



The
University
Of
Sheffield.

Understanding the Development and Structure of Bipolar Disorders

Adele Gnadte

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Clinical Psychology

The University of Sheffield

Faculty of Science

Clinical Psychology Unit, Department of Psychology

Submission Date: May 2021

The results, discussions and conclusions presented herein are identical to those in the printed version. This electronic version of the thesis has been edited solely to ensure conformance with copyright legislation and all excisions are noted in the text. The final, awarded and examined version is available for consultation via the University Library.

Declaration

I declare that this work has not been submitted for any other degree at the
University of Sheffield or any other institution

Structure and Word Counts

Lay Summary: 500

Section I: Literature Review

Excluding references and tables: 7,784

Including references and tables: 13,195

Section II: Empirical Study

Excluding references and tables: 7,421

Including references and tables: 9,977

Total

Excluding references and tables: 15,705

Including references and tables: 23,672

Lay Summary

Bipolar disorders are characterised by recurrent periods of depressive (feeling intensely low) and manic (feeling extremely happy) symptoms with disturbances in cognitive and motor activity of varying severity, duration, and frequency. The chronicity associated with bipolar disorders often leads to poor psychological outcomes with an elevated risk of suicide and psychiatric comorbidity. Bipolar disorders are highly heritable, and a positive family history remains the strongest predictor for the development of a bipolar disorder. Clinically, bipolar disorders have proven difficult to identify which is partly due to a lack of consensus around diagnostic definitions and difficulties differentiating bipolar subtypes from one another. Current categorical approaches are limited in their ability to capture the heterogenous nature of bipolar disorders or their relationships with other symptoms of psychopathology. To contribute to our understanding of bipolar disorders, this project examined the structure of bipolar disorders and processes that may lead to affective disorders.

The first section of this thesis reviews the literature concerned with Akiskal's model of affective temperament and its relationship to mood disorders. Temperament is defined as an individual's predisposition towards patterns of emotional reactivity, which remain stable over time and are inheritable. Akiskal's theory describes five affective temperaments (depressive, hyperthymic, irritable, cyclothymic and anxious) that are thought to constitute the behavioural phenotypes in the pre-morbid course of affective disorders. The review provides partial support for a continuum model of temperament in which individuals with a bipolar disorder generally yielded higher scores on a measure of temperament than genetically at-risk first-degree relatives across depressive, irritable, cyclothymic and anxious subtypes. Although mixed, the findings have

implications for identifying those at risk or within prodromal phases of bipolar disorder and could provide important insights into the clinical evolution of mood disorders. However, this review was based on a limited number of studies, of varying quality, and should therefore be interpreted with caution.

To investigate the structure and development of bipolar disorders the second section of the thesis includes a network analysis of bipolar disorder symptoms and common psychiatric diagnoses to explore comorbidity and pathways from psychopathological states. According to network models, mental disorders (clusters of symptoms) occur, not because the symptoms have common underlying causes, but because symptoms are connected in a network of causal relationships. Networks were constructed using symptoms scores of 7076 participants from a general population sample. Results revealed symptoms of energy and activity as core features of bipolar disorders given their centrality and connectivity within the bipolar network. In addition, community analyses revealed four communities including a 'pure mania' community and a 'mixed' community consisting of irritability, distractibility and racing thoughts amongst depression and anxiety disorders. Analyses highlighted 'racing thoughts' as a possible bridge between communities, suggesting the presence of racing thoughts as a risk factor for convergence or comorbidity. The findings have clinical implications such as the development of interventions that target key connections to decouple strongly connected symptoms and 'deactivate' networks. However, as symptoms were assessed cross-sectionally they provide limited insight into how networks may change over time.

Acknowledgements

I would like to thank my supervisors, Professor Richard Bentall and Dr Georgina Rowse, for all their guidance and support throughout this project, as well as their words of encouragement and responsiveness at times when I thought I wouldn't quite make it. I would also like to thank Alba Contreras for all her R and network expertise, as well as Dr Zaynah Arshad for kindly appraising the literature review studies.

A thank you to my fellow trainees, Ellie, Georgia, Zaynah and Charlotte, for all their reflections and containment, but more importantly their friendship, and making my journey so memorable.

To my family and friends for their help, support, and interest (even if that meant asking me what my thesis was about every time I saw them). My mum for being my mentor and always pushing me to do my best. My daughter Aniyah, for revealing a new level of strength and determination that I did not know I had (and taking extra-long naps, sleeping all-night-long and giving me endless reasons to smile). As well as Mammam and Nanny Jan for all the childcare! My partner Aaron, for all the reassurance, love, and patience, throughout this whole process (and for sacrificing his desk to make my life easier!). Finally, to my late Nanna, who brightened me up after long days and kept me company during relentless commutes. We did it!

List of Contents

Access to Thesis Form	ii
Declaration	iv
Structure and Word Counts	v
Lay Summary	vi
Acknowledgments	viii
Section I: Literature Review	
<i>A Systematic Review to Explore Akiskal's Model of Affective Temperament in Bipolar Disorder and First-degree Relatives</i>	
	1
Abstract	2
Introduction	4
Method	12
Results	15
Discussion	36
References	45
Appendix A: The TEMPS	54
Appendix B: Search Terms	58
Appendix C: JBI Critical Appraisal Checklist	59
Appendix D: Quality Appraisal Table	60

Section II: Empirical Study***A Network Analysis of Bipolar Disorders and Common Psychiatric***

<i>Diagnoses</i>	65
Abstract	66
Introduction	68
Method	74
Results	81
Discussion	90
References	99
Appendix A: Ethical Approval	109
Appendix B: Mood Disorder Questionnaire	110
Appendix C: Network Accuracy Analyses (Aim 1)	111
Appendix D: Network Accuracy Analyses (Aim 2)	115

Section I: Literature Review

A Systematic Review to Explore Akiskal's Model of Affective Temperament in
Bipolar Disorder and First-degree Relatives

Abstract

Objectives

Bipolar disorders are highly heritable, and a positive family history remains the strongest predictor for the development of a bipolar disorder. Affective temperament, defined as an individual's predisposition towards certain patterns of emotional reactivity, are thought to represent the phenotypic expression of the genes underlying bipolar disorders. The current review aimed to examine whether there are differences in Akiskal's five affective temperament subtypes (depressive, hyperthymic, irritable, cyclothymic and anxious) for those at genetic risk of developing a bipolar disorder as measured by the Temperament Evaluation Memphis, Pisa, Paris, and San Diego Auto-questionnaire (TEMPS).

Methods

A systematic search of MEDLINE, PsychINFO, Web of Science, and Google Scholar was conducted. Studies were included if they administered the TEMPS to those at genetic risk of developing a bipolar disorder and compared scores to relatives with a bipolar disorder or healthy controls. The methodology of eligible studies was quality appraised, and a narrative synthesis of the data was conducted.

Results

Sixteen studies were included in the review which revealed mixed results in relation to a continuum model of temperament. Generally, studies reported higher TEMPS scores within bipolar disorder groups when compared to relatives, who had higher scores than healthy controls across four temperament domains (depressive, irritable, cyclothymic and anxious). In contrast several

studies reported higher scores amongst healthy controls on the hyperthymic subscale.

Conclusions

The review provides partial support for theories of an affective temperament continuum from extreme temperament to affective illness, that suggests a genetic basis for bipolar disorders.

Practitioner Points:

- The assessment of temperament in those at risk or within prodromal phases of bipolar disorder could provide important insights into the long-term course of affective disorders i.e. significantly higher affective temperament scores have been associated with suicide attempts.
- Screening for temperament within clinical settings could help clinicians more accurately assess treatment options, for instance cyclothymic temperament has been associated with poor treatment adherence.
- Psychological formulation could be enhanced by models of affective temperament, which focus on both strengths and difficulties and can help individuals to understand their predisposition and maximise positive attributes to overcome difficulties

Keywords: Bipolar disorder, affective temperament, TEMPS, genetic-risk, first-degree relatives.

Introduction

Bipolar disorders are a series of chronic mental health difficulties associated with poor psychological outcomes, inter-episodic dysfunction, with an elevated risk of suicide and psychiatric comorbidity (McIntyre, 2018). Bipolar disorders are thought to be one of the most heritable mental illnesses (Gordovez & McMahon, 2020) and a positive family history remains the strongest predictor for the development of a bipolar disorder (Loftus et al., 2016). Studies have shown that 7% of first-degree relatives of those with bipolar disorders, develop a bipolar disorder themselves, suggesting a sevenfold increase in risk compared to the general population (Kelsoe, 2003). Similarly, meta-analyses have shown that, compared with children of healthy controls, the child of a parent with a bipolar disorder has a 33% risk of developing any serious mental health condition, which is more than twice that of controls (Rasic et al., 2013). As such, it can be argued that studies of first-degree relatives provide one of the most reliable and valid means of identifying a sample of individuals at high risk for developing a severe mental disorder and can provide important information on prodromal signs, symptoms and rates of transmission (Loftus et al., 2016).

Nonetheless, a variety of mood symptoms and mental health difficulties have been observed in the relatives of those with bipolar disorders, suggesting a complex interplay between both biological and environmental factors (Kelsoe, 2003). Numerous environmental factors have been found to contribute to the development of bipolar disorders including urban up-bringing, stressful life events, substance abuse, prenatal infections and complications which may interfere with brain development (Misiak et al., 2017). Current methods of diagnosis have proven to be limited in their ability to capture the variation

observed in the families of those with bipolar disorders (Evan et al., 2005).

