents an excellent opportunity to conduct MRI studies to uncover neurobiological changes which may be amenable to early intervention.

Meta-analyses have been conducted of cross-sectional MRI studies examining structural brain changes in FEM patients. Vita et al. analyzed 11 primary research articles and found significant decreases in intracranial and white matter volumes, but no difference in grey matter volumes of FEM BD-I patients vs healthy controls (HCs) (Vita et al., 2009). A second meta-analysis, conducted by De Peri et al. analyzed studies which examined both BD-I and schizophrenia patients. Specifically, five of the included studies studied BD-I patients only, nine incorporated both BD-I and schizophrenia patients, while the remaining 31 included only schizophrenia patients. This meta-analysis found that while both patient groups exhibited loss in

Figure 1-1 A representative 3 Tesla (3T) T1-weighted magnetic resonance imaging (MRI) brain scan image. Cerebrospinal fluid (CSF) appears as dark, grey matter (GM) as next darkest, and white matter (WM) as brightest due to their different relaxation times.
global grey matter volume at disease onset, these deficits were more marked in the schizophrenia patients (De Peri et al., 2012). Finally, a recent review and meta-analysis examined 12 FEM VBM studies. While four of the included studies found no regional differences in GMV between BD-I patients and HCs, the remaining 11 studies revealed differences in various frontal and temporal regions. The study’s meta-analysis found reduction in bilateral pregenual anterior cingulate cortex (ACC) in the FEM patients compared with HCs (Keramatian et al., 2020).

Smaller cross-sectional FEM studies present more conflicting results. A recent study conducted by Keramatian et al. found no significant grey matter differences between FEM patients and HCs (Keramatian et al., 2021). A study looking at regional subcortical volume (SCV) in FEM patients found decreased amygdala GMV (Rosso et al., 2007), while two other studies noted enlarged subcortical regions such as the thalamus and basal ganglia (Chen et al., 2012; Kozicky et al., 2013); Kozicky et al.’s study noted that the increase in caudate volume was associated with poorer performance on executive function tasks in the BD-I cohort only. Two studies looked at temporal lobe volume and temporal structures in FEM patients found conflicting results. First, Hirayasu et al. reported no baseline differences when looking at the Heschl Gyrus and Planum Temporale between groups (Hirayasu et al., 2000), while Bond et al. observed a relationship between increased body mass index (BMI) and decreased overall temporal lobe volumes in FEM BD-I patients (Bond et al., 2011). Although Keramatian et al.’s meta-analysis reports findings of decrease pregenual ACC volumes in FEM patients, smaller left subgenual ACC volumes have also been found (Singh et al., 2012).

### 1.3.2 Cross-sectional Neuroimaging Studies

There are many cross-sectional MRI studies of patients with longer duration BD-I, with these studies more consistently showing structural changes than what has been shown in FEM. A variety of systematic reviews and meta-analyses have been conducted to synthesize the results from these individual studies. The following sections will examine these reviews/meta-analyses, as well as individual cross-sectional studies. Frontal lobe, temporal lobe, and subcortical region findings are reported separately as these are in-line with regions examined in this thesis.

#### 1.3.2.1 Frontal Regions

Frontal lobe regions have received special anatomical consideration in BD studies due to their associations with cognition and emotional processing, especially the ACC, which has received additional consideration due to its participation in the fronto-limbic circuit (Beckmann et
Figure 1 presents a visual representation of frontal brain regions and subregions of the ACC.

Meta-analyses and reviews report decreased prefrontal cortex (PFC) and ACC volumes in BD-I patients. Reviews of early imaging studies (circa the early 2000s) identified PFC and ACC decreases in patients, and reported preserved volumes in these regions with lithium use, suggesting a neuroprotective effect for this mood stabilizer (Haldane et al., 2004; Monkul et al., 2005). A meta-analysis of BD VBM studies found GM reduction in the left rostral ACC, and additionally found that longer illness duration was associated with GMV decline in the subgenual ACC. This study also reported ACC enlargement associated with lithium use (Bora et al., 2010). Another meta-analysis of 15 studies similarly found reductions in the subgenual and perigenual ACC (Ellison-Wright et al., 2010). Finally, a meta-analysis of 32 VBM studies corroborated smaller volumes in the ACC of patients with BD (Wise et al., 2017). Reductions in these regions are equally associated with psychosis (Wang et al., 2018) and suicidality (Bani-Fatemi et al., 2018; Hozer et al., 2016) in BD-I patients. Interestingly, one meta-analysis of VBM studies noted volumetric reductions in the PFC alongside enlargements in the cingulate, positing that PFC deficits is a key pathophysiological maker, whereas the ACC enlargement is a compensatory mechanism (Lu et al., 2019).

A large-scale study conducted by the ENIGMA Consortium pooled neuroimaging data from 2447 Patients with BD and 4056 HCs from world-wide research centers. When looking at cortical thickness (CT) measures from these participants, they found notably thinner CT in the left pars opercularis and rostral middle frontal cortex (parts of the inferior frontal cortex and PFC, respectively) in the patient population. They further observed that a longer illness duration was associated with reduced CT in broad frontal regions, but this was especially pronounced in the left rostral ACC (Hibar et al., 2018).

Large-scale studies allow for well-powered analyses to detect group level differences. However, smaller primary cross-sectional research studies provide insight into relationships between frontal brain changes and aspects of BD such as frequency of mood episodes, age, and cognitive functioning. GMV reductions in the ACC seem to be especially pronounced in patients with BD currently experiencing depressive episodes, with significant relationships between lower ACC volumes and Hamilton Depression Rating Scale (HAMD) scores (Cai et al., 2015; Matsuo et al., 2019). Cai et al.’s study noted that these decreases were unique to patients with BD (i.e. not
seen in major depressive disorder (MDD) patients; however, other studies have shown ACC reductions in both BD and MDD patients (MacMaster et al., 2014). Smaller subgenual ACC regions were associated with a history of past suicide attempts (Jabbi et al., 2020). Reductions in ACC volume have also been shown to correlate with deficits in cognitive tasks (Sun et al., 2020), and GMV in the rostral and subgenual subregions of the ACC were able to predict patient performance on the Wisconsin Card Sorting Task, a test for executive function (Zimmerman et al., 2006). Lower ACC volume was also found to be significantly associated with number of hospital visits, especially in male patients (Delvecchio et al., 2019).

DLPFC volumes/densities are also smaller in patients with BD (Dickstein et al., 2005) and appear to be associated with depression (Brooks et al., 2009; Matsuo et al., 2019). More recently, relationships have been observed between this region and suicidality (Jabbi et al., 2020; Zhao et al., 2021). Associations between the number of lifetime manic episodes and DLPFC volume have also been shown (Ekman et al., 2010). Decreased GMV in the inferior frontal regions are similarly noticed in BD-I patients and are associated with lifetime number of manic episodes (Ekman et al., 2010) and advanced age (Beyer et al., 2009; Stanfield et al., 2009).
Figure 1-2 A) Frontal brain regions as parcellated by the Desikan-Killiany Atlas (Desikan et al., 2006). Overlay of larger groupings include the prefrontal cortex (in blue), the inferior frontal cortex (in green) and the premotor area (in pink). Adapted with permission from R.S. Desikan et al. Neuroimage 31, 968-980 (2006); B) Subregions of the anterior cingulate cortex (ACC) include the subgenual ACC (yellow), pregenual ACC (pink), and anterior mid ACC (blue).
1.3.2.2 Temporal Lobes

The temporal lobes have received considerably less attention than the frontal lobes in BD, but numerous studies have included them in the past 15 years. Figure 1-3 presents a visual of temporal brain regions.

As with frontal regions, temporal regions were largely found to be smaller in patients with BD. When combining 16 VBM studies, a meta-analysis by Selvaraj et al. found decreased temporal GMV in patients (Selvaraj et al., 2012). Cortical thinning was also found in these regions (Lu et al., 2019; Phillips & Swartz, 2014). One systematic review reported GMV reductions in the superior temporal gyrus in patients with BD, and these reductions were associated with suicidality in the patient population (Bani-Fatemi et al., 2018). ENIGMA’s large-scale study found thinner CT in the fusiform gyrus (Hibar et al., 2018). This region’s main function is facial recognition and expression processing, which is necessary for proper interaction in social settings (Uppal et al., 2013).

Some smaller primary cross-sectional research studies provide insight into relationships between temporal brain changes and aspects of BD such as substance use disorder (SUD), cognitive function, and psychosis. A 2017 study looking at the effects of SUD found reduced GMV in the bilateral temporal cortices regardless of SUD status in BD-I patients, while frontal regions were associated with SUD only (Altamura et al., 2017). Decreased GMV in the right superior temporal region was also demonstrated associations with deficits in working memory (Sun et al., 2020).

Several studies have examined the association between psychotic symptoms, antipsychotic use and temporal lobe volumes. Two separate publications studying temporal regions, including the fusiform (Ekman et al., 2017), and temporo-parietal regions (Altamura et al., 2018), found GMV deficits in these regions in patients with BD with psychosis compared with patients with BD without psychosis and healthy controls, who showed no differences from each other. Another study found decreased GMV in the middle temporal gyrus but increased GMV in the parahippocampal gyrus of patients. Psychotic symptoms, lithium use, and family history were associated with these abnormalities (Chen et al., 2007). Specifically, hallucinations and other positive psychosis symptoms are related to GM reductions in the middle temporal region (Stanfield et al., 2009). Finally, a study analyzing changes in the temporal cortex and antipsychotic use in patients with BD found that current antipsychotic use was associated with larger bilateral temporal white matter
volume, but not grey matter volume (Jones et al., 2009). Overall, the temporal cortex pathology may be associated with the presence of psychosis symptoms, possibly due to the localization of the auditory cortex within the temporal lobes.
Figure 1-3 Temporal brain regions as parcellated by the Desikan-Killiany Atlas (Desikan et al., 2006). Adapted with permission from R.S. Desikan et al. Neuroimage 31, 968-980 (2006)
1.3.2.3 Subcortical Regions

Due to their involvement in cognition, emotion, and memory circuitry, subcortical structures have also received considerable attention in bipolar studies. Figure 1-4 presents an illustration of the subcortical regions.

Systematic reviews and meta-analyses have lent particular attention to the amygdala in BD-I neuroimaging studies. The amygdala is a key region in emotional processing circuitry. Decreased volume in the amygdala is a replicated finding in BD-I, with these decreases being especially marked during depressive episodes (Foland-Ross et al., 2012; Phillips et al., 2014). Amygdala alterations are seen in multiple age groups in patients with BD, including adults and pediatric populations (Haldane et al., 2004; Monkul et al., 2005). In another study conducted by the ENIGMA consortium, brains scans acquired in 1710 patients with BD and 2594 healthy controls were used to examine each of the eight subcortical structures. Across all patients with BD (including both BD-I & BD-II populations), this study showed consistently decreased GMV in patients with BD in the bilateral hippocampi and thalami. When looking at only the BD-I patients, significantly smaller GMV was observed in the hippocampus and amygdala. Patients taking lithium exhibited significantly larger thalami than patients not on lithium (Hibar et al., 2016). Smaller studies looking at other subcortical structures have observed smaller subcortical volume (SCV) in the putamen (Almeida et al., 2009) and the nucleus accumbens (Dickstein et al., 2005) of BD-I patients versus HCs.

Moreover, the thalamus has been extensively studied in patients with BD, with smaller GMV compared to HCs reported in studies using voxel-based morphometry (VBM) analysis methods (Lee et al., 2020; Nery et al., 2015; Sani et al., 2016) and reduced GMV and shape abnormalities in frontal-projecting regions of the thalamus (Skåtun et al., 2018). In fact, lower thalamus GMV has been found in patients who experienced adverse childhood events (Poletti et al., 2016), including both emotional and physical neglect (Duarte et al., 2016), and in patients with a history of SUD (Altamura et al., 2017). Conversely, two studies have demonstrated increased thalamus volumes in patients with BD (Adler et al., 2007; Chen et al., 2012); Adler et al. posit that this volume increase in their FEM population is likely due to elevated neuron count from an increase in projective excitatory neurons interacting with frontal regions, while Chen et
al. hypothesize that the enlargement of the thalamus of their male-only study participants is likely due to these patients being on lithium. Indeed, previous studies have shown that thalamus volume is increased in patients on lithium (Hibar et al., 2016), while lithium-naive patients had decreased volume relative to their medicated counterparts, who had similar thalamic volumes as HCs (Radenbach et al., 2010). The finding of larger subcortical structures in patients taking lithium versus those who do not was also found in the hippocampus and amygdala (Sani et al., 2018).

Some SCV neuroimaging studies have taken the age of patients with BD into account. A 2011 study found reduced GM concentrations in elderly euthymic patients in the caudate, nucleus accumbens and putamen (Haller et al., 2011). Congruently, discoveries of older patients with smaller hippocampal and amygdala volumes have been noted, demonstrating that the decrease in hippocampus volume was associated with longer duration of mood episodes (Wijeratne et al., 2013). Furthermore, in a study looking at psychotic BD, age-related GMV deficits were observed in the thalamus, putamen, and amygdala in the right hemisphere (Altamura et al., 2018). Meanwhile, another study found no difference in SCV of older patients with BD versus HC, but did note that these elderly patients had worse performance on tests of all cognitive domains (Rej et al., 2014). On the opposite end of the age spectrum, reduced left hippocampus and bilateral putamen GMV was observed in adolescent patients (MacMaster et al., 2014).
1.3.3 Longitudinal Neuroimaging

While cross-sectional studies provide important insights into structural brain changes in BD-I, they are confounded by several factors including long-term illness and medication effects, mood episode recurrence, and substance use, to name a few. For this reason, longitudinal studies have been conducted to better understand neuroprogression in bipolar disorder. To date, two reviews summarizing longitudinal studies in BD-I have been conducted; however, they incorporate findings from both studies starting at FEM and longitudinal studies starting at some later time-point in the disease course. This section focuses instead on the individual studies not starting at FEM.

A variety of frontal lobe structures display decreases over time. Over a period of approximately 2.5 years, one study saw bilateral GMV loss in ventral and rostral PFC, as well as in the ACC; this was especially prevalent in younger patients (Kalmar et al., 2009). Another study noted greater amounts of GMV reduction in the orbitofrontal cortex (OFC), rostral and dorsolateral PFC of adolescents with BD-I versus HC over a 2 year period (Najt et al., 2016). In patients who
attempted suicide during the 3-year follow-up period, Lippard et al. observed lower ventral and rostral PFC GMV at baseline, suggesting a predictive capability in these “future attempters” (Lippard et al., 2019). However, this study was conducted in a mixed cohort of BD-I and MDD patients, so the results are less generalizable to the larger BD-I population. When examining the effects of recurrent manic episodes, Abé et al. found decreased CT in the DLPFC and inferior frontal cortex in patients with recurrent manic episodes over a six year period, whereas patients who did not experience episodes actually showed increased CT in these regions (Abé et al. 2015).

Two studies specifically looked at medication effects on frontal regions over time. First, Moore et al. noted greater global GMV, especially in the PFC, of patients treated with lithium over a one-year follow-up period, while these same regions declined in patients who experienced depression during that time (Moore et al., 2009). Second, a comprehensive, longitudinal study by Lisy et al. noted differences both at baseline and over time: patients had smaller GMV in widespread frontal regions at baseline but showed increased ventrolateral PFC over the one-year follow-up. The use of atypical antipsychotics and anticonvulsants was associated with increased GMV in the left medial frontal region, but there were no associations with lithium use. The number of depressive episodes during the follow-up period correlated with increased GMV in the right cingulate cortex (Lisy et al., 2011).

