RELAPSE IN THE LONGITUDINAL TRAJECTORY OF FUNCTIONING IN BIPOLAR DISORDER TYPE I

RELAPSE IN THE LONGITUDINAL TRAJECTORY OF FUNCTIONING IN BIPOLAR DISORDER TYPE I

By MAYA KOVACHEFF, B.Sc. (Hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2024) Hamilton, Ontario (Psychology)

TITLE: RELAPSE IN THE LONGITUDINAL TRAJECTORY OF FUNCTIONING IN BIPOLAR DISORDER TYPE I

AUTHOR: Maya Kovacheff, B.Sc. (Hons.) (University of Toronto)

SUPERVISOR: Dr. Nicholas Bock

NUMBER OF PAGES: xi, 69

Lay Abstract

Bipolar disorder type I (BD-I) is a debilitating mood disorder involving manic episodes (e.g. heightened mood, excessive energy, impulsivity) and oftentimes depressive episodes as well (e.g. low mood, little interest in activities, sleep problems). BD-I is associated with high rates of functional impairment. Some models suggest that BD-I is a progressive illness wherein a longer duration of illness, a higher number of mood episodes, and lingering impairment at times outside of mood episodes can be indicators of the illness getting worse overtime and affecting individuals more negatively. Using data from a larger study, the current study aimed to investigate functional impairment in individuals with BD-I who experienced an episode relapse (BDR) compared to those who did not (BDNR) and healthy controls (HC). Two measures of function were used: psychosocial functioning was measured by the Functional Assessment Short Test (FAST) and subjective cognitive functioning was measured by the Cognitive Failures Questionnaire (CFQ). Participants completed up to 3 visits over 2 years, spaced approximately 1 year apart. Differences between both BD groups and the HC group were found for both scales, suggesting a sustained functional deficit in BD over time. No differences between the BDR and BDNR groups were found, but the BDNR group demonstrated improvement in subjective cognitive functioning over the 2-year period. These findings suggest that BD-I shows impairment in psychosocial and subjective cognitive functioning as compared to HC but that relapse status did not have an effect. This research suggests that perhaps mood episode relapse may not influence functioning negatively, but a lack of relapse may have positive effects on subjective cognitive functioning.

Abstract

Introduction: Bipolar disorder (BD) is a recurrent and chronic mood disorder. BD type I (BD-I) is associated with high disability, lower quality of life, and excess mortality. Importantly, BD is also associated with severe functional impairment. Staging models suggest BD is a progressive illness and use episode recurrence and functional impairment in euthymia as main proxy measures. Research has identified deficits in functioning in BD compared to healthy controls (HC) and suggest that impairment may be sustained in periods of euthymia and related to episode recurrence. The current research uses data from a larger longitudinal neuroimaging study to investigate psychosocial and subjective cognitive function, as measured by the Functional Assessment Short Test (FAST) and the Cognitive Failures Questionnaire (CFQ), in individuals with BD-I who experienced an episode relapse (BDR) compared to those who did not (BDNR) and healthy controls (HC).

Methods: Participants completed up to 3 visits over 2 years, that took place approximately 1 year apart. The final sample consisted of 61 HC, 21 BDR, and 26 BDNR participants. Three analyses were conducted to explore between and within-subject differences: mixed-effects analysis of variance (ANOVA), one-way Kruskal-Wallis tests, and simple linear growth analyses. Missing data precluded using three time points for some analyses, so the mixed-effects ANOVA and one-way Kruskal-Wallis tests used two re-binned timepoints (baseline and follow-up) and the growth curve analysis used all three timepoints.

Results: Significant differences were found between the HC group and both BD groups (BDR and BDNR) for both the FAST and CFQ at baseline and follow-up visits. No significant differences were found between the BDR and BDNR groups, neither differences at timepoints

nor differences in change across timepoints. The BDNR group demonstrated a significant decrease in CFQ scores over the 3 timepoints.

Conclusion and Future Directions: The results suggest that individuals with BD-I experience sustained impairment in psychosocial and subjective cognitive function over time compared to HC, but relapse does not have a significant effect on this impairment. Since the BDNR group demonstrated a decrease in CFQ scores over time, not experiencing relapse may be implicated in the improvement of subjective cognitive functioning. Future studies with longer measurement windows and larger sample sizes could further clarify these findings.

Acknowledgements

I am very fortunate that I have many people to acknowledge as integral in the completion of this thesis. I would like to first thank my supervisor and committee members - Dr. Nicholas Bock, Dr. Benicio Frey, and Dr. Andrea Gonzalez - for their support. I will always appreciate their expertise and continued guidance throughout this program.

Next, I would like to recognize several friends and colleagues that helped me considerably throughout this journey. I would like to first thank Mazen Elkhayat for all his time and generosity with endless troubleshooting support as I truly would not have been able to complete this thesis without him. I would also like to thank Niousha Gazor for her continued encouragement, even after the completion of her own degree. In addition, thank you to my affectionately named Free Support Group: Reshna Ninan, Grace Arthur, Renata Soljmosi, Jackie Chau, and Dania Daanish. They have been there since the first day I started this program, and I could not imagine a better group of friends to have had over the past two years. In addition, although I did not use all our work in the final analyses, I would like to thank Dr. Krysta Andrews for her time and support with learning growth analyses, which I had no knowledge of previously. Importantly, I would also like to thank all of the research coordinators who collected data for this study as well as all the participants that took part.

Last and certainly not least, I would like to thank my mother Dr. Allison Bramwell, my father Dr. Jonathon Kovacheff, my sister Dr. Chloe Kovacheff, and my husband Nicolas Mertens who were instrumental in my ability to complete this program. Their endless patience, generosity, support, and love was the difference. Words cannot express how grateful I am, and I can truly say that I would not be where I am today without them.

Table of Contents

Lay Abstract ..... **iii**
Abstract ..... **iv**
Acknowledgements ..... **vi**
List of Tables ..... **ix**
List of Figures ..... **x**
List of Abbreviations ..... **xi**
Declaration of Academic Achievement ..... **xii**
1. Introduction **1**
1.1 Bipolar Disorder Overview 1
1.1.1 Diagnostic Criteria 2
1.1.2 Disease Burden 4
1.2 Functional Impairment 5
1.3 Illness Severity 7
1.3.1 Mood Episode Relapse 9
1.3.2 Impairment During Remission 9
1.4 Current Research 11
1.4.1 Primary Goals 11
1.4.2 Significance 12
2. Methods **12**
2.1 Overall Study Procedure 13
2.2 Participant Overview 13
2.3 Measures 15
2.3.1 Demographics 15
2.3.2 Structured Clinical Interview for DSM-5 (SCID-5) 15
2.3.4 Montgomery-Asberg Depression Rating Scale (MADRS) 16
2.3.5 Functional Assessment Short Test (FAST) 17
2.3.6 Cognitive Failures Questionnaire (CFQ) 17
2.3.7 Mood Episode Relapse for BD Participants 18
2.4 Quality Control Procedure 18
2.4.1 Demographics 19
2.4.2 Episode Relapse 19

2.4.3 Scale Data 21
2.5 Data Analyses 21
2.5.1 Demographics at Baseline 21
2.5.2 Medication Status for BD Groups 22
2.5.3 Main Analyses 22
2.5.3.1 Mixed-Effects ANOVA: Between and Within Group Differences in Scores 23
2.5.3.2 One-Way Kruskal-Wallis: Between Group Differences in Delta Scores 24
2.5.3.3 Growth Curve Analysis (GCA) 24
3. Results 25
3.1 Demographics at Baseline 25
3.2 Demographics of Dropouts 27
3.3 Medication Status of BD Subjects Throughout Study 29
3.4 Mixed-Effects ANOVA 31
3.4.1 FAST 31
3.4.2 CFQ 33
3.5 One-Way Kruskal-Wallis 35
3.5.1 FAST 35
3.5.2 CFQ 37
3.6 Growth Curve Modeling 38
3.6.1 FAST 38
3.6.2 CFQ 40
4. Discussion 43
4.1 Between Group Differences in Scale Scores 43
4.2 Between Group Differences in Change in Scores 45
4.3 Within Group Differences in Change in Scores 46
4.4 General Discussion 48
4.5 Limitations 50
4.6 Future Directions 52
5. Conclusion 53
References 55
Supplementary Material 68

List of Tables

Table 1: Demographic and Clinical Characteristics of HC and BD at Baseline

Table 2: Baseline Demographics of BDR and BDNR Groups

Table 3: Baseline Demographics of BD Dropouts

Table 4: Baseline Demographics of HC Dropouts

Table 5: Medication Status of BD Groups at All Visits

Table 6: Robust ANOVA Results: FAST

Table 7: Post-hoc Pairwise Comparisons: FAST

Table 8: Robust ANOVA Results: CFQ

Table 9: Post-hoc Pairwise Comparisons: CFQ

Table 10: One-Way Kruskal-Wallis Results: FAST

Table 11: Pairwise Comparisons Using the Dunn Test: FAST

Table 12: One-Way Kruskal-Wallis Results: CFQ

Table 13: Pairwise Comparisons Using the Dunn Test: CFQ

Table 14: Model Comparisons for FAST Data

Table 15: Parameter Estimates for Model 2: FAST

Table 16: Model Comparisons for CFQ Data

Table 17: Parameter Estimates for Model 3: CFQ

List of Figures

Figure 1: Relapse Trajectory of BD Participants with Medication Status

Figure 2: Results for Robust ANOVA: FAST Scores

Figure 3: Results for Robust ANOVA: CFQ Scores

Figure 4: Results for One-Way Kruskal-Wallis Test: Delta FAST Scores

Figure 5: Results for One-Way Kruskal-Wallis Test: Delta CFQ Scores

Figure 6: Linear Model of Best Fit for FAST Data

Figure 7: Linear Model of Best Fit for CFQ Data

Figure 8: Checking Growth Curve Analysis Model: FAST Data

Figure 9: Checking Growth Curve Analysis Model: CFQ Data

List of Abbreviations

BD	Bipolar Disorder
AUD	Alcohol Use Disorder
SUD	Substance Use Disorder
BD-I	Bipolar Disorder Type I
BD-II	Bipolar Disorder Type II
QoL	Quality of Life
HC	Healthy Control
FAST	Functional Assessment Short Test
CFQ	Cognitive Failures Questionnaire
BDR	Bipolar Disorder Relapse
BDNR	Bipolar Disorder No Relapse
YMRS	Young Mania Rating Scale
MADRS	Montgomery-Asberg Depression Rating Scale
SCID-5	Structured Clinical Interview for DSM-5
MRI	Magnetic Resonance Imaging
REDCap	Research Electronic Data Capture
CRF	Case Report Form
ANOVA	Analysis of Variance
GCA	Growth Curve Analysis
MLE	Maximum Likelihood Estimation
LL	Loglikelihood Estimation

Declaration of Academic Achievement

Along with Dr. Nicholas Bock, I was responsible for conceptualizing the research questions explored in this thesis. All study data were collected by myself and past research staff from participating sites. I completed all quality control for the clinical data presented in this thesis. Niousha Gazor was responsible for the quality control of basic demographic information. Dr. Nicholas Bock, with support from Dr. Benicio Frey and Dr. Andrea Gonzalez, assisted me with conceptualizing the statistical analyses. I completed statistical analyses with the support of Dr. Nicholas Bock, Dr. Krysta Andrews, and Mazen Elkhayat and I completed the interpretation of the results.

1. Introduction

1.1 Bipolar Disorder Overview

Similar to a Thalia and Melpomene theatre mask presenting both an exaggerated expression of happiness and sadness, bipolar disorder (BD) is often perceived as a simple dichotomy of emotional states. Yet, this illustration is an inaccurate way to conceptualize this complex and debilitating mood disorder. BD is a chronic disorder (Fagiolini et al., 2013) characterized by recurring mood episodes that range from depressive (e.g. low mood, sadness) to manic (e.g. excessive energy, impulsive behaviour, irritability) and can differ in length and severity. Individuals with BD also experience periods of stability free of mood symptoms, termed euthymia. BD affects approximately 2.2% of the Canadian population (Schaffer et al., 2006) and has an estimated lifetime prevalence of 2.4% globally (Merikangas et al., 2011).

Bipolar disorder is highly comorbid with other mental disorders, namely anxiety disorders, alcohol use disorder (AUD), and substance use disorders (SUD) (Hirschfield & Vornik, 2005). There are no established sex differences in the prevalence of BD, but there are some differences in rates and types of comorbid disorders. Specifically, men have higher rates of AUD and SUD, while women have higher rates of anxiety, and eating disorders as well as Post-Traumatic Stress Disorder (PTSD) (Azorin et al., 2013; Diflorio & Jones, 2010; Loftus et al., 2020; Messer et al., 2017).

Age of onset is becoming an important factor in understanding the course and outcome of BD as individuals who report an earlier age of onset could be at greater risk of recurrence and poorer functional outcomes compared to those with a later onset (Perlis et al., 2009). According to a recent meta-analysis, a trimodal age-at-onset model may work best to described individuals

with BD where there are three onset periods: early-onset (17.3 years), mid-onset (26 years), and late-onset (41.9) (Bolton et al., 2021). Conceptualizing age of onset this way may help to inform clinical trajectories and treatment choices.

There are many types of pharmacological treatments used for bipolar disorder. Patients are often given mood stabilizers prophylactically and sometimes anti-depressants to help with depressive symptomology (Baldessarini et al., 2019). Common mood stabilizer treatments include lithium (which can help to manage suicidality), antipsychotics, and anticonvulsants (Baldessarini et al., 2019). Worryingly, untreated bipolar disorder is associated with longer duration of illness and a higher frequency of suicide attempts, which emphasizes the importance of proper diagnosis and treatment (Altamura et al., 2010).

1.1.1 Diagnostic Criteria

The two main classifications of bipolar disorder are type I (BD-I) and type II (BD-II), differentiated by the experience of mania, which is also what distinguishes BD from other mood disorders. The estimated lifetime prevalence of BD-I and BD-II is 1.1% and 1.2%, respectively (Clemente, 2015).

Although those with BD-I may also experience other types of mood episodes, an individual must have experienced at least one manic episode to meet criteria for BD-I (American Psychiatric Association, 2013; Kaltenboeck et al., 2016) Conversely, a diagnosis of BD-II is determined by the experience of at least one hypomanic and one depressive episode (American Psychiatric Association, 2013; Kaltenboeck et al., 2016). It is important to note that although the episodes and courses of BD-I and BD-II may present differently, that does not mean that either type is more severe than the other (Kaltenboeck et al., 2016; Karanti et al., 2020; Tondo et al., 2022). For example, BD-I has been associated with higher rates of hospitalization, higher body

mass index (BMI), and lower general functioning (Karanti et al., 2020; Tondo et al., 2022), but BD-II has been associated with higher suicide attempts as well as higher rates and longer periods of depression (Karanti et al., 2020; Tondo et al., 2022).

Mania is defined by a state of extremely elevated mood and energy or extreme irritability lasting a period of one week or more (unless hospitalization is required earlier) (American Psychiatric Association, 2013). Diagnostically, at least three of the following seven criteria must be met in addition to elevated mood and energy: inflated self-esteem or grandiosity, decreased need for sleep, marked increase in talkativeness, racing thoughts, distractibility, increase in goal-directed activity or psychomotor agitation, and involvement in risky behaviour (American Psychiatric Association, 2013). Four of these seven criteria must be met if the initial presentation is extreme and abnormal irritability as opposed to elevated mood (American Psychiatric Association, 2013). Mania can also involve psychotic features such as delusions and hallucinations (American Psychiatric Association, 2013).

In contrast, depressive episodes are characterized by a period lasting a minimum of two weeks with a marked low mood and decreased interest in most activities (anhedonia) (American Psychiatric Association, 2013). Diagnostically, a depressive episode must meet at least three of the following seven criteria in addition to depressed mood and anhedonia: weight loss/gain and/or increase/decrease in appetite, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness, diminished concentration, and recurrent thoughts of death (American Psychiatric Association, 2013). Importantly, the endorsed symptoms must be present nearly every day during the depressive episode.

The other types of mood episodes that can be experienced in BD include hypomanic and mixed episodes. Hypomania involves the same diagnostic criteria as mania but is less severe as

these episodes never require hospitalization and do not cause severe impairment in functioning. In addition, hypomania must last at least four days as opposed to one week (American Psychiatric Association, 2013). Mixed episodes are characterized by the co-occurrence of manic and depressive symptoms wherein symptom threshold for a manic or hypomanic episode is met with at least three depressive symptoms present (American Psychiatric Association, 2013; Substance Abuse and Mental Health Services Administration, 2016).