Thus, it has been suggested that bipolar disorder may be better explained by a polygenic model, in which many genes, each with small effects, interact and contribute to the development of the disorder (Kelsoe, 2003). Within this model it is thought that the relevant genes produce a continuous variation of affective phenotypes that also penetrate the realm of 'normal' behaviour (Evan et al., 2005). Akiskal and Akiskal (2005a) have suggested that the concept of 'affective temperament' can be used to explain the spectrum of affective disturbances from healthy emotional reactivity to major affective disorders (Vazquez & Gonda, 2013).

Affective temperament is defined as an individual's predisposition towards certain patterns of emotional sensitivity and reactivity which remains stable over time and is inheritable (Goldsmith et al., 1987). Affective temperaments are thought to represent the phenotypic expression of the genes underlying bipolar disorders (Kelsoe, 2003), providing a link between predisposing familial factors and affective illness (Akiskal et al., 2005a). Four basic affective dispositions (depressive, manic, cyclothymic and irritable) were first described by Kraepelin (1921 as cited in Vazquez et al., 2008) who considered these temperaments to be subclinical forms of what was then known as 'manic-depressive' illness. Kraepelin and colleagues identified that affective temperaments were present not only in their patients but also in the relatives of those affected by mood disorders. It is thought that, whilst temperaments may predispose an individual to a mood disorder, the presence of a dominant temperament could also be considered a variation of normal affectivity which may or may not lead to an affective illness (Vazquez & Gonda 2013).

Drawing on these long-standing ideas and more recent scientific observations Akiskal and colleagues (2005a) developed a contemporary model of affective temperament to encompass the whole spectrum of affective disturbances. The model includes the four original temperamental types described by Kraeplin (with 'manic' redefined as "hyperthymic"), as well as a new "anxious" temperament type. Whilst there is limited research on the continuum between 'normal' and extreme temperaments, there is some evidence to suggest that temperamental traits are continually distributed (Akiskal & Akiskal, 2005a) and include both pathological and adaptive features (Akiskal et al., 2005a).

Types of temperament

The hyperthymic temperament

The hyperthymic temperament has typically been characterised by exuberant, upbeat, overenergetic, and overconfident lifelong traits (Akiskal & Akiskal, 2005a). However, other authors have described the behavioural pattern of those with a hyperthymic temperament as extroverted, verbally aggressive, risk-taking and sensation-seeking (Possl & von Zerssen, 1990), as well as scheming, tireless and meddlesome (Gardner, 1982).

Akiskal and colleagues (2005a) have developed an operational definition of the hyperthymic temperament which was revealed following the analysis of 110-items in a self-rated form (Akiskal et al., 2005b). Those with a hyperthymic temperament have been described as cheerful and overoptimistic, warm, and extroverted, high energy levels, typically with short length of sleep, uninhibited, sensation-seeking with promiscuous tendency, impulsive and overinvolved with

several activities. The factor analysis revealed close to opposite factor loading to the depressive temperament.

The depressive temperament

Individuals with a depressive or dysthymic temperament have been described as self-denying, often dedicating themselves to others, with a sensitivity to suffering (Akiskal & Akiskal, 2005a). The depressive temperament is characterised by low energy levels, harm avoidance and a dislike of change, with a desire for harmony and security. The depressive temperament is thought to have an associated mood dominated by dejection and unhappiness, and a self-concept that includes low self-esteem and beliefs of worthlessness (Vanquez & Gonda, 2013). At the extreme end, this affective temperament can be a risk factor for the development of major depression (Akiskal & Akisal, 2005a).

The irritable temperament

The irritable temperament classically defined by Kraepelin (1921 as cited in Vazquez & Gonda, 2013) as the lifelong combination of both hyperthymic and depressive temperaments at the same time. The irritable temperament is characterised by depressive mood with periods of irritability, impulsivity, restlessness, and unhappiness (Akiskal & Akiskal, 2005b). It is thought that those with an irritable temperament are hypercritical and could be described as confrontational and aggressive, with a bitter disposition.

The cyclothymic temperament

The cyclothymic temperament was historically described by Kraepelin as the alternation of hyperthymic and depressive temperaments (Kraepelin, 1921 as

cited in Vazquez & Gonda, 2013) and is now thought to be a pattern of alternation between hypomanic or irritable, and depressive subclinical moods, cognitions, and behaviours. The cyclothymic temperament is characterised by extreme lability which can include periods of hypersomnia alternating with a decreased need for sleep, introverted self-absorption alternating with uninhibited people-seeking (which may lead to hypersexuality), being more talkative than usual including inappropriate laughing and joking or being less talkative with unexplained crying or tearfulness, periods of mental confusion and apathy alternating with restless pursuit of activities and periods of sharpened and creative thinking, unstable self-esteem alternating between low self-confidence and grandiose overconfidence as well as times of over-optimism or exaggeration of achievements alternating with a pessimistic view of the future. It has been proposed that those with a cyclothymic temperament may use alcohol or drugs to manage their mood or to enhance excitement (Akiskal 1992; Perugi, 2003).

The anxious temperament

The anxious temperament is a new addition to the modern concept of affective temperaments and has been described by Akiskal and Akiskal (2005a) as an exaggerated personality disposition toward worrying. It is thought that the anxious temperament could represent a predisposition to generalised anxiety disorder and is characterised by harm avoidance, hypervigilance, dependency, tension, and an inability to relax. The anxious temperament has been associated with gastrointestinal and muscular symptoms which from an evolutionary perspective functions to prevent relaxation to aid survival. It has been suggested that the anxious temperament was not identified historically as

it may have been overshadowed by the greater emotionality and intensity of the other affective temperaments (Akiskal, 1998).

The measurement and clinical significance of temperament

Temperament is related to many of the processes implicated in bipolar disorder, including emotion regulation, arousal, and affect (Youngstrom et al., 2011). It is proposed that when any of the affective temperaments, are expressed in a marked form, they represent the subclinical manifestation of affective illnesses, and can be considered precursor states and represent high-risk conditions (Vanquez & Gonda (2013). Current evidence suggests that bipolarity may lie along a continuum from extreme temperament to affective illness (Akiskal & Akiskal, 2005a). Some older hypotheses are also consistent with this notion. As quoted in Akiskal and Akiskal (2005a), Kretschmer (1936) states “endogenous psychoses are nothing but exaggerated forms of normal temperament.” Based on this theoretical framework it has been hypothesised that affective temperament has a key role in the development of affective disorders (Akiskal & Akiskal, 2005a) and temperamental dysregulation could constitute the link between predisposing familial-genetic factors and bipolar disorders (DiFlorio et al., 2010).

Theories of temperament are typically measured by questionnaire, with the most recent review comparing the validity of eleven different personality inventories (Grucza & Goldbery, 2007). However, most questionnaires focus on personality more broadly, with only two questionnaires specifically considering theories of temperament, the Tridimensional Cloninger Inventory (TCI; Cloninger et al., 1993) and the Temperament Evaluation of the Memphis, Pisa, Paris, and San Diego (TEMPS; Akiskal & Akiskal, 2005b). The TCI (Cloninger

et al., 1993) takes a dimensional approach to personality assessment and measures two components of personality traits, temperament, and character. Cloninger and colleagues (1993) propose the existence of four dimensions of temperament: novelty seeking, harm avoidance, reward dependence and persistence, alongside three dimensions of character. Each dimension within Cloninger's model is thought to reflect normally distributed quantitative traits, accounting for both normal and maladaptive individual differences which present in varying degrees in everyone (Fountoulakis, 2019; Fruyt 2006).

The TEMPS developed by Akiskal and Akiskal (2005b), focuses specifically on the affective components of temperament, encompassing the five temperament profiles derived from theoretical and clinical considerations (Akiskal et al., 2005a). The TEMPS is the most widely used set of instruments to measure affective temperaments (Elias et al., 2017) and exists in several versions. The TEMPS was originally designed to be administered by semi-structured interview (TEMPS-I; Akiskal et al., 1998), but has been developed into a self-rated version, known as the TEMPS-A (Akiskal et al., 2005c), and a short version intended for clinical use (Akiskal et al., 2005b; see appendix A).

The TEMPS-A contains 110 questions across depressive, cyclothymic, hyperthymic, irritable and anxious subscales, requiring "yes" (score 1) or "no" (score 0) responses. Each subscale is then scored by dividing the sum of the subscale by the number of questions within the subscale. The TEMPS characterises the dominant temperament of an individual based on the severity of the traits ranging from 0 to 1. The TEMPS has been validated for use in both healthy controls and those with a psychiatric diagnosis (Akiskal et al., 1998). The TEMPS has been translated into several languages and has consistently

demonstrated acceptable psychometric properties across different settings and populations (Akiskal et al., 2005b).