Longitudinal research studying temporal regions are fewer and have more conflicting results. Moorhead et al. found a progressive loss in fusiform GMV over a two year period, and this temporal volume loss was associated with decreased verbal IQ measures (Moorhead et al., 2007). Contrarily, Lisy et al. observed an increase in medial temporal and parahippocampal volume (Lisy et al., 2011). Delaloye et al. found a steady entorhinal volume in BD-I patients over the 2-year follow-up period (Delaloye et al., 2011). A four-year longitudinal study noted a decrease in fusiform GM density over time in patients with BD, and those reductions were correlated with decline in intellectual functions and the number of recurrent episodes during that time (William et al., 2007).

Longitudinal analyses of subcortical regions in BD-I are also conflicting. When looking at the amygdala, lower volumes were noted (Blumberg et al., 2005), but changes were not progressive, while neither Delaloye et al. nor Moorhead et al. noted any change in amygdala volume (Delaloye et al., 2011; Moorhead et al., 2007). When examining the hippocampus, Delaloye et al. observed no change, but Yucel et al. observed increased hippocampus volume over
time in BD-I patients receiving lithium treatment (Delaloye et al., 2011; Yucel et al., 2007). Finally, Lisy et al. observed smaller right amygdala and bilateral caudate GMV in patients with BD at baseline, but then saw an increase in right caudate volume over time. They also observed a greater increase in right basal ganglia (including the caudate) and bilateral amygdala in patients with BD than in HCs over the one-year follow-up period, while HC SCV significantly decreased over time (Lisy et al., 2011). This study also attributes these SCV increases to lithium use among their patients. In contrast to the above studies, Emsell et al. found no group differences over time in grey matter regions in euthymic BD-I patients (Emsell et al., 2013).

1.4 The Systematic Treatment Optimization Program for Early Mania

As part of an ongoing, naturalistic prospective study, the Systematic Treatment Optimization Program for Early Mania (STOP-EM), clinical, cognitive, and neuroimaging data has been collected on a cohort of patients with BD-I and age- and sex-matched HCs since 2004. This case-control study allows for the examination of these data immediately following BD-I diagnosis in addition to over the long-term. Patients are recruited within three months of their diagnosis (FEM), receive clinical assessment, undergo cognitive tests using the CANTAB battery, and an MRI scan. Patients receive regular clinical and cognitive follow-up at 6-month intervals, and MRI scans are conducted at baseline (BL), years 1, 3, 5, 7, 10, 13, 15, 17, & 20. The protocol has been briefly described elsewhere (Yatham et al., 2009).

BD-I and HC volunteers participating in STOP-EM are used as the population of interest in the study in Chapter 3, with further description of the relevant protocol provided.

1.5 Outstanding Questions, Overview, and Research Goals

While existing longitudinal studies in BD-I suggest progressive changes over time, these have used patients at varying stages of illness, so it is not clear what changes happen early in the disease course. The more consistent and larger magnitude GM changes seen in cross-sectional studies in longer duration BD-I, compared to cross-sectional FEM studies, suggests that progressive GM changes evolve after FEM, but there has not been a systematic or rigorous examination of longitudinal changes starting at FEM to show early and neuroprogressive changes more clearly. Furthermore, cross-sectional and longitudinal studies in BD-I cohorts with longer duration of illness have shown that certain clinical and demographic factors, such as number and
type of mood episodes, illness duration, medication use, presence of psychotic features, and comorbidities, impact brain morphological changes. It remains unclear how these factors may influence neuroprogressive changes early in the disease course.

Lastly, few studies exist which examine CT and surface area (SA), either cross-sectionally or longitudinally. CT and SA are thought to be mediated by different ontogenetic processes (Panizzon et al., 2009); thus, examining these measures separately can give more precise information about the nature and etiology of morphological GM changes in BD-I patients. Therefore, the longitudinal study conducted in Chapter 3 examines changes in CT, SA, and SCV in BD-I patients during the first three years of the disorder, comparing these measures with those of healthy controls.

This thesis examines the regional neuroprogression of grey matter structures in BD-I patients early during the disease course. A systematic review of longitudinal MRI studies following FEM in BD-I can be found in Chapter 2. In addition to reviewing primary outcomes from extant longitudinal studies starting at FEM, this systematic review also examines the potential impact of clinical and demographic factors such as mood episode recurrence, psychosis, age of onset and medication use on reported results. In Chapter 3, cortical thickness and surface area in five cortical regions, and volumes in four sub-cortical a priori ROIs, are examined in a longitudinal experiment, following patients over the first three years of BD-I. These regions include the caudal ACC, the fusiform gyrus, the pars opercularis, the rostral ACC, and the rostral middle frontal gyrus, as well as the amygdala, caudate, putamen and thalamus.

This thesis seeks to address the following aims:

- **Aim 1**: Provide a qualitative overview of longitudinal GM neuroimaging studies of BD-I patients following FEM to provide guidance for future studies in the field.
- **Aim 2**: Assess trajectory of a priori ROIs in BD-I patients and healthy controls, over the first three years of disease, and compare measures over time between the two groups
  - We hypothesize that, over the 3-year period, a priori regions in BD-I patients will exhibit more loss of CT/SA/SCV than healthy controls. These differences are unlikely to be observed at baseline but will progress thereafter.
Chapter 2: Longitudinal Grey Matter Changes Following First Episode Mania in Bipolar I Disorder: A Systematic Review

2.1 Brief Introduction

BD-I is a chronic disease where patients experience mood episodes interspersed with periods of euthymia, during which psychosocial and cognitive impairments persist (Zarate et al., 2000). The course of BD-I is characterized by progressive worsening of clinical and functional outcomes (Solé et al., 2017), thus forming the basis for a clinical staging model, describing early, intermediate, and late stages of the disorder (Berk et al., 2017; Kapczinski et al., 2009; Yatham et al., 2018). Such staging models re-conceptualize BD-I as a neuroprogressive disorder and suggest that effective intervention at an early stage may help arrest further neurodegeneration. However, literature characterizing brain pathophysiology in the early stages of BD-I remains sparse, and studies following patients from FEM are even more rare. It remains unclear at what point during the illness these brain changes begin to manifest. Therefore, we conducted a systematic review of longitudinal studies examining BD-I patients starting at FEM and following-up with them over time. Our review consolidates findings from 15 longitudinal studies, focusing on GM structures in the brain. Noting how brain regions change longitudinally early in the disease course may be key to elucidating the mechanisms behind clinical progression in BD-I. These observations also have the potential to form the basis for developing treatment strategies that arrest neuroprogression, ameliorating clinical and functional outcomes in affected patients (Barbosa et al., 2014).

2.2 Methods

This systematic review was conducted in compliance with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) (Moher et al., 2010). The protocol for this project has been registered (PROSPERO registration No. CRD42red 20146562). As a novel examination of early structural changes in BD-I, we were purposively inclusive of studies spanning a wide range of imaging methodologies and structural indices, with a primary aim of examining the state of the literature. Thus, a meta-analysis was not planned in the protocol.

2.2.1 Search Strategy

Searches were conducted for articles in English or French in MEDLINE, EMBASE and Web of Science with no restriction of publication date. The primary search used keywords “bipolar
disorder” OR “mania” OR “manic” AND “magnetic resonance imaging” OR “MRI” OR “neuroimaging” AND “first episode” OR “early episode” OR “single episode” OR “recent onset” in all possible combinations. A secondary search for relevant articles was conducted by reviewing references of articles found in the primary search. Searches were conducted up to September 17, 2020.

2.2.2 Eligibility Criteria

The following inclusion criteria were applied: first, studies had to include a BD-I cohort with a current or recent FEM, with diagnosis established using standardized assessment criteria. If the cohort included BD-II patients, the data must have been reported or available separately for BD-I patients. In the case of studies using first episode affective psychosis (FEAF) cohorts, BD-I patients must have comprised at least 70% of the cohort. Second, studies were eligible if they included structural MRI of GMV (cortical or subcortical), CT, or SA longitudinally after FEM. Third, only longitudinal studies with a minimum one-year follow-up and which included a baseline comparison between FEM patients and a HC group, were included. We did not include studies that had baseline and follow-up patient groups that were comprised of different participants. Only peer-reviewed studies were included, and the grey literature was not searched. Reviews and meta-analyses were not included.

2.2.3 Study Selection, Data Collection, and Data Extraction

One reviewer (A.C) completed the entire search process. After removal of duplicate references, two reviewers (A.C. and T.C.) independently examined titles and abstracts to select studies for full text review. These same two authors independently reviewed full texts to determine eligibility for inclusion in the review. Consensus was reached between the two reviewers for all studies selected. Data from eligible studies were independently extracted and entered into a standardized template spreadsheet by two authors (A.C. and C.F.) (Microsoft Excel for Windows, version 16.0, Microsoft Corporation). The data extracted from these studies included sample size, participants’ clinical and demographic variables, the type of imaging and processing used on the images, the scanner strength (in Teslas (T)), slice thickness (in mm), length of longitudinal follow-up and relevant imaging findings (i.e. region, direction of volumetric/thickness/area change compared to healthy control group, level of significance and effect sizes, if reported).

Structural findings were considered relevant if they showed significant longitudinal structural change at a level of p≤0.01 for ROI studies, p<0.05 family-wise error (FWE) or false
discovery rate (FDR) corrected for VBM studies, OR if findings were based on well-established *a priori* knowledge in the field. The relevant structural findings were summarized in terms of change from baseline to follow-up, as well as between patients and controls at those time points. Authors of original studies were contacted to obtain missing data.

### 2.2.4 Subgroup Analysis

In addition, we extracted data such as: image processing type, age of onset (childhood/adolescent versus adult onset), episode recurrence (any episode recurrence versus no recurrence in the study time frame) and substance use comorbidities. Where study number permitted, we qualitatively compared results between these different subgroups.

### 2.2.5 Risk of Bias

Two reviewers (A.C. and C.F.) independently assessed the risk of bias of studies using a modified Newcastle-Ottawa Assessment Scale (Wells et al., 2013). We included “representativeness of the exposed cohort”, “ascertainment of exposure”, “comparability of cohorts”, “assessment of outcome”, “follow-up length”, and “adequacy of follow-up cohorts” items from the original scale for a total of seven (7) possible stars. “Selection of non-exposed cohort” and “outcome of interest not present at start” were excluded as they were not applicable to the types of studies included in our review. Studies were classified as having low risk of bias (6-7), high risk of bias (4-5), or very high risk of bias (0-3) based on a modified rating scale found in Ka-Lok Lo et al. 2014 (Lo, Mertz, & Loeb, 2014). Disagreements were resolved by consensus.

### 2.3 Results

#### 2.3.1 Study Selection

A total of 980 publications were originally identified following our primary search (Figure 2-1). Thirty-three studies remained for full text review after removal of duplicates and studies not otherwise meeting inclusion criteria. Eighteen articles were further excluded following full text review for the following reasons: not an original research article (1), baseline imaging not conducted at FEM (5), clinical follow-up only with no repeat MRI scan (2), bipolar spectrum diagnosis (1), and inclusion only of patients with primary psychotic disorders (7). No eligible studies were found in secondary search. Thus, 15 original research articles were included in this systematic review (Arango et al., 2012; Berk et al., 2017; Bitter et al., 2011; Bond et al., 2019; Castro-Fornieles et al., 2018; de Castro-Manglano et al., 2011; Farrow et al., 2005; Kasai et al.,
2.3.2 Study Characteristics

Of the 15 studies included, 7 examined only BD-I and HC cohorts and the other 8 included cohorts with FEAF and first episode psychotic disorder (see Table 2-1 for study characteristics). In FEAF studies, the percentage with a BD-I diagnosis ranged from 71%-93% (de Castro-Manglano et al., 2011; Kasai et al., 2003a; Kasai et al., 2003b; Lee et al., 2016; Nakamura et al., 2003a; Kasai et al., 2003b; Koo et al., 2008; Kozicky et al., 2016; Lee et al., 2016; Nakamura et al., 2007; Ohtani et al., 2018; Salisbury et al., 2007).
Results were extracted exclusively for BD-I or FEAF cohorts, and HC. Twelve studies’ exclusively enrolled FEM patients displaying psychotic features.

Participants aged 7-55 years were included in the studies. The average age of patients was 21.2 ± 3.72 years and 20.4 ± 3.38 for HC. Two studies exclusively recruited pediatric patients (Arango et al., 2012; Bitter et al., 2011), ranging from 7-17 years. Four studies had mixed pediatric and adult samples (Berk et al., 2017; Bond et al., 2014; de Castro-Manglano et al., 2011; Farrow et al., 2005), with ages ranging from 11 to 35 years. Study sample sizes ranged from 8 to 55 for patients and 17 to 70 for controls. The average duration from FEM was 1.86 ± 5.54 months before the baseline MRI (Arango et al., 2012; Berk et al., 2017; Castro-Fornieles et al., 2018; de Castro-Manglano et al., 2011; Farrow et al., 2005; Ohtani et al., 2018). Follow-up times ranged from 1-3 years, with a mean of 1.6 ± 0.51 years.

Two sets of studies had overlapping patient samples. First, Kozicky et al. (2016) and Bond et al. (2019) used patients from the same overarching study. Both papers were included as the sample sizes were different and they used different image analysis techniques – Kozicky et al. used VBM, while Bond et al. used Freesurfer v6.0. Kasai et al. (2003a, 2003b), Salisbury et al. (2007) and Nakamura et al. (2007) similarly used overlapping participants. However, while both of Kasai et al.’s papers used the same participants, each study reported on different brain structures (Kasai et al., 2003a, 2003b). Salisbury et al., while looking at similar regions as Kasai et al. (2003a), overlapped in less than half of their FEAF patient cohorts. Finally, while participants used by Nakamura et al. (2007) greatly overlapped with the three other studies, they examined neocortical gray matter regions separated by lobes, whereas the other studies examined more targeted structures.