1.1.2 Disease Burden

Bipolar disorder leads to disability and a decreased quality of life (QoL), as well as high mortality rates and increases in economic burden. There is evidence that BD is among the top five mental illnesses with the highest disability burden in Ontario (Ratnasingham et al., 2013). Globally, BD is one of the top twenty causes of disability due to its severity and chronicity, with population growth and aging contributing to an increased burden since 1990 (Ferrari et al., 2016). Qualitative research revealed that when individuals were asked about the impact of BD of their QoL, they reported difficulties navigating areas in life including routine, independence, stigma and disclosure, identity, and social support (Michalak et al., 2006). Along with lower QoL scores compared to healthy controls (HC), decreased QoL is associated with deficits in cognitive functioning in BD (Brissos et al., 2008), suggesting that functional impairment may be implicated in poorer QoL in this disorder.

Even more consequential, BD is a disorder with excess mortality, and suicide is a primary contributor. It is estimated that the rate of suicide in BD is 20-30 times greater than the general population (Miller & Black, 2020). Along with longer periods of untreated illness, depressive and mixed episodes are the most influential on suicidality (Altamura et al., 2010; Miller & Black,

2020). Previous suicide attempts are also related to poorer quality of life compared to those with no previous attempts (Miller & Black, 2020).

Bipolar disorder is also associated with a high economic burden both on medical systems and individuals. In the United States, BD-I has an estimated total cost of $202.1 billion, averaging approximately $81,559 per diagnosed individual (Cloutier et al., 2018). The costs contributing most to this economic burden are unemployment (36%), caregiving (25%), and direct healthcare costs (e.g. emergency care, inpatient and outpatient services, and pharmacy costs) (23%) (Cloutier et al., 2018). Other countries with different healthcare systems such as France and Sweden still suggest economic burden associated with BD, with an average cost of € 6910 and € 28,000 per year respectively and the cost of BD-I in France is on average 2.1 times higher (Laidi et al., 2022). The indirect costs of BD are also important to consider. Many individuals with BD have trouble staying employed and among those who are employed, there are high rates of occupational impairment due to absenteeism and decreased productivity (Fagiolini et al., 2013).

1.2 Functional Impairment

BD is associated with severe functional impairment in work, family, and social life (Sanchez-Moreno et al., 2009) and high levels of cognitive impairment (Depp et al., 2012). Some common scales used to measure certain types of functioning including the Functional Assessment Short Test (FAST) and the Cognitive Failures Questionnaire (CFQ).

The FAST has strong psychometric properties as scores are lower for individuals with BD in euthymic states as compared to manic and depressive states (Rosa et al., 2007), making it a useful tool to measure psychosocial functioning in this population. The FAST measures functioning in six domains including autonomy, occupational functioning, cognitive functioning,

financial issues, interpersonal relationships, and leisure time. A meta-analysis has demonstrated that during euthymia, BD shows a marked impairment in all domains of the FAST compared to HC (Léda-Rêgo et al., 2020). More specifically, 58.6% of BD participants demonstrated global functional impairment with the occupational functioning domain being the most affected (Léda-Rêgo et al., 2020). The FAST has been further used to capture changes in functioning in BD. Three distinct functional trajectories were established in a sample of BD subjects in various mood states measured over three years: stable mild functional impairment (72%), stable severe functional impairment (20%) and moderate impairment that improved over time (8%) (Godin et al., 2020). Among many baseline factors, the severe impairment trajectory was associated with more frequent unemployment, longer duration of illness, a higher number of lifetime hospitalizations, more residual depressive symptoms, and more comorbidities compared to the mild impairment group (Godin et al., 2020).

The CFQ assesses subjective cognitive functioning by asking questions related to cognitive errors in daily activities (Broadbent et al., 1982). The items of the CFQ yield three component scores clustered by related items: forgetfulness, distractibility, and false triggering. Self-reported psychosocial challenges have been associated with subjective cognitive complaints in BD (Demant et al., 2015). Further, a sample of individuals with BD-I demonstrated that subjective cognitive functioning was related to psychosocial functioning, suicidal ideation, and occupational functioning (Luo et al., 2020). Approximately 87% of depressed participants and 78% of euthymic subjects demonstrated subjective cognitive dysfunction, and both groups demonstrated more dysfunction than HC (Luo et al., 2020). There is evidence that subjectively reported cognitive function is not correlated with objective cognitive functioning in BD (i.e. formally measured using neurocognitive tests) (Demant et al., 2015; van der Werf-Eldering et

al., 2011). Even in healthy populations, correlations between cognitive failures and objective functioning are weak (Carrigan & Barkus, 2016). But importantly, reported cognitive failures relate to real-life outcomes (e.g. car accidents, university entrance grades, and spousal observations) (Carrigan & Barkus, 2016). Hence, subjective reporting may relate to cognitive capacity if not cognitive performance (Carrigan & Barkus, 2016) and may still have important implications for functional impairment. Although this research is cross-sectional, it offers some evidence of the validity of exploring subjective cognitive impairment as a form of functional impairment in BD-I.

1.3 Illness Severity

To understand the relationship between illness severity and functional outcomes in BD, staging models have been proposed. These models suggest illness progression in BD and include various clinical stages where later stages represent more severe progression. Further, they suggest that this disorder changes over time and earlier intervention may lead to better outcomes (Berk et al., 2007; Berk et al, 2017; Frank et al., 2015).

The various staging models primarily use two different proxy measures to define illness progression: episode recurrences and functioning during euthymia (Berk et al., 2007; Kapczinski et al., 2009; Martino et al., 2016). The stages involved in models using mood episodes as main measures include: 0, asymptomatic and at risk; 1a, non-specific distress or symptoms; 1b, high-risk and sub-threshold; 2, first episode; 3a, recurrence of symptoms; 3b, first relapse; 3c, multiple relapses; 4, treatment resistance or unremitting illness (Berk et al., 2007; Berk et al., 2017). The stages involved in models using impairment in euthymia as main measures include: 0, at-risk with latent symptoms; 1, well-defined periods of euthymia; 2, symptoms in interepisodic periods;

3, marked impairment in cognition and functioning; 4, unable to live autonomously (Kapczinski et al., 2009).

Longitudinal research conducted over three years clustered bipolar participants initially ranging from least to most-severe based on clinical characteristics, physical health, cognition, real-world functioning and health-related QoL (de la Fuente-Tomas et al., 2020). They found that 49.6% of patients remained at the same stage, 20.9% progressed one stage (i.e. worsened) and 23.3% regressed one stage (i.e. improved) (de la Fuente-Tomas et al., 2020). Furthermore, 85% of patients who stayed euthymic during that period remained at the same stage or regressed to previous stages (i.e. improved) (de la Fuente-Tomas et al., 2020). Using retrospective longitudinal life chart data of individuals in the first 5 years after onset, van der Markt et al. (2019) found the majority (72%) of BD participants progressed to stage 3 (marked impairment in cognition and functioning), as defined by episode recurrence. Although these studies did not exclude for comorbid disorders, they offer interesting insight into how clinical staging models can be applied to this population.

There is extensive research in the areas of neuroimaging, neurocognition, biomarkers, and functional outcomes that has found evidence of progression in BD. For example, longitudinal neuroimaging research found that manic episodes were associated with faster frontocortical thinning in BD, which suggests that symptomatology may have a role in these structural brain changes (Abé et al., 2022). In terms of cognitive impairment, the number of episodes and number of hospitalizations seem to be related to a decline in cognitive abilities cross-sectionally and over the course of BD (Cardoso et al., 2015). Conversely, first-treatment BD-I participants who did not experience relapse over 1-year demonstrated cognitive improvement (Demmo et al., 2017). Although much of the research done using staging models of

illness progression, including studies on neuroprogression and cognitive functioning, are beyond the scope of this thesis, the main proxy measures of severity – mood episodes and impairment in euthymia – are important in the methodology and conceptualization of this research.

1.3.1 Mood Episode Relapse

Recurrence of mood episodes is associated with indicators of clinical illness progression in bipolar disorder. BD is a recurrent disorder with a 25-50% probability of symptom recurrence in the first year of follow up (Etain et al., 2021) and approximately 50% of individuals experience a mood episode after a period of remission (Perlis et al., 2006). A seminal review demonstrated evidence that an increased number of episodes is associated with an increased risk of episode recurrence, duration of subsequent episodes, and symptomatic severity of episodes, supporting a progressive clinical course of illness (Kessing & Andersen, 2017).

Mood episodes may have implications for functional outcomes. Individuals who have experienced multiple mood episodes (i.e. late stage) exhibit worse functioning in autonomy, occupation, cognition, interpersonal relationships, and leisure time compared to those who have experienced only one episode (i.e. early stage) (Rosa et al., 2012). A longitudinal study exploring psychosocial functioning of individuals with BD-I found that 49.4% of participants were characterized by a progression in functional impairment (Lopez-Villarreal et al., 2020). This group of subjects had a higher number of episode relapses and hospitalizations during the 5-year follow up and exhibited worse neurocognitive functioning compared to those who did not have functional impairment over time as well as healthy controls (Lopez-Villarreal et al., 2020).

1.3.2 Impairment During Remission

Remission (or euthymia), which is the return to mood stability with no symptoms of a current mood episode, also offers an opportunity to explore the concept of progression.

Psychosocial and cognitive deficits have been implicated during these periods. It is estimated that between 30-60% of individuals with BD do not regain full functioning in occupational and social settings in periods of remission from mood episodes (MacQueen et al., 2001). Cross-sectional research found evidence of impaired cognitive functioning compared to HC, and impairment in working memory, and verbal learning that were related to poorer functional outcomes (Martínez-Arán et al., 2004). Participants with recurring episodes had worse cognitive functioning in euthymia compared to participants after only one manic episode, although impairments were still observed in the single-episode group (Elshahawi et al., 2011). Longitudinally, euthymic BD patients showed executive functioning and processing speed deficits that were sustained over 2 years, which was not associated with episode relapse (Mur et al., 2008). Euthymic subjects demonstrate better functioning than subjects in an episode, as measured by the FAST, but exhibit worse overall functioning compared to HC in all domains (Fagiolini et al., 2005; Rosa et al., 2009; Rosa et al., 2010).

Although there is a growing argument for sustained impairment in bipolar disorder, it should be understood with caution as there is some contradictory evidence, particularly in the domain of cognitive deficits. After comparing cross-sectional and longitudinal research, a review argues that available evidence does not support the natural course of BD contributing to cognitive decline (Strejilevich et al., 2015), while another more recent review suggests that the findings are mixed due to limited longitudinal studies (Van Rheenen et al., 2020). A six-year longitudinal study including mostly euthymic participants at baseline, found no significant differences in the trajectory of cognitive function between HC and BD participants, regardless of sub-type and manic episodes experienced during the study (Sparding et al., 2021). A meta-analysis of longitudinal studies spanning from 2 to 4.6 years also suggests a lack of sustained

cognitive impairment in BD (Samamé et al., 2014). Last, a sweeping review of longitudinal studies including episode recurrence, cognitive function, functional impairment and response to treatment in BD argues that there are still many gaps in the literature and a lack of consensus around the concept of neuroprogression and more longitudinal studies are needed to clarify this (Martino et al., 2016). Due to the complex nature and therefore conceptualization of the course of bipolar disorder, it becomes important to be clear about the measures being used to investigate various outcomes. In addition, longitudinal study design becomes very important for testing the concept of progression in the course of BD.

1.4 Current Research

The current research involves data from a multi-site study titled The Longitudinal Course of Intracortical Myelination and Cognitive Function in Bipolar Disorder, which was funded by the Canadian Institutes of Health Research. This study recruited healthy controls (HC) and individuals with bipolar disorder type I (BD) for a two-year longitudinal study primarily exploring the effect of relapse in BD on timepoint-to-timepoint trajectories of myelin in the cerebral cortex as measured by magnetic resonance imaging (MRI). This study also collected various clinical, and cognitive parameters as well as blood.

For the present thesis, these longitudinal data were used to explore and compare trajectories of functioning based on episode relapse status using reported psychosocial functioning and subjective cognitive functioning as measured by the FAST and CFQ, respectively.

1.4.1 Primary Goals

Using data from the two-year longitudinal study, the primary goals of this thesis are twofold. The first goal is to characterize the study population using clinical and demographic variables of

interest. The second and primary goal, is to compare trajectories and potential changes in functioning between healthy controls (HC) and individuals with bipolar disorder type I who experienced episode relapse (BDR) and those who did not (BDNR), by using their total FAST and CFQ scores.

1.4.2 Significance

This research provides several methodological strengths that could help to further understand functional impairment in bipolar disorder. First, this research will investigate psychosocial and subjective cognitive impairment in BD as compared to HC over time to support evidence of sustained deficits in BD. Second, evaluating BD subjects in a state of euthymia, coupled with grouping BD subjects by relapse status (i.e. experiencing more recent recurrence), could provide helpful insights as to whether models of illness progression in functional impairment are substantiated in BD. Third, using longitudinal data allows for a direct comparison of those who experienced relapse with those who did not as opposed to comparing late stage and early stages of BD cross-sectionally. With such high rates of burden in BD, this research could add to the existing literature regarding how episode relapse may be implicated in functional impairment during euthymia.

2. Methods

The primary research study titled The Longitudinal Course of Intracortical Myelination and Cognitive Function in Bipolar Disorder was approved by the Hamilton Integrated Research Ethics Board (Project Number: 2456) and was funded by the Canadian Institutes of Health Research. Five research sites participated in subject recruitment including McMaster University (Hamilton, Ontario), Queen's University (Kingston, Ontario), Dalhousie University (Halifax,

Nova Scotia), the University of Calgary (Calgary, Alberta), and the University Health Network (Toronto, Ontario). All participating sites followed the same study protocols.

2.1 Overall Study Procedure

This thesis analyzes data from the larger longitudinal study, which was conducted over two years with a total of three visits. First, informed consent was obtained, and participants were subsequently screened for eligibility. Next, after participants were deemed eligible, a baseline visit took place. The baseline visit was followed by a second visit conducted approximately 1 year later, and a final visit conducted approximately 2 years from baseline. Visits took place as closely to 1 year apart as possible.

At the baseline visit, subjects completed a variety of clinical and cognitive measures as well as an MRI scan and a blood draw. The baseline visit took approximately 8 hours, with breaks, to complete the 2.5-3 hour cognitive battery and 1 hour MRI scan. At the follow up visits, which took approximately 4 hours, participants completed a condensed version of the baseline visit with fewer measures. Specifically, IQ measures were not administered as part of the cognitive battery and fewer clinical assessments were required. Participants were compensated $100 at each study visit.

2.2 Participant Overview

Subjects were recruited directly from clinics operated at the participating sites and through local advertisements. All participants were between 16 and 45 years old. A total of 84 healthy controls (HC) and 85 individuals with bipolar disorder type I (BD) were recruited and enrolled across all sites. 15 subjects failed to meet eligibility criteria for the study, such that 83 HC and 71 BD completed a baseline visit. Of these 154 participants, 59 HC and 42 BD completed the year 1 follow-up visit and 45 HC and 36 BD completed the year 2 follow-up visit.

Study dropout between visits was primarily due to subjects being lost to follow up but some participants withdrew from the study due to having moved away or not being able to take the time to attend the study visits.

To meet eligibility criteria for the study at baseline, HC subjects did not have any current or lifetime psychiatric conditions according to the Structured Clinical Interview for DSM-5 (SCID-5) (American Psychiatric Association, 2013). Conversely, BD subjects had a diagnosis of BD type-I according to the SCID-5 Axis I Disorders and were in a state of euthymia, meaning they did not meet criteria for a mood episode within the past month. We chose to recruit participants during euthymia both to reduce complications during the cognitive and neuroimaging portion of the study, but also to more easily capture relapse between study visits and reduce variability in the study. The BD subjects were required to have had no medication changes for the previous 2 months before enrolment, although psychiatric medication use could change at follow up visits. Similarly, the BD subjects were required to have had no presence of any current co-morbid psychiatric disorders including substance/alcohol use disorders within the past 6 months, with the exception of current anxiety disorders. Due to the high prevalence of co-morbid anxiety disorders with BD-I, these disorders were the only accepted comorbidity to help with recruitment. Alcohol and substance use patterns could change status at follow up visits.

Exclusion criteria applied to all participants included 1) presence of unstable medical conditions; 2) presence of any MRI contraindications; 3) history of neurological disorders including head trauma resulting in loss of consciousness and severe migraines; 4) pregnancy. These criteria were important to ensure safety during the MRI scan as well as ensuring that any findings from the imaging data were not due to unstable medical conditions or past head trauma.

2.3 Measures

Several clinical assessments and cognitive measures were used at the study visits but only those relevant to this thesis will be detailed below. These measures were administered by trained research staff.