Although the TEMPS does not have any intrinsic psychopathological predictive value, the predominant affective temperament maps onto the spectrum of bipolar disorders from subclinical presentations through bipolar disorder type I (BD-I), type-II (BD-II), and major depressive disorder (MDD) at the opposite end (Solmi et al., 2016). Unlike the TCI, the TEMPS is rooted in evolutionary theory and its validity has been supported by genetic studies which have found significant associations between TEMPS scores and serotonin transporters (Gonda et al., 2005). Despite some indication that the TEMPS may be applicable within clinical settings (Vöhringer et al., 2011), temperament itself is not typically assessed outside of research settings (Youngstrom et al., 2011). There are however important clinical and therapeutic implications associated with affective temperament including the early identification of those with poorer prognostic features (Vöhringer et al., 2011). Research has suggested that interventions can be more effective when designed and delivered in accordance with an individual's temperament (Akiskal & Akiskal, 2005a). For example, hyperthymic temperaments prefer action-orientated approaches and anxious temperaments can benefit from meditation. Furthermore, some temperaments are associated with lower levels of trait conscientiousness making it more likely that there will be difficulties following through with taking medication, keeping appointments, or completing tasks from therapy (Youngstrom et al., 2011).

Importantly, if temperamental dysregulation does constitute a link between predisposing genetic factors and bipolar disorders, the assessment of temperament, could assist in identifying and providing appropriate early interventions for those at risk of developing a mood disorder. Several studies

have utilised the TEMPS over the years to examine affective temperament in those with bipolar disorders, relatives, and control groups. There has been some research to suggest that cyclothymic temperament is highly prevalent in the children of parents with bipolar disorders, and the trait increases the risk of developing a bipolar disorder above that associated with heritability or temperament alone. In addition, young people with depression and cyclothymic temperament are much more likely to convert to bipolar disorder than young people with depression alone (Youngstrom et al., 2011).

Despite numerous studies assessing the link between affective temperament and bipolar disorders, very few have attempted to review the current evidence (Solmi, 2016; Vazquez, 2013). Solmi and colleagues (2016) conducted a meta-analysis to investigate a metric continuum of TEMPS scores in healthy controls, mood disorder patients and individuals with other psychiatric diagnoses. The results provided support for a continuum model of temperament, however, data concerning relatives was only available in four of the 26 studies that were included in the review, limiting the validity and generalisability of the findings.

Thus, the current research aimed to address this gap and review the evidence concerned with first-degree relatives to examine whether there are differences in affective temperament for those at genetic risk for a bipolar disorder. If Akiskal's theoretical model of affective temperaments is correct, then we expect the TEMPS to reveal intermediate scores for unaffected relatives who share some but not all the genes of first-degree relatives with a bipolar disorder, with relatively higher scores for affected relatives (relatives with any diagnosable mental health disorder). Likewise, we would anticipate the highest

scores for probands with a diagnosis of bipolar disorder and the lowest scores for healthy controls without a family history of bipolar disorders.

Method

The present review adhered to the Manual for Evidence Synthesis by the Joanna Briggs Institute (JBI; Aromataris & Munn, 2020) and the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Page et al., 2020), following a predetermined, but unpublished protocol.

Search strategy

A systematic literature search was conducted in the databases MEDLINE, PsycINFO, Web of Science and Google Scholar to identify eligible articles published in English. The search period was from the beginning of the databases to April 2021, with the final search conducted on 13 April 2021. The Boolean operators AND and OR were used to combine search terms associated with “temperament”, “bipolar disorder” and “first-degree relatives” (see Appendix B for a full list of search terms). The author performed keyword searches and a citation search using Web of Science. The reference lists of included papers and relevant reviews were checked for potentially eligible studies. Figure 1 outlines the process and outcome of the literature search in a PRISMA diagram (Page et al., 2020).

The titles and abstracts of studies were scanned for relevance and then assessed for eligibility. Papers were eligible for inclusion if they (1) administered the TEMPS (any version), (2) to those at genetic risk of developing a bipolar disorder (i.e. any first-degree relative of an individual with BD-I, BD-II or bipolar disorder not otherwise specified (BD-NOS), (3) compared scores to probands

with a bipolar disorder or healthy controls and (4) published in peer-reviewed journals. Studies were excluded if they (1) did not include first-degree relatives, (2) utilised measures of affective temperament that did not include the TEMPS, (3) were not available in English and (4) were not original published articles including reviews or summaries.

Data extraction

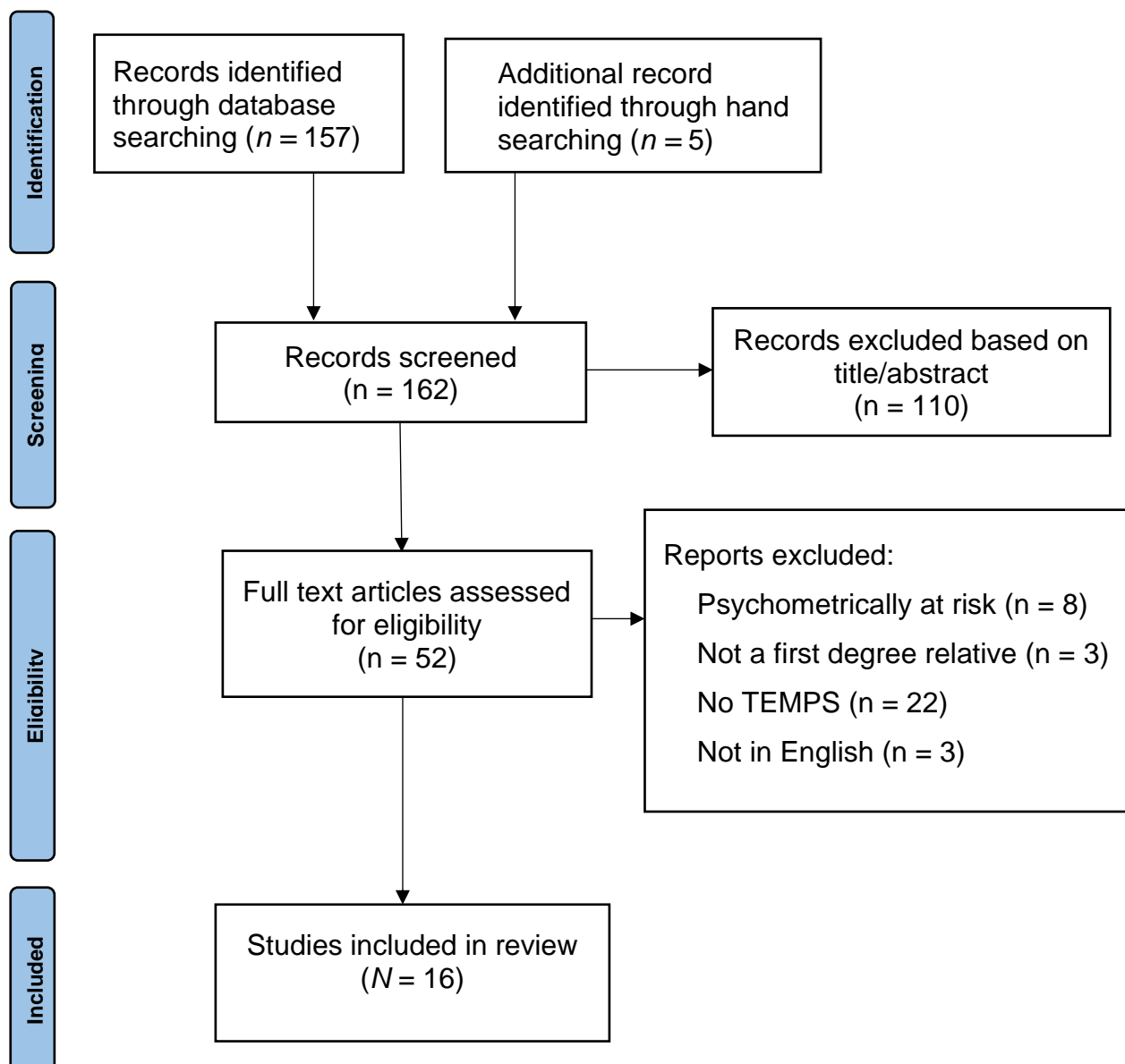
The following study characteristics were extracted from eligible studies: author, publication year, country, demographic characteristics for subjects and comparison groups (mean age, diagnoses, sex, sample size), TEMPS version and mean and SD of TEMPS scores in each group. The primary outcome measure was the standardised mean difference of each TEMPS score between first-degree relatives, probands with a bipolar disorder and any other available control group.

Quality assessment

To assess the methodological quality of the research articles, the strengths and weaknesses of each paper were systematically reviewed. Considering the limited number of studies in this area, studies were not excluded because of their quality appraisal score. The studies included in this review were all cross-sectional studies. Therefore, study quality was assessed using the JBI Critical Appraisal Checklist for Cross-sectional studies (Moola et al., 2020; see Appendix C).

Figure 1

PRISMA diagram of study selection process



Each paper was assigned an overall quality score based on how many of the criteria they met on the checklist. Two points were awarded when the paper fully met the criteria for an item, one point when it was unclear whether the paper met the criteria or the criteria was partially met, and no points when the paper did not meet the criteria, with a maximum score of 16 available. To aid comparison between papers, a total quality percentage score was calculated for each paper by dividing the overall quality score by its maximum possible score and multiplying by 100. Since the JBI has not published a categorisation system for its checklists, the author created an arbitrary categorisation system and categorised papers as either: *poor* (<59%), *fair* (60%-69%), *good* (70%-79%) or *excellent* (>80%) quality.

To assess interrater reliability, a peer researcher randomly selected 50% of the eligible papers ($n=8$) and conducted an independent assessment of quality using the JBI checklist. The second rater was blind to the first rater's scoring. Discrepancies in ratings were discussed and resolved.