Ten of the 15 studies reported using 1.5T MRI scanners (Arango et al., 2012; de Castro-Manglano et al., 2011; Farrow et al., 2005; Kasai, Shenton, Salisbury, Hirayasu, Lee, et al., 2003; Kasai, Shenton, Salisbury, Hirayasu, Onitsuha, et al., 2003; Koo et al., 2008; Lee et al., 2016; Nakamura et al., 2007; Ohtani et al., 2018; Salisbury et al., 2007), three studies using 3.0T (Berk et al., 2017; Bond et al., 2014; Kozicky et al., 2016), one using 3.0T and 4.0T scanners (Bitter et al., 2011) (see Table 2-1). Data was not available for one study (Castro-Fornieles et al., 2018). Most studies used a slice thickness of 1.5mm, except for three which used 1mm (Berk et al., 2017; Bitter et al., 2011; Bond et al., 2014) for their T1 images and 3.0mm for their T2 images, save for one study that used 3.5mm (Arango et al., 2012). Tissue measurements were obtained using four
different methods: one study used semiautomatic segmentation based on the Talairach atlas (Arango et al., 2012), five used voxel-based morphometry (M Berk et al., 2017; Castro-Fornieles et al., 2018; de Castro-Manglano et al., 2011; Farrow et al., 2005; Kozicky et al., 2016), eight used manual ROI placements (Bitter et al., 2011; Kasai et al., 2003a; Kasai, et al., 2003b; Koo et al., 2008; Lee et al., 2016; Nakamura et al., 2007; Ohtani et al., 2018; Salisbury et al., 2007) and one study used the longitudinal automatic pipeline in Freesurfer v6.0 (Bond et al., 2014). All studies examined gray matter volume, with none reporting on cortical thickness or surface area. This information is summarized in Table 2-1.
<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design</th>
<th>Number of Participants in longitudinal analysis</th>
<th>Average Age (years) ± SD</th>
<th>Average months after FEM ± SD</th>
<th>Follow-Up Time (Months) ±SD</th>
<th>Scanner Strength (Teslas)</th>
<th>Slice Thickness (mm)</th>
<th>Image Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arango, 2012</td>
<td>PN-HC</td>
<td>16, 6F</td>
<td>16.6±0.9</td>
<td>15.3±1.5</td>
<td>2.5±2.4</td>
<td>25.6±2.2</td>
<td>1.5T</td>
<td>1.5 (T1), 3.5 (T2)</td>
</tr>
<tr>
<td>Berk, 2017</td>
<td>Single blind RCT of lithium vs quetiapine; HC included for longitudinal comparison</td>
<td>30, 7F</td>
<td>21.5±2.4</td>
<td>21.4±2.5</td>
<td>0.4±0.6</td>
<td>12.0±NR</td>
<td>3T</td>
<td>1.0 (T1)</td>
</tr>
<tr>
<td>Bitter, 2011</td>
<td>PN-HC</td>
<td>17, 11F</td>
<td>15.0±1.2</td>
<td>NR</td>
<td>NR</td>
<td>12.8±1.1</td>
<td>3T &amp; 4T</td>
<td>1.0 (T1)</td>
</tr>
<tr>
<td>Bond, 2019</td>
<td>PN-HC</td>
<td>55, 30F</td>
<td>22.6±4.4</td>
<td>22.4±4.0</td>
<td>Within 3 mos</td>
<td>Within 12 mos</td>
<td>3T</td>
<td>1.0 (T1)</td>
</tr>
<tr>
<td>Castro-Fornieles, 2019</td>
<td>PN-HC</td>
<td>15, 6F</td>
<td>16.5±0.7</td>
<td>15.3±1.5</td>
<td>2.8±2.4</td>
<td>15.6±2.4</td>
<td>Not Reported</td>
<td>VBM</td>
</tr>
<tr>
<td>de Castro-Manglano, 2011</td>
<td>PN-HC</td>
<td>10, F NR</td>
<td>18.5±4.0</td>
<td>19.7±6.3</td>
<td>9.9±18.2</td>
<td>32.8±12.7</td>
<td>1.5T</td>
<td>1.5 (T1), 3.0 (T2)</td>
</tr>
<tr>
<td>Study</td>
<td>Study Design</td>
<td>Number of Participants in longitudinal analysis</td>
<td>Average Age (years) ± SD</td>
<td>Average months after FEM ± SD</td>
<td>Follow-Up Time (Months) ±SD</td>
<td>Scanner Strength (Teslas)</td>
<td>Slice Thickness (mm)</td>
<td>Image Processing</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
<td>----------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Farrow, 2005</td>
<td>Prospective, naturalistic; no longitudinal comparison with HC</td>
<td>8, 4F</td>
<td>22, 9F</td>
<td>17.5±2.0</td>
<td>20.5±4.0</td>
<td>0.4±0.4</td>
<td>1.5T</td>
<td>1.0 (T1)</td>
</tr>
<tr>
<td>Kasai, 2003a</td>
<td>PN-HC</td>
<td>13, 1F</td>
<td>22, 2F</td>
<td>21.8±2.9</td>
<td>25.0±4.3</td>
<td>NR</td>
<td>17.7±8.1</td>
<td>1.5T</td>
</tr>
<tr>
<td>Kasai, 2003b</td>
<td>PN-HC</td>
<td>13, 1F</td>
<td>22, 2F</td>
<td>21.8±2.9</td>
<td>25.6±4.7</td>
<td>NR</td>
<td>17.7±8.1</td>
<td>1.5T</td>
</tr>
<tr>
<td>Koo, 2008</td>
<td>PN-HC</td>
<td>17, 3F</td>
<td>18, 3F</td>
<td>22.8±4.5</td>
<td>23.0±3.2</td>
<td>NR</td>
<td>18.3±3.9</td>
<td>1.5T</td>
</tr>
<tr>
<td>Kozicky, 2016</td>
<td>PN-HC</td>
<td>41 (21 BD-R, 20 BD-W), 21F</td>
<td>BD-I only</td>
<td>25</td>
<td>22.9±4.0</td>
<td>22.0±4.0</td>
<td>NR</td>
<td>12.6±1.2</td>
</tr>
<tr>
<td>Lee, 2016</td>
<td>PN-HC</td>
<td>20, 3F</td>
<td>23, 4F</td>
<td>22.7±5.1</td>
<td>24.2±3.9</td>
<td>NR</td>
<td>16.6±6.1</td>
<td>3T</td>
</tr>
<tr>
<td>Nakamura, 2007</td>
<td>PN-HC</td>
<td>NR</td>
<td>26, 4F</td>
<td>22.4±3.2</td>
<td>24.6±4.1</td>
<td>NR</td>
<td>18.7±9.6</td>
<td>1.5T</td>
</tr>
<tr>
<td>Ohtani, 2018</td>
<td>PN-HC</td>
<td>22, 4F</td>
<td>23, 5F</td>
<td>23.3±4.9</td>
<td>23.0±3.6</td>
<td>0.2±0.3</td>
<td>18.6±10.2</td>
<td>1.5T</td>
</tr>
<tr>
<td>Salisbury, 2007</td>
<td>PN-HC</td>
<td>13, 3F</td>
<td>32, 10F</td>
<td>21.8±5.0</td>
<td>24.1±3.7</td>
<td>Mean 1.7 mos</td>
<td>18</td>
<td>1.5T</td>
</tr>
</tbody>
</table>
Abbreviations: BD-I, Bipolar I Disorder; HC, Healthy Controls; BD-R, bipolar I disorder patients with recurrent mood episodes; BD-W, bipolar I disorder patients without recurrent mood episodes; FEAF, First Episode Affective Psychosis patients; F, Females; NR, Not Reported; SD, Standard Deviation; ROI, Region of Interest; SPM, Statistical Parametric Mapping; VBM, Voxel-Based Morphometry

† studies whose reported findings pertain to all FEAF patients combined i.e., do not report BD-I findings separately. The total number of FEAF participants is indicated.

‡ studies that noted that conducting analysis with BD-I patients only did not alter results.

N.B. Patient numbers are reported for BD-I sample only. Patient ages are reported for BD-I groups, or for the total FEAF cohort when individual data was not available.
2.3.3 Risk of Bias within Studies

All studies were assigned a “low risk of bias” according to the Newcastle-Ottawa Scale (6-7/7). Castro-Fornieles et al. (2018), Farrow et al. (2005), Kasai et al. (2003b), Koo et al. (2008), and Kozicky et al. (2016) received a score of 6. The remainder of the low-risk studies were assigned a score of 7.

2.3.4 Baseline GM Comparison between FEM Patients and Controls

A majority of the studies did not detect differences in GMV between FEM patients and HC at baseline (study results summarized in Table 2-2). Five studies that conducted whole-brain analyses using VBM or other semiautomatic segmentation methods (Arango et al., 2012; Farrow et al., 2005; Kozicky et al., 2016, Bond et al., 2019; de Castro-Manglano et al., 2011) failed to demonstrate baseline differences. Similarly, six ROI studies – examining the Heschl gyrus and planum temporale (Kasai, et al., 2003a; Salisbury et al., 2007), the superior temporal gyrus and amygdala-hippocampal complex (Kasai et al., 2003b), the cingulate (Koo et al., 2008), insula and temporal pole (Lee et al., 2016), and multiple frontal lobe regions (Ohtani et al., 2018) - detected no baseline differences between FEM patients and HC.

However, four studies reported volume decreases in the FEM cohort compared to HCs at baseline. Using VBM, Berk et al (2017) found smaller volumes in the right orbitofrontal cortex, the bilateral ACC, inferior frontal gyri, and bilateral cerebellar GMV (Berk et al., 2017) in FEM patients. Farrow et al.’s (2005) VBM study noted smaller volumes in right frontal, bilateral temporal and left parietal areas. Although a significant difference was not noted in the full cingulate, Koo et al. (2008) observed smaller left and right subgenual subregions (Koo et al., 2008) in patients using manually placed ROIs. Finally, Nakamura et al. (2007) noted a 3.9% decrease in total neocortical grey matter in BD-I patients compared with HCs (Nakamura et al., 2007) using a manual ROI tracing method. All four of the studies reporting baseline GMV differences compared to HC exclusively examined FEM patients with psychosis, initially suggesting the presence of psychotic features as contributory to GMV baseline alterations. However, nine of the 11 (73%) studies not detecting baseline differences similarly only recruited FEM patients with psychosis. No salient differences in illness duration, comorbidity or other clinical/demographic features were apparent between studies detecting baseline differences versus those that did not.
2.3.5 Longitudinal GM Changes in BD-I Patients Following FEM

Results regarding longitudinal GM changes from the 15 included studies were mixed (Table 2-2). Some studies observed no longitudinal changes between baseline and follow-up volumes (Arango et al., 2012; Berk et al., 2017; Kasai et al., 2003a; Kasai et al., 2003b; Lee et al., 2016; Ohtani et al., 2018; Salisbury et al., 2007), while others did in a variety of areas. All studies compared longitudinal changes with that of a HC group over the same time period, save for Farrow et al. (2005) (Farrow et al., 2005).

2.3.5.1 Longitudinal Frontal GM Changes

Four studies demonstrated a loss in cingulate volume from baseline to follow-up in FEM patients versus HC. Kozicky et al. (2016) observed decreases in ACC volume in FEM patients who experienced a recurrent mood episode in the one year follow up period, a change that was not seen in patients with BD who remained well or in HCs (Kozicky et al., 2016). Similarly, Farrow et al. (2005) observed progressive decreases in ACC volumes over two years in FEM patients (although this study only included 8 BD-I patients and had no longitudinal comparison with an HC group) (Farrow et al., 2005). Both Kozicky et al. (2016) and Farrow et al. (2005) were VBM studies; however, it should be noted that the three other VBM studies included in this review did not detect longitudinal ACC changes (Berk et al. 2017; Castro-Fornieles et al. 2019; de Castro-Manglano et al. 2011). Both ROI studies that parcellated the ACC reported longitudinal GM changes in this region. Koo et al. (2008) noted greater volume loss in bilateral subgenual ACC in FEM patients over 1.5 years compared with HC. Bond et al. (2019) found that left ACC volume loss in the year following FEM was specific to BD-I patients who experienced clinically significant weight gain over the study period, while patients without clinically significant weight gain showed a similar increase in ACC volumes as HC.

Additionally, Kozicky et al. (2016) noted significant GMV loss in the left precentral gyrus, inferior frontal gyrus and operculum, but again only in BD-I patients who had a recurrence of mood episodes (Kozicky et al., 2016). Over a period of two years, Castro-Fornieles et al. (2018) observed greater GMV loss in a cluster that encompassed the right rectus gyrus and right midline anterior part of the superior frontal gyrus in BD-I patients compared with controls (Castro-Fornieles et al., 2018). Contrarily, the study by Nakamura et al. (2007), which used a manual segmentation protocol to isolate neocortical grey matter, observed a significantly greater increase in frontal lobe neocortical GM of BD-I patients over a 1.5-year duration versus HC. Further
analyses in this cohort revealed that the use of mood stabilizing medications in combination with an antipsychotic may have been driving this increase – FEAF patients in this cohort treated with a mood stabilizer and antipsychotic combination showed a 6.1% increase in overall neocortical GM, compared to a -0.2% change in unmedicated patients.

### 2.3.5.2 Longitudinal Parietal-Occipital GM Changes

Two studies reported longitudinal changes in parietal areas. Kozicky et al. (2016) observed significantly greater GMV loss in the right inferior parietal and supramarginal gyrus in BD-I patients with recurrent mood episodes compared with HC (Kozicky et al., 2016). De Castro-Mangalo et al. (2011) revealed trends for longitudinal loss in the whole parietal and occipital lobes over a three-year period (de Castro-Manglano et al., 2011). Castro-Fornieles et al. (2018) showed a greater decrease in left superior occipital GMV over time in BD-I patients compared with HC, but at follow-up, the measurements did not differ significantly ( Castro-Fornieles et al., 2018). Nakamura et al. (2007) detected an increase in neocortical GMV in the combined parietal and occipital lobes in BD-I patients versus HC over a 1.5-year period, which again appeared to be secondary to medication effects.

### 2.3.5.3 Longitudinal Temporal GM Changes

Three studies noted changes in temporal areas over their respective follow-up periods. Bond et al. (2019) observed a decrease in the left inferior temporal gyrus over one year, again only in patients who experienced clinically significant weight gain (Bond et al., 2019). In patients with episode recurrence, Kozicky et al.’s (2016) study revealed progressive loss in the right superior temporal gyrus, as well as greater loss than HC in the left superior and middle temporal gyri (Kozicky et al., 2016). Whole temporal lobe GMV showed trends in reduction over three years in de Castro-Mangalo et al. (2011) study (de Castro-Manglano et al., 2011).

### 2.3.5.4 Longitudinal Subcortical GM Changes

Left and right amygdala trajectory significantly differed between BD-I patients and HC in Bitter et al.’s (2011) adolescent mania study. No change was seen in bilateral amygdala volumes in the BD-I group, whereas the HC group showed a significant increase in amygdala volumes over one year. Both the left and right amygdala were significantly larger in the HC versus patient group at follow-up (Bitter et al., 2011). Castro-Fornieles et al. (2018) noted a larger decrease in BD-I patients versus HC in the left caudate and right putamen over two years, but measurements were not significantly different at follow-up ( Castro-Fornieles et al., 2018).
Table 2-2 Results for differences at baseline and during longitudinal follow up from the 15 studies included in the systematic review