2.3.1 Demographics

Basic demographic information was collected at baseline including age, sex, ethnicity, years of education, employment status, smoking status, and past and current health conditions. In addition, non-psychiatric medications and family history of mental illness data were collected. For the BD group, age of onset of first episode, duration of illness, lifetime episodes (total, depressive, manic, hypomanic, and mixed), and current psychiatric medications were collected. At follow-up visits, all participants were asked to report any changes in smoking status, health condition status, and non-psychiatric medication use. BD subjects were asked about mood episode relapse (described below) and any changes in psychiatric medication use including discontinued medication, new medication type(s), and dose(s).

2.3.2 Structured Clinical Interview for DSM-5 (SCID-5)

The complete SCID-5 was administered at baseline to determine eligibility for all participants. Completing the entire SCID-5 ensured the HC group did not have any psychiatric diagnoses and confirmed the BD group presented with symptoms of BD-I with no current mood state symptoms (depressive, (hypo)manic, mixed) or co-morbid disorders other than anxiety disorders. At follow-up visits, only the mood, alcohol use, and substance use modules of the SCID-5 were administered. These specific modules were administered at follow-up visits for several reasons. First, they ensured the BD participants were euthymic at the time of the visit. Second, the past mood episode modules verified episodes within the previous year by asking BD

participants about symptoms exhibited since the last research visit as opposed to their lifetime experience. Third, the alcohol and substance use modules captured any changes in alcohol and substance use severity for all participants. Last, administering these modules to the HC participants ensured they were not presenting with symptoms of BD or other potential mood disorders.

2.3.3 Young Mania Rating Scale (YMRS)

The Young Mania Rating Scale (YMRS) is an 11-item clinician-administered questionnaire that was administered at each visit to all participants. The YMRS assesses the presence and severity of (hypo)manic symptoms within the previous week. Items include: elevated mood, increased energy, sexual interest, reduced sleep, irritability, speech, language – thought disorder, content, disruptive behaviour, appearance, and insight, with some items scored on a scale of 4 and some on a scale of 8. This instrument ensured no significant (hypo)manic symptoms were present at the time of the research visits. The total score is the sum of all item scores. Linking the YMRS to the Clinical Global Impressions Scale (CGI-S), recently proposed cut-off scores are as follows: 6 corresponds to ‘borderline mentally ill’; 12 to ‘mildly ill’; 20 to ‘moderately ill’; 30 to ‘markedly ill’; 40 to ‘severely ill’; and 52 to ‘among the most extremely ill’ (Samara et al., 2023). The YMRS has high internal consistency (r= 0.91) and inter-rater reliability (0.93) (Young et al., 1978).

2.3.4 Montgomery-Asberg Depression Rating Scale (MADRS)

The Montgomery-Asberg Depression Rating Scale (MADRS) is a 10-item clinician-administered questionnaire that was also administered at each visit to all participants. The MADRS assesses the presence and severity of depressive symptoms in the past week. Items include: apparent sadness, reported sadness, inner tension, reduced sleep, reduced appetite,

concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts, each scored on a scale of 0 to 6. This instrument ensured that no significant depressive symptoms were present at the time of the research visits. The total score is the sum of all item scores. The cut-off scores for various levels of severity are as follows: 0 to 6 represents recovery; 7 to 19 represents mild depressive symptoms; 20 to 24 is moderate; and 35 to 60 is severe (Snaith et al., 1986). The MADRS has high internal consistency (r= 0.95) and inter-rater reliability (0.89-0.97) (Montgomery et al., 1979).

2.3.5 Functional Assessment Short Test (FAST)

The Functional Assessment Short Test (FAST) is a 24-item scale that assesses general psychosocial functioning across various domains. The areas of functioning measured include: autonomy, occupational functioning, cognitive functioning, financial issues, interpersonal relationships, and leisure time. Participants were asked to rate the level of difficulty experienced/associated with each item over the last 15 days, on a scale of 0 to 3, with 0 representing no difficulty and 3 representing great difficulty. If an item did not apply (e.g. living alone), participants were instructed to think hypothetically about the item and respond accordingly. The total FAST score is a sum of all item scores. The FAST was administered at each visit to all participants. Suggested threshold scores include: 0 to 11 represents no impairment; scores from 12 to 20 represent mild impairment; scores from 21 to 40 represent moderate impairment; and scores above 40 represent severe functional impairment (Bonnìn et al., 2018).

2.3.6 Cognitive Failures Questionnaire (CFQ)

The Cognitive Failures Questionnaire (CFQ) is a 25-item self-report questionnaire assessing subjective cognitive functioning. The CFQ was administered at each visit to all

participants. Participants were asked to report how often each of the items happened to them over the past 6 months on a scale of never to very often. If an item did not apply (e.g. "do you fail to notice signposts on the road?"), participants were instructed to think hypothetically about the item and respond accordingly. To generate a total score for the CFQ, each response was converted to a numerical value ranging from 0 to 4, respectively. The total CFQ score is a sum of all converted numerical item scores. These numerical values also yield three component scores clustered by related items: forgetfulness, distractibility, and false triggering. There are no reliable suggested threshold scores for the CFQ, but higher scores indicate more cognitive failures.

2.3.7 Mood Episode Relapse for BD Participants

To capture mood episode relapse between timepoints, BD participants were asked to report the number of episode relapses (depressive, manic, hypomanic, or mixed) that had occurred since the last visit (baseline or year 1 timepoint). These values were then verified using the past mood episode modules of the SCID-5, wherein the questions were asked about symptoms experienced since the last visit (approximately 1 year previously) instead of lifetime symptoms. To further verify episode relapse, participants were asked to complete The National Institute of Mental Health Life-Chart Methodology (NIMH-LCM) between study visits, which is a self-reported mood scale that helps to track mood episodes over time (Denicoff et al., 2000). Unfortunately, very few participants filled out the NIMH-LCM logs, therefore this measure did not generate robust data for our study. Subsequently, the participants were categorized for analyses as a subject who experienced relapse at any point in the study (BDR) or a subject who did not experience any relapse during the study (BDNR).

2.4 Quality Control Procedure

Due to the longitudinal and multi-site nature of the study, a rigorous quality control (QC) procedure was necessary to ensure that the data quality was consistent across sites. All the study data including neuroimaging and cognitive data went through a QC process, but for the purposes of this thesis, only the process used for the demographics and clinical data will be detailed here. Clinical and neurocognitive data (except computer-based tests) were captured and documented on paper and subsequently entered through Research Electronic Data Capture (REDCap) after study visits were completed. All data detailed below are housed through a study-specific REDCap managed by McMaster University.

2.4.1 Demographics

All demographic data were verified but to complete the demographic table, age, years of education, and duration of illness (for BD participants only) went through a QC procedure. Age was verified using the date of the study visit and the subjects' birthdate to get a precise age. This was primarily done for neurocognitive scoring QC but to keep data consistent for all analyses, this value was used for the present research as well. For duration of illness, values were documented inconsistently with some noted as time since diagnosis while others were time since first episode (age at onset). Since duration of illness represents the number of years since the first mood episode, age of onset was used to verify this value to ensure consistency. Therefore, duration of illness was verified by using the QC age at the study visit and age of onset. Years of education were standardized based on what was selected as the highest level of education on the Case Report Form (CRF) if the years of education did not match up with the selected highest level of education.

2.4.2 Episode Relapse

To verify mood episode relapse between visits for BD subjects, several steps were involved. First, the demographic form administered at baseline and the follow-up form administered at the second and third visits were verified for number of lifetime episodes listed (depressive, hypomanic, manic, and mixed). At follow-up visits, if an episode relapse occurred and was verified by the SCID-5 (detailed below), this value was added to the lifetime episodes on the CRF at each visit. The QC process therefore started with these CRFs as, reported relapse would change the value of lifetime episodes at follow-up visits after being verified by the SCID-5 past mood episode modules. Participants were not asked to report lifetime episodes at each visit but rather, reported relapses were documented and then added to lifetime episodes. Hence, lifetime mood episodes were first verified to see if there were any discrepancies between timepoints (e.g. value decreasing instead of increasing or remaining the same). Although number of episodes is a variable that is not consistently defined or measured, it is very commonly used (Tremain et al., 2020).

Next, to ensure that reported relapses were accurate, the SCID-5 mood modules were verified, as this was used to verify episodes between follow-up visits. At baseline lifetime episodes were verified using the SCID-5 past mood modules and at follow-up visits, participants were asked if they had an episode relapse since the last visit, which was also verified using the SCID-5. Therefore, since the SCID-5 was used to verify the increase in lifetime episodes or consistent value of lifetime episodes, it was verified again to ensure it was consistent with what was reported on the CRFs.

Last, if a subject brought NIMH-LCM logs with them, the logs were reviewed with the participant at the time of the visit to ensure their reporting was accurate. During the QC process, these logs were reviewed again as a secondary measure to validate relapses status.

2.4.3 Scale Data

The quality control procedure for the FAST and CFQ data across all sites involved first ensuring data entry on REDCap was accurate. Since there was a possibility of data entry errors when inputting scale data into the electronic database, a double data entry system was set up wherein the data were entered twice by the same person, with some time between entries. The two entries for the FAST and CFQ for each participant at each visit were compared to find any discrepancies. If any discrepancies were found, the source documents were reviewed to fix the error and update the value(s) accordingly.

Next, it was necessary to address missing items on the two scales for the purpose of the current analyses. Only 5.6% (n= 6, HC) of participants used in the analyses were partial respondents across both scales (FAST and CFQ). Only 2 HC participants were missing the FAST in its entirety at the third timepoint, and this was only relevant for the simple growth curve analysis. The remaining participants were considered full respondents with responses to all constructs on the scales. To reduce listwise deletion and therefore maximize useable data, person-mean imputation was used. This process involves using the mean of available items for a given participant's scale as the value for any missing items. This procedure is considered appropriate for item-level missing data (Newman, 2014).

2.5 Data Analyses

All statistical analyses were conducted using R Studio version 2023.12.0+369. An alpha level of 0.05 was used for all analyses.

2.5.1 Demographics at Baseline

Means and standard deviations of age, years of education, YMRS score, and MADRS score, were calculated for both HC and BD at baseline and age of onset and duration of illness

for BD subjects only. Levene and Shapiro-Wilk tests were conducted for age, years of education, sex, smoking status, YMRS scores, and MADRS scores at baseline. From the results of the Levene and Shapiro-Wilk tests, Mann-Whitney U tests were conducted for age, years of education, YMRS scores, and MADRS scores to compare the HC and BD groups baseline. Chi-squared tests were conducted for sex and smoking status to compare the groups at baseline. The same process was completed for the BDR and BDNR groups at baseline with the addition of age at onset and duration of illness. In addition, this process was also completed to compare participants dropouts to participants who did not drop out, with the BD and HC groups compared separately.

2.5.2 Medication Status for BD Groups

BD participants were asked at each visit for any changes in medication status, type, and dose. The medication status and different types of medication use for the BD participants is listed in a table comparing the BDR and BDNR groups. To compare the medication status between the BDR and BDNR groups throughout the study, a chi-squared test was conducted.

2.5.3 Main Analyses

Three data-driven approaches were used to analyze the longitudinal FAST and CFQ data. The first two re-binned data across the 3 timepoints into baseline and follow-up timepoints to account for missing timepoint data. The third analysis used data across all three timepoints. First, mixed-effects Analysis of Variances (ANOVA) were conducted using 2 timepoints to explore any differences in FAST and CFQ scores between and within the HC, BDR, and BDNR groups at baseline and follow-up. Next, one-way Kruskal Wallis tests using delta FAST and CFQ scores were conducted to compare any differences in the change in scores between groups. Last, a simple linear growth curve analysis was conducted to explore any differences in the change in

FAST and CFQ scores over 3 timepoints between and within groups using all available data. The final sample used for all analyses was 61 HC, 21 BDR, and 26 BDNR (n=108).

For the mixed-effects and one-way ANOVAs, each participant was assigned a Baseline (B) and Follow-up (F) visit for a total of 2 visits. Due to dropouts between visits for both HC and BD subjects and therefore incomplete data for many participants, these types of analyses could not be conducted by using all available data for 3 timepoints. To capture as many participants as possible and to be able to run the ANOVAs and Kruskal-Wallis tests, follow-up visits were re-binned depending on available data. Therefore, the initial baseline visit 1 was used for all participants as the B visit. Visit 2 was used as the F visit for 102 participants (60 HC, 20 BDR, 22 BDNR) and visit 3 was used as the F visit for 6 participants (1 HC, 1 BDR, 4 BNDR). Visit 2 was used as the F visit for the majority of participants due to dropout between Visit 2 and 3, as well as many BD participants experiencing a relapse between baseline and visit 2. Since these analyses are exploring the potential effect of relapse on FAST and CFQ scores as measures of functioning, it was logical to use the follow-up visit closest to a recent relapse as possible. Visit 3 was used for the 6 remaining participants due to them not being able to complete a visit 2 in the study. The final sample for these analyses were 61 HC, 21 BDR, and 26 BDNR.

For the growth curve analyses, all available data were used for the same sample used in the other analyses. The final sample used was also 61 HC, 21 BDR, and 26 BDNR (n=108), with an additional 16 visit 3s for the BDNR group, 15 for the BDR group, and 42 for the HC group included.

2.5.3.1 Mixed-Effects ANOVA: Between and Within Group Differences in Scores

To explore if the 3 groups (HC, BDR, BDNR) differed significantly at B and F visits in their FAST and CFQ scores, a mixed-effects ANOVA was conducted. The Shapiro-Wilk test was

conducted to assess normality of the FAST and CFQ data and the Levene test was used to assess the homogeneity of variances. From the outcomes of these tests, a robust ANOVA on trimmed means was performed using the WRS2 package in R (Mair & Wilcox, 2020). Post-hoc non-parametric Mann-Whitney U and Wilcoxon Signed Rank tests were conducted to identify between and within group differences at B and F visits. Mann-Whitney U tests are the non-parametric equivalent of an independent samples t-test and were therefore used to explore between-group differences, whereas Wilcoxon Signed Ranked tests are the non-parametric equivalent to dependent samples t-tests and were used to explore the within-group differences. The models used for this analysis were as follows:

FAST Total Score ~ Group + Visit + Group*Visit

CFQ Total Score ~ Group + Visit + Group*Visit

2.5.3.2 One-Way Kruskal-Wallis: Between Group Differences in Delta Scores

To explore if the 3 groups (HC, BDR, BDNR) differed significantly on the change in scores between the B and F visits, one-way Kruskal-Wallis tests were conducted. The delta value for the FAST and CFQ scores between B and F were calculated for each subject. The Shapiro-Wilk test was conducted to assess normality of the delta FAST and delta CFQ data and the Levene test was used to assess the homogeneity of variances. From the outcomes of these tests, one-way Kruskal-Wallis and post-hoc Dunn tests were conducted. The models used for this analysis were as follows:

Delta FAST Score ~ Group

Delta CFQ Score ~ Group

2.5.3.3 Growth Curve Analysis (GCA)

To explore any differences in the trajectories of change in FAST and CFQ scores between and within the 3 groups (HC, BDR, BDNR), simple growth curve analyses (Mirman, 2014) were also conducted. To compare the trajectories of scores across all available data and 3 timepoints, three linear models were run for both the FAST and CFQ data using the lme4 package in R to compare model fit. The first model included timepoints as the fixed effect with subject as the random effect. The second model added a fixed effect of group to test any differences between the groups' intercepts. The third model added the effect of group to the linear term to test any differences between the groups' slopes. Although the reported results used the HC group as the reference group, all 3 groups were used as reference groups in order to compare all potential group differences in intercepts and slopes. The three models were compared using maximum likelihood estimation *(MLE)* (Mirman, 2014). Improvements in model fit were evaluated using the change in the log-likelihood (LL) value as measured by the chi-squared test (χ^2) (Mirman, 2014).

3. Results

3.1 Demographics at Baseline

Basic demographic and clinical characteristics of the HC and BD groups at baseline are listed in the demographic table below. This includes the entire sample measured at baseline. The HC group had a mean age of 27 (SD = 7) and 32 (SD = 8) for BD, with the BD group being significantly older (p=<0.001) (Table 1). Similarly to age, the groups were not matched for years of education at baseline, with the HC group having more years of education (p= 0.026) (Table 1). The HC and BD groups were not matched for smoking status (p=0.01) but were matched for sex (p=0.95) (Table 1). The HC and BD groups differed significantly on the MADRS and YMRS

scores at baseline with the BD group having higher scores across both scales (p=<0.001) (Table 1).