Results

Characteristics of included studies and a summary of results are reported in Table 1. Sixteen studies were included in the systematic review (Chiaroni, 2005; Evans, 2005; Ferencztajn, 2015; Ferreira, 2013; Gandotra, 2011; Greenwood, 2013; Higier, 2014; Kesebir, 2005; Kesebir, 2020; Liu, 2021; Mahon, 2013; Mendlowicz, 2005; Saguem, 2019; Savitz, 2008; Vázquez, 2008; Whalley, 2011). The included studies contained 4442 participants, these included 931 with a diagnosis of bipolar disorder (BDI: 359; BDII: 120; BD-NOS: 3), 409 with MDD, 143 with a diagnosis of schizophrenia, 1396 healthy controls

and 1431 relatives of subjects with a bipolar disorder (healthy relatives: 1069; relatives with a diagnosable mental health disorder: 60).

Four studies were carried out in the USA, two in Brazil and one each in France, Poland, India, Sweden, Turkey, China, Tunisia, South Africa, Argentina and the UK. Eight of the studies described the relationship between the proband and family member (i.e. parent, child or sibling), five studies stated that participants were first-degree relatives, however, three studies only described subjects as family or relatives (Evans, 2005; Mendlowicz, 2005; Savitz, 2008).

Fourteen of the studies compared relatives to healthy controls, eleven studies included bipolar disorder subjects in the analysis and one study included a comparison of BDI vs BDII (Savitz et al., 2008). Five studies included other psychiatric diagnoses (MDD and schizophrenia) as comparison groups (Evans, 2005; Higier, 2014; Gandotra, 2011; Greenwood, 2013; Savitz, 2008). Eight of the studies that included bipolar disorder subjects reported the stage of bipolar disorder as euthymic or in remission, two of the studies reported that not all of the subjects were in remission (Savitz, 2008; Vázquez, 2008), one study did not report the status of bipolar disorder probands (Evans et al., 2005). All the studies included outpatients.

The full version of the TEMPS-A questionnaire was used in five studies (Evans, 2005; Ferensztajn, 2015; Greenwood, 2013; Mahon, 2013; Savitz, 2008), and three studies administered the TEMPS-A short or clinical version (Ganotra, 2011; Mendlowicz, 2005; Whalley, 2011). The other seven studies used the TEMPS- Rio de Janeiro (Ferreira et al., 2013), the Turkish version of the TEMPS-A (Kesebir, 2005; Kesebir, 2020), the Chinese version of the TEMPS-A (Liu et al., 2021), the Arabic version of the TEMPS-A (Saguem et al.,

2021), the TEMPS-A Buenos Aires (Vázquez et al., 2008) and two studies used modified or only selected items from the TEMPS-A (Chiaroni, 2005; Higier, 2014).

Quality appraisal

The results of the quality appraisal are presented in Appendix D. The calculation of the intraclass correlation co-efficient (ICC) indicated that there was good interrater reliability (ICC = 0.752, 95% CI [0.237, 0.950], $F(7,7) = 4.038$, $p < 0.05$) (Koo & Li, 2016). Most of the studies ($n = 10$) were considered by the author to be of fair to good quality with ratings falling between 62% and 75%. Five studies were rated to be of poor quality (Evans, 2005; Ferreira, 2013; Greenwood, 2013; Mendlowicz, 2005; Savitz, 2008) and one study was rated as excellent quality (Saguem, 2021). Studies generally benefitted from the standardised assessment of individuals with a bipolar disorder, with most utilising qualified and experienced clinicians to decrease the risk of bias. However, only half of the included studies attempted to control for confounding factors such as the phase of bipolar disorder. In addition, several studies failed to clearly define the inclusion or exclusion criteria and whether this was determined prior to recruitment.

Table 1*Study characteristics and main findings*

Authors (year)/ country	Design	Participant characteristics	Illness phase BD	Exclusion criteria	TEMPS version	Main finding	Quality (%)
Chiaroni (2005)/ France	CS	177 participants BDR [$n=40$; 62% female; mean age = $8.2(\pm 16.2)$] HC [$n=100$; 52% female; mean age = $38.3(\pm 9.1)$] HC-P [$n=37$; 64% female; mean age = $7.8(\pm 11.2)$]	-	Proband relatives with affective disorder HC with a history of an affective disorder	TEMPS-A (8 items [CT only])	Significant difference in cyclothymic temperament scores between BDR and control groups with gradation (higher scores for BDR and lower scores for HC) Cyclothymic temperament more prevalent in HC-P than HC	62%
Evans (2005)/ USA	CS	443 participants Group 1 (BD): BD-I ($n=109$) BD-II ($n=46$) SA-BD ($n=4$) Group 2 (AR): MDD-recurrent ($n=69$) MDD-single ep ($n=31$) Group 3 (UR; $n=124$) Group 4 (HC; $n=63$)	-	-	TEMPS-A	Significant difference in temperament scores with expected average order (BD [highest], AR, UR and HC [lowest]) except for the hyperthymic scale in which controls scored highest.	43%
Ferensztajn (2015)/ Poland	CS	50 AR and UR grouped by proband lithium response [54% female; mean age = $34(\pm 9)$] Group 1: ELR	-	-	TEMPS-A	Distinct temperament profiles for relatives by response to lithium.	62%

		[54% female; mean age = 34(±9)] Group 2: PR [62% female; mean age = 37(±8)] Group 3: NR [51% female; mean age = 33(±9)]				Higher scores for AR of ELM, except for the hyperthymic scale, in which UR scored higher.	
Ferreira (2013)/ Brazil	CS	446 participants BD-I/BD-II [<i>n</i> = 90; 64.5% female; mean age = 38.38 (±11.97) years] MDD [<i>n</i> = 88; 75% woman; mean age = 46.78 (±11.91) years] UR [<i>n</i> = 132; 65.90% female; mean age = 36.55 (±12.47) years] HC [<i>n</i> = 136; 64.5% female; mean age = 33.43 (±12.18) years]	Remission	BD patients excluded with score above 8 on HDRS-17 or YMRS UR excluded if criteria for psychiatric diagnosis was met according to MINI-PLUS HC excluded if there was any personal or first-degree family history of mental health disorders	TEMPS-Rio de Janeiro	All temperament variables showed significant between-group differences. Scores were higher in the clinical groups compared to controls on all temperaments except for the hyperthymic scale	56%
Gandotra (2011)/ India	CS	198 participants BD proband [<i>n</i> = 33; 27.3% female; mean age = 30.90 (±9.93) years] BDR [<i>n</i> = 33; 18.2% female; mean age = 43.18 (±14.69) years] Matched controls (characteristics not reported): Sch (<i>n</i> = 33) SchR (<i>n</i> = 33) HC (<i>n</i> = 33) HC-R (<i>n</i> = 33)	Remission	Relatives of probands with major psychiatric illness Probands with any comorbidities or history of head injury	TEMPS-A (Short version)	Hyperthymic temperament was the predominant temperament type with a gradual decremental gradient (significantly higher in the BD and lower in HC)	75%

Authors (year)/country	Design	Participant characteristics	Illness phase BD	Exclusion criteria	TEMPS version	Main finding	Quality (%)
Greenwood (2013)/USA	CS	670 participants from 101 families [62% female; mean age = 45(±17)] BD group: BD-I (<i>n</i> = 128; 29.9%) BD-II (<i>n</i> = 40; 9.3%) SA-BD (<i>n</i> = 9; 2.1%) MDD group: MDD-Recurrent (<i>n</i> = 100; 23.4%) MDD-Single episode (<i>n</i> = 18; 4.2%) UR (<i>n</i> = 148; 28.5%) HC (<i>n</i> = 53)	-	-	TEMPS-A	Significant group differences across all five temperaments. All except hyperthymic showing the expected trend of more pathological scores for BD subjects, followed by MDD, then unaffected relatives, and finally controls except for hyperthymic.	50%
Higier (2014)/Sweden	CS	258 twin pairs Control twins [<i>n</i> = 63; 52.4% female; mean age = 47 (±11.5) years] Sch proband twins [<i>n</i> = 55; 49.1% female; mean age = 49.7 (±10.2) years] Sch UR-twins [<i>n</i> = 31; 51.6% female; mean age = 51.6 (±9.9) years] BD proband twins [<i>n</i> = 64; 60.9% female; mean age = 48.8 (±10.7) years] BD UR-twins [<i>n</i> = 45; 60% female; mean age = 49.1 (±9.8) years]	Clinically stable	Any subject with neurological disorder, history of significant head injury with loss of consciousness, mental retardation, substance use disorder within the previous 6 months, and inability to read or comprehend spoken and written Swedish. HC twin pairs with personal/family history of BD or Sch.	TEMPS-A short form (33 items) split into "positive" and "negative" scales	Bipolar UR-twins showed elevated scores on a "positive" temperament scale compared with controls and bipolar probands, while bipolar probands scored higher on a "negative" scale compared with their UR-twins and controls, who did not differ	75%
Kesebir (2005)/USA	CS	638 participants BD probands [<i>n</i> = 100; 57% female; mean age = 36.7(±14.5) years]	Euthymic	Illiteracy or serious physical or mental disability (i.e. dementia) or	TEMPS-A (Turkish version)	Hyperthymic temperament was significantly more common in the patient-group, FDR, and FDR with BD than in the controls	68%