<table>
<thead>
<tr>
<th>Study</th>
<th>Region(s) examined</th>
<th>Baseline Differences</th>
<th>Longitudinal Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Significant Regions</td>
<td>Direction</td>
<td>p-value</td>
</tr>
<tr>
<td>Arango et al., 2012</td>
<td>Total and bilateral frontal, parietal, and temporal lobes</td>
<td>None</td>
<td>N/A</td>
</tr>
<tr>
<td>Berk et al., 2017</td>
<td>Whole brain VBM</td>
<td>Right OFC, bilateral ACC, bilateral IFG</td>
<td>Patients smaller than HC</td>
</tr>
<tr>
<td>Bitter et al., 2011</td>
<td>Bilateral amygdala</td>
<td>None</td>
<td>N/A</td>
</tr>
<tr>
<td>Bond et al., 2019</td>
<td>Bilateral OFC, cingulate gyri, superior/middle/ inferior temporal gyri, fusiform gyri, parahippocampal gyri</td>
<td>None</td>
<td>N/A</td>
</tr>
<tr>
<td>Study</td>
<td>Region(s) examined</td>
<td>Baseline Differences</td>
<td>Longitudinal Changes</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Castro-Fornieles et al., 2019</td>
<td>Whole brain VBM</td>
<td>None</td>
<td>Significant Regions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direction</td>
<td>Direction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Effect Size/ Magnitude of difference</td>
<td>Effect Size/ Magnitude of difference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>1. Right putamen, rectus gyrus, midline anterior part of SFG 2. Left caudate 3. Left SOG</td>
<td>Larger decrease over time in BD-I patients in all regions compared to HC</td>
<td>1. p&lt;0.001 2. p=0.003 3. p=0.006 FWE corrected</td>
</tr>
<tr>
<td>De Castro-Mangalano et al., 2011</td>
<td>Whole brain VBM</td>
<td>None</td>
<td>No difference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No difference</td>
<td>No difference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>1. Right IFG/PCG 2. Bilateral ITG/uncus, insula, posterior ITG 3. Left PCC</td>
<td>Smaller than HC</td>
<td>Significant decrease in BD-I patients over time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All p&lt;0.001 uncorrected</td>
<td>p&lt;0.001 uncorrected</td>
</tr>
<tr>
<td>Farrow et al., 2005</td>
<td>Whole brain VBM</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No difference</td>
<td>No longitudinal changes in patients compared to HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Kasai et al., 2003a</td>
<td>1. Bilateral Heschl gyrus 2. Bilateral PT</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>No longitudinal changes in patients compared to HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Kasai et al., 2003b</td>
<td>1. Bilateral STG 2. Bilateral amygdala-hippocampal complex</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>No longitudinal changes in patients compared to HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Study</td>
<td>Region(s) examined</td>
<td>Baseline Differences</td>
<td>Longitudinal Changes</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------</td>
<td>-----------------------------------------------------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Significant Regions</td>
<td>Significant Regions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direction</td>
<td>Direction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p-value</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Effect Size/Magnitude of difference</td>
<td>Effect Size/Magnitude of change</td>
</tr>
<tr>
<td><strong>Koo et al., 2008</strong></td>
<td>Bilateral subgenual, anterorostral, anterodorsal and posterior cingulate cortex</td>
<td>Bilateral subgenual ACC</td>
<td>Greater bilateral volume loss in patients compared to HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smaller than controls</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L: p&lt;0.001</td>
<td>L subgenual ACC -5.67% vs -0.36% in HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: p=0.002</td>
<td>R subgenual ACC -5.29% vs -0.27% in HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L: d=0.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: d=0.67</td>
<td></td>
</tr>
<tr>
<td><strong>Kozicky et al., 2016</strong></td>
<td>Whole brain VBM</td>
<td>None</td>
<td>Significant loss in volume over time in BD-I patients with episode recurrence contrasted to HC in regions 1-3. Significant volume loss in BD-I patients with episode recurrence in regions 4-6, with similar changes not seen in HC. No changes in BD-I patients without episode recurrence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>1. p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>2. p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>3. p&lt;0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>4. p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>5. p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>6. p&lt;0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>FWE corrected</td>
</tr>
<tr>
<td><strong>Lee et al., 2016</strong></td>
<td>1. Insula</td>
<td>None</td>
<td>No longitudinal changes in patients compared to HC</td>
</tr>
<tr>
<td></td>
<td>2. Temporal Pole</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Nakamura et al., 2007</strong></td>
<td>Total and lobar NCGM</td>
<td>Total NCGM</td>
<td>BD-I patients increase in NCGM in all regions compared to HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smaller than controls</td>
<td>1. +3.6% vs 0.05% HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p=0.004</td>
<td>2. +2.8% vs -0.3% HC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.9% less</td>
<td>3. +4.6% vs +0.2% HC</td>
</tr>
<tr>
<td>Study</td>
<td>Region(s) examined</td>
<td>Baseline Differences</td>
<td>Longitudinal Changes</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td></td>
<td>Significant Regions</td>
<td>Direction</td>
<td>p-value</td>
</tr>
<tr>
<td>Ohtani et al., 2018</td>
<td>1. Frontal pole 2. SFG 3. MFG 4. IFG</td>
<td>None</td>
<td>N/A</td>
</tr>
<tr>
<td>Salisbury et al., 2007</td>
<td>Heschl Gyrus</td>
<td>None</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Abbreviations: ACC, Anterior Cingulate Cortex; BD-I, Bipolar I Disorder patients; CSWG, Clinically-significant weight gain; CG, cingulate gyrus; d, Cohen’s d; FWE, Family-Wise Error; HC, healthy controls; IFG, Inferior Frontal Gyrus; IPG, Inferior Parietal Gyrus; ITG, Inferior Temporal Gyrus; MFG, Middle Frontal Gyrus; MTG, Middle Temporal Gyrus, NCGM, Neocortical Grey Matter; OFC, Orbitofrontal Cortex; PCC, Posterior Cingulate Cortex; PCG, Precentral Gyrus; PT, Planum Temporale; SFC, Superior Frontal Cortex; SMG, Supramarginal Gyrus; SOG, Superior Occipital Gyrus; STG, Superior Temporal Gyrus; VBM, Voxel-Based Morphometry
2.3.6  **Subgroup Analyses**

2.3.6.1  **VBM vs ROI**

Two major classes of imaging analysis used in included studies were voxel-based morphometry VBM and manual ROI segmentation. There was no pattern in findings (i.e. studies detecting significant longitudinal changes versus not) that distinguished one type of analysis from the other. Two VBM studies reported no significant longitudinal changes (Berk et al., 2017; de Castro-Manglano et al., 2011), while three studies noted changes over time in a variety of areas, especially frontal regions (Castro-Fornieles et al., 2018; Farrow et al., 2005; Kozicky et al., 2016). Similarly, five manual ROI studies showed no significant longitudinal changes (Kasai et al., 2003a; Kasai et al., 2003b; Lee et al., 2016; Ohtani et al., 2018; Salisbury et al., 2007), but three did, specifically in the amygdala (Bitter et al., 2011), the subgenual ACC (Koo et al., 2008), and neocortical gray matter. (Nakamura et al., 2007)

2.3.6.2  **Episode Recurrence and Relationship to Longitudinal GM Change**

One study (Kozicky et al. 2016) evaluated the relationship between recurrences of mood episode during the follow-up period and longitudinal structural brain changes, which found that patients with any episode recurrence experienced longitudinal changes compared to HC, while patients that remained well followed a similar trajectory to HC (see Table 2-2). Additionally, Bond et al. found that more days spent in a depressive episode in the one-year follow up period was associated with greater volume loss in the left middle temporal gyrus (Bond et al., 2019). In a post-hoc analysis, Bitter et al. (2011) noted that there was no difference in amygdala volume change between the 7 BD-I adolescent patients who experienced an additional mood episode during the one year follow up and the 10 who did not. The remaining 12 studies did not report on the association between episode recurrence and longitudinal GM changes.

2.3.6.3  **Substance Use**

Two studies reported on the possible effect of substance use on longitudinal GM morphology. Bitter et al.’s (2011) adolescent study found no significant baseline or longitudinal GMV differences between the six patients who developed a substance use disorder during the one year follow up and the 11 who did not (Bitter et al., 2011). Kozicky et al. (2016) noted a higher rate of substance abuse/dependence in patients with BD who experienced a recurrence (the patient cohort which showed longitudinal changes in GMV compared to HC) compared to patients with
BD who remained well. The remainder of studies either excluded participants with a history of substance use or did not report on any associations between substance use and longitudinal GM changes.

2.3.6.4 Psychotic Features in FEM

Twelve of the reviewed studies only included BD-I patients experiencing psychotic features during FEM. Of the remaining three studies, none reported separate data in patients with or without psychotic features.

2.3.6.5 Pediatric vs Adult-Onset BD-I

Two studies exclusively recruited pediatric patients. Arango et al.’s (2012) study, which included participants aged 7-17, did not find any longitudinal differences between youth with BD-I and age matched HC. Indeed, patients’ negative lobar volume change was in line with healthy development. However, this study only looked at whole lobe volumes, and so some changes in more specific regions could have been missed. In contrast, Bitter et al. (2011), who included adolescents aged 12-17, observed that while the amygdala volumes of HC increased over one year, amygdala volumes decreased in BD-I youth. There were no detectable baseline differences, suggesting that negative trajectory in amygdala volume in BD-I patients began after FEM.

Four additional studies which included pediatric participants (Berk et al. [15-25 years] (Berk et al., 2017), Bond et al. [15-25] (Bond et al., 2019), Farrow et al. [13-25] (Farrow et al., 2005), and de Castro-Mangalo [11-29]) (de Castro-Manglano et al., 2011) did not report separate data for pediatric versus adult subgroups.

2.3.6.6 Depressive Episodes Prior to FEM

None of the reviewed studies included prior depressive episodes in their analysis.

2.4 Discussion

We conducted a systematic review of longitudinal GM changes in BD-I patients following the first episode of mania. To the best of our knowledge, there has not been a systematic review examining longitudinal structural brain changes, with an exclusive focus on BD-I patients, in the early stages of the disorder following FEM. Therefore, our review singularly adds to the existing body of literature by exclusively examining longitudinal studies of BD-I patients starting at FEM. Of the 15 papers meeting our inclusion criteria, the most commonly reported finding was of lower ACC and inferior frontal gyrus volumes in patients at baseline (Berk et al., 2017; Farrow et al.,
2005; Koo et al., 2008) and longitudinal decreases in ACC following FEM (Bond et al., 2019; Farrow et al., 2005; Koo et al., 2008; Kozicky et al., 2016), consistent with findings in prior meta-analysis of baseline FEM changes (Bora et al., 2010; Keramatian et al., 2020).

Though these results suggest that pathologic structural changes in brain regions involved in emotional processing and regulation evolve in the months following FEM, it should be noted that any replicated positive findings were counterbalanced by a number of studies which did not report changes in these regions. For instance, longitudinal changes reported in the ACC and inferior frontal gyrus are suggestive, given the critical role these regions play in emotional regulation and decision making (Hiser and Koenigs, 2018; Howard et al., 2019; Nusslock et al., 2015). Both the ACC and the inferior frontal gyrus in particular have been shown to be involved in emotional regulation and emotional perception (Dixon et al., 2017; Tabei, 2015; Whalen et al., 1998), as well as having important connections with the amygdala (Banks et al., 2007; Dolcos et al., 2006). Previous studies have also reported smaller volumes and patterns of abnormal activation in the ACC in patients with BD with longer duration of illness (Bürger et al., 2017; Keramatian et al., 2020), and a recent meta-analysis found evidence for decreased ACC volumes in FEM patients versus healthy comparators (Keramatian et al., 2020). Both studies that included the ACC as a ROI, as well as two VBM studies, found longitudinal decreases in FEM patients compared to HC, suggesting that structural changes in the ACC evolve in the early phase of the illness following FEM (Berk et al., 2017; Bond et al., 2019; Koo et al., 2008; Kozicky et al., 2016). However, three VBM studies did not detect changes in the ACC, raising questions regarding the reliability of this finding. Similarly, inferior frontal gyrus volumes were found to be smaller at baseline in two VBM studies (Berk et al. 2017; Farrow et al. 2005). This finding is in line with a large-scale, cross-sectional study of BD-I, which found significantly smaller inferior frontal cortical thickness when compared with HC (Hibar et al., 2018), and studies finding decreased volumes and abnormal connectivity in this region (Roberts et al., 2017; Townsend et al., 2012). However, three VBM studies and one ROI study of the inferior frontal gyrus did not detect any baseline differences (Castro-Fornieles et al. 2019; De Castro-Mangalano et al. 2011; Kozicky et al. 2016; Ohtani et al. 2018), and longitudinal reductions in this region were only reported during the follow up period in one study (Kozicky et al. 2016). Though our review furthers evidence that pathology in the ACC, in particular, in BD-I may be present by the first episode and become more pronounced in the early progression of the disorder, thus representing a potential biological substrate for clinical
and functional deterioration, findings across studies were in general too inconsistent to be regarded as definitive.

Interestingly, there were few consistent findings in subcortical structures such as the amygdala, hippocampus and basal ganglia, despite their known role in emotional processing (Schultz, 2016; Sergerie et al., 2008). While two of the reviewed studies examined amygdala volume (Bitter et al., 2011; Kasai et al., 2003a), only one detected longitudinal volumetric differences compared to HC. Bitter et al.’s 2011 study, which assessed adolescent amygdala volumes following FEM, was particularly interesting (Bitter et al., 2011). This study showed that while both BD-I patients and HC had similar amygdala volumes at baseline, the trajectory of volume change in BD-I youth differed significantly from HC as well as a cohort with attention deficit-hyperactive disorder. While BD-I patients displayed a significant decrease in amygdalae volumes over one year following FEM, volumes in the two other groups increased. This provided interesting preliminary evidence that a lack of developing appropriate amygdala volume may be a specific, early pathophysiologic characteristic of BD-I, aligning with the amygdala’s known role in emotional processing and connections with emotional frontal circuitry. Though suggestive, however, this finding was not replicated by any other study reviewed here.

The lack of consistently detected changes across the studies reviewed may itself be a point of interest. Longitudinal studies following first episode schizophrenia fairly consistently show progressive decrements in cortical GMV, albeit with variability in terms of specific regions affected (Vita et al. 2012). The number of negative studies (n=7 out of 15) in this review, as well as the localized nature of regional changes in those studies that do report longitudinal alterations following FEM, suggest that neuroprogression in early BD-I may be less marked compared to early schizophrenia. Indeed, most reviewed studies that also included a first episode schizophrenia arm reported more pronounced and widespread longitudinal GM changes compared to first episode BD-I (Arango et al. 2012; Castro-Fornieles et al. 2019; Farrow et al. 2005; Kasai et al. 2003a; Kasai et al. 2003b; Koo et al. 2008; Lee et al. 2016; Nakamura et al. 2007; Ohtani et al. 2017; Salisbury et al. 2007). Larger comparative studies can validate differences in neuroprogression early in non-affective psychotic versus affective disorders, and clarify whether these may be due to pathophysiologic differences or other clinical (i.e. medication) factors.

Though we provide a novel synthesis of GM structural changes (or lack thereof) in the early stages of BD-I, significant limitations in the literature base should be acknowledged. First,
few studies fit the inclusion criteria, and included studies had small sample sizes (8-55 BD-I patients). Samples were mixed in terms of primary diagnosis as well, with some studies examining an “affective psychosis” group including BD-I patients and major depressive disorder patients with psychosis - although studies included in this review were required to have at least 70% of participants with BD-I (de Castro-Manglano et al., 2011; Kasai et al., 2003a, 2003b; Koo et al., 2008; Lee et al., 2016; Nakamura et al., 2007; Ohtani et al., 2018). While some of these papers indicated that restricting the analysis to only BD-I patients did not change the results, not all of these studies reported BD-I specific results. Furthermore, many studies included only BD-I patients displaying psychotic features in FEM, limiting the generalizability of the results, and there was wide variability/limited information in other potentially relevant patient factors such as age, substance use and episode recurrence (discussed further below). Finally, the studies reviewed looked at many different brain regions, such that their findings differed greatly from one another. Due to these limitations, results were highly heterogenous with little replicability of specific brain regions amongst studies. Thus, the limitations in the extant literature preclude drawing any definitive conclusions.

Planned subgroup analyses, including clinical factors such as episode recurrence, psychosis, substance use, age, and prior depressive episodes which may impact longitudinal GM changes, were not completed due to limited reporting on these variables in the primary literature. Only one reviewed study examined the difference in brain morphology between patients who experienced episode recurrence and those who did not during the follow up period, finding that BD-I patients who experienced recurrence experienced greater volume loss than their no-episode counterparts, the latter being indiscernible from HC at follow-up (Kozicky et al., 2016). Previous studies in patients with longer duration BD found that recurrent manic episodes were associated with longitudinal decreases in dorsolateral prefrontal and inferior frontal cortex, supporting the neurotoxic nature of manic episodes (Abé et al., 2015; Barbosa et al., 2014). These studies highlight the need for further examination of the neurobiological impact of mood episode recurrence shortly following FEM.

Similarly, substance use (reviewed by Fowler et al., 2007), presence of psychotic features (Ekman et al., 2017; Radaelli et al., 2013; Hibar et al., 2018b) pediatric versus adult onset (Lu et al., 2012; Serafini et al., 2014) and number of depressive episodes (Strakowski et al., 2002) prior to FEM have shown evidence for impacting brain morphology in BD-I. However, there was very
limited reporting of the impact of these factors on longitudinal GM changes following FEM. Despite substance use being a common comorbidity in BD-I (Salloum and Brown, 2017) and being linked with decreased grey matter (Unterrainer et al., 2019), only two studies reviewed here accounted for substance use comorbidity in their analysis of longitudinal grey matter changes (Bitter et al., 2011; Kozicky et al., 2016). Furthermore, there was significant under-representation of BD-I patients without psychotic features in the pooled sample, as 11 out of 15 studies exclusively recruited BD-I patients with psychotic features. Previous studies have shown that patients with BD with psychosis exhibit smaller GM volumes in the inferior frontal gyri, and reduced surface area in the left inferior temporal gyrus and right caudal anterior cingulate, compared with patients with BD without psychosis (Ekman et al., 2017). Thus, patients with BD with psychosis may experience a distinct trajectory of structural changes compared to those without. Only two studies reported exclusively on youth with BD-I aged 18 or younger, limiting our understanding of how those with pediatric onset of BD-I differ from normal neurodevelopmental trajectories. Future longitudinal studies could examine different age brackets of onset to determine if the findings differ by age groupings. Although not a planned analysis, some studies revealed a potential effect of medication on longitudinal changes – for example, Nakamura et al. (2007) finding significantly greater neocortical GM increase in patients treated with a combination of mood stabilizer and antipsychotic. Conversely, Berk et al. (2017) did not find any changes in longitudinal GMV with one year of lithium or quetiapine monotherapy treatment. The role of specific mono- and combination treatment strategies in altering the trajectory of brain structural changes early in BD-I requires further examination.