Table 1

Demographic and Clinical Characteristics of HC and BD at Baseline

	BD (n=71) *M(SD)*	HC (n=83) *M(SD)*	Test statistic	p-value
Age	32.3 (7.6)	27.3 (7.3)	W = 4023.5	<0.001*
Female *N* (%)	38 (54%)	44 (53%)	χ^2=0.004	0.95
Race/Ethnicity *N* (%)				
White South Asian Latin American Mixed Race Arab Black Chinese East Asian Filipino Korean Southeast Asian Unspecified	61 (86%) 3 2 5	53 (64%) 6 6 1 2 3 4 2 1 2 2 1		
Education (years)	14.5 (2.6)	15.3 (2.5)	W = 2347	0.026*
Smoking *N* (%)				
Yes	14 (20%)	5 (6%)	χ^2=6.635	0.01*
No	57 (80%)	78 (94%)		
Age of onset	18.3 (5.7)	-	-	-
Duration of illness	14.5 (8.7)	-	-	-
YMRS score	1.9 (2.4)	0.4 (0.8)	W=4349	<0.001*
MADRS score	4.9 (4.3)	1 (1.9)	W= 4794	<0.001*

*p=<0.05

Baseline demographics for the BDR and BDNR groups are compared in the table below (Table 2). The two groups were matched on all demographic variables except sex, although this *p-value* was approaching non-significance.

Table 2

Baseline Demographics of BDR and BDNR Groups

	BDR (n=21) *M(SD)*	BDNR (n=26) *M(SD)*	Test statistic	p-value
Age	30.7 (8.6)	34.8 (7.4)	t = -1.7704	0.08344
Female *N* (%)	15 (71 %)	11 (42%)	χ^2=3.9857	0.04589*
Race/Ethnicity *N* (%)				
White South Asian Latin American Mixed Race	19 (90%) 1	24 (92%) 1 1		
Education (years)	14.5 (2.1)	15 (2.4)	W = 236.5	0.4269
Smoking *N* (%)				
Yes	3 (14%)	7 (27%)	χ^2= 1.1076	0.2926
No	18 (86%)	19 (73%)		
Age of onset	17.2 (4.2)	19 (6.8)	W = 239	0.8626
Duration of illness	14.5 (9.5)	15.9 (8.9)	t = -0.50411	0.6168
YMRS score	1.9 (1.5)	1.56 (2.2)	W = 182	0.07276
MADRS score	4.1 (3.6)	3.88 (3.6)	W = 269.5	0.885

*p=<0.05

3.2 Demographics of Dropouts

Basic demographic and clinical characteristics at baseline of the BD and HC subjects who dropped out between visits are detailed in the tables below (Table 3; Table 4). These tables represent the participants who dropped out between baseline and visit 2 (year 1) (BD, *n*=24; HC, *n*=22) or between visit 2 (year 1) and visit 3 (year 2) (BD, *n*=11; HC, *n*=16), compared to the participants who did not dropout throughout the study (BD, *n*=36; HC, *n*=45). It is important to note that the sample of non-dropouts does not line up with the final study sample since the dropouts between visit 2 and visit 3 were included in the analyses. At baseline, the BD dropouts and non-dropouts were matched for age, sex, years of education, smoking status, age of onset,

duration of illness, MADRS scores, and YMRS scores. The HC dropouts and non-dropouts were also matched for age, sex, years of education, smoking status, MADRS scores, and YMRS scores.

Table 3

Baseline Demographics of BD Dropouts

	Dropouts *(n*=35) *M(SD)*	Non-dropouts (*n*= 36) *M(SD)*	Test statistic	p-value
Age	31.8 (7)	32.9 (8.1)	W = 580.5	0.573
Female *N* (%)	21 (60%)	17 (47%)	χ^2= 1.1648	0.2805
Race/Ethnicity *N* (%)				
White South Asian Southeast Asian Latin American Mixed Race	28 (80%) 1 1 1 4	33 (92%) 1 1 1		
Education (years)	14 (2)	15 (2.3)	W = 475.5	0.0676
Smoking *N* (%)				
Yes	7 (20%)	7 (19%)	χ^2= 0.00346	0.9531
No	28 (80%)	29 (81%)		
Age of onset	17.3 (5.2)	19.1 (7)	W= 427	0.279
Duration of illness	14.9 (8.3)	14.1 (9.4)	W= 557	0.6785
YMRS score	2.2 (3)	1.5 (1.4)	W= 657.5	0.7486
MADRS score	5.6 (4.8)	4.2 (3.6)	W = 710	0.2518

Table 4

Baseline Demographics of HC Dropouts

	Dropouts *(n=38) M(SD)*	Non-dropouts (*n*= 45) *M(SD)*	Test statistic	p-value
Age	27.6 (7)	27.1 (7.6)	W= 899.5	0.6875
Female *N* (%)	18 (47%)	26 (58%)	χ^2= 0.8962	0.3438
Race/Ethnicity				

N (%)				
White	22 (58%)	31 (69%)		
Arab	1	1		
Black	3	2		
Chinese	2	4		
South Asian	2	3		
Southeast Asian	3			
Latin American	4	2		
Korean	1			
East Asian				
Filipino				
Mixed Race		1		
Unknown		1		
Education (years)	15 (2.3)	15.5 (2.7)	$W = 772.5$	0.4403
Smoking *N* (%)				
Yes	3 (8%)	2 (4%)	$\chi^2 = 0.43323$	0.5104
No	35 (92%)	43 (96%)		
YMRS score	0.4 (0.9)	0.3 (0.6)	$W = 858$	0.9747
MADRS score	1.2 (2.3)	0.9 (1.6)	$W = 914.5$	0.5334

3.3 Medication Status of BD Subjects Throughout Study

The medication status (medicated versus unmedicated) of the BDR and BDNR groups throughout the study are detailed in the table below along with the types of medication used (Table 5). Although some types and doses of medications changed for medicated participants, no BD subjects changed medication status throughout the study. There were no differences in medication status between the BDR and BDNR groups throughout the study (χ^2=0.025, p=0.875).

Table 5

Medication Status of BD Groups at All Visits

	BDR	BDNR
Baseline	(n= 21)	(n=26)
Medicated	17	21
Unmedicated	4	5
Lithium	10	9

Anticonvulsants	7	10
Antidepressants	4	5
Antipsychotics	11	14
Anxiolytics	4	4
Other (CNS Stimulant, Sedative)	2	1,1
Visit 2 (Year 1)	(*n*=20)	(*n*=22)
Medicated	16	17
Unmedicated	4	5
Lithium	9	9
Anticonvulsants	9	6
Antidepressants	6	4
Antipsychotics	11	8
Anxiolytics	3	3
Other (CNS Stimulant, Sedative)	2	1,1
Visit 3 (Year 2)	(*n=16*)	(*n*=20)
Medicated	13	14
Unmedicated	3	6
Lithium	5	8
Anticonvulsants	7	6
Antidepressants	5	3
Antipsychotics	10	10
Anxiolytics	3	2
Other (CNS Stimulant)	2	2

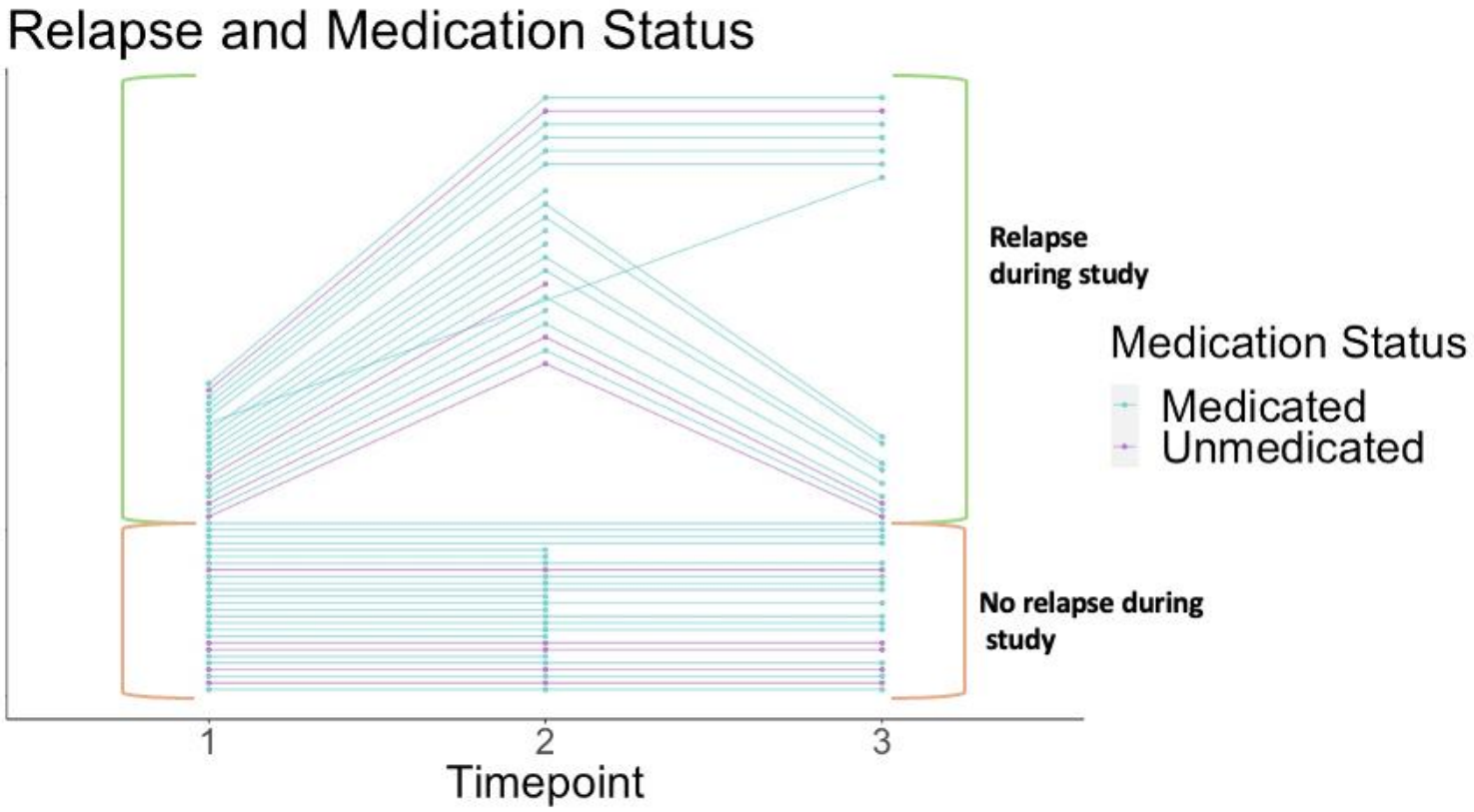
Note: Subjects with only baseline visits excluded since could not determine relapse group status. Values of *n* not same for all visits due to missing data/dropout between visits. Values for types of medications are total *n* participants using at least one of each type.

**p= <0.05*

Figure 1 demonstrates the relapse trajectories of the BD participants throughout the study and is colour-coded for medication status.

Figure 1

Relapse Trajectory of BD Participants with Medication Status (n=47)



Note: Values on y-axis are arbitrary. They are only meant to demonstrate relapse vs. non-relapse group trajectories across timepoints.

3.4 Mixed-Effects ANOVA

3.4.1 FAST

The FAST data failed to meet assumptions because the HC group were not normally distributed, and the homogeneity of variance assumption was violated. The omnibus robust ANOVA revealed a significant group by visit interaction (p=0.015) (Table 6). Post-hoc Mann-Whitney tests revealed significant differences between the BDNR and HC groups at B (p=<0.001) and F (p=<0.001) and between the BDR and HC groups at (p=<0.001) and F (p=<0.001), with the HC group having lower scores at both visits (Table 7; Figure 2). No

significant difference was found between the BDNR and BDR groups at B or F. Wilcoxon Signed Rank tests revealed a significant difference for the BDR (p=0.0043) and BDNR (p=0.0018) groups between B and F, but not the HC group (Table 7; Figure 2).

Table 6

Robust ANOVA Results: FAST

	Value	*df*	*p-value*
Group	35.4261	2	0.000*
Visit	17.4774	1	0.0002*
Group*Visit	5.0132	2	0.0154*

Note: Group was the between-subjects factor and visit was the within-subjects factor, with subject as a random factor. The interaction between the factors "Group" and "Visit" is shown as "Group*Visit". *df,* degrees of freedom.

* $p < 0.05$.

Table 7

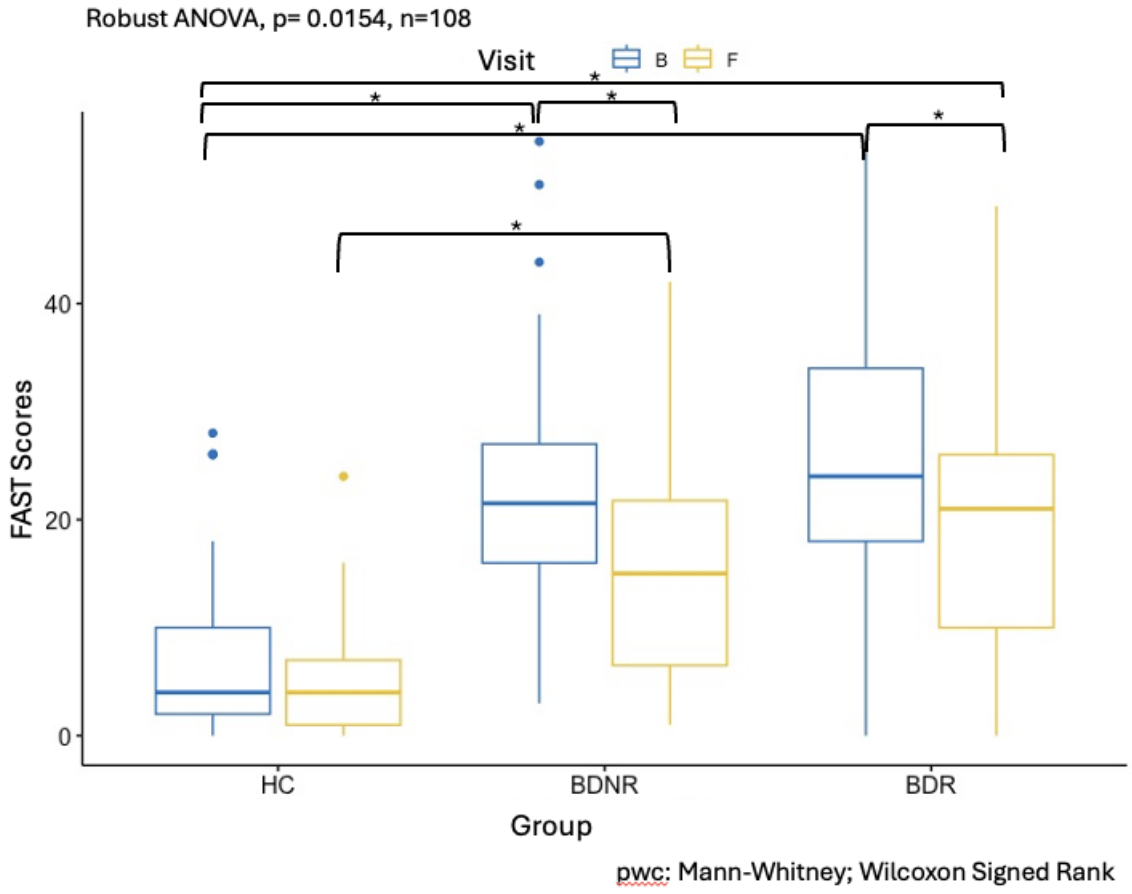
Post-hoc Pairwise Comparisons: FAST

Pairwise comparison	*Test statistic*	*p-value*
BDNR B – BDR B	$W = 247.5$	0.5924
BDNR B – HC B	$W = 1392.5$	<0.001*
BDR B – HC B	$W = 1122.5$	<0.001*
BDNR F – BDR F	$W = 218$	0.2431
BDNR F – HC F	$W = 1320.5$	<0.001*
BDR F – HC F	$W = 1093$	<0.001*
BDNR B – BDNR F	$V = 259.5$	0.001827*
BDR B – BDR F	$V = 151.5$	0.004305*
HC B – HC F	$V = 853.5$	0.01656

Note: *$p < 0.0056$ (due to multiple comparisons)

Figure 2

Results for Robust ANOVA: FAST Scores



*p=<0.05

3.4.2 CFQ

The CFQ data failed to meet all assumptions because the BDR and HC groups were not normally distributed. The omnibus robust ANOVA revealed a significant group by visit interaction (p=0.016) (Table 8). Post-hoc Mann-Whitney tests revealed significant differences between the BDNR and HC groups at B (p=<0.001) and F (p=< 0.0009) and between the BDR and HC groups at (p=<0.001) and F (p=<0.001), with the HC group having lower scores at both visits (Table 9, Figure 3). No significant difference was found between the BDNR and BDR groups at B or F. Wilcoxon Signed Rank tests revealed a significant difference for the BDR (p=0.0053) and BDNR (p=0.0018) groups between B and F, but not the HC group (Table 9; Figure 3).