		BDR [$n = 219$; 54.3% female; mean age = 43.2 (± 16.3) years] Age and gender-matched controls for both groups		younger than 16 years of age			
Kesebir (2020)/ Turkey	CS	75 participants BD proband [$n = 25$; 72% female; mean age = 32.8 (± 5.7) years] UR [$n = 25$; 88% female; mean age = 36.4 (± 4.1) years] HC [$n = 25$; 80% female; mean age = 40.1 (± 5.3) years]	Euthymic	Physical or neurological illness	TEMPS-A (Turkish version)	Cyclothymic and hyperthymic temperament scores were similar between BD probands and UR but higher than HC	62%
Liu (2021)/ China	CS	112 BDR AR [$n = 60$; 48% female; mean age = 15.57 (± 5.77) years] UR [$n = 52$; 55% female; mean age = 16.83 (± 5.48) years] HC [$n = 48$; 54% female; mean age = 15.63 (± 4.50) years]	NA	DSM-IV-TR Axis I disorders Alcohol or drug dependence Pregnant women history of brain trauma Illiteracy, any family history of psychiatric disorder for HC	TEMPS-A (short - Chinese version)	AR had more predominant cyclothymic, irritable, depressive and anxious temperaments than UR and HCs. UR and HCs showed similar affective temperament traits.	75%
Mahon (2013)/ USA	CS	221 participants BD probands [$n = 55$ (BDI $n=47$; BDII $n=5$; BD-NOS $n=3$) 65% female; mean age = 39.4 (± 12.2) years] UR (sibling) [$n = 53$; 66% female; mean age = 40.7 (± 11.7) years] HC [$n = 113$; 46% female; mean age = 38.0 (± 11.6) years]	Clinically stable	UR Axis I mood or psychotic disorder HC with any lifetime Axis I diagnoses or first-degree relative with Axis I diagnoses substance abuse or dependence	TEMPS-A	Those at genetic risk for BD (i.e. siblings of probands) show elevated levels of traits associated with the clinical manifestation of BD. Scores for UR were typically higher than HC but lower than probands.	62%

Authors (year)/country	Design	Participant characteristics	Illness phase BD	Exclusion criteria	TEMPS version	Main finding	Quality (%)
Mendlowicz (2005)/Brazil	CS	75 participants Recovered BD probands [$n = 23$ (BDI $n=18$; BDII $n=5$) 47% female; mean age = 43.03 (± 16.61 years)] UR [$n = 52$; 50% female; mean age = 50.93 (± 19.29) years] HC [$n = 102$; 37% female; mean age = 36.47 (± 12.33) years]	Clinically recovered	UR with current or history of Axis I or II diagnosis More than 2 relatives with a mood disorder	TEMPS-A (short version)	Depressive and irritable temperament scores differentiated between recovered BD probands, UR and HC. UR exhibit cyclothymic scores halfway between this of BD and HC. Hyperthymic temperament highest in HC although this was not significant.	56%
Saguem (2021)/Tunisia	CS	180 participants BD proband [$n = 60$; sex ratio = 0.33, mean age = 43 (± 10) years] UR (siblings [$n = 60$; sex ratio = 0.5, mean age = 41 (± 12) years]) HC [$n = 60$; sex ratio = 0.42, mean age = 44 (± 10) years]	Remission	More than 2 symptoms of mania, MDD or hypomania no sibling UR and HC with a history of mood disorder HC with a family history of mood disorder	TEMPS-A (Arabic version)	BD probands showed significantly higher scores in hyperthymic, cyclothymic and depressive temperament dimensions compared to HC. Temperamental dysregulation UR who showed higher scores in cyclothymic and hyperthymic temperament dimensions compared to HC.	87%
Savitz (2008)/South Africa	CS	296 participants BDI ($n = 57$) BDII ($n = 24$) MDD recurrent ($n = 58$) MDD single episode ($n = 45$) UR ($n=86$)	Euthymic (most)	-	TEMPS-A	The BDI, BDII, and to a lesser extent the MDD recurrent group scored higher, on average than UR on cyclothymic and irritability temperaments	56%
Vázquez (2008)/Argentina	CS	229 participants UR [$n = 114$, 68.4% female, mean age = 43 (± 17.2) years] HC [$n = 115$, 61.7% female, mean age = 39 (± 13.9) years]	NA	Diagnosis of BD	TEMPS-A	UR of bipolar probands exhibit higher levels of temperamental traits than HC. Cyclothymic and anxious temperament scores differentiated between UR and HC	62%

Whalley (2011)/ UK	CS	163 participants UR [$n = 93$, 51% female, mean age = 21.01 (± 2.83) years] HC [$n = 70$, 54% female, mean age = 20.89 (± 2.30) years]	NA	History of MDD, mania, hypomania, psychosis, or any major neurological or psychiatric disorder; a history of substance dependence; learning disability, head injury that included loss of consciousness	TEMPS-A (short version)	No main conclusions for TEMPS-A data – no significant differences found. Increases in activation in those at familial risk of bipolar disorder in the left amygdala versus comparison subjects.	68%
--------------------------	----	--	----	---	-------------------------	--	-----

Note. BDR: bipolar disorder relative; CS: cross-sectional; HC: healthy control; HC-P: healthy control positive family history of psychiatric illness; HC-R: relatives of healthy controls; BD-I: bipolar disorder type 1; BD-II: bipolar disorder type 2; SA-BD: schizoaffective bipolar disorder; AR: affected relative; UR: unaffected relative; MDD: major depressive disorder; ELR: excellent lithium responder; PR: partial responder; NA: not applicable; NR: non-responder; Sch: schizophrenia; Sch-R: schizophrenia relative; RR: relative risk;

Comparison of affective temperament scores

Table 2 presents a summary of the main findings.

First-degree Relatives vs Bipolar Disorder groups

Ten studies in total compared temperament scores of first-degree relatives and bipolar disorder groups across the five subscales (Evans, 2005; Ferreira, 2013; Greenwood, 2013; Higier 2014; Kesebir, 2005; Kesebir, 2020; Mahon, 2013; Mendlowicz, 2005; Saguem, 2021; Savitz, 2008).

Table 2

Summary of results

TEMPS-A Subscale	Comparison	Expected direction of effect	Total studies	Studies reporting expected significant differences	Studies reporting opposing significant differences	Studies reporting non-significant findings
Hyperthymic	FDR vs BD	BD > FDR	9	2	6	1
	AR vs UR	AR > UR	5	3	-	2
	FDR vs HC	FDR > HC	10	0	7	3
Depressive	FDR vs BD	BD > FDR	8	6	-	2
	AR vs UR	AR > UR	4	3	-	1
	FDR vs HC	FDR > HC	11	4	-	7
Irritable	FDR vs BD	BD > FDR	9	7	-	2
	AR vs UR	AR > UR	5	4	-	1
	FDR vs HC	FDR > HC	11	7	1	3
Cyclothymic	FDR vs BD	BD > FDR	9	7	-	2
	AR vs UR	AR > UR	5	4	-	1
	FDR vs HC	FDR > HC	12	7	1	4
Anxious	FDR vs BD	BD > FDR	8	6	-	2
	AR vs UR	AR > UR	3	1	-	2
	FDR vs HC	FDR > HC	11	6	1	4

Note. AR: affected relative; BD: bipolar disorder; HC: healthy control; UR: unaffected relative

Hyperthymic temperament. Nine studies compared hyperthymic temperament in subjects with a bipolar disorder and first-degree relatives. Six studies found higher hyperthymic TEMP scores in first-degree relatives compared to individuals with a bipolar disorder, but significant differences were only observed in two of the studies (Evans, 2005; Greenwood, 2013). Both studies were of poor quality and, importantly, neither reported the illness phase of bipolar disorder subjects or the familial relationship of the relatives. Of the three studies that observed higher hyperthymic temperament scores in the bipolar disorder group, two revealed significant differences between groups (Ferreira, 2013; Savitz, 2008). Similarly, both studies were of poor quality due to an increased risk of bias.

Depressive temperament. Eight studies compared depressive temperament scores of relatives and subjects with bipolar disorder. Six found that subjects with a bipolar disorder had significantly higher TEMPS scores on the depressive subscale than unaffected relatives (Evans, 2005; Ferreira, 2013; Greenwood, 2013; Mahon, 2013; Mendlowicz, 2005; Saguem, 2021) and affected relatives (Evans et al., 2005). Three of the studies did not control for the illness phase of the bipolar disorder group and were of poor quality (Evans, 2005; Ferreira, 2013; Greenwood, 2013). However, the remaining three studies reported that all subjects with a bipolar disorder were clinically recovered (Mendlowicz et al., 2005) or were in remission (Mahon, 2013; Saguem, 2021) and were considered fair and excellent quality respectively.

Irritable temperament. Nine studies compared irritable temperament in subjects with a bipolar disorder and first-degree relatives. Seven studies found that TEMPS scores for the irritable subscale were significantly higher in the bipolar disorder group compared to first-degree relatives. Most of the studies

were of poor quality (Evans, 2005; Ferreira, 2013; Greenwood, 2013; Mendlowicz, 2005; Savitz, 2008), and the remaining two studies were fair (Kesebir et al., 2005; Mahon, 2013). Although not significant, the remaining two studies did find higher irritable TEMPS scores amongst the bipolar disorder group. One of these studies was of excellent quality (Saguem et al., 2021) and attempted to control for mood state, by only including bipolar disorder participants once remission had been confirmed and gathering data on significant life events for participants six months before participation.