In addition to participant and clinical factors, scanning protocols and image analysis techniques differed between the reviewed studies. As above, there were no discernible differences in findings between studies that used manual ROI placement versus VBM. A previous study comparing these two analysis methods on grey matter structures in patients with schizophrenia concluded that neither method posed any significant advantage over the other when analyzing subcortical or cortical structures, and should thus be used in tandem (Giuliani et al., 2005). A more recent study aimed to compare the two methods’ abilities to measure subcortical regions over a range of psychiatric disorders found strong agreement between the two methods for some, but not all, subcortical structures (Focke et al., 2014). Further research into the comparison of automated VBM versus manual ROI placement might need to be explored in the context of BD-I longitudinal
brain changes. All but four studies reviewed used a 1.5T MRI scanner, as opposed to a 3T scanner (Berk et al., 2017; Bitter et al., 2011; Bond et al., 2019; Kozicky et al., 2016). This limits the interpretation of the findings as 3T has superior spatial resolution, allowing for better detection of minute differences in brain structures (Wood et al., 2011). Therefore, differences in structural findings, or lack thereof, must be interpreted with caution. Furthermore, the use of 1.5mm slice thickness in all but five studies (Berk et al., 2017; Bitter et al., 2011; Bond et al., 2019; Farrow et al., 2005; Kozicky et al., 2016) further limits the interpretability of their results. It should also be noted that a meta-analysis was not planned or conducted; if a meta-analysis had been part of the original protocol, the wide variation in imaging methodologies and follow up periods, as well as the relatively small number of studies and sample size, would have rendered any quantitative estimates of effect size questionable.

2.4.1 Conclusions and Implications

The results of this review are equivocal regarding neuroprogressive grey matter brain changes in the early phase of BD-I following FEM. In particular, volume loss in the ACC, a key region involved in emotional processing, was reported by four studies. This, in concert with prior cross-sectional BD-I studies finding morphological abnormalities in this region, suggests that the ACC should be an area of interest in early-stage BD-I. However, this was not found in all studies that would have examined the ACC as part of a whole brain analysis, and numerous limitations in the existing literature lessen the homogeneity and generalizability of the results. These limitations suggest future directions in furthering our understanding of early-stage BD-I. First, longitudinal BD-I studies with larger sample sizes are necessary. While large-scale projects such as the ENIGMA consortium are aiming to uncover more reproducible areas of brain morphology differences between patients with BD and controls (Hibar et al., 2018; Schmaal et al., 2016), these analyses have thus far been cross-sectional. Uniformity in imaging techniques (i.e. in parcellation atlases and longitudinal imaging pipelines) would also allow for more robust and reliable results.

Furthermore, potential clinical sources of heterogeneity such as age of onset, psychotic features and substance use need to be further explored and considered when examining brain changes over time. Importantly, preliminary evidence suggests that episode recurrence in the first year following FEM is associated with more pronounced grey matter atrophy, suggesting a vital need to aggressively prevent mood episodes in the early stages to slow neuroprogression. Further
elucidating the impact of mood episode recurrence in the early stage of BD-I is thus of critical clinical importance.

Finally, all the longitudinal studies examining brain differences starting at FEM in BD-I populations included specifically examined volumetric differences. Regional volume is affected by both cortical thickness and surface area, which are separately influenced by different developmental and genetic factors (Winkler et al., 2010). For this reason, large-scale studies such as ENIGMA have preferentially looked at cortical thickness and surface area, foregoing cortical volumetric analysis altogether (Hibar et al., 2018). It would be useful for future longitudinal BD-I studies to examine these metrics as well.

Further longitudinal BD-I studies following FEM, taking these clinical and methodological factors into account, and following a more unified approach to image analysis, are needed to discover more robust and replicable neurodegenerative aspects of bipolar disorder in the early stages. This unified approach, also incorporating molecular, psychological, and behavioral indices, may be key to elucidating the pathophysiology of early BD and clarifying the role of early intervention in preventing detrimental brain changes.
Chapter 3: Comparison of Grey Matter Brain Measures Over Time in Bipolar I Disorder Patients versus Healthy Controls

3.1 Brief Introduction

As noted in Chapter 2, few studies longitudinally assess GM change in BD-I patients starting at FEM. We also found that no extant longitudinal studies which look at multiple follow-up time points, nor any longitudinal studies looking at longitudinal changes in CT or SA following FEM. Thus, we conducted a longitudinal analysis examining GM CT/SA/SCV change over the first three years of disease in patients with BD-I. In addition, longitudinal scan data was acquired in a cohort of age- and sex-matched HCs.

The aims of this experiment were to determine how a priori GM structures change over time in BD-I patients in the three years following diagnosis. Specifically, five cortical structures (caudal and rostral ACC, the fusiform gyrus, the pars opercularis and the rostral middle frontal gyrus) and four subcortical structures (amygdala, caudate, putamen, and thalamus) were chosen as they were found in multiple studies to show longitudinal differences in GM volume change in FEM patients compared to HC (specifically, the ACC – see Chapter 2), or exhibited large effect size differences in BD-I patients with longer duration illness compared with HC in a large, cross-sectional multi-center study (Hibar et al., 2018). We aimed to answer the following question: how does the trajectory of change in cortical thickness, surface area and subcortical volumes in these ROIs change in BD-I patients compared to HCs during the three-years following FEM? Based on previous literature (Castro-Fornieles et al., 2018; Hibar et al., 2018, 2016; Koo et al., 2008; Kozicky et al., 2016), we hypothesized that over the three-year period, these five cortical and four subcortical a priori GM structures would exhibit more loss in BD-I patients than HC.

3.2 Methods

We conducted a longitudinal investigation of a priori ROI measures compared between BD-I patients and HCs. Clinical data were obtained as described below. All clinical data were collected by trained researchers and clinicians.

3.2.1 Participants

The data for this analysis was collected as a part of the STOP-EM at the University of British Columbia (UBC). All study procedures for STOP-EM were approved by the UBC Clinical
Research Ethics Board and complied with ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, revised in 2008. All participants gave written informed consent prior to enrollment. Clinically stable patients meeting DSM-IV-TR criteria for BD-I were enrolled in the STOP-EM from UBC Hospital (UBCH)/affiliated sites, and community referrals. The complete protocol has been previously described (Yatham et al., 2009). Briefly, board-certified psychiatrists confirmed diagnoses using clinical and Mini International Neuropsychiatric Interview (M.I.N.I.) (Sheehan et al., 1998). Inclusion criteria were 14-35 years old, experienced the first manic/mixed episode in the preceding 3 months, and on mood stabilizing treatment at enrollment. Exclusion criteria included history of neurological disorder (e.g., seizure disorder or severe traumatic brain injury) and poor English. Comorbidities or substance abuse were not exclusionary in order to capture the range of patients seen in clinical practice.

HCs matched for age, sex, and premorbid IQ were recruited through word of mouth and advertisements at UBCH/affiliated sites. HCs had no family or personal history of psychiatric disorder, confirmed via the M.I.N.I. Participants completed a structural MRI scan at baseline (BL) and subsequently at years 1 (Y1) and 3 (Y3). Data from 62 BD-I patients (35 females; 22.7±4.5 years, range:15-34) and 47 age- and sex-matched HCs (23 females; 23.7±4.9 years, range:16-37) were used from the larger STOP-EM dataset to complete this study. These participants were chosen based on having undergone MRI scans during the first three years of the STOP-EM study, and whose MRI scans passed quality control. 61 patients had successful BL scans (one scan failed quality control), 45 at Y1, and 32 at Y3. 47 HCs had successful scans at BL, 35 at Y1, and 18 at Y3.

The baseline assessment was conducted within 3 months of diagnosis of FEM. Information including a sociodemographic interview, the M.I.N.I., diagnosis of comorbid disorder, physical examination, medical investigation, and neurocognitive assessment using the Cambridge Neuropsychological Test Automated Battery (CANTAB) was acquired. Patients participated in follow-up visits every six months to ascertain clinical information during that time (medication, mood episodes, etc.). At baseline and at follow-up appointments, mood symptoms were assessed with the MADRS and the Young Mania Rating Scale (YMRS) (Montgomery et al., 1979; Young et al., 1978). The CANTAB and the M.I.N.I., alongside collection of National Institute of Mental Health Life Charts (Denicoff et al., 2000) for episode recurrence, were conducted at baseline and
follow-up appointments. Patients continued to receive regular treatments from their own clinicians and received open-label treatment as per the Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines. Full demographic information for participants included in the study can be found in Table 3-1.
Table 3-1 Demographic and clinical features of STOP-EM study participants

<table>
<thead>
<tr>
<th>Baseline Measures</th>
<th>BD-I</th>
<th>Mean (SD)</th>
<th>HC</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL Recruitment n</td>
<td></td>
<td>62</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td>22.7 (4.5)</td>
<td>23.7 (4.9)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td>35F/27M</td>
<td>23F/24M</td>
<td></td>
</tr>
<tr>
<td>Education (years)</td>
<td></td>
<td>14.0 (2.2)</td>
<td>14.8 (2.7)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity (n)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td></td>
<td>48</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td>9</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>FEM Psychotic Features (%)</td>
<td></td>
<td>81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEM No Psychotic Features (%)</td>
<td></td>
<td>15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL Lifetime Drug Abuse (n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(including alcohol)</td>
<td></td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Longitudinal Measures</th>
<th>BL</th>
<th>Y1</th>
<th>Y3</th>
<th>BL</th>
<th>Y1</th>
<th>Y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants with scans (n)</td>
<td>61†</td>
<td>45</td>
<td>32</td>
<td>47</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td>YMRS Score</td>
<td>3.5 (5.7)</td>
<td>1.1 (3.7)</td>
<td>0.8 (1.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADRS Score</td>
<td>6.6 (8.4)</td>
<td>2.2 (4.0)</td>
<td>4.2 (6.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAF Score</td>
<td>67.1 (13.0)</td>
<td>76.8 (10.1)</td>
<td>75.2 (17.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mood Stabilizer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithium</td>
<td>19</td>
<td>19</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valproate</td>
<td>18</td>
<td>18</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antipsychotic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risperidone</td>
<td>23</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olanzapine</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quetiapine</td>
<td>11</td>
<td>9</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antidepressant</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Two BD-I patients and three HC participants have missing ethnicity data.
† One BD-I scan failed quality control

Abbreviations: BD-I, Bipolar I Disorder Patients; BL, Baseline; GAF, Global Assessment of Functioning Scale; FEM, F, Female; First Episode of Mania; HC, Healthy Controls; M, Male; MADRS; Montgomery Depression Rating Scale; SD, Standard Deviation; YMRS, Young Mania Rating Scale; Y1, Year 1; Y3, Year 3
3.2.2 Neuroimaging

Whole brain, 3D T1 MRI scans were acquired on a 3.0T Philips Achieva whole-body MRI (Philips Healthcare, Andover, MD) at the University of British Columbia Magnetic Resonance Imaging Research Scanner. Images were acquired using an 8-channel head coil. 3D Magnetization Prepared-Rapid Gradient Echo (MPRAGE) scans with repetition time/echo time (TR/TE) 7.6/3.5ms, 180 contiguous axial slices, slice thickness=1mm, field of view (FOV)=25.6cm, and flip angle=8°.

3.2.3 Image Analysis

All post-processing and analyses were carried out in FreeSurfer v7.1.0 (Fischl, 2012). Images from participants with only a baseline scan underwent the standard “recon-all” FreeSurfer pipeline (Fischl, 2012), while those with longitudinal scans were automatically processed using the FreeSurfer longitudinal pipeline (Reuter et al., 2012). Specifically, an unbiased within-subject template space and image was created using robust, inverse consistent registration. Skull stripping, Talairach transforms, atlas registration and spherical surface maps and parcellations were then initialized with common information from the within-subject templates. The longitudinal time-points were then constructed from these within-subject templates. This longitudinal pipeline has been shown to reduce within-subject variability and improve sensitivity to detect subtle changes (Reuter et al., 2012). CT and SA measures were then collected based on the Desikan-Killiany atlas (Desikan et al., 2006), and GMV for subcortical volumes.

Prior to statistical analysis, the imaging variables were adjusted for imaging-protocol effects using ComBat in RStudio (Fortin et al., 2018) (https://github.com/Jfortin1/ComBatHarmonization) for participants with a single scan, and Longitudinal ComBat for patients with longitudinal scans (Beer et al., 2020) (https://github.com/jcbeer/longCombat).

3.2.4 Statistical Analysis

Repeated measures mixed linear models were used to compare change in brain measures between groups, with post hoc tests for significant main effects, using SPSS (IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp). Time, group, and time by group (time*group) interactions were included in the model as fixed effects, and a random intercept for subject to account for within-subject correlations, giving the following equation for GM measure values:
\[ GM \text{ measure} = Time + Group + Time \ast Group + \text{Random Effect (subject)} \]

In order to choose the best model fit for the data, CT, SA, and SCV measures for all ROIs were run through both compound symmetry and unstructured maximum likelihood mixed models. A likelihood ratio test was then conducted to determine the best model fit for the data (Casella and Berger, 2001a). The lesser -2 Restricted Log Likelihood value was subtracted from the greater, as was the lesser degree of freedom (df) value from the greater. The two resulting values were run through a chi-squared test to determine if the two models were significantly different (RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/). If the difference was found to be non-significant, the model with the lesser df value was used (compound symmetry, df=9), as it is a less complex model; otherwise, the unstructured model (df=13) was used. All regions used the compound symmetry model except the following: left rostral middle frontal and right fusiform CT, left and right pars opercularis and rostral middle frontal SA, and left and right putamen SCV.

Age and sex were included as covariates in all models, in addition to intracranial volume (ICV) for SA and SCV comparisons. ICV is corrected for in this instance because volume scales with head size, which is mostly due to changes in SA, whereas CT scales to a lesser degree (https://surfer.nmr.mgh.harvard.edu/fswiki/eTIV).

The Estimated Marginal (EM) means of the data are calculated by SPSS as part of the mixed linear model analysis. Briefly, using Maximum Likelihood Estimation methods, the EM means are the group means for the variables of interest (here, the CT/SA/SCV of the ROIs) once adjusted for the covariates in the model. The EM means are then compared across and between groups, rather than using the raw data, to calculate findings. Statistical significance was defined as \( p<0.05 \), and the FDR method (Benjamini and Hochberg, 1995) was used to control for multiple comparisons.

3.2.4.1 Missing Data

Originally, 114 patients and 68 controls were recruited as part of the STOP-EM project. 92 BD-I patients consented to BL scans while 22 did not and 64 HCs consented to BL scans while 4 did not. Using unpaired t-tests, we found no significant differences in sex, age, YMRS and
MADRS scores at BL between patients who consented to BL scans and those who did not. Similarly, there were no differences in these measures between patients who continued at Y1 versus dropouts, nor when comparing patients who continued at Y3 versus dropouts.

BL and subsequent brain scans were excluded from 29 BD-I patients and 17 HCs due to scan time being out of the first episode window (within three months of FEM) or having been scanned on a different MRI protocol. One BD-I and four HC participants’ scans were excluded for failing quality control. Thus, scans from 62 BD-I patients and 47 HC were used in the analysis. Among the study participants included in this analysis, there were no significant differences in sex, age, baseline YMRS or MADRS scores between patients who continued to Y1 versus those who dropped out, nor between patients who continued to Y3 versus those who dropped out. Little’s Missing Completely at Random (MCAR) test, taking age, sex, MADRS and YMRS scores at BL into account, was non-significant, providing further evidence that data were missing at random (Kang, 2013).