Table 8

Robust ANOVA Results: CFQ

	Value	*df*	*p-value*
Group	25.5712	2	0.000*
Visit	13.0986	1	0.0011*
Group*Visit	4.8926	2	0.0155*

Note: Group was the between-subjects factor and timepoint was the within-subjects factor, with subject as a random factor. The interaction between the factors "Group" and "Visit" is shown as "Group*Visit". *df,* degrees of freedom.

* $p < 0.05$.

Table 9

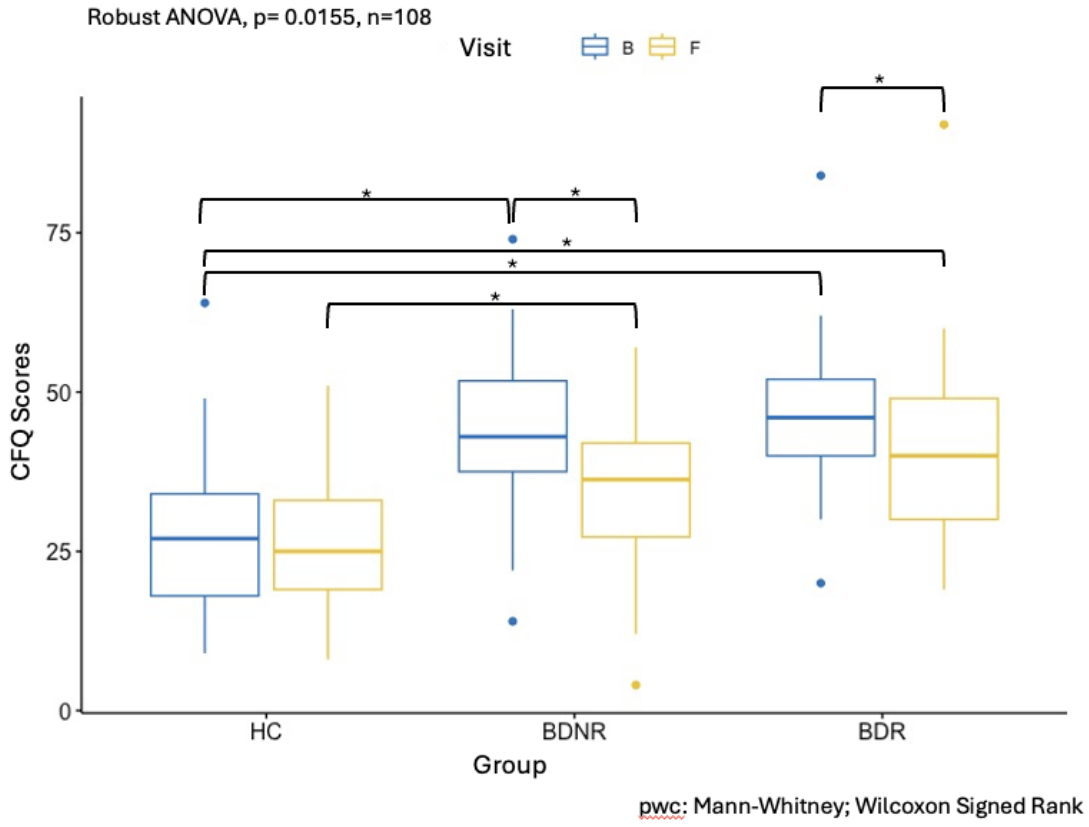
Post-hoc Pairwise Comparisons: CFQ

Pairwise comparison	*Test statistic*	*p-value*
BDNR B – BDR B	$W = 230.5$	0.3684
BDNR B – HC B	$W = 1325.5$	<0.001*
BDR B – HC B	$W = 1150.5$	<0.001*
BDNR F – BDR F	$W = 215.5$	0.222
BDNR F – HC F	$W = 1151.5$	<0.0008953*
BDR F – HC F	$W = 1036.5$	<0.001*
BDNR B – BDNR F	$V = 299$	0.001764*
BDR B – BDR F	$V = 196$	0.005384*
HC B – HC F	$V = 871$	0.9074

Note: *$p < 0.0056$ (due to multiple comparisons)

Figure 3

Results for Robust ANOVA: CFQ Scores



*p=<0.05

3.5 One-Way Kruskal-Wallis

3.5.1 FAST

The delta FAST data failed to meet assumptions because the BDNR and HC groups were not normally distributed, and the homogeneity of variance assumption was violated. Therefore, a non-parametric Kruskal-Wallis test was performed and the significant effect of group was then decomposed using the Dunn test. The omnibus test revealed a significant effect of group (p=0.017) on delta FAST scores with a small effect size (η^2= 0.059) (Table 10). A post-hoc Dunn test was conducted to look at pairwise differences between groups. After p-value corrections, no significant differences were found (Table 11; Figure 4).

Table 10

One-Way Kruskal-Wallis Result: FAST

Dependent variable	*n*	*Statistic*	*df*	*p-value*	η^2
Delta FAST	108	8.15	2	0.017*	0.0586

*$p<0.05$; η^2, eta-squared

Table 11

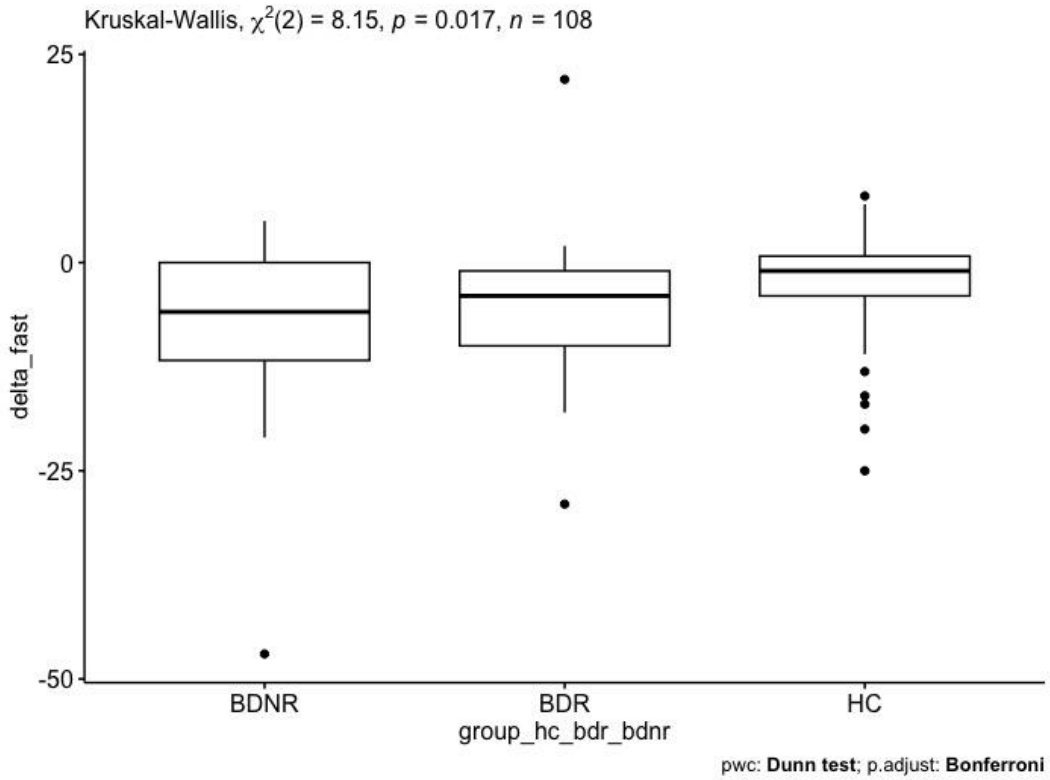
Pairwise Comparisons Using the Dunn Test: FAST

Pairwise comparison	*Statistic*	*p-value*	*Adjusted p-value*
BDNR - BDR	-0.269	0.788	1
BDNR - HC	2.21	0.0274*	0.0822
BDR - HC	2.35	0.0186*	0.0558

*$p<0.05$

Figure 4

Results for One-Way Kruskal-Wallis Test: Delta FAST Scores



3.5.2 CFQ

The delta CFQ data failed to meet all assumptions because the BDNR group's scores were not normally distributed. Therefore, a non-parametric Kruskal-Wallis test was performed and the significant effect of group was then decomposed using the Dunn test. The omnibus test revealed a significant effect of group (p=0.0031) on delta CFQ scores with a moderate effect size (η^2= 0.091) (Table 12). A post-hoc Dunn test was conducted to look at pairwise differences between groups. The Dunn test revealed a significant difference in delta CFQ between the HC and BDNR groups only (p=0.00724) (Table 13; Figure 5).

Table 12

One-Way Kruskal-Wallis Results: CFQ

Dependent variable	*n*	*Statistic*	*df*	*p-value*	η^2
Delta CFQ	108	11.6	2	0.0031*	0.091

*p<0.05; *df,* degress of freedom; η^2, eta-squared

Table 13

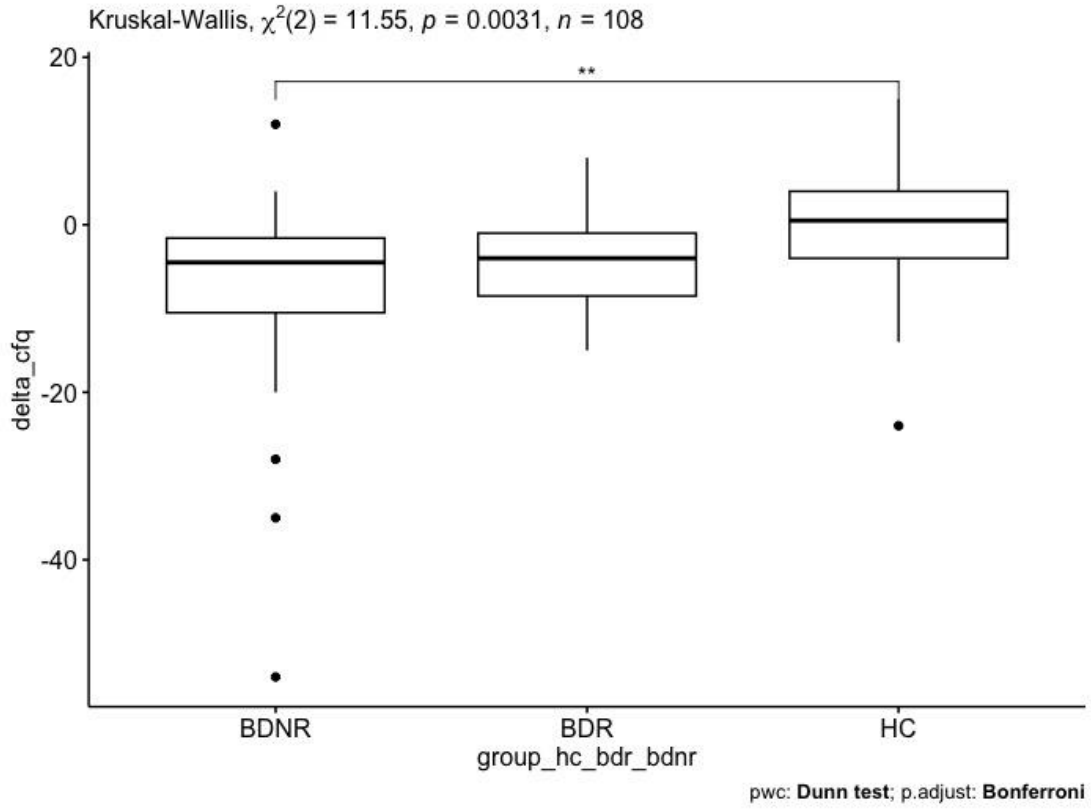
Pairwise Comparisons Using the Dunn Test: CFQ

Pairwise comparison	*Statistic*	*p-value*	*Adjusted p-value*
BDNR - BDR	0.392	0.695	1
BDNR - HC	3.03	0.00241*	0.00724*
BDR - HC	2.31	0.0211*	0.0632

*p=<0.05

Figure 5

Results for One-Way Kruskal-Wallis Test: Delta FAST Scores



3.6 Growth Curve Modeling

Three models were compared for goodness of fit as measured using changes in the LL using the chi-squared test (χ^2). The model comparisons for the FAST models and CFQ models are listed in the tables below (Table 14; Table 16). The assumptions of the linear models were checked and included in the supplementary material (Figure 8; Figure 9). Some assumptions were violated, but this is discussed later in the discussion section. The parameter estimates are listed in Table 15 and Table 17.

3.6.1 FAST

The second model was the most appropriate fit to the FAST data (Table 13). The effect of group on the intercept improved model fit ($\chi^2(2) = 62.7$, p=<0.0001), but the effect of group on the linear term did not improve model fit ($\chi^2(2) = 4.5$, p=0.105) (Table 14).

Table 14

Model Comparisons for FAST Data

Model	*AIC*	*BIC*	*logLik*	*Deviance*	*Test statistic*	*df*	*p-value*
Model 1	2104.3	2126.3	-1046.2	2092.3			
Model 2	2045.7	2075	-1014.8	2029.7	χ^2= 62.672	2	<0.0001*
Model 3	2045.2	2081.8	-1012.6	2025.2	χ^2=4.5067	2	0.105

**p*=<0.05

The second model summary is demonstrated in Table 15, with the HC group acting as the reference group. The model is plotted in Figure 6. This model suggests that there were some differences between the group intercepts but not between the group slopes. There were significant differences between the intercepts of the HC and BDNR groups (*p*=<0.001) and the HC and BDR groups (*p*=0.033) (Table 15). The model was then run with the BDR group as the reference group to test differences between the BDNR and BDR group intercepts, but no significant difference was found (*p*=0.08).

Table 15

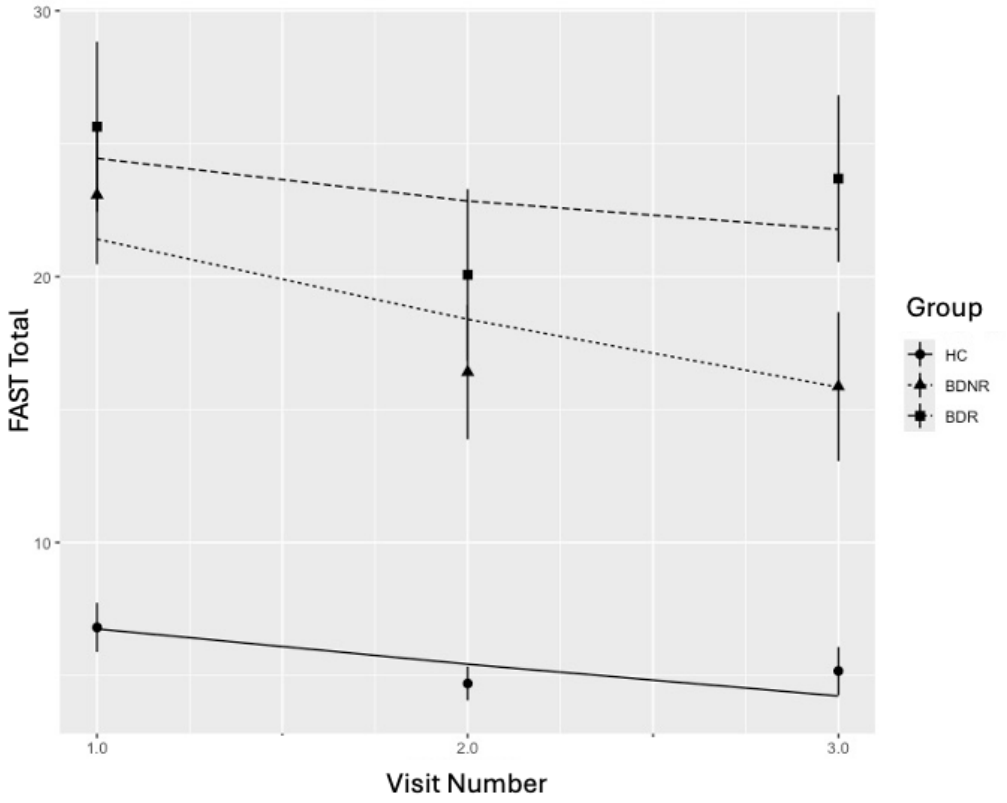
Parameter Estimates for Model 2: FAST

Fixed effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	19.061	1.4173	112.5011	13.449	<0.001*
Visit number	-1.7129	0.5194	96.5051	-3.298	0.00137*
Group BDNR	-10.1519	1.1014	106.0613	-9.217	<0.0001*
Group BDR	2.8564	1.3200	106.0357	2.164	0.03272*

*p=<0.05. Note: The intercept and visit number term refer to the HC group. Significant *p-values* for these terms suggest that the HC group intercept is significantly different from zero and that the effect of visit number is significantly different from zero.