Cyclothymic temperament. Nine studies compared cyclothymic temperament in subjects with a bipolar disorder and first-degree relatives. Comparably to the irritable temperament findings, seven studies found that TEMPS scores for the irritable subscale were significantly higher in the bipolar disorder group compared to first-degree relatives. As previously discussed, two of the studies were deemed to be fair quality (Kesebir et al., 2005; Mahon, 2013), however, most of the studies were poor (Evans, 2005; Ferreira, 2013; Greenwood, 2013; Mendlowicz, 2005; Savitz, 2008).

Anxious temperament. Eight studies compared anxious temperament in subjects with a bipolar disorder and first-degree relatives. Similarly, to both irritable and cyclothymic temperaments, six studies found significantly higher TEMPS scores for the anxious subscale in the bipolar disorder group compared to first-degree relatives. Four of the studies were poor (Evans, 2005; Ferreira, 2013; Greenwood, 2013; Mendlowicz, 2005) whilst two were deemed to be fair (Kesebir et al., 2005; Mahon, 2013). Saguem et al. (2021) conducted an excellent quality study, and whilst similar results were observed, the differences between groups was not significant.

One final study (Higier et al., 2014) compared TEMPS scores between subjects with bipolar disorder and first-degree relatives. Whilst the study was deemed to be of good quality, the findings were grouped by “positive” and “negative” items making it difficult to interpret and integrate findings. Nonetheless, the results showed significantly higher TEMPS scores for “positive” items for first-degree relatives compared to the bipolar disorder group. Similarly, relatives scored significantly lower on the “negative” scale compared to the bipolar disorder group.

Affected Relatives vs Unaffected Relatives

Five studies in total compared temperament scores of affected relatives and unaffected relatives across the five subscales. Affected relatives included first degree relatives meeting the diagnostic criteria for MDD (Evans, 2005; Savitz, 2008) bipolar spectrum disorder (Gandotra et al., 2011) or any mood disorder (Ferensztajn, 2015; Liu, 2021).

Hyperthymic temperament. Five studies compared hyperthymic temperament scores between affected relatives and unaffected relatives. All five studies reported higher scores for unaffected relatives, although only three found significant differences between the groups. Two of the studies were poor; one did not describe the familial relationship of relatives (Evans et al., 2005) and the other experienced a significant rate of study refusal from relatives which could have biased the sample (Savitz, et al. 2008). The third study was fair in quality (Ferensztajn, 2015), although only relatives of ‘excellent lithium responders’ were included in the analyses which limit the generalisability of results. All the relatives included in the studies were assessed by qualified professionals using standardised interview procedures.

Depressive temperament. Four studies compared depressive temperament scores between affected relatives and unaffected relatives. Three of the four studies found significantly higher depressive temperament scores in affected relatives compared to unaffected relatives. The studies were poor (Evan et al., 2005), fair (Ferenztajn et al., 2015) and good (Liu et al., 2021) quality.

Irritable temperament. Similarly, to the hyperthymic comparison, five studies compared irritable temperament scores between affected relatives and unaffected relatives. Four studies found significantly higher irritable TEMPS scores for affected relatives compared to unaffected relatives (Evans, 2005; Gandotra, 2011; Liu, 2021; Savitz, 2008). Two of the studies were of good quality (Gandotra, 2011; Liu, 2021) with strengths in limiting possible bias through the systematic recruitment of relatives and taking into account the duration of contact with bipolar disorder probands (Gandotra et al, 2011). Ferensztajn et al. (2015) conducted a study of 'fair' quality, however, the greater TEMPS scores for affected relatives were not significant.

Cyclothymic temperament. Five of the studies compared cyclothymic temperament scores between affected relatives and unaffected relatives. Four studies found significantly higher cyclothymic TEMPS scores for affected relatives compared to unaffected relatives which were poor (Evans, 2005; Savitz, 2008), fair (Ferenztajn et al., 2015) and good (Liu, et al., 2021) quality. Gandotra et al. (2011) conducted a 'good' quality and found that the greater TEMPS scores for affected relatives, were not significant.

Anxious temperament. Three of the studies (Evans, 2005; Ferensztajn, 2015; Liu, 2021) compared anxious temperament scores between affected

relatives and unaffected relatives. All three studies observed higher anxious TEMPS scores within the affected relative group. However only one of the studies revealed significant differences between the two groups (Evans, 2005).

First-degree Relatives vs Healthy Controls

Thirteen studies in total compared temperament scores of first-degree relatives and healthy controls across the five subscales (Chiaroni, 2005; Evans, 2005; Ferreira, 2013; Greenwood, 2013; Higier 2014; Kesebir, 2005; Kesebir, 2020; Liu, 2021; Mahon, 2013; Mendlowicz, 2005; Saguem, 2021; Vázquez, 2008; Whalley, 2011).

Hyperthymic temperament. Ten studies compared hyperthymic temperament scores between first-degree relatives and healthy controls. Significant differences in hyperthymic TEMPS scores were observed in three studies (Evans, 2005; Ferreira, 2013; Greenwood; 2013). All three studies revealed higher hyperthymic temperament scores in healthy controls. However healthy controls were recruited through advertisements that specified that the research was concerned with psychiatric diagnoses which could have biased the sample. All three studies were considered poor quality due to limited attempts to control for confounding variables. Although the remaining seven studies did not find significant differences between the groups, there was a slight trend towards higher hyperthymic scores in healthy controls compared to first-degree relatives (Mahon 2013; Mendlowicz, 2005; Vázquez, 2008; Whalley, 2011).

Depressive temperament. Eleven studies compared depressive temperament scores between first-degree relatives and healthy controls. Four found significantly higher TEMPS scores amongst relatives compared to healthy

controls. The studies were generally of poor quality (Evans, 2005; Ferreira, 2013; Greenwood, 2013), however, Vázquez and colleagues (2008) conducted a study of slightly better quality (fair) study that included well-matched relative and control groups. The remaining studies found no significant differences between depressive scores although scores were generally higher amongst the relatives. The studies were poor (Mendlowicz et al., 2005) fair (Kesebir, 2020; Mahon, 2013; Whalley, 2011), good (Kesebir, 2005; Liu, 2021) and excellent (Saguem et al., 2021) quality.

Irritable temperament. Eleven studies compared irritable temperament scores between first-degree relatives and healthy controls. Eight studies observed significant differences in mean TEMPS scores between the two groups. Seven of the studies found significantly higher irritable TEMPS scores amongst first-degree relatives when compared to controls. The studies varied in quality and were rated as poor (Evans, 2005; Greenwood, 2013) fair (Ferreira, 2013; Mahon, 2013; Vázquez, 2008) and good (Liu, 2021; Mendlowicz, 2005). In contrast, Kesebir and colleagues (2005) found higher TEMPS scores for healthy controls compared to relatives and the study was 'fair' in quality. One study that was of excellent quality (Saguem et al., 2021) did not find any significant differences in irritable TEMPS scores between relatives and healthy controls. The study included well-matched groups and consideration of recent life events that could impact temperament scores.

Cyclothymic temperament. Twelve studies compared cyclothymic temperament between first-degree relatives and healthy controls. Eight studies observed significant differences in TEMPS scores between the groups. In seven of the studies, cyclothymic TEMPS scores were significantly higher for first-degree relatives compared to healthy controls (Chiaroni, 2005; Evans, 2005;

Ferreira, 2013; Greenwood, 2013; Mahon, 2013; Mendlowicz, 2005; Vázquez, 2008). In contrast, Liu et al. (2021) found significantly higher cyclothymic scores amongst healthy controls. Four studies did not observe any significant differences between relatives and healthy controls and were fair (Kesebir, 2005; Kesebir, 2020; Whalley, 2011) and excellent (Sauguem et al., 2021) quality.

Anxious temperament. Eleven studies compared anxious temperament scores between first-degree relatives and healthy controls. Seven studies reported significant differences between relatives and healthy controls on anxious TEMPS scores. Scores were significantly higher amongst relatives in six of the studies (Evans, 2005; Ferreira, 2013; Greenwood, 2013; Mahon, 2013; Mendlowicz, 2005; Vázquez, 2008), in contrast to Kesebir and colleagues (2005) who again observed higher scores amongst healthy controls. Three studies of fair (Vázquez, 2008), good (Liu, 2021) and excellent (Sauguem, 2021) quality did not observe any significant differences between groups.

Table 3 presents mean scores across temperament subtypes for each comparison group. Where studies did not provide mean scores, each group was given a rank to reflect the temperament continuum (highest temperament = 1 to lowest temperament scores = 5).