We used all available data and Maximum Likelihood Estimation to generate the model parameters that best fit the observed data. This presents an advantage over the more widely known repeated measures ANOVA, which only includes participants with data at all time points (i.e listwise deletion), thus allowing for a more well-powered analysis.
Figure 3-1 Visual representation of regions of interest (ROIs) analyzed. A) Five cortical ROIs analyzed. Cortical parcellation based on the Desikan-Killiany Atlas. Adapted with permission from R.S. Desikan et al. Neuroimage 31, 968-980 (2006); B) Four subcortical ROIs analyzed. Subcortical segmentation originally described by Fischl et al. (2002). Image adapted from: Xueyi Shen, et al., CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons.
3.3 Results

3.3.1 Demographic Variables at Baseline

There were no significant differences between patients and healthy controls in age, sex, years of education, or ethnicity.

3.3.2 Cortical Thickness Measures

For cortical thickness, there were significant main time effects prior to FDR correction in the left caudal ACC (F=6.5, p=0.032), left rostral ACC (F=3.2, p=0.045), right pars opercularis (F=3.5, p=0.035), right rostral middle frontal (F=3.3, p=0.042), and the right fusiform (F=3.5, p=0.035). Specifically, both groups demonstrated a decrease over time in the left caudal ACC (BL-Y3: -0.039mm, p=0.009), left rostral ACC (BL-Y3: -0.035mm, p=0.014), right pars opercularis (BL-Y3: -0.036mm, p=0.016; Y1-Y3: -0.035mm, p=0.02), right rostral middle frontal (BL-Y3: -0.024, p=0.033; Y1-Y3: -0.028mm, p=0.015), and the right fusiform (BL-Y3: -0.036mm, p=0.011) (Table 3-2). However, none of the time effects survived FDR correction. Moreover, there were no significant group or time*group interaction effects.
Table 3-2 Cortical Thickness Estimated Marginal Means measures and Mixed Linear Model results

<table>
<thead>
<tr>
<th>CT Region</th>
<th>BD-I EM Means mm (SEM)</th>
<th>HC EM Means mm (SEM)</th>
<th>Mixed Linear Model</th>
<th>Group</th>
<th>Time</th>
<th>Time*Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BL</td>
<td>Y1</td>
<td>Y3</td>
<td>BL</td>
<td>Y1</td>
<td>Y3</td>
</tr>
<tr>
<td><strong>Left Caudal ACC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.64 (0.018)</td>
<td>2.62 (0.019)</td>
<td>2.61 (0.021)</td>
<td>2.64 (0.020)</td>
<td>2.63 (0.022)</td>
<td>2.60 (0.027)</td>
</tr>
<tr>
<td><strong>Right Caudal ACC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.46 (0.021)</td>
<td>2.44 (0.022)</td>
<td>2.42 (0.024)</td>
<td>2.45 (0.024)</td>
<td>2.45 (0.0260)</td>
<td>2.43 (0.029)</td>
</tr>
<tr>
<td><strong>Left Fusiform</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.51 (0.015)</td>
<td>2.51 (0.017)</td>
<td>2.50 (0.019)</td>
<td>2.50 (0.017)</td>
<td>2.47 (0.019)</td>
<td>2.45 (0.025)</td>
</tr>
<tr>
<td><strong>Right Fusiform</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.51 (0.014)</td>
<td>2.60 (0.020)</td>
<td>0.57 (0.020)</td>
<td>0.61 (0.016)</td>
<td>2.59 (0.023)</td>
<td>2.57 (0.025)</td>
</tr>
<tr>
<td><strong>Left Parsopercularis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.67 (0.016)</td>
<td>2.68 (0.017)</td>
<td>2.70 (0.019)</td>
<td>2.72 (0.019)</td>
<td>2.70 (0.020)</td>
<td>2.71 (0.023)</td>
</tr>
<tr>
<td><strong>Right Parsopercularis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.96 (0.018)</td>
<td>0.70 (0.019)</td>
<td>2.66 (0.021)</td>
<td>2.72 (0.020)</td>
<td>2.72 (0.022)</td>
<td>2.68 (0.026)</td>
</tr>
<tr>
<td><strong>Left Rostral ACC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.70 (0.021)</td>
<td>2.68 (0.022)</td>
<td>2.67 (0.024)</td>
<td>0.67 (0.024)</td>
<td>2.65 (0.025)</td>
<td>2.63 (0.029)</td>
</tr>
<tr>
<td><strong>Right Rostral ACC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.61 (0.025)</td>
<td>2.62 (0.026)</td>
<td>2.61 (0.027)</td>
<td>2.60 (0.028)</td>
<td>0.57 (0.030)</td>
<td>0.56 (0.033)</td>
</tr>
<tr>
<td><strong>Left Rostral Middle Frontal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.45 (0.014)</td>
<td>2.45 (0.013)</td>
<td>2.45 (0.017)</td>
<td>2.46 (0.016)</td>
<td>2.47 (0.015)</td>
<td>2.45 (0.021)</td>
</tr>
<tr>
<td><strong>Right Rostral Middle Frontal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.47 (0.013)</td>
<td>2.48 (0.014)</td>
<td>2.46 (0.015)</td>
<td>2.48 (0.015)</td>
<td>2.48 (0.016)</td>
<td>2.45 (0.020)</td>
</tr>
</tbody>
</table>
Abbreviations: ACC, Anterior Cingulate Cortex; BD-I, Bipolar I Disorder patients; BL, Baseline; CT, Cortical Thickness; df, Degrees of Freedom; EM, Estimated Marginal; HC, Healthy Controls; SEM, Standard Error of the Mean; Y1, Year 1; Y3, Year 3. *p<0.05. † main effect findings that did not survive False Discovery Rate (FDR) correction.
EM Means were adjusted for age and sex
Table 3-3 Pairwise comparison of Cortical Thickness measures at Baseline, Year 1, and Year 3 between first episode BD-I patients and healthy controls

<table>
<thead>
<tr>
<th>CT Region</th>
<th>BL ∆ (SEM) mm HC-BDI</th>
<th>F(df1,df2)</th>
<th>p</th>
<th>Y1 ∆ (SEM) mm HC-BDI</th>
<th>F(df1,df2)</th>
<th>p</th>
<th>Y3 ∆ (SEM) mm HC-BDI</th>
<th>F(df1,df2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Caudal ACC</strong></td>
<td>0.008 (0.027)</td>
<td>0.08(1,148)</td>
<td>0.78</td>
<td>0.013 (0.03)</td>
<td>0.18(1,179)</td>
<td>0.67</td>
<td>-0.012 (0.034)</td>
<td>0.13(1,227)</td>
<td>0.72</td>
</tr>
<tr>
<td><strong>Right Caudal ACC</strong></td>
<td>-0.008 (0.032)</td>
<td>0.06(1,131)</td>
<td>0.81</td>
<td>0.003 (0.034)</td>
<td>0.01(1,154)</td>
<td>0.92</td>
<td>0.003 (0.038)</td>
<td>0.01(1,198)</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Left Fusiform</strong></td>
<td>-0.01 (0.023)</td>
<td>0.19(1,176)</td>
<td>0.66</td>
<td>-0.033 (0.026)</td>
<td>1.62(1,207)</td>
<td>0.20</td>
<td>-0.045 (0.032)</td>
<td>2.01(1,238)</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>Right Fusiform</strong></td>
<td>0.009 (0.021)</td>
<td>0.17(1,109)</td>
<td>0.68</td>
<td>-0.014 (0.03)</td>
<td>0.23(1,91)</td>
<td>0.63</td>
<td>0.003 (0.032)</td>
<td>0.01(1,77)</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>Left Parsopercularis</strong></td>
<td>0.046 (0.025)</td>
<td>3.44(1,133)</td>
<td>0.07</td>
<td>0.026 (0.027)</td>
<td>0.96(1,159)</td>
<td>0.33</td>
<td>0.042 (0.03)</td>
<td>2.02(1,207)</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>Right Parsopercularis</strong></td>
<td>0.023 (0.027)</td>
<td>0.74(1,149)</td>
<td>0.39</td>
<td>0.02 (0.029)</td>
<td>0.49(1,180)</td>
<td>0.49</td>
<td>0.017 (0.034)</td>
<td>0.25(1,227)</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Left Rostral ACC</strong></td>
<td>-0.029 (0.032)</td>
<td>0.86(1,123)</td>
<td>0.36</td>
<td>-0.031 (0.034)</td>
<td>0.83(1,148)</td>
<td>0.36</td>
<td>-0.041 (0.037)</td>
<td>1.22(1,198)</td>
<td>0.27</td>
</tr>
<tr>
<td><strong>Right Rostral ACC</strong></td>
<td>-0.01 (0.038)</td>
<td>0.07(1,121)</td>
<td>0.79</td>
<td>-0.044 (0.039)</td>
<td>1.27(1,140)</td>
<td>0.26</td>
<td>-0.046 (0.043)</td>
<td>1.18(1,179)</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Left Rostral Middle Frontal</strong></td>
<td>0.014 (0.021)</td>
<td>0.44(1,109)</td>
<td>0.51</td>
<td>0.02 (0.02)</td>
<td>1.02(1,103)</td>
<td>0.32</td>
<td>0.001 (0.027)</td>
<td>&lt;0.001(1,77)</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Right Rostral Middle Frontal</strong></td>
<td>0.016 (0.019)</td>
<td>0.69(1,151)</td>
<td>0.41</td>
<td>0.004 (0.021)</td>
<td>0.04(1,184)</td>
<td>0.84</td>
<td>-0.011 (0.025)</td>
<td>0.20(1,231)</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Abbreviations: ACC, Anterior Cingulate Cortex; BD-I, Bipolar I Disorder patients; BL, Baseline; CT, Cortical Thickness; df, Degrees of Freedom; HC, Healthy Controls; SEM, Standard Error of the Mean; Y1, Year 1; Y3, Year 3; ∆, delta: change measures (HC-BDI)
Figure 3-2 Change in cortical thickness (CT) over time in five bilateral cortical regions. Significant time effects were noted in the left caudal anterior cingulate cortex (ACC), right fusiform, right pars opercularis and right rostral middle frontal cortex. There were no significant group nor time*group effects. None of the CT findings survived False Discovery Rate (FDR) corrections. *p<0.05, **p<0.01; † main effect findings that did not survive False Discovery Rate (FDR) correction.
3.3.3 Surface Area Measures

For surface area, there were significant main time effects prior to FDR correction in the left caudal ACC ($F=3.6$, $p=0.029$), right caudal ACC ($F=3.7$, $p=0.027$), left pars opercularis ($F=24.5$, $p<0.001$), right pars opercularis ($F=11.5$, $p<0.001$), left rostral middle frontal ($F=5.5$, $p=0.006$), right rostral middle frontal ($F=10.1$, $p<0.001$), and the right fusiform ($F=6.8$, $p=0.002$). Specifically, both groups demonstrated a decrease over time in the left caudal ACC ($BL$-$Y3$: $-16.3$ mm$^2$, $p=0.011$; $Y1$-$Y3$: $-14.8$ mm$^2$, $p=0.022$), right caudal ACC ($BL$-$Y3$: $-14.5$ mm$^2$, $p=0.011$; $Y1$-$Y3$: $-13.7$ mm$^2$, $p=0.018$), the left pars opercularis ($BL$-$Y3$: $-35.2$ mm$^2$, $p<0.001$; $BL$-$Y1$: $-13.9$ mm$^2$, $p=0.013$; $Y1$-$Y3$: $-21.3$ mm$^2$, $p<0.001$), the right pars opercularis ($BL$-$Y3$: $-26.0$ mm$^2$, $p<0.001$; $Y1$-$Y3$: $-26.3$ mm$^2$, $p=0.003$), the left rostral middle frontal ($BL$-$Y3$: $-70.5$ mm$^2$, $p=0.002$; $Y1$-$Y3$: $50.3$ mm$^2$, $p=0.019$), the right rostral middle frontal ($BL$-$Y3$: $-69.6$ mm$^2$, $p<0.001$; $Y1$-$Y3$: $-57.4$ mm$^2$, $p=0.003$), and the right fusiform ($BL$-$Y3$: $-40.7$ mm$^2$, $p<0.001$; $Y1$-$Y3$: $-35.2$ mm$^2$, $p=0.003$) (Table 3-3). Significant time effect findings survived FDR correction in the right fusiform, left, and right pars opercularis, and left and right rostral middle frontal, but not in the left and right caudal ACC. There were no significant group or time*group effects in surface area measurements.
Table 3-4 Surface Area Estimated Marginal Means measures and Mixed Linear Model results

<table>
<thead>
<tr>
<th>SA Region</th>
<th>BD-I EM Means mm² (SEM)</th>
<th>HC EM Means mm² (SEM)</th>
<th>Mixed Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BL</td>
<td>Y1</td>
<td>Y3</td>
</tr>
<tr>
<td><strong>Left Caudal ACC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1220.16(25.31)</td>
<td>1220.55(25.54)</td>
<td>1205.53(25.82)</td>
</tr>
<tr>
<td><strong>Right Caudal ACC</strong></td>
<td>872.85(23.07)</td>
<td>872.01(23.28)</td>
<td>861.04(23.52)</td>
</tr>
<tr>
<td><strong>Left Fusiform</strong></td>
<td>2775.93(42.31)</td>
<td>2773.57(42.74)</td>
<td>2759.01(42.27)</td>
</tr>
<tr>
<td><strong>Right Fusiform</strong></td>
<td>2731.45(42.84)</td>
<td>2723.38(43.28)</td>
<td>2689.75(43.83)</td>
</tr>
<tr>
<td><strong>Left Parsopercularis</strong></td>
<td>1586.84(30.94)</td>
<td>1594.23(31.82)</td>
<td>1606.43(32.88)</td>
</tr>
<tr>
<td><strong>Right Parsopercularis</strong></td>
<td>3914.43(70.05)</td>
<td>3862.56(71.30)</td>
<td>3776.07(72.82)</td>
</tr>
<tr>
<td><strong>Left Rostral ACC</strong></td>
<td>1252.31(28.31)</td>
<td>1249.40(28.54)</td>
<td>1245.01(28.82)</td>
</tr>
<tr>
<td><strong>Right Rostral ACC</strong></td>
<td>881.89(20.80)</td>
<td>867.16(21.03)</td>
<td>863.13(21.30)</td>
</tr>
<tr>
<td><strong>Left Rostral Middle Frontal</strong></td>
<td>1788.19(30.78)</td>
<td>1794.19(31.75)</td>
<td>1772.14(32.93)</td>
</tr>
<tr>
<td><strong>Right Rostral Middle Frontal</strong></td>
<td>7725.75(105.05)</td>
<td>7701.72(109.42)</td>
<td>7646.25(114.73)</td>
</tr>
</tbody>
</table>

Abbreviations: ACC, Anterior Cingulate Cortex; BD-I, Bipolar I Disorder patients; BL, Baseline; df, Degrees of Freedom; EM, Estimated Marginal; HC, Healthy Controls; SA, Surface Area; SEM, Standard Error of the Mean; Y1, Year 1; Y3, Year 3. *p<0.05, **p<0.01, ***p<0.001; † main effect findings that did not survive False Discovery Rate (FDR) correction

EM Means were adjusted for age, sex, and intracranial volume (ICV)
### Table 3-5 Pairwise comparison of Baseline, Year 1, and Year 3 separated by group for Surface Area measures