Figure 6

Linear Model of Best Fit for FAST Data



3.6.2 CFQ

The third model was the most appropriate fit to the CFQ data (Table 16). The effect of group on the intercept improved model fit ($\chi^2(2) = 42.59$, p=<0.0001), and the effect of group on the linear term further improved the model ($\chi^2(2) = 7$, p=0.03) (Table 16).

Table 16

Model Comparisons for CFQ Data

Model	*AIC*	*BIC*	*logLik*	*Deviance*	*Test statistic*	*p-value*
Model 1	2183.5	2205.5	-1085.8	2171.5		
Model 2	2144.9	2174.2	-1064.5	2128.9	χ^2=42.5949	<0.0001*
Model 3	2141.9	2178.6	-1061	2121.9	χ^2=7	0.0302*

**p*=<0.05

The third model summary is demonstrated in Table 17, with the HC group acting as the reference group. This model is plotted in Figure 7. This model suggests that there were some differences between the group intercepts and slopes. There were significant differences between the intercepts of the HC and BDNR groups (*p*=<0.001) and the HC and BDR groups (*p*=<0.001). There was also a significant difference between the slope of the BDNR group compared to the HC group (*p*=0.009), but not between the BDR and HC group (Table 17). The model was then run with the BDR group as the reference group to test differences between the BDNR and BDR group intercepts and slopes, but no significant differences were found (*p*=0.74; *p*=0.242). Further, the HC and BDR groups' slopes were not significant (*p*=0.63; *p*=0.119), suggesting no change over the timepoints but the BDNR group's slope was significant (*p*= 0.0007).

Table 17

Parameter Estimates for Model 3: CFQ

Fixed effects	*Estimate*	*Std. Error*	*df*	*t-value*	*p-value*
Intercept	26.9092	2.0674	106.2642	13.016	<0.001*
Visit number	-0.3848	0.8079	107.5800	0.476	0.63484
Group BDNR	19.3417	3.7701	105.1937	5.130	<0.0001*
Group BDR	20.9106	4.0709	104.9877	5.137	<0.0001*

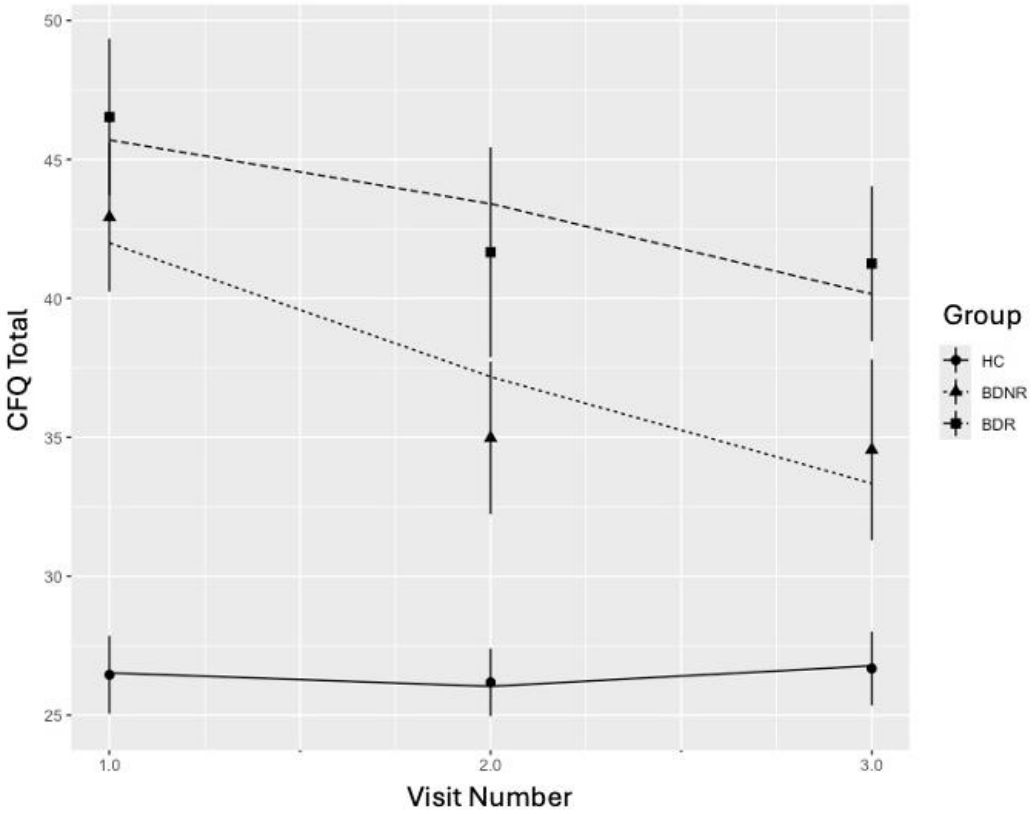
Visit*Group BDNR	-3.8735	1.4563	104.4490	-2.660	0.00905*
Visit*Group BDR	-1.7385	1.5744	104.6107	-1.104	0.27203

Note: The intercept and visit number term refer to the HC group. Significant *p-values* for these terms suggest that the HC group intercept is significantly. different from zero. The interaction between the specific factors "Group" and "Visit" is shown as "Group*Visit" with the relevant group listed. Interactions represented slope of specified group compared to the HC group.

*p=<0.05

Figure 7

Linear Model of Best Fit for the CFQ Data



4. Discussion

The present thesis investigated the trajectories of functioning of healthy controls, individuals with BD-I who experienced episode relapse and those who did not, as measured by the FAST and CFQ. Using the FAST total score offered an opportunity to measure global psychosocial functioning while the CFQ allowed us to measure subjective cognitive functioning. Due to the longitudinal nature of this study, the fact that BD subjects were evaluated during periods of euthymia, and that relapse status was the grouping variable for the BD subjects, we were able to investigate potential evidence of progression in functional impairment in BD. In addition, this study involved individuals with BD-I with no comorbidities besides anxiety, which helps with the confidence in our findings about this population.

Before discussing the results of the main analyses, it is worth noting some important results regarding the characteristics of the sample. Past research has found that an estimated 50% of individuals with BD experience a mood episode after a period of remission (Perlis et al., 2006). Our study numbers were close to this approximation as 45% of participants in the study experienced relapse between periods of stability during the study. In addition, along with no BD participants changing medication status throughout the study, the BDR and BDNR groups were matched on medication status across visits. No other characteristics of the BD groups are of note as they were matched on all demographic variables at baseline.

Overall, some significant differences were found between the BD groups and HC group for the FAST and CFQ. No significant differences were found between the BDR and BDNR groups. A significant change over time was found in the BDNR group only for the CFQ. All of these results will be discussed below.

4.1 Between Group Differences in Scale Scores

The mixed-effects ANOVA first explored whether the 3 groups (HC, BDR, BDNR) differed significantly at B and F visits in their FAST and CFQ scores. Significant differences were found between the HC and both BD groups at B and F visits for both the FAST and the CFQ. At both visits, the HC group had lower overall scores than the BD groups. To further understand these between-group differences and to maximize data and timepoints used, the simple linear growth curve analysis provided insight into differences between the group intercepts of a linear model. Similarly to the mixed-effects analysis, both the BDR and BDNR groups' intercepts differed from the HC group for both the FAST and CFQ. These findings support the established functional deficit in BD as compared to HC and further suggests that this deficit is sustained over time.

Cross-sectional research in BD-I has identified established differences in psychosocial functioning as measured by the FAST between BD and HC in euthymia (Fagiolini et al., 2005; Léda-Rêgo et al., 2020; Rosa et al., 2010). In Lopèz- Villarrel et al.'s (2022) longitudinal study using the FAST, only the BD-I group who demonstrated progressive functional impairment showed a psychosocial deficit compared to HC whereas both our BD groups demonstrated worse functioning. A potential explanation for this is the group that demonstrated progressive impairment also had greater residual depressive symptoms, and the measurement window of this other study was longer, so perhaps more change could be captured with a longer time period. Although a different scale was used, differences in subjective cognitive functioning between BD and HC have also been established cross-sectionally (Luo et al., 2020), which would also support our findings. To our knowledge, there is limited longitudinal research comparing functioning of BD and HC using the FAST and CFQ specifically. These results suggest that the difference in impairment between HC and BD is likely sustained over time.

No significant differences were found between the BDR and BDNR group at B and F visits for either the FAST or the CFQ nor did the groups' intercepts differ from each other in the growth analysis. These findings suggest that relapse was not associated with worse impairment in euthymia compared to the group who did not experience relapse. This is contradictory to Lopèz-Villarrel et al.'s (2022) findings wherein the 49.4% of individuals with BD-I characterized by a worsening in functional impairment had a higher number of episode relapses and hospitalizations during the follow up period. But again, this finding could be due to this group also having greater residual depressive symptoms as well as being followed for 5 years (Lopèz-Villarrel et al.'s 2022), as our measurement window was less than half as long. Study subjects in the current research were monitored using the MADRS for current depressive symptomology. If subjects did not endorse a current depressive episode with the criteria from the SCID-5 but had a moderate to high score on the MADRS, this would have been flagged as a participant potentially not being in a euthymic state. The concept of residual and subsyndromal symptoms will be further discussed in the within-group and general discussion section, but it is worth noting here as residual depressive symptoms were associated with progressive impairment in past research and therefore may have been an influencing factor in their findings.

4.2 Between Group Differences in Change in Scores

The between-group differences of the mixed-effects analysis and the differences in intercepts of the growth analysis help us to see the group differences in scores at certain timepoints, but not the group difference in change in scores across timepoints. To further explore the difference in change between groups, we can look to the results from the one-way Kruskal Wallis test and the differences in slopes between groups in the growth analysis. These analyses both revealed that only the CFQ was implicated in significant change between visits. Further, the

only difference was between the HC and BDNR group. Lopèz-Villarrel et al. (2022) established that the trajectory of psychosocial functioning, as measured by the FAST, was stable in the BD group as a whole compared to the control group (HC). This finding is consistent with ours since we did not detect any significant difference in the change in FAST scores in the BD groups compared to the HC group.

The Kruskal-Wallis test demonstrated a significant difference in the delta CFQ score between the HC and BDNR group, with the delta scores of the BDNR group being lower, suggesting more negative change (i.e. improvement). The growth analysis demonstrates that the slopes only differ between the HC and BDNR group for the CFQ. Again, we see no significant difference in change between the BDR and BDNR groups, neither in delta scores nor in slopes over all timepoints for the FAST or CFQ. This suggests that there was no significant difference in the trajectories of scores over time between these groups. The difference between the HC and BDNR in the CFQ group suggests that there is evidence of change in the BDNR group but to determine if this was relevant, it is necessary to look at within-group change.

4.3 Within Group Differences in Change in Scores

The within group component of the analyses reveals which group had meaningful change across timepoints. The mixed-effects analysis explored whether the within groups' total scores differed significantly between the B and F visits. Significant differences were found between the B and F visits for the BDNR and BDR groups but not the HC group for both the FAST and CFQ data. These differences also demonstrated that the scores decreased between the B and F visits, suggesting that the BDR and BDNR groups' scores improved significantly.

These findings were somewhat in conflict with the results from the growth analysis, which revealed that only the BDNR group demonstrated a significant change in slope over time

and only for the CFQ. This slightly contradictory finding might be due to the mixed-effects analysis looking at primarily 2 timepoints (and mostly using visit 2 as a follow-up) whereas the growth analysis was using all 3 available timepoints for subjects. There seem to have been a significant decrease in both scores for the BDR group between B and F visits but this did not hold for this group across 3 timepoints, nor did it for the BDNR group and their FAST scores. The growth analysis suggests that only the BDNR had meaningful change across the 3 timepoints with an overall decrease in CFQ scores, suggesting that there was a significant improvement in scores for this group. There is no robust longitudinal data on subjective cognitive functioning as measured by the CFQ, but some related literature might help to clarify this finding. There is correlational evidence of subjective cognitive complaints being related to depressive symptomology, (Demant et al., 2015; van der Werf-Eldering et al., 2011). Perhaps it is the lack of depressive symptoms during the follow-up period that influenced an improvement in the BDNR group.

Relatedly and as mentioned earlier, it is possible that these findings relate to residual and subsyndromal symptoms. Much research supports that residual or subsyndromal symptoms can persist after mood episodes (Joffe et al., 2004; Judd et al., 2008; MacQueen et al., 2003), some even suggesting that a substantial portion of time is spent with subsyndromal symptoms, even more so in BD-I (Joffe et al., 2004). Individuals with subsyndromal symptoms are also associated with more comorbid disorders and worse global functioning compared to euthymic patients (MacQueen et al., 2003). In the current research, the YMRS and MADRS were used to ensure current euthymia. Any extreme scores on these measures would have indicated that perhaps the subject was not in a euthymic state and the visit would therefore not have been conducted. In addition, participants had no comorbid mental disorders besides anxiety. The

findings from the mixed-effects analysis suggests that both the BDNR and BDR groups improved between B and F visits, and the BDNR improved on the CFQ over all visits, so perhaps since participants seemed to be in a confirmed euthymic state, the absence of residual symptoms influenced their scores positively. We cannot be certain of this since we did not include the YMRS and MADRS as part of our analyses, but participants did not present with high scores on these at study visits as part of the study protocol for ensuring euthymia.

4.4 General Discussion

In sum, we found significant differences between the HC and BD groups for both FAST and CFQ scores at B and F, suggesting impairment in psychosocial functioning and subjective cognitive functioning in BD as compared to HC, that is sustained over time. The other main significant finding of interest is the BDNR group's CFQ scores decreasing over time, suggesting an improvement in subjective cognitive functioning. The last important finding is the lack of established difference between BDNR and BDR groups at B and F visits or differences in trajectories, which suggests that our current research does not support that episode recurrence was related to worse psychosocial and subjective cognitive impairment in euthymia.

There are several potential reasons why we did not find a significant difference between the BDNR and BDR groups. First, it is possible that the BD participants that dropped out had more recurrences than the final sample. Many participants were lost to follow up and we therefore do not know why they dropped out of the study. It is possible that those who dropped out were more severely and frequently ill and therefore too impaired to continue to take part in the study. For example, the lack of insight associated with mania may lead to higher levels of dropout for BD-I patients with more severe manic courses (Judd et al., 2003). If this was the case and individuals with more severe courses had stayed in the study, their trajectories may have

influenced the BDR group's outcomes. In addition, it is difficult to gage the illness severity of the dropouts compared to the final sample as we have their data for lifetime episodes but this measure is not the most reliable. There is little concordance with recall of episodes and clinician-rated mood episodes, and the SCID-5 only asks to report lifetime depressive episodes as opposed to (hypo)manic episodes (Tremain et al., 2020). Although we asked about lifetime episodes of all types during the study, lifetime reporting may have been inaccurate, especially if an individual has had BD for a long period of time.

This also brings up the second issue of potential inaccurate reporting of relapse during the study period. We are more confident in our relapse data throughout the study, since reporting relapse within a shorter period is easier than reporting lifetime episodes, but there is still a chance that episodes were not reported properly. We were not able to verify relapses beyond using self-report and the SCID-5 past mood episodes modules using history within the past year as opposed to lifetime history. Since very few participants filled out their NIMH-LCM logs, we cannot be certain about the numbers reported. Therefore, grouping participants by BDR and BDNR was logical. So, even if the BDR group was over-reporting number of relapses, we can be somewhat confident that a relapse still occurred if not about the specific number of relapses, but we have no other ways to confirm these data more carefully.

Third, we studied a group of BD-I participants with no comorbidities besides anxiety disorders. Cross-sectional studies exploring functional outcomes in euthymia claim impairment in periods of remission but do not exclude for comorbidities (Fagiolini et al., 2005; Rosa et al., 2009; Rosa et al., 2010; Rosa et al., 2012). Further, it does not seem as though many studies take comorbidity into account in their analyses, which could influence findings. Although BD is a highly comorbid disorder and these studies may be representative of the BD population as a

whole, comorbidities such as AUD or SUD could influence functional outcomes, so it is difficult to claim that impairment is primarily due to BD-I group membership. Since only anxiety disorders were accepted as comorbidities in our sample, there is more confidence that we are exploring outcomes related to BD-I in isolation and perhaps past results were influenced by comorbid conditions.

Fourth, there could be many other variables that affect psychosocial and subjective cognitive functioning, both positively and negatively, beyond relapse group status. For example, there could be influences like unemployment, stigma, and lack of social support affecting functional outcomes (Martino et al., 2016). On the other hand, if participants were having positive experiences such as social support or consistent employment, it is foreseeable that this could neutralize or improve functional outcomes.