Table 3*Mean temperament scores for each group within the studies*

Study	Groups	Hyperthymic temperament		Depressive temperament		Irritable temperament		Cyclothymic temperament		Anxious temperament	
		M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences
Chiaroni (2005)	FDR	-	-	-	-	-	-	±15% [1]	FDR vs [HC and HC-P], $\chi^2 = 6.84$; $p = 0.017$	-	-
	HC-P							±0.81% [3]			
	HC							±1% [2]	HC vs. HC-P vs. FDR, $\chi^2 = 11.30$; $p = 0.001$		
Evans (2005)	BDI	9.22 (0.43[3])	$F = 5.8, p = 0.0007$	9.21 (0.43[1])	$F = 32.8, p < 0.00001$	6.75 (0.44[1])	$F = 59.6, p < 0.00001$	11.08 (0.58[1])	$F = 91.0, p < 0.00001$	11.95 (0.75[1])	$F = 53.4, p < 0.00001$
	BDII	8.79 (0.66[3])		9.78 (0.68[1])		7.22 (0.72[1])		11.10 (0.76[1])		11.43 (1.00[1])	
	AR (MDD)	[4]		[2]		[2]		[2]		[2]	
	UR	[2]		[3]		[3]		[3]		[3]	
	HC	[1]		[4]		[4]		[4]		[4]	
Ferensztajn (2015)	AR	0.34 (0.24)	$p < 0.05$ (Mann-Whitney U)	0.54 (0.18)	$p < 0.05$ (Mann-Whitney U)	0.38 (0.20)	-	0.56 (0.24)	-	0.53 (0.25)	$p < 0.05$ (Mann-Whitney U)
	UR (ELR)	0.51 (0.20)		0.32 (0.25)		0.24 (0.22)		0.31 (0.27)		0.32 (0.24)	
Ferreira (2013)	BD	[1]	$F = 10.41, p < 0.001$	[2]	$F = 72.65, p < 0.001$	[1]	$F = 24.70, p < 0.001$	[1]	$F = 33.08, p < 0.001$	[2]	$F = 42.02, p < 0.001$
	MDD-R	[4]		[1]		[2]		[2]		[1]	
	UR	[3]		[3]		[3]		[3]		[3]	
	HC	[2]		[4]		[4]		[4]		[4]	
Gandotra (2011)	AR	5.67 (3.91)	-	3.00 (2.45)	-	3.44 (2.46)	-	6.66 (3.54)	-	-	-
	UR	6.33 (3.74)		3.38 (2.22)		1.38 (1.50)		4.83 (3.12)			
	Sch-AR	6.33 (4.51)		2.33 (0.58)		0.67 (0.58)		5.33 (2.08)			
	Sch-UR	6.37 (3.85)		3.07 (1.53)		1.30 (1.32)		5.50 (2.27)			

Study	Groups	Hyperthymic temperament		Depressive temperament		Irritable temperament		Cyclothymic temperament		Anxious temperament	
		M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences	M (SD) / [rank]*	Group differences
Greenwood (2013)	BD	[3]	$F=10.0$ $p<0.001$	[1]	$F=37.6$ $p<0.001$	[1]	$F=55.0$ $p<0.001$	[1]	$F=97.3$ $p<0.001$	[1]	$F=58.0$ $p<0.001$
	MDD	[4]		[2]		[2]		[2]		[2]	
	UR	[2]		[3]		[3]		[3]		[3]	
	HC	[1]		[4]		[4]		[4]		[4]	
Higier (2014)		<i>"Positive scale"</i>	$(F=6.32,$ $df=4,$ $104,$ $p<0.01)$	<i>"Negative scale"</i>	$(F=12.71,$ $df=4, 104,$ $p<0.01)$						
	BD	[2]		[2]							
	UR	[1]		[4]							
	Sch	[5]		[1]							
	Sch-UR	[4]		[5]							
	HC	[3]		[3]							
Kesebir (2005)	BD	11.6 (1.1)	BD vs BD (c) $t=0.43$ $p=NS$	5.0 (0.52)	BD vs (c) $t=-2.72$ $p=0.007$	5.2 (0.8)	BD vs (c) $t=1.38$ $p<0.001$	11.2 (1.1)	BD vs (c) $t=2.2$ $p<0.001$	11.5 (0.7)	BD vs (c) $t=0.43$ $p=NS$
	FDR	12.2 (1.4)		5.2 (0.6)		3.6 (0.7)		7.9 (1.2)		5.8 (0.8)	
	BD(c)	11.9 (1.7)		6.2 (0.6)		4.8 (1.6)		6.7 (0.6)		12.8 (1.8)	
	FDR(c)	12.0 (2.7)		FDR vs FDR(c) $t=1.15$ $p=NS$		6.7 (1.2)		FDR vs (c) $t=-6.31$ $p<0.001$		4.0 (2.1)	
		BD vs FDR $t=-1.58$ $p=NS$		BD vs FDR $t=0.527$ $p=NS$		BD vs FDR $t=4.569$ $p<0.001$		BD vs FDR $t=6.33$ $p<0.001$		BD vs FDR $t=3.22$ $p<0.001$	
Kesebir (2020)	BD	17.1 (1.6)	-	16.5 (2.5)	-	19.1 (1.1)	-	15.9 (1.3)	-	19.6 (1.2)	-
	UR	17.4 (1.2)		11.1 (1.4)		16.4 (1.2)		14.8 (1.2)		14.7 (2.3)	
	HC	10.2 (2.3)		10.6 (1.8)		13.5 (2.7)		7.5 (2.4)		14.3 (1.9)	
Liu (2021)		§	$\chi^2=1.12$ $p=0.570$	§	$\chi^2=8.29$ $p=0.016$	§	$\chi^2=14.06$ $p=0.001$	§	$\chi^2=11.22$ $p=0.004$	§	$\chi^2=6.04$ $p=0.49$
	AR	46.24		50.22		54.66		51.95		51.70	
	UR	49.70		45.90		43.94		44.58		44.03	
	HC	49.17		45.90		42.77		48.26		44.03	

Study	Groups	Hyperthymic temperament		Depressive temperament		Irritable temperament		Cyclothymic temperament		Anxious temperament	
		M (SD) / [rank] _†	Group differences	M (SD) / [rank] _†	Group differences	M (SD) / [rank] _†	Group differences	M (SD) / [rank] _†	Group differences	M (SD) / [rank] _†	Group differences
Mahon (2013)	BD	14.80 (4.8)	$p=.714$	7.77 (4.5)	$p<0.001$ (Mann-Whitney <i>U</i>)	4.70 (3.4)	$p<0.001$ (Mann-Whitney <i>U</i>)	7.28 (3.8)	$p<0.001$ (Mann-Whitney <i>U</i>)	10.79 (5.8)	$p<0.001$ (Mann-Whitney <i>U</i>)
	UR	13.59 (5.3)		3.31 (3.6)		3.00 (2.5)		2.62 (2.5)		5.98 (4.32)	
	HC	13.88 (5.1)		2.00 (2.38)		1.75 (1.8)		1.88 (2.3)		3.68 (3.3)	
Mendlowicz (2005)	BD	3.48 (0.48)	$F=2.32, p=0.10$	1.05 (0.20)	$F=4.14, p=0.02$	1.61 (0.23)	$F=7.98, p=0.001$	3.66 (0.39)	$F=26.283, p=0.001$	0.80 (0.12)	$F=8.40, p=0.001$
	UR	3.86 (0.34)		0.42 (0.14)		0.78 (0.16)		1.41 (0.28)		0.61 (0.08)	
	HC	4.50 (0.24)		0.45 (0.10)		0.58 (0.11)		0.48 (0.19)		0.32 (0.06)	
Saguem (2021)	BD	10.63 (4.48)	Intergroup p value: $p < 10$ intragroup p value: $p = 0.003$ (BD & HC) $p < 10$ (UR & HC) $p = 0.425$ (BD & UR)	10.83 (4.11)	Intergroup p value: $p < 10$ intragroup p value: $p < 10$ (BD & HC) $p = 0.882$ (UR & HC) $p = 0.002$ (BD & UR)	5.63 (4.81)	Intergroup p value: $p = 0.136$ intragroup p value: $p = 0.307$ (BD & HC) $p = 0.896$ (UR & HC) $p = 0.137$ (BD & UR)	10.27 (5.69)	Intergroup p value: $p < 10$ intragroup p value: $p < 10$ (BD & HC) $p < 10$ (UR & HC) $p = 0.537$ (BD & UR)	9.68 (6.11)	Intergroup p value: $p = 0.201$ intragroup p value: $p = 0.195$ (BD & HC) $p = 0.891$ (UR & HC) $p = 0.410$ (BD & UR)
	UR	11.58 (3.64)		8.62 (3.17)		4.20 (3.70)		9.28 (4.57)		8.38 (5.21)	
	HC	8.13 (4.34)		8.30 (3.57)		4.53 (3.63)		5.48 (4.87)		7.92 (5.35)	
Savitz (2008)	BDI	[1]	$F=2.43, p<0.038$	-	-	[2]	$F=5.52, p<0.001$	[1]	$F=5.84, p<0.001$	-	-
	BDII	[2]		[1]		[2]					
	MDD-R	[5]		[3]		[3]					
	MDD-S	[4]		[4]		[4]					
	UR	[3]		[5]		[5]					
Vázquez (2008)	UR	8.97 (0.408)	NS	8.92 (0.376)	$p < 0.001$ (<i>t</i> -test)	4.91 (0.376)	$p < 0.01$ (<i>t</i> -test)	6.26 (0.448)	$p < 0.01$ (<i>t</i> -test)	9.25 (0.547)	$p < 0.001$ (<i>t</i> -test)
	HC	9.13 (0.409)		6.44 (0.306)		3.66 (0.284)		4.69 (0.378)		6.19 (0.416)	
Whalley (2011)	UR	¶ 2 (2)	$p = .44$ (Mann-Whitney <i>U</i>)	¶ 0 (1)	$p = .81$ (Mann-Whitney <i>U</i>)	¶ 1 (2)	$p = .20$ (Mann-Whitney <i>U</i>)	¶ 2 (4)	$p = .06$ (Mann-Whitney <i>U</i>)	¶ 0 (1)	$p = .49$ (Mann-Whitney <i>U</i>)
	HC	2.5 (3)		0 (2)		1(2)		0 (3)		1 (2)	