<table>
<thead>
<tr>
<th>SA Region</th>
<th>BL ∆ (SEM) mm² HC-BDI</th>
<th>Y1 ∆ (SEM) mm² HC-BDI</th>
<th>Y3 ∆ (SEM) mm² HC-BDI</th>
<th>p</th>
<th>F(df1,df2)</th>
<th>p</th>
<th>F(df1,df2)</th>
<th>p</th>
<th>F(df1,df2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Caudal ACC</td>
<td>11.31 (38.36)</td>
<td>7.44 (38.70)</td>
<td>7.98 (39.41)</td>
<td>0.85</td>
<td>0.04(1,115)</td>
<td>0.85</td>
<td>0.04(1,115)</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Caudal ACC</td>
<td>64.94 (34.98)</td>
<td>64.90 (35.27)</td>
<td>59.53 (35.89)</td>
<td>0.07</td>
<td>3.39(1,114)</td>
<td>0.07</td>
<td>2.75(1,122)</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Fusiform</td>
<td>88.19 (64.14)</td>
<td>84.47 (65.60)</td>
<td>81.36 (66.95)</td>
<td>0.20</td>
<td>1.66(1,116)</td>
<td>0.20</td>
<td>1.48(1,125)</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Fusiform</td>
<td>79.40 (64.94)</td>
<td>91.25 (64.78)</td>
<td>81.33 (66.09)</td>
<td>0.16</td>
<td>1.98(1,115)</td>
<td>0.16</td>
<td>1.51(1,124)</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Parsopercularis</td>
<td>47.53 (45.93)</td>
<td>56.62 (45.28)</td>
<td>55.47 (45.92)</td>
<td>0.21</td>
<td>1.56(1,109)</td>
<td>0.21</td>
<td>1.46(1,107)</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Parsopercularis</td>
<td>32.95 (42.07)</td>
<td>33.49 (43.75)</td>
<td>22.72 (41.01)</td>
<td>0.45</td>
<td>0.61(1,108)</td>
<td>0.45</td>
<td>0.31(1,109)</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Rostral ACC</td>
<td>10.37 (42.92)</td>
<td>8.81 (43.26)</td>
<td>30.70 (43.95)</td>
<td>0.84</td>
<td>0.06(1,110)</td>
<td>0.84</td>
<td>0.49(1,121)</td>
<td>0.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Rostral ACC</td>
<td>22.73 (31.54)</td>
<td>37.59 (31.87)</td>
<td>34.70 (32.54)</td>
<td>0.24</td>
<td>0.52(1,111)</td>
<td>0.24</td>
<td>1.14(1,125)</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Rostral Middle Frontal</td>
<td>139.14 (131.88)</td>
<td>132.40 (128.32)</td>
<td>150.47 (127.52)</td>
<td>0.30</td>
<td>1.11(1,108)</td>
<td>0.30</td>
<td>1.39(1,110)</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Rostral Middle Frontal</td>
<td>199.56 (131.57)</td>
<td>190.52 (132.76)</td>
<td>157.76 (125.02)</td>
<td>0.15</td>
<td>2.3(1,108)</td>
<td>0.15</td>
<td>1.59(1,109)</td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ACC, Anterior Cingulate Cortex; BDI, Bipolar I Disorder patients; BL, Baseline; df, Degrees of Freedom; HC, Healthy Controls; SA, Surface Area; SEM, Standard Error of the Mean; Y1, Year 1; Y3, Year 3; ∆, delta: change measures (HC-BDI)
Figure 3-3 Change in Surface Area (SA) over time in five bilateral cortical regions. Significant time effects were noted in the left and right caudal anterior cingulate cortices (ACC), right fusiform, the left and right pars opercularis and right rostral middle frontal cortex, and the left and right rostral middle frontal cortices. The findings in the left and right caudal ACC did not survive False Discovery Rate (FDR) correction. There were no significant group nor time*group effects. *p<0.05, **p<0.01, ***p<0.001; † main effect findings that did not survive False Discovery Rate (FDR) correction.
3.3.4 Subcortical Volume Measures

For subcortical volumes, there were significant time effects prior to FDR correction in the left (F=9.7, p<0.001) and right (F=5.3, p=0.006) caudate, and right putamen (F=5.8, p=0.005). These time effects all survived FDR correction. Specifically, both groups experienced a significant decrease in left caudate (BL-Y3: -105.8 mm$^3$, p<0.001; Y1-Y3: -88.3 mm$^3$, p=0.001), right caudate (BL-Y3: -93.0 mm$^3$, p=0.002), and right putamen volumes (BL-Y3: -135.7 mm$^3$, p=0.001; Y1-Y3: -99.4 mm$^3$, p=0.011) over time (Table 3-4). No SCV regions experienced significant change from BL to Y1. There was an additional significant group effect in the right thalamus where, over the three time points, controls had significantly greater volume than BD-I patients (281.8 mm$^3$, p=0.039). When looking at pairwise comparisons at the individual time points, BD-I patients showed a significantly lower right thalamus volume at Y3 (HC-BL: 322.68mm$^3$, F=4.18, p=0.04, effect size = 0.44) (Table 3-7). The group effect and time-point difference at year 3 in the right thalamus did not survive FDR. There were no significant time*group interactions in subcortical volume measurements.
Table 3-6 Subcortical Volume Estimated Marginal Means measures and Mixed Linear Model results

<table>
<thead>
<tr>
<th>SCV Region</th>
<th>BD-I EM Means mm³ (SEM)</th>
<th>HC EM Means mm³ (SEM)</th>
<th>Mixed Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BL</td>
<td>Y1</td>
<td>Y3</td>
</tr>
<tr>
<td>Left Amygdala</td>
<td>1586.84 (30.94)</td>
<td>1594.23 (31.82)</td>
<td>1606.43 (32.88)</td>
</tr>
<tr>
<td>Right Amygdala</td>
<td>1788.19 (30.78)</td>
<td>1793.19 (31.75)</td>
<td>1772.14 (32.93)</td>
</tr>
<tr>
<td>Left Caudate</td>
<td>3914.43 (70.05)</td>
<td>3862.56 (71.30)</td>
<td>3776.07 (72.82)</td>
</tr>
<tr>
<td>Right Caudate</td>
<td>4140.26 (71.48)</td>
<td>4061.87 (73.17)</td>
<td>3998.05 (78.23)</td>
</tr>
<tr>
<td>Left Putamen</td>
<td>5534.45 (78.54)</td>
<td>5465.56 (78.24)</td>
<td>5426.66 (79.42)</td>
</tr>
<tr>
<td>Right Putamen</td>
<td>5744.48 (82.49)</td>
<td>5694.27 (75.23)</td>
<td>5607.92 (81.31)</td>
</tr>
<tr>
<td>Left Thalamus</td>
<td>7725.75 (9105.05)</td>
<td>7701.72 (109.42)</td>
<td>7646.25 (114.73)</td>
</tr>
<tr>
<td>Right Thalamus</td>
<td>7676.28 (90.94)</td>
<td>7635.52 (95.44)</td>
<td>7545.56 (100.89)</td>
</tr>
</tbody>
</table>

Abbreviations: ACC, Anterior Cingulate Cortex; BD-I, Bipolar I Disorder patients; BL, Baseline; df, Degrees of Freedom; EM, Estimated Marginal; HC, Healthy Controls; SCV, Subcortical Volume; SEM, Standard Error of the Mean; Y1, Year 1; Y3, Year 3. *p<0.05, **p<0.01, ***p<0.001; † main effect findings that did not survive False Discovery Rate (FDR) correction

EM Means were adjusted for age, sex, and intracranial volume (ICV)
Table 3-7 Pairwise comparison of Baseline, Year 1, and Year 3 separated by group for Subcortical Volume measures

<table>
<thead>
<tr>
<th>SCV Region</th>
<th>BL Δ (SEM) mm³ HC-BDI</th>
<th>F(df1,df2)</th>
<th>p</th>
<th>Y1 Δ (SEM) mm³ HC-BDI</th>
<th>F(df1,df2)</th>
<th>p</th>
<th>Y3 Δ (SEM) mm³ HC-BDI</th>
<th>F(df1,df2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left Amygdala</strong></td>
<td>56.37 (46.91)</td>
<td>1.44(1,115)</td>
<td>0.23</td>
<td>24.72 (48.19)</td>
<td>0.26(1,127)</td>
<td>0.61</td>
<td>45.55 (50.83)</td>
<td>0.8(1,152)</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Right Amygdala</strong></td>
<td>16.68 (46.65)</td>
<td>0.13(1,118)</td>
<td>0.72</td>
<td>-0.13 (48.08)</td>
<td>&lt;0.001(1,131)</td>
<td>1.00</td>
<td>29.17 (51.02)</td>
<td>0.33(1,159)</td>
<td>0.57</td>
</tr>
<tr>
<td><strong>Left Caudate</strong></td>
<td>-80.07 (106.19)</td>
<td>0.57(1,113)</td>
<td>0.45</td>
<td>-11.21 (108.03)</td>
<td>0.01(1,121)</td>
<td>0.92</td>
<td>-14.88 (111.83)</td>
<td>0.02(1,137)</td>
<td>0.89</td>
</tr>
<tr>
<td><strong>Right Caudate</strong></td>
<td>-118.02 (108.36)</td>
<td>1.19(1,116)</td>
<td>0.28</td>
<td>-34.52 (110.84)</td>
<td>0.1(1,126)</td>
<td>0.76</td>
<td>-19.64 (115.96)</td>
<td>0.03(1,147)</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>Left Putamen</strong></td>
<td>43.61 (119.06)</td>
<td>0.13(1,108)</td>
<td>0.72</td>
<td>127.81 (118.43)</td>
<td>1.16(1,104)</td>
<td>0.28</td>
<td>65.11 (124.76)</td>
<td>0.27(1,86)</td>
<td>0.60</td>
</tr>
<tr>
<td><strong>Right Putamen</strong></td>
<td>14.83 (125.04)</td>
<td>0.01(1,108)</td>
<td>0.91</td>
<td>42.61 (113.95)</td>
<td>0.14(1,104)</td>
<td>0.71</td>
<td>16.55 (126.50)</td>
<td>0.02(1,94)</td>
<td>0.90</td>
</tr>
<tr>
<td><strong>Left Thalamus</strong></td>
<td>271.51 (159.25)</td>
<td>2.91(1,121)</td>
<td>0.09</td>
<td>310.02 (165.67)</td>
<td>3.5(1,138)</td>
<td>0.06</td>
<td>319.74 (178.79)</td>
<td>3.2(1,174)</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Right Thalamus</strong></td>
<td>263.56 (137.86)</td>
<td>3.66(1,125)</td>
<td>0.06</td>
<td>259.06 (144.46)</td>
<td>3.22(1,145)</td>
<td>0.08</td>
<td>322.68 (157.91)</td>
<td>4.18(1,185)</td>
<td>0.04*†</td>
</tr>
</tbody>
</table>

Abbreviations: ACC, Anterior Cingulate Cortex; BDI, Bipolar I Disorder patients; BL, Baseline; d, Cohen’s d; df, Degrees of Freedom; HC, Healthy Controls; SCV, Subcortical Volume; SEM, Standard Error of the Mean; Y1, Year 1; Y3, Year 3; Δ, delta: change measures (HC-BDI), *p<0.05, † pairwise findings that did not survive False Discovery Rate (FDR) correction.
Figure 3-4 Change in Subcortical Volume (SCV) over time in four bilateral subcortical regions. Significant time effects were noted in the left and right caudate, and right putamen. A significant group effect was found in the right thalamus as well as a significant pairwise comparison between groups at year 3 in the right thalamus. The finding in the right thalamus did not survive False Discovery Rate (FDR) correction. There were no significant group nor time*group effects. *p<0.05, **p<0.01, ***p<0.001; † main effect and pairwise findings that did not survive False Discovery Rate (FDR) correction.
3.4 Discussion

In this study, both BD-I patients and HCs showed significant decreases over the three-year study period in SA and SCV ROIs; while both groups demonstrated reductions in CT over the study period, CT time effects were not significant after FDR correction. Significant time effects were found in SAs of the right fusiform gyrus, left and right rostral middle frontal gyrus, and the left and right pars opercularis, in addition to volumes of the bilateral caudate and right putamen. However, no time*group terms were significant, indicating that both HC and BD-I patients showed a similar degree of reduction in SA and SCV in these ROIs. This would indicate that the changes in the patient group reflect normal time-related progression, rather than evidence of longitudinal disease effects.

Decreases in these GM measures over time in both HC and BD-I patients is expected. Prior studies have shown that brain GM structures are prone to decreases in GMV, CT and SA in healthy populations (Lemaitre et al., 2012; Storsve et al., 2014). Indeed, studies examining longitudinal cortical thinning in healthy populations exist, especially concerning older (Shaw et al., 2016) and younger (Zhou et al., 2015) populations. As noted in some of the studies included in the systematic review in Chapter 2, as well as other studies, both patient and control cohorts exhibited total GMV decrease over time in both adult and pediatric populations (Castro-Fornieles et al., 2018; de Castro-Manglano et al., 2011; Fraguas et al., 2016).

While decrements in GM structures may suggest atrophy, it is possible that a decrease in these measures may also signal increased efficiency and maturation of connections, especially in the age range of the included participants in this study. For example, Zhou et al. found accelerated cortical thinning in adolescents, compared with children and young adults (Zhou et al., 2015), corroborating adolescence being a time of great learning and emotional development. Moreover, the mechanisms driving SA and CT fluctuations are thought to be mediated by two genetically-unrelated processes (Panizzon et al., 2009) – while cortical column generation increases SA, neuronal genesis within the columns increases CT (Rakic, 1988). A 2014 study by Storsve et al. explored age-related changes in cortical volume, thickness, and area in healthy adults. Scanning participants aged 20-84 at baseline and subsequently three years later, the researchers found that while longitudinal change in CT and volume have a positive relationship, both volume and thickness change have a negative relationship with SA change (Storsve et al., 2014). Therefore, it
is possible that while changes in one feature, e.g., CT may signal a developmental milestone, changes in other features, e.g., SA could suggest disease-related atrophy.

Our data did not show any significant group or time*group interactions, suggesting that BD-I patients and HC do not follow differentiated GM trajectories in these brain regions during the first three years of disease. This is contrary to prior findings of greater longitudinal SCV loss in regions such as the caudate and putamen following FEM (Castro-Fornieles et al., 2018), and large cross-sectional analyses from the ENIGMA consortium finding significant reductions amygdala and thalamic volume (Hibar et al., 2016) and smaller CT in the pars opercularis, rostral middle frontal and fusiform regions (Hibar et al., 2018) in patients with longer duration BD-I compared to healthy controls. This discrepancy may be due to differences in disease duration; while the patients in our study were in the early stages of BD-I, patients in the ENIGMA study were at varying illness durations. When contrasted against prior studies, our findings suggest that perhaps the three years following FEM is too early to see pronounced changes, and that BD-I follows a neuroprogressive course rather than differentiated neurodevelopment prior to disease onset. The lack of significant time*group interactions could also be due to the naturalistic aspect of the STOP-EM study; when a patient comes into the clinic, every effort is made to treat them optimally starting at disease onset and over time. Long-term use of neuroprotective medications (more below) could explain the lack of differentiation of brain regions in BD-I patients from HCs over the study period. We did, however, find a significant group effect in the right thalamus, in addition to significantly higher SCV in the same region at Y3 in the HC compared with BD-I, but neither finding survived FDR. The magnitude of the difference was moderate, as evidenced by an effect size of 0.44, suggesting that a larger sample size may have been needed to for this difference to reach the a priori threshold for significance. This may suggest the thalamus as a region that experiences change early in the disease course.

Additionally, the ACC has been highlighted as an important neuro-marker in BD (Delvecchio et al., 2019), showing decreased GMV over time (Abé et al., 2020; Bond et al., 2019; Koo et al., 2008; Kozicky et al., 2016), and being prominently smaller in drug-naïve, psychotic patients (Atmaca et al., 2007; Keramatian et al., 2016; Yatham et al., 2007). However, we did not find any BD-I specific effects on the ACC in our study, likely due to a majority of our patients being on mood stabilizers and antipsychotic medication for the duration of the study (as seen in Table 3-1).
Some limitations and potential confounding factors need to be considered, including the effect of mood episodes, substance abuse, and medication use on brain morphology. As there are few longitudinal studies beginning at FEM, as seen in Chapter 2, it is difficult to characterize the relationship between these considerations and brain structure. Limitations of this study are discussed in Chapter 4.