4.5 Limitations

Although this research offers some compelling evidence, there are several limitations to address. First, as with many studies of this nature, the sample size was limiting. Due to a generally low sample size in addition to dropouts from both the HC and BD groups between visits, many potential factors of interest were not included in the analyses. Demographic variables such as sex and age as well as variables related to relapse such as type of relapse could not be included due not having enough statistical power. There is some evidence that specific types of episodes may have differential effects on functional outcomes, but we were not able to explore this. Additionally, there was not enough power to explore the sub-sections of the scales to see if certain areas of functioning were more implicated than others. This could have been helpful both to look at specific areas that differed between the HC and BD groups as well as within the BDNR group to see if certain sub-sections improved more than others.

Second, the way that relapse-status grouping was conceptualized in this study may have been limiting to the growth analyses. Relapse status was quantified as a grouping variable across the whole study. If instead relapse was modeled as a time-varying covariate wherein relapse status could change at each timepoint, we could have modeled these data in a slightly more nuanced way. For example, if a participant relapsed only between baseline and their second visit but not between their second and third visit, we could model this varying covariate across time to see how it may have affected scores. It is important to note as well that this would be more realistic with a larger study sample and a larger study window with more measurement instances. This would allow for models that could account for more complexity and heterogeneity in the relapse trajectories over time.

Third and related to the second limitation, the growth results should be interpreted with caution. These results were based on the linear model of best fit to the data. The two models for the FAST and CFQ data were checked based on assumptions of linear models (Figure 8; Figure 9) and some violations were identified. Specifically, both linear models seem to violate the assumptions of homogeneity of variance and normality of the residuals as well as influential observations. These violations are similar to those of the other analyses, but more difficult to deal with in the modeling of our data as they may suggest that there is another factor of importance that the model is not accounting for. Alternatives include non-linear modeling or generalized linear modeling (GLMM), but when attempting these, it seemed as though there were not enough predictors in our model to properly run them. Since one of our findings is based on this modeling and suggests the BDNR group's CFQ scores improved over time, it should likely be interpreted with caution.

Fourth and last, there is some question about the validity of self-report in cognitive functioning. As mentioned previously, there is evidence that subjective cognitive impairment does not or weakly correlates with objective measures (Demant et al., 2015; van der Werf-Eldering et al., 2011), even in healthy populations (Carrigan & Barkus, 2016). Although this does not mean that self-reports cannot be useful, a concern arises about if participants may be over or under-stating their complaints. Burdick et al. (2005) argue that BD subjects may actually under-report cognitive impairments, which may be indicative of scales not capturing functioning appropriately. This is important to note since the BDNR group's CFQ scores improved over time and perhaps this group was under-reporting over time. It would be necessary to explore the CFQ scores in conjunction with cognitive measures to clarify this. In the meantime, this might further suggest that we approach this finding with caution.

4.6 Future Directions

These findings and limitations offer opportunities for future research. Overall, replicating this research will require a longer window and a larger sample size to further validate these findings. A longer window with closer and more measurement periods may allow for relapse to be captured more reliably and consistently over time. A larger sample size would allow analyses to include potential factors of interest such as type of relapse in addition to looking at sub-scale sections of the measures. Next, some other and potentially more objective variables of interest that may affect functional outcomes could be included to see if there is a correlation. For example, participants' employment status, life stressors, and social supports could be documented as they may have an effect on scale scores. These could be documented within the same period as the scales by asking about these factors occurring with the last 15 days and over the past 6 months.

Next, it would be interesting to compare the CFQ results with the objective cognitive measures completed in this study. There is evidence of neurocognitive functioning improving after one year for first-treatment BD-I subjects who did not experience episode relapse compared to those who did and healthy controls (Demmo et al., 2017). These participants also demonstrated better global and occupational functioning (Demmo et al., 2017). Although this study sample was slightly different and we did not find improved global functioning in the BDNR group or worse functioning in the BDR group, we did find improved subjective cognitive functioning in the BDNR group. Perhaps this improvement in subjective functioning is related to cognition in our sample. In addition, with the lack of correlation established between subjective and objective cognitive function, this could be further elucidated using other study data. Last, with a longer window, sample size, and more frequent measurements, it might be easier to model the data using relapse as a time-varying covariate in growth modeling. This could offer an opportunity for a further nuanced exploration of how relapse status affects functional outcomes in euthymia in BD-I.

5. Conclusion

Overall, this thesis aimed to clarify the relationship between episode relapse in euthymic BD-I and functional impairment. Despite several limitations, this study has strengths including the longitudinal study design, the inclusion of episode relapse, the control over comorbidities and residual symptoms, and measuring individuals with BD-I in euthymia. We found a sustained difference between both BD subjects who experienced relapse and those who did not compared to the HC group over time for both psychosocial and subjective cognitive functioning. We did not find a difference in functional impairment between the BDR and BDNR in either domain, neither in differences at timepoints nor in change in across timepoints. But, there is some

evidence of improvement in scale scores over 2 timepoints, which may have been due to our sample not having any evidence of comorbid mental disorders or residual symptoms. We also found evidence of an improvement in subjective cognitive functioning in the BDNR group across all 3 timepoints. Since we are not aware of many longitudinal studies looking at subjective cognitive functioning as measured by the CFQ, this offers some evidence that not experiencing episode relapse might be implicated in some type of buffering effect, but further studies are needed to clarify this. This analysis was meant to offer an initial investigation and potential foundation on which to build further exploration of psychosocial and subjective cognitive impairment in BD-I. Although we could not include more factors into our models, these findings offer additional information about how functional impairment might present during euthymia.

References

Abé, C., Ching, C. R. K., Liberg, B., Lebedev, A. V., Agartz, I., Akudjedu, T. N., Alda, M., Alnæs, D., Alonso-Lana, S., Benedetti, F., Berk, M., Bøen, E., Bonnin, C. D. M., Breuer, F., Brosch, K., Brouwer, R. M., Canales-Rodríguez, E. J., Cannon, D. M., Chye, Y., Dahl, A., … ENIGMA Bipolar Disorder Working Group (2022). Longitudinal Structural Brain Changes in Bipolar Disorder: A Multicenter Neuroimaging Study of 1232 Individuals by the ENIGMA Bipolar Disorder Working Group. *Biological psychiatry*, *91*(6), 582–592. https://doi.org/10.1016/j.biopsych.2021.09.008

Altamura, A. C., Dell'Osso, B., Berlin, H. A., Buoli, M., Bassetti, R., & Mundo, E. (2010). Duration of untreated illness and suicide in bipolar disorder: a naturalistic study. *European Archives of Psychiatry and Clinical Neuroscience*, *260*(5), 385–391. https://doi.org/10.1007/s00406-009-0085-2

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.)

Azorin, J. M., Belzeaux, R., Kaladjian, A., Adida, M., Hantouche, E., Lancrenon, S., & Fakra, E. (2013). Risks associated with gender differences in bipolar I disorder. *Journal of affective disorders*, *151*(3), 1033–1040. https://doi.org/10.1016/j.jad.2013.08.031

Baldessarini, R. J., Tondo, L., & Vázquez, G. H. (2019). Pharmacological treatment of adult bipolar disorder. *Molecular Psychiatry*, *24*(2), 198–217. https://doi.org/10.1038/s41380-018-0044-2

Berk, M., Hallam, K. T., & McGorry, P. D. (2007). The potential utility of a staging model as a course specifier: a bipolar disorder perspective. *Journal of affective disorders*, *100*(1-3), 279–281. https://doi.org/10.1016/j.jad.2007.03.007

Berk, M., Post, R., Ratheesh, A., Gliddon, E., Singh, A., Vieta, E., Carvalho, A. F., Ashton, M. M., Berk, L., Cotton, S. M., McGorry, P. D., Fernandes, B. S., Yatham, L. N., & Dodd, S. (2017). Staging in bipolar disorder: from theoretical framework to clinical utility. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *16*(3), 236–244. https://doi.org/10.1002/wps.20441

Bolton, S., Warner, J., Harriss, E., Geddes, J., & Saunders, K. E. A. (2021). Bipolar disorder: Trimodal age-at-onset distribution. *Bipolar disorders*, *23*(4), 341–356. https://doi.org/10.1111/bdi.13016

Bonnín, C. M., Martínez-Arán, A., Reinares, M., Valentí, M., Solé, B., Jiménez, E., Montejo, L., Vieta, E., & Rosa, A. R. (2018). Thresholds for severity, remission and recovery using the functioning assessment short test (FAST) in bipolar disorder. *Journal of affective disorders*, *240*, 57–62. https://doi.org/10.1016/j.jad.2018.07.045

Brissos S, Dias VV, Kapczinski F. Cognitive Performance and Quality of Life in Bipolar Disorder. *The Canadian Journal of Psychiatry*. 2008;53(8):517-524. https://doi.org/10.1177/070674370805300806

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *The British Journal of Clinical Psychology*, *21*(1), 1–16. https://doi.org/10.1111/j.2044-8260.1982.tb01421.x

Cardoso, T., Bauer, I. E., Meyer, T. D., Kapczinski, F., & Soares, J. C. (2015). Neuroprogression and Cognitive Functioning in Bipolar Disorder: A Systematic Review. *Current psychiatry reports*, *17*(9), 75. https://doi.org/10.1007/s11920-015-0605-x

Carrigan, N., & Barkus, E. (2016). A systematic review of cognitive failures in daily life: Healthy populations. *Neuroscience and biobehavioral reviews*, *63*, 29–42.

https://doi.org/10.1016/j.neubiorev.2016.01.010

Clemente, A. S., Diniz, B. S., Nicolato, R., Kapczinski, F. P., Soares, J. C., Firmo, J. O., & Castro-Costa, É. (2015). Bipolar disorder prevalence: a systematic review and meta-analysis of the literature. *Revista brasileira de psiquiatria (Sao Paulo, Brazil : 1999)*, *37*(2), 155–161. https://doi.org/10.1590/1516-4446-2012-1693

Cloutier, M., Greene, M., Guerin, A., Touya, M., & Wu, E. (2018). The economic burden of bipolar I disorder in the United States in 2015. *Journal of affective disorders*, *226*, 45–51. https://doi.org/10.1016/j.jad.2017.09.011

de la Fuente-Tomás, L., Sierra, P., Sanchez-Autet, M., Arranz, B., García-Blanco, A., Safont, G., & García-Portilla, M. P. (2020). A clinical staging model for bipolar disorder: longitudinal approach. *Translational psychiatry*, *10*(1), 45. https://doi.org/10.1038/s41398-020-0718-9

Demant, K. M., Vinberg, M., Kessing, L. V., & Miskowiak, K. W. (2015). Assessment of subjective and objective cognitive function in bipolar disorder: Correlations, predictors and the relation to psychosocial function. *Psychiatry research*, *229*(1-2), 565–571. https://doi.org/10.1016/j.psychres.2015.05.022

Demmo, C., Lagerberg, T. V., Kvitland, L. R., Aminoff, S. R., Hellvin, T., Simonsen, C., Haatveit, B., Andreassen, O. A., Melle, I., & Ueland, T. (2018). Neurocognitive functioning, clinical course and functional outcome in first-treatment bipolar I disorder patients with and without clinical relapse: A 1-year follow-up study. *Bipolar disorders*, *20*(3), 228–237. https://doi.org/10.1111/bdi.12569

Denicoff, K. D., Leverich, G. S., Nolen, W. A., Rush, A. J., McElroy, S. L., Keck, P. E., Suppes,

T., Altshuler, L. L., Kupka, R., Frye, M. A., Hatef, J., Brotman, M. A., & Post, R. M. (2000). Validation of the prospective NIMH-Life-Chart Method (NIMH-LCM-p) for longitudinal assessment of bipolar illness. *Psychological Medicine*, *30*(6), 1391–1397. https://doi.org/10.1017/s0033291799002810

Depp, C. A., Mausbach, B. T., Harmell, A. L., Savla, G. N., Bowie, C. R., Harvey, P. D., & Patterson, T. L. (2012). Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder. *Bipolar disorders*, *14*(3), 217–226. https://doi.org/10.1111/j.1399-5618.2012.01011.x

Diflorio, A., & Jones, I. (2010). Is sex important? Gender differences in bipolar disorder. *International review of psychiatry (Abingdon, England)*, *22*(5), 437–452. https://doi.org/10.3109/09540261.2010.514601

Elshahawi, H. H., Essawi, H., Rabie, M. A., Mansour, M., Beshry, Z. A., & Mansour, A. N. (2011). Cognitive functions among euthymic bipolar I patients after a single manic episode versus recurrent episodes. *Journal of affective disorders*, *130*(1-2), 180–191. https://doi.org/10.1016/j.jad.2010.10.027

Etain, B., Bellivier, F., Olié, E., Aouizerate, B., Aubin, V., Belzeaux, R., Courtet, P., Dubertret, C., Schwan, R., Roux, P., Polosan, M., Leboyer, M., Godin, O., FondaMental Advanced Center of Expertise for Bipolar Disorders (FACE-BD), & FACE-BD Clinical Sites and Principal Collaborators in France (2021). Clinical predictors of recurrences in bipolar disorders type 1 and 2: A FACE-BD longitudinal study. *Journal of psychiatric research*, *134*, 129–137. https://doi.org/10.1016/j.jpsychires.2020.12.041

Fagiolini, A., Forgione, R., Maccari, M., Cuomo, A., Morana, B., Dell'Osso, M. C., Pellegrini, F.,

& Rossi, A. (2013). Prevalence, chronicity, burden and borders of bipolar disorder. *Journal of affective disorders*, *148*(2-3), 161–169. https://doi.org/10.1016/j.jad.2013.02.001

Fagiolini, A., Kupfer, D. J., Masalehdan, A., Scott, J. A., Houck, P. R., & Frank, E. (2005). Functional impairment in the remission phase of bipolar disorder. *Bipolar disorders*, *7*(3), 281–285. https://doi.org/10.1111/j.1399-5618.2005.00207.x

Ferrari, A. J., Stockings, E., Khoo, J. P., Erskine, H. E., Degenhardt, L., Vos, T., & Whiteford, H. A. (2016). The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar disorders*, *18*(5), 440–450. https://doi.org/10.1111/bdi.12423

Frank, E., Nimgaonkar, V. L., Phillips, M. L., & Kupfer, D. J. (2015). All the world's a (clinical) stage: rethinking bipolar disorder from a longitudinal perspective. *Molecular psychiatry*, *20*(1), 23–31. https://doi.org/10.1038/mp.2014.71

Godin, O., Leboyer, M., Mazroui, Y., Aouizerate, B., Azorin, J. M., Raoul, B., Bellivier, F., Polosan, M., Courtet, P., Dubertret, C., Henry, C., Kahn, J. P., Loftus, J., Olié, E., Passerieux, C., Costagliola, D., Etain, B., & FondaMental Advanced Centers of Expertise in Bipolar Disorders (FACE-BD). (2020). Trajectories of functioning in bipolar disorders: A longitudinal study in the FondaMental Advanced Centers of Expertise in Bipolar Disorders cohort. *The Australian and New Zealand journal of psychiatry*, *54*(10), 985–996. https://doi.org/10.1177/0004867420945796

Hirschfeld, R. M., & Vornik, L. A. (2005). Bipolar disorder--costs and comorbidity. *The American journal of managed care*, *11*(3 Suppl), S85–S90.