Currently, the relationship between mood episode recurrence and GM anatomy is not well characterized and there is a paucity of literature on the topic. It has been suggested that manic episodes are neurotoxic (Barbosa et al., 2014), and longitudinal studies have shown decreased GMV in BD-I patients having recurrent episodes relative to patients who remained well during the follow-up period (Abé et al., 2015; Kozicky et al., 2016). One study found volume loss in the left middle temporal gyrus as a predictor of more days in depression during the follow-up period (Bond et al., 2019), while another was unable to show any relationships between episode recurrence and GMV (Frey et al., 2008). Further longitudinal neuroimaging investigations taking mood episode recurrence into account are needed.

BD and SUD present high comorbidity rates with an estimated 60% lifetime prevalence of substance abuse (Cassidy et al., 2001). A recent study explored the relationship between baseline GMV in BD-I patients who later went on to develop SUD. They found that patients who developed SUD had lower baseline GMV in a variety of prefrontal areas (Lippard et al., 2017), suggesting a predisposition to SUD. However, follow-up scans were not acquired in this study. Other studies have looked at more specific substance use in patients with BD. For example, research studying cannabis use found limited brain effects of the drug, showing that patients with BD exhibited lower GM density than HC regardless of cannabis use (Abush et al., 2018), or displayed greater cortical thinning in frontal regions only in patients who used prior to onset as opposed to those who began afterwards (Hartberg et al., 2018), while others showed reduced fusiform GMV and increased caudate volume in BD using cannabis (Jarvis et al., 2008). Two studies showed cortical thinning in frontal regions, including the ACC, showing a dose-dependent effect regardless of diagnosis (Lange et al., 2017) and even when patients had been in remission for a few years (Nery et al., 2011). Finally, in a study examining the effect of cigarette smoking in patients with BD, a decrease in left caudal ACC thickness was observed in patients who smoked versus those who did not (Jørgensen et al., 2015). Given this evidence, it would be safe to say that substance use plays a role in GM morphology, both in patients with BD and in HC, and so this should be taken into
consideration in future studies. To the best of our knowledge, no longitudinal study looking at substance use in BD exists; thus, the FEM dataset is perfect to fill in this knowledge gap. Unfortunately, we were unable to look at the effects of substance use in this thesis as SUD data was acquired only at baseline, and at time points greater than Y3.

Medication effects are additionally an important factor for consideration in BD neuroimaging studies. A variety of relationships between medication use and brain measures exist. Lithium is well-known as neuroprotective and neuroproliferative in BD cohorts; showing increase in GMV and density in both the ACC (Bearden et al., 2007; Moore et al., 2000) and the amygdala (Chang et al., 2005b; Moore et al., 2000). Similarly, valproate and valproate+quetiapine combinations have shown increased GMV in the same regions (Atmaca et al., 2007; Chang et al., 2005b). Contrary findings have also been studied, where psychotropic medications were found to have no effect in pediatric populations (Blumberg et al., 2003; Chang et al., 2005a; Sanches et al., 2005). While excluding patients on medication can help limit this confound, it would not be truly representative of the chronic illness progression seen in most patients (Phillips et al., 2008). We were unable to examine medication effects in this study as it was out of the scope of this thesis.
Chapter 4: Discussion

4.1 Overview of Findings

This thesis examined the longitudinal trajectory of GM structures in the early stages of BD-I. While GM brain changes are well documented in cross-sectional studies, literature looking at longitudinal investigations of the same structures is scarce, and findings are generally inconclusive. Furthermore, longitudinal GM studies have only explored volume in the past. Therefore, this study aimed to add to this body of literature by looking instead at different attributes of GM: SA and CT, in addition to SCV. First, in the systematic review in Chapter 2, we summarized findings of 15 longitudinal studies and found the most common finding of reduced GMV in the ACC both at baseline and longitudinally. Next, in Chapter 3, we conducted a longitudinal analysis of CT, SA, and SCV of select GM structures and found a pattern of decreased SA and SCV over time in both BD-I subjects and HC.

4.2 Longitudinal Gray Matter Changes Following First Episode Mania in Bipolar I Disorder: A Systematic Review

To the best of our knowledge, this is the first systematic review exclusively surveying longitudinal structural brain changes in BD-I patients in the early stages of the disorder. Lower ACC and inferior frontal gyrus volumes were commonly reported both at baseline (Berk et al., 2017; Farrow et al., 2005; Koo et al., 2008) and longitudinally (Bond et al., 2019; Farrow et al., 2005; Koo et al., 2008; Kozicky et al., 2016). While the finding of decreased volume in these frontal regions has the potential to be a neural biomarker for early disease progression, this finding was only observed in a small proportion of the studies included. Thus, there is evidence for pathophysiology in these regions present at FEM and progressing alongside illness duration, but this should be interpreted with caution.

In addition to reviewing the general structural findings of these studies, we conducted subgroup analyses looking at a variety of important factors such as analysis method, impact of episode recurrence, psychosis symptoms, substance use, age, and prior depressive episodes. Kozicky et al. showed that patients who experienced episode recurrence during a one-year follow up have demonstrably smaller GM volumes, and a steeper rate of volume decline, than their episode-free BD-I and HC counterparts, of which the latter two were indistinguishable. Bond et
al. noted that more days spent in a depressive episode during the follow-up period was associated with greater volume loss in the left middle temporal gyrus, while Bitter et al. found no differences in the amygdala volume between BD-I patients with and without episode recurrence. The remainder of the studies did not look at this variable. Similarly, Kozicky et al. and Bitter et al. reported substance use in their populations – while the former study found a higher rate of substance abuse in the episode-recurrence population, the latter reported no difference in findings between patients with SUD and those without. Finally, studies investigating psychosis and age did not report any differences that might be specific depending on age or presence/absence of psychosis symptoms. None of the reviewed studies included prior depressive episodes in their analysis.

The studies included used three different methods of image analysis: VBM, manual ROI, and FreeSurfer. This makes it hard to compare findings between all the studies for multiple reasons, including the inherent difference of the methods (Popescu et al., 2016) in addition to the specific regions examined. VBM allows for whole-brain analysis, and involves the use of group-wise comparisons, on a voxel-by-voxel basis, to localize focal changes across subject’s brains, while FreeSurfer conducts automatic compartmentalization of the brain into multiple regions to quantify tissue type and measures. These are both in contrast to manual ROI methods, which use an a priori, hypothesis-driven approach, where researchers manually delineate structure boundaries and the analysis program used provides measures for these structures. Studies that utilized ROI methods looked exclusively at certain regions, including temporal regions (Kasai et al., 2003a; Kasai et al., 2003b; Lee et al., 2016; Salisbury et al., 2007), subregions of the ACC (Koo et al., 2008), and the amygdala (Bitter et al., 2011), all obviously reporting very different findings. While these are important results in and of themselves, it makes it impossible to compare findings between numerous studies.

This systematic review is limited by the small number of studies that met inclusion criteria, with relatively short follow-up times, largely heterogenous findings, and all these studies only considered GMV. The 15 studies included used a variety of analysis methods, thus looking at a wide variety of regions and reported very different findings. This lack of replicability makes it difficult to make definitive conclusions about GM neuroprogression in BD-I, especially since there are not many studies against which to compare the results. As noted above, SA and CT development are dependent on different measures (Panizzon et al., 2009; Rakic, 1988), and have
different relationships with GMV (Storsve et al., 2014b). Thus, future longitudinal studies should consider these measures, providing a more in-depth understanding of GM progression in the early stages of BD-I.

4.3 Comparison of GM Brain Measures over Time in BD-I Patients vs Healthy Controls

This study aimed to observe the trajectory of GM structures over time in both BD-I patients and HCs, and to see how these trajectories might differ. While we found a common pattern of decrease in SA and SCV over time, we were unable to find any group differences or differences in trajectory. The decrease in GM SA and SCV over time in both groups follows prior developmental and healthy aging studies (Lemaitre et al., 2012; Shaw et al., 2016; Storsve et al., 2014; Zhou et al., 2015), as well as prior longitudinal studies that observed decrease in GMV over time (Castro-Fornieles et al., 2018; de Castro-Manglano et al., 2011; Fraguas et al., 2016).

There were no significant group nor time*group interactions in our results which is at odds with prior longitudinal studies showing differing trajectories with a shorter time period between scans (Bitter et al., 2011; Castro-Fornieles et al., 2018; Koo et al., 2008). However, our analyses differ from these three studies. First, Castro-Fornieles et al. used a VBM method whereas our study used FreeSurfer. While their study follows a whole brain, voxel-by-voxel methodology, ours used an a priori, hypothesis-driven approach. Second, Koo et al. looked at the whole cingulate cortex, including the posterior cingulate, in addition to parcellating it into specific subregions. Their findings showed differed trajectory in the subgenual subregion but showed no differences in the cingulate as a whole; our study examined larger and different subdivisions of anterior portions of the cingulate: the caudal and rostral ACCs. Moreover, these studies analyzed cortical volumes, whereas ours focused on CT and SA. Finally, Bitter et al.’s study specifically investigated amygdala trajectory during teenage years (12-17 years) and observed a decrease in volume over time, whereas the HC and ADHD patients in the same study remained level. We found no results in the amygdala in our study, including no significant decrease in volume over time, but our study encompasses a much larger and later age population (15-35 years). Furthermore, most of our patients were receiving therapeutic intervention during this study, which could have saved these structures from pathophysiological damage.

Our study adds to the literature in three important ways. First, while prior longitudinal GM studies following FEM in BD-I only looked at changes in GMV, our study examined longitudinal
trajectories of CT and SA in addition to SCV. As mentioned above, SA and CT follow different developmental and pathophysiological processes than GMV, and so recording their changes separately can provide greater insight than what we have now. Our study noted significant time effects in both SA and SCV measures, but not CT, which could imply that aging-related GM changes occur via the decrease of the number of cortical columns rather than the amount of neurons within them (Rakic, 1988). Second, the addition of the intermediary time-point at Y1; while three years is not a long time, and apparently is not sufficiently long enough to see marked group-specific differences, the addition of the intermediary time point could provide a more specific reference for when brain changes occur during illness or natural aging progression. As the STOP-EM study continues to scan patients and fill in the much-needed sample size, further time-points can be compared. Finally, the lack of baseline differences points to BD being a neuroprogressive, rather than a neurodevelopment, disorder.

4.4 Strength and Limitations

The strengths and limitations of the systematic review are well defined in Chapter 2, so this section will focus on those of the original research conducted in Chapter 3.

This study characterized the progressive trajectories of a variety of GM structures in both BD-I patients and HCs. Patients included in the study were naturalistically observed, with clinical and cognitive assessments completed at regular intervals. This study was strengthened by using CT and SA measures, providing new insight into the development of these regions, in addition to the intermediary time point at Y1.

While this study provides a strong basis upon which to observe GM changes over the first three years of disease, and seeing how they compare with HCs, this study is limited primarily by its small sample size and low retention rate. We were able to start off with a decent sample size, but numbers declined steadily due to attrition, especially in the HC cohort. This is unfortunate as it would have been useful to look at time points past Y3 in the longitudinal analysis, but the sample size at those time-points is not large enough to conduct a well-powered analysis. Moreover, the large amount of missing data at subsequent time points (27% and 25% missing at Y1, 48% and 62% missing at Y3, for BD-I and HC, respectively) has the potential be a cause for concern (Clark & Altman, 2003; Jakobsen et al., 2017), leading to bias in the study results. We tested for random patterns of missingness and ascertained that clinical variable had no bearing on participants who
dropped out versus those who did not (see Section 3.2.1.4); our non-significant findings in the Little’s test is suggestive of a random missing pattern in our data. The smaller sample sizes at the later time points could potentially account for the inability to observe significant longitudinal group effects at the timepoints following BL that has been observed in prior studies (e.g. (Castro-Fornieles et al., 2018)). Prior literature has suggested that missing data greater than 40% is problematic (Jakobsen et al., 2017; Overall et al., 2006); our large gaps at Y3 are therefore inconclusive and could rather be used as hypothesis-generating results for future studies.

There exist a variety of ways in which to deal with missing data, and these include: pairwise deletion (AKA: all available data; using all data available but not corrected for missingness), listwise deletion (AKA: complete case; excluding participants who do not have complete datasets), single imputation (such as last value carried forward), Maximum Likelihood approaches (which we used), and multiple imputation. Prior literature has cautioned against the use of pairwise and single imputation (Little et al., 2012; Newman, 2014), and listwise deletion is only appropriate if data is Missing Completely at Random, and less than 5% of the data is missing (Jakobsen et al., 2017). This left us with two options: Maximum Likelihood methods or multiple imputation (Jakobsen et al., 2017). Unlike multiple imputation, maximum likelihood estimation does not substitute missing values, but estimates the model parameters that are most likely to have resulted in the observed data. When there is a non-negligible amount of missing data and data are missing at random, maximum likelihood and multiple imputation are the two most commonly used methods used to generate unbiased models (Jakobsen et al. 2017). We opted for the Maximum Likelihood Model as it is included in the SPSS Mixed Model analysis package, and furthermore, our missing data at Y3 exceeded 40%; it is cautioned against using multiple imputation with such a large proportion of missing data (Jakobsen et al., 2017).

Finally, this thesis did not include any sex or gender-based analyses. This was mainly due to the small sample size. There are few prior neuroimaging studies in BD that investigate differences based on biological sex, and results have been conflicting. The ENIGMA study found a sex by diagnosis interaction in adolescent/young female BD patients, whereby these females showed less cortical thinning than could be explained by sex and diagnosis alone; the strongest effect was noted in the right pars triangularis. Other examples include sex*diagnoses interactions in the ventral PFC and hippocampal-amygdala complex, where male patients had decreased volumes, which was inversely correlated with the number of adverse life events (Jogia et al., 2012).
Contrarily, a different study found trend-level decreases in left hippocampal volumes in schizophrenia and BD patients, where this decrease was especially pronounced in the female patients with early-onset BD (Frazier et al., 2008). Aside from biological sex, there has been even less examination of the impact of gender on the clinical and neurobiological expression of BD. In any instance, inclusion of sex and gender-based analyses is necessary in the future in order to understand how this disorder might affect males and females separately, and allowing for more personalized treatment options.

For these reasons, future directions of this study could include the continuation of scanning participants at further time points and taking mood episode recurrence into consideration. To the best of our knowledge, alongside de Castro-Mangalo’s study (de Castro-Manglano et al., 2011), this work represents the longest longitudinal follow-up period in a neuroimaging BD-I study following FEM (3 years). It is possible that three years may not be long enough to detect substantial differences, or that the lack of significant group effects was potentially caused by patients following a strict medication regime. Either way, this study provides new insights into GM changes, especially considering the inclusion of the intermediary time point (Y1), and of CT and SA measures. Another feasible future study could be to recruit patients with established BD-I at various time points of their disorder and clump them into illness duration groups (ex. 0-5 years, 5-10 years, etc) and see how GM structure changes across diagnostic and sex groups, considering factors such as medication, substance abuse, and episode recurrence.

4.5 Conclusions

The systematic review in this thesis provides an in-depth summary of existing longitudinal imaging studies of BD-I patients following FEM, guiding the original research conducted thereafter. Thus, this thesis provides a solid starting point for examining longitudinal GM structures over time in FEM BD-I patients. It opens a conversation around the advantage of including SA and CT measurements in future longitudinal studies and suggests that GM trajectories early in the disease course of optimally treated BD-I patients are similar to that of HCs. This is suggestive of three years not being long enough to see pronounced neuroprogressive differences, or that the medications the patients take are serving their neuroprotective purposes. The continuation of the STOP-EM study is therefore at the utmost importance, as is multi-centre
collaboration to acquire the largest possible and most encompassing sample of patients with BD-I.
References


Lippard, Elizabeth T Cox, Johnston, J. A. Y., Spencer, L., Quatrano, S., Fan, S., Sankar, A.,


Merikangas, K. R., Jin, R., He, J. P., Kessler, R. C., Lee, S., Sampson, N. A., Viana, M. C.,