Joffe, R. T., MacQueen, G. M., Marriott, M., & Trevor Young, L. (2004). A prospective,

longitudinal study of percentage of time spent ill in patients with bipolar I or bipolar II disorders. *Bipolar disorders*, *6*(1), 62–66. https://doi.org/10.1046/j.1399-5618.2003.00091.x

Judd, L. L., Akiskal, H. S., Schettler, P. J., Coryell, W., Maser, J., Rice, J. A., Solomon, D. A., & Keller, M. B. (2003). The comparative clinical phenotype and long term longitudinal episode course of bipolar I and II: a clinical spectrum or distinct disorders?. *Journal of affective disorders*, *73*(1-2), 19–32. https://doi.org/10.1016/s0165-0327(02)00324-5

Judd, L. L., Schettler, P. J., Akiskal, H. S., Coryell, W., Leon, A. C., Maser, J. D., & Solomon, D. A. (2008). Residual symptom recovery from major affective episodes in bipolar disorders and rapid episode relapse/recurrence. *Archives of general psychiatry*, *65*(4), 386–394. https://doi.org/10.1001/archpsyc.65.4.386

Kaltenboeck, A., Winkler, D., & Kasper, S. (2016). Bipolar and related disorders in DSM-5 and ICD-10. *CNS spectrums*, *21*(4), 318–323. https://doi.org/10.1017/S1092852916000079

Kapczinski, F., Dias, V. V., Kauer-Sant'Anna, M., Frey, B. N., Grassi-Oliveira, R., Colom, F., & Berk, M. (2009). Clinical implications of a staging model for bipolar disorders. *Expert review of neurotherapeutics*, *9*(7), 957–966. https://doi.org/10.1586/ern.09.31

Karanti, A., Kardell, M., Joas, E., Runeson, B., Pålsson, E., & Landén, M. (2020). Characteristics of bipolar I and II disorder: A study of 8766 individuals. *Bipolar disorders*, *22*(4), 392–400. https://doi.org/10.1111/bdi.12867

Kessing, L. V., & Andersen, P. K. (2017). Evidence for clinical progression of unipolar and bipolar disorders. *Acta psychiatrica Scandinavica*, *135*(1), 51–64. https://doi.org/10.1111/acps.12667

Léda-Rêgo, G., Bezerra-Filho, S., & Miranda-Scippa, Â. (2020). Functioning in euthymic

patients with bipolar disorder: A systematic review and meta-analysis using the Functioning Assessment Short Test. *Bipolar disorders*, *22*(6), 569–581. https://doi.org/10.1111/bdi.12904

Loftus, J., Scott, J., Vorspan, F., Icick, R., Henry, C., Gard, S., Kahn, J. P., Leboyer, M., Bellivier, F., & Etain, B. (2020). Psychiatric comorbidities in bipolar disorders: An examination of the prevalence and chronology of onset according to sex and bipolar subtype. *Journal of affective disorders*, *267*, 258–263. https://doi.org/10.1016/j.jad.2020.02.035

López-Villarreal, A., Sánchez-Morla, E. M., Jiménez-López, E., Martínez-Vizcaíno, V., Aparicio, A. I., Mateo-Sotos, J., Rodriguez-Jimenez, R., Vieta, E., & Santos, J. L. (2020). Progression of the functional deficit in a group of patients with bipolar disorder: a cluster analysis based on longitudinal data. *European archives of psychiatry and clinical neuroscience*, *270*(8), 947–957. https://doi.org/10.1007/s00406-019-01050-9

Luo, X., Zhu, Y., Lu, D., Zong, K., & Lin, X. (2020). Subjective cognitive dysfunction in patients with bipolar disorder: The prevalence, related factors and effects on predicting psychosocial functioning and suicidal ideation. *Psychiatry research*, *284*, 112669. https://doi.org/10.1016/j.psychres.2019.112669

MacQueen, G. M., Marriott, M., Begin, H., Robb, J., Joffe, R. T., & Young, L. T. (2003). Subsyndromal symptoms assessed in longitudinal, prospective follow-up of a cohort of patients with bipolar disorder. *Bipolar disorders*, *5*(5), 349–355. https://doi.org/10.1034/j.1399-5618.2003.00048.x

MacQueen, G. M., Young, L. T., & Joffe, R. T. (2001). A review of psychosocial outcome in patients with bipolar disorder. *Acta psychiatrica Scandinavica*, *103*(3), 163–170. https://doi.org/10.1034/j.1600-0447.2001.00059.x

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior research methods*, *52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Martínez-Arán, A., Vieta, E., Colom, F., Torrent, C., Sánchez-Moreno, J., Reinares, M., Benabarre, A., Goikolea, J. M., Brugué, E., Daban, C., & Salamero, M. (2004). Cognitive impairment in euthymic bipolar patients: implications for clinical and functional outcome. *Bipolar disorders*, *6*(3), 224–232. https://doi.org/10.1111/j.1399-5618.2004.00111.x

Martino, D. J., Samamé, C., Marengo, E., Igoa, A., & Strejilevich, S. A. (2016). A critical overview of the clinical evidence supporting the concept of neuroprogression in bipolar disorder. *Psychiatry research*, *235*, 1–6. https://doi.org/10.1016/j.psychres.2015.12.012

Merikangas, K. R., Jin, R., He, J. P., Kessler, R. C., Lee, S., Sampson, N. A., Viana, M. C., Andrade, L. H., Hu, C., Karam, E. G., Ladea, M., Medina-Mora, M. E., Ono, Y., Posada-Villa, J., Sagar, R., Wells, J. E., & Zarkov, Z. (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of general psychiatry*, *68*(3), 241–251. https://doi.org/10.1001/archgenpsychiatry.2011.12

Messer, T., Lammers, G., Müller-Siecheneder, F., Schmidt, R. F., & Latifi, S. (2017). Substance abuse in patients with bipolar disorder: A systematic review and meta-analysis. *Psychiatry research*, *253*, 338–350. https://doi.org/10.1016/j.psychres.2017.02.067

Michalak, E. E., Yatham, L. N., Kolesar, S., & Lam, R. W. (2006). Bipolar disorder and quality of life: a patient-centered perspective. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, *15*(1), 25–37.

https://doi.org/10.1007/s11136-005-0376-7

Miller, J. N., & Black, D. W. (2020). Bipolar Disorder and Suicide: a Review. *Current psychiatry reports*, *22*(2), 6. https://doi.org/10.1007/s11920-020-1130-0

Mirman, D. (2014). *Growth Curve Analysis and visualization using R* (H. Wickham, T. Hothorn, D. T. Lang, & J. M. Chambers, Eds.). CRC press.

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry : The Journal of Mental Science*, *134*, 382–389. https://doi.org/10.1192/bjp.134.4.382

Mur, M., Portella, M. J., Martínez-Arán, A., Pifarré, J., & Vieta, E. (2008). Long-term stability of cognitive impairment in bipolar disorder: a 2-year follow-up study of lithium-treated euthymic bipolar patients. *The Journal of clinical psychiatry*, *69*(5), 712–719.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, *17*(4), 372-411. https://doi.org/10.1177/1094428114548590

Perlis, R. H., Dennehy, E. B., Miklowitz, D. J., Delbello, M. P., Ostacher, M., Calabrese, J. R., Ametrano, R. M., Wisniewski, S. R., Bowden, C. L., Thase, M. E., Nierenberg, A. A., & Sachs, G. (2009). Retrospective age at onset of bipolar disorder and outcome during two-year follow-up: results from the STEP-BD study. *Bipolar disorders*, *11*(4), 391–400. https://doi.org/10.1111/j.1399-5618.2009.00686.x

Perlis, R. H., Ostacher, M. J., Patel, J. K., Marangell, L. B., Zhang, H., Wisniewski, S. R., Ketter, T. A., Miklowitz, D. J., Otto, M. W., Gyulai, L., Reilly-Harrington, N. A., Nierenberg, A. A., Sachs, G. S., & Thase, M. E. (2006). Predictors of recurrence in bipolar disorder: primary outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *The American journal of psychiatry*, *163*(2), 217–224.

https://doi.org/10.1176/appi.ajp.163.2.217

Ratnasingham, S., Cairney, J., Manson, H., Rehm, J., Lin, E., & Kurdyak, P. (2013). The burden of mental illness and addiction in ontario. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *58*(9), 529–537. https://doi.org/10.1177/070674371305800908

Rosa, A. R., González-Ortega, I., González-Pinto, A., Echeburúa, E., Comes, M., Martínez-Àran, A., Ugarte, A., Fernández, M., & Vieta, E. (2012). One-year psychosocial functioning in patients in the early vs. late stage of bipolar disorder. *Acta Psychiatrica Scandinavica*, *125*(4), 335–341. https://doi.org/10.1111/j.1600-0447.2011.01830.x

Rosa, A. R., Reinares, M., Franco, C., Comes, M., Torrent, C., Sánchez-Moreno, J., Martínez-Arán, A., Salamero, M., Kapczinski, F., & Vieta, E. (2009). Clinical predictors of functional outcome of bipolar patients in remission. *Bipolar disorders*, *11*(4), 401–409. https://doi.org/10.1111/j.1399-5618.2009.00698.x

Rosa, A. R., Reinares, M., Michalak, E. E., Bonnin, C. M., Sole, B., Franco, C., Comes, M., Torrent, C., Kapczinski, F., & Vieta, E. (2010). Functional impairment and disability across mood states in bipolar disorder. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, *13*(8), 984–988. https://doi.org/10.1111/j.1524-4733.2010.00768.x

Rosa, A. R., Sánchez-Moreno, J., Martínez-Aran, A., Salamero, M., Torrent, C., Reinares, M., Comes, M., Colom, F., Van Riel, W., Ayuso-Mateos, J. L., Kapczinski, F., & Vieta, E. (2007). Validity and reliability of the Functioning Assessment Short Test (FAST) in bipolar disorder. *Clinical practice and epidemiology in mental health : CP & EMH*, *3*, 5. https://doi.org/10.1186/1745-0179-3-5

Samamé, C., Martino, D. J., & Strejilevich, S. A. (2014). Longitudinal course of cognitive deficits in bipolar disorder: a meta-analytic study. *Journal of affective disorders*, *164*, 130–138. https://doi.org/10.1016/j.jad.2014.04.028

Samara, M. T., Levine, S. Z., & Leucht, S. (2023). Linkage of Young Mania Rating Scale to Clinical Global Impression Scale to Enhance Utility in Clinical Practice and Research Trials. *Pharmacopsychiatry*, *56*(1), 18–24. https://doi.org/10.1055/a-1841-6672

Sanchez-Moreno, J., Martinez-Aran, A., Tabarés-Seisdedos, R., Torrent, C., Vieta, E., & Ayuso-Mateos, J. L. (2009). Functioning and disability in bipolar disorder: an extensive review. *Psychotherapy and psychosomatics*, *78*(5), 285–297. https://doi.org/10.1159/000228249

Schaffer, A., Cairney, J., Cheung, A., Veldhuizen, S., & Levitt, A. (2006). Community survey of bipolar disorder in Canada: lifetime prevalence and illness characteristics. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *51*(1), 9–16. https://doi.org/10.1177/070674370605100104

Snaith, R. P., Harrop, F. M., Newby, D. A., & Teale, C. (1986). Grade Scores of the Montgomery—Åsberg Depression and the Clinical Anxiety Scales. *British Journal of Psychiatry*, *148*(5), 599–601. https://doi.org/10.1192/bjp.148.5.599

Sparding, T., Joas, E., Clements, C., Sellgren, C. M., Pålsson, E., & Landén, M. (2021). Long-term trajectory of cognitive performance in people with bipolar disorder and controls: 6-year longitudinal study. *BJPsych open*, *7*(4), e115. https://doi.org/10.1192/bjo.2021.66

Strejilevich, S. A., Samamé, C., & Martino, D. J. (2015). The trajectory of neuropsychological dysfunctions in bipolar disorders: a critical examination of a hypothesis. *Journal of affective disorders*, *175*, 396–402. https://doi.org/10.1016/j.jad.2015.01.018

Substance Abuse and Mental Health Services Administration. (2016). *Impact of the DSM-IV to DSM-5 Changes on the National Survey on Drug Use and Health*. Substance Abuse and Mental Health Services Administration (US).

Tondo, L., Miola, A., Pinna, M., Contu, M., & Baldessarini, R. J. (2022). Differences between bipolar disorder types 1 and 2 support the DSM two-syndrome concept. *International journal of bipolar disorders*, *10*(1), 2. https://doi.org/10.1186/s40345-022-00268-2

Tremain, H., Fletcher, K., & Murray, G. (2020). Number of episodes in bipolar disorder: The case for more thoughtful conceptualization and measurement. *Bipolar disorders*, *22*(3), 231–244. https://doi.org/10.1111/bdi.12872

van der Markt, A., Klumpers, U. M., Draisma, S., Dols, A., Nolen, W. A., Post, R. M., Altshuler, L. L., Frye, M. A., Grunze, H., Keck, P. E., Jr, McElroy, S. L., Suppes, T., Beekman, A. T., & Kupka, R. W. (2019). Testing a clinical staging model for bipolar disorder using longitudinal life chart data. *Bipolar disorders*, *21*(3), 228–234. https://doi.org/10.1111/bdi.12727

van der Werf-Eldering, M. J., Burger, H., Jabben, N., Holthausen, E. A., Aleman, A., & Nolen, W. A. (2011). Is the lack of association between cognitive complaints and objective cognitive functioning in patients with bipolar disorder moderated by depressive symptoms?. *Journal of affective disorders*, *130*(1-2), 306–311. https://doi.org/10.1016/j.jad.2010.10.005

Van Rheenen, T. E., Lewandowski, K. E., Bauer, I. E., Kapczinski, F., Miskowiak, K., Burdick, K. E., & Balanzá-Martínez, V. (2020). Current understandings of the trajectory and emerging correlates of cognitive impairment in bipolar disorder: An overview of evidence. *Bipolar disorders*, *22*(1), 13–27. https://doi.org/10.1111/bdi.12821

Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for mania: Reliability, validity and sensitivity. *The British Journal of Psychiatry : The Journal of Mental Science*, *133*, 429–435. https://doi.org/10.1192/bjp.133.5.429

Supplementary Material

Figure 8

Checking Growth Curve Analysis Model: FAST Data

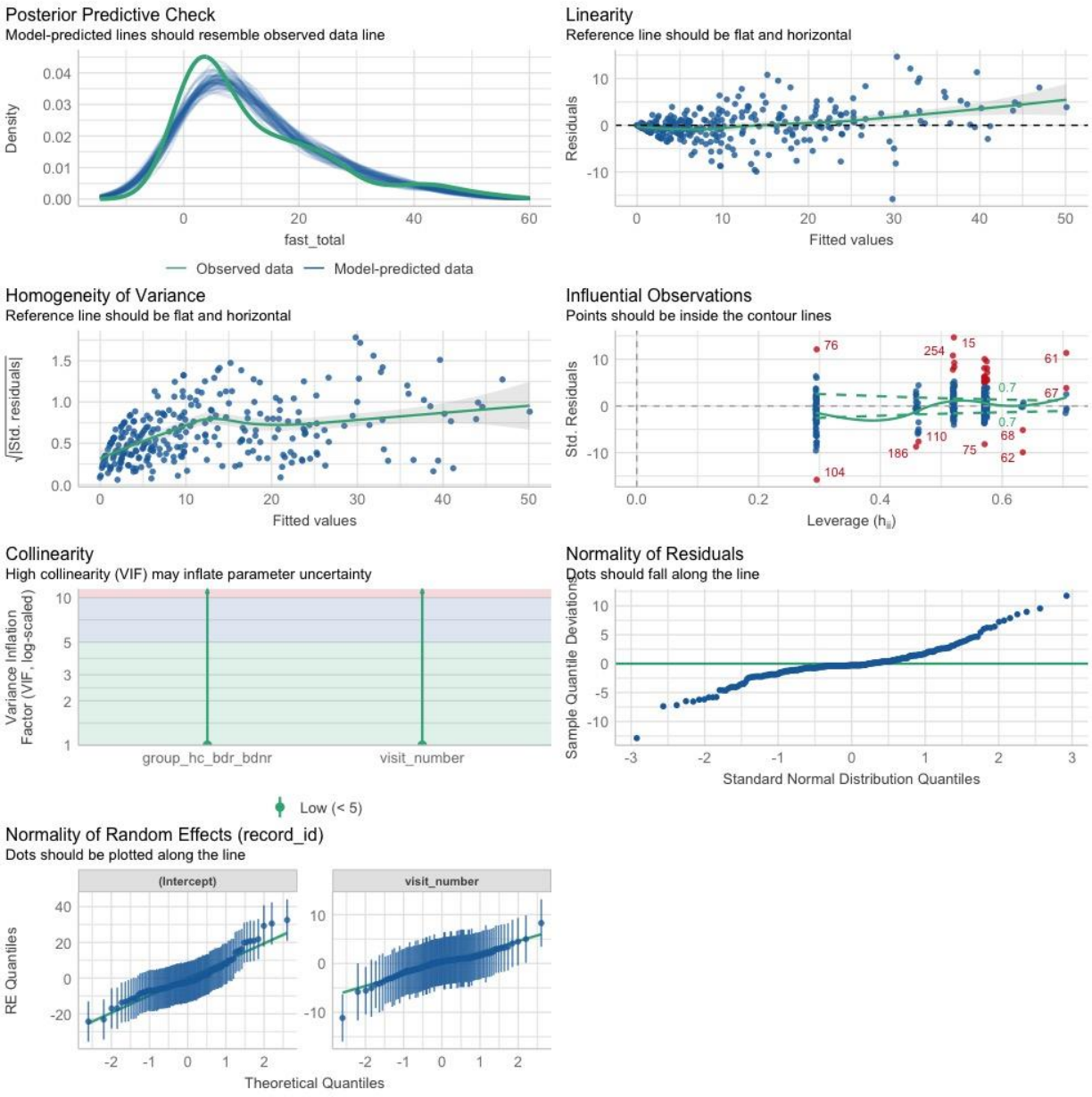


Figure 9

Checking Growth Curve Analysis Model: CFQ Data