Note. AR: affected relative; BD-I: bipolar disorder type 1; BD-II: bipolar disorder type 2; BDR: bipolar disorder relative; ELR: excellent lithium responder; HC: healthy control; HC-P: healthy control positive family history of psychiatric illness; HC-R: relatives of healthy controls; MDD: major depressive disorder; MDD-R: major depressive disorder relatives; NA: not applicable; NS: not significant; SA-BD: schizoaffective bipolar disorder; Sch: schizophrenia; Sch-R: schizophrenia relative; UR: unaffected relative

†Groups ranked by mean score on temperament scale [1] = highest scores [4] = lowest scores (means scores unreported)

‡Percentage of subjects with temperament score above the cut-off indicating presence of temperament

§Presented as Z scores

¶Median score (interquartile range)

Discussion

Affective temperament is related to many of the processes associated with bipolar disorder (Youngstrom et al., 2011) and it has been proposed that bipolarity may lie along a continuum from extreme temperament to affective illness (Akiskal & Akiskal, 2005a). The TEMPS (Akiskal, 1998; Akiskal, 2005b; Akiskal, 2005c) has been used widely to assess the affective temperament of individuals with a bipolar disorder, their families and those without any psychiatric history (Elias et al., 2017). There is some evidence to suggest that affective temperament could constitute a genetic link between predisposing familial factors and affective illness (Akiskal et al., 2005a). However, it is unclear whether the continuum model of affective temperaments is reflected in those who are genetically at risk of developing a bipolar disorder. Therefore, this review aimed to examine whether there are differences in temperament between individuals with a bipolar disorder and their first-degree relatives.

Sixteen studies of varying quality were included in this review. The overall results of TEMPS affective scores across subjects with a bipolar disorder, first-degree relatives, and healthy controls provide partial support for a continuum model of temperament, spanning from lower scores in healthy controls through unaffected and affected first-degree relatives to bipolar disorder. However, the expected trend was not observed within the hyperthymic temperament domain, in which scores were typically higher for healthy controls followed by unaffected relatives, with lower scores for affected relatives and those with a bipolar disorder. These findings must be interpreted with caution as only a limited number of studies were available to draw conclusions and the methodological quality of most of the studies was only fair-

to-good. Whilst one of the included studies was rated as excellent quality, significant differences were rarely observed between the included groups.

The findings of the current review are consistent with previous research that has found partial support for hypotheses of a gradient in affective temperament across the bipolar disorder spectrum. Di Florio and colleagues (2010), analysed TEMPS-A scores across the bipolar disorder spectrum including BDI, BDII, MDD and healthy controls including 927 subjects. Although relatives were not included in the analysis, bipolar disorder groups generally scored higher than MDD and healthy control groups, except for the hyperthymic subscale in which healthy controls had the highest TEMPS scores. Similarly, Solmi and colleagues (2016), conducted a meta-analysis of TEMPS scores for individuals affected by mood disorders and other psychiatric conditions. The analysis found some evidence of a continuum trajectory across cyclothymic, irritable, and anxious temperament domains, and again observed a similar pattern of high hyperthymic scores for healthy controls.

It has been suggested that the hyperthymic subscale on the TEMPS-A, which was utilised by all studies included in the current review, could be measuring an 'idealised normality' rather than hyperthymia (Evans et al., 2005). Several items within the hyperthymic scale have potentially positive connotations for example "I am always on the go", "I have great confidence in myself" or "I'm the kind of person who believes everything will eventually turn out all right" which are limited to a "true" or "false" response. In several of the included studies, healthy controls were aware of the research topic when recruited. Thus, it is possible that items were more readily endorsed by those who perceive themselves as 'healthy' or 'positive'. Previous research has also suggested that item endorsement could be impacted by an individual's

“optimistic” or “pessimistic” state of mind. Studies that have utilised the interview version of the TEMPS found higher scores for BD subjects compared to controls (Evans et al., 2005) and within the current review, Saguem and colleagues (2021) found no significant difference between healthy controls and bipolar disorder subjects after administering the TEMPS, with slightly higher scores within the bipolar disorder group.

Further, methodological limitations highlighted in some of the included studies could account for inconsistencies in the observed results. Whilst several studies attempted to control for the phase of bipolar disorder, several studies reported that not all subjects were euthymic during participation or that remission could not be guaranteed for bipolar disorder patients. Patients with bipolar disorders typically spend predominantly more time experiencing depressive symptoms, Judd et al. (2002). As such hyperthymic scores could have been impacted by those with bipolar disorders experiencing mixed or depressive episodes which is more common than experiences of ‘pure’ hypomania or mania (Judd et al., 2005). In addition, previous research has highlighted a trend towards higher hyperthymic scores in BDI and lower in BDII (Solmi et al., 2016), however, the majority of studies in the current review were unable to distinguish between BDI and BDII participants for analyses due to limited sample size.

On the contrary, it has been hypothesised that we should expect hyperthymic temperament to be higher amongst controls and healthy relatives as this could have a protective role against the development of a mood disorder (Ferensztajn, 2015). In a national epidemiological study to determine the possible role of temperament in mental disorders, the hyperthymic temperament had a uniquely protective effect on most mental disorders, except separation

anxiety, bipolar disorders, substance abuse and impulse control disorders (Karam et al., 2010). Studies that highlighted no significant differences between healthy controls and individuals with a bipolar disorder would support this theory, inferring that the protective nature of hyperthymia is redundant when an individual has a predisposition towards the development of a bipolar disorder (Ferreira, 2013).

Nonetheless, the current findings support previous hypotheses that the relatives of individuals with a bipolar disorder exhibit more marked affective temperament traits than healthy control groups. The continuum of affective temperament features from bipolar disorders through relatives and to the wider population, support theories of a possible endophenotype (Di Florio, 2010; Solmi, 2016) and suggest that traits of affective temperament could represent a vulnerability marker for mood disorders (Ferreira, 2012). Based on the results of the review temperamental dysregulation can be present not only in individuals with a bipolar disorder but also in unaffected relatives. This would suggest that relatives of bipolar patients are at a higher risk for developing not only bipolar disorder but also temperamental dysregulation (Kelsoe, 2003).

Limitations

Only a small number of studies were eligible for inclusion within this review and very few were of 'excellent' quality. Several inconsistencies were highlighted between studies including sample size and definition of comparison groups making it difficult to reliably synthesise the findings of affective temperament across the bipolar spectrum. Several of the included studies grouped all bipolar disorder subjects despite differences in scores between BDI and BDII subjects. Studies generally included outpatients, which could lead to a

bias in the sample and limit the generalisability of findings. Further only half of the studies attempted to control for the current mood state of participants which can make it difficult to separate temperament and psychopathology or even scar effects from recent mood shifts (Chang et al., 2003).

Several studies failed to report the inclusion and exclusion criteria of first-degree relatives which made it difficult to determine the familial relationship of some of the participants in line with the aims of the review. Importantly whilst the current review provides support for theories of a possible endophenotype (Di Florio, 2010; Solmi, 2016), these studies alone do not evidence that bipolar disorders or temperament dysregulation are genetic. Family studies typically involve families that share both environment and genes and cannot control for non-genetic transmission (Kesloe, 2003). Several studies have found that specific genes implicated in bipolar disorders (COMT, BDNF and FKBP5) may interact with early life stressors and cannabis use (Misiak et al., 2018). Although some have argued that genetic factors could make individuals more prone to engaging in high-risk environments that could lead to substance use (Van Os et al., 2008). Thus, the development of bipolar disorders remains complicated and both genetic and environmental factors should be considered when exploring high heritability rates.

Differences in the methodologies of the included studies limited the scope of comparisons in the current review. For instance, some studies compared mean TEMPS scores, whilst others considered the prevalence of affective temperament based on pre-determined cut-off scores, some studies utilised z-scores and others calculated relative risk. In addition, some of the studies within the review employed different versions of the TEMPS. Whilst different versions of the TEMPS have demonstrated consistent reliability (Solmi

et al., 2016), some of the versions have been developed for use in specific cultures and have demonstrated culture-specific findings upon examination (Elias et al., 2017), which could limit the comparability of studies included in the review.

The review only included studies that investigated affective temperament using the TEMPS. Whilst the TEMPS is the most widely used instrument to assess affective temperament and has been validated across 15 countries (Elias et al. 2017), greater insight could be gained from combining different measures such as the TCI (Cloninger et al., 1993). As well as character dimensions, the TCI measures specific temperament dimensions such as novelty seeking, harm avoidance, reward dependence and persistence. Research has shown significant correlations between TCI domains and affective temperament categories. For example, novelty-seeking correlated positively with the cyclothymic temperaments, harm avoidance with anxious, depressive, and cyclothymic temperaments and negatively with the hyperthymic temperament. Reward dependence correlated with anxious temperament and persistence with the hyperthymic and irritable scales (Rózsa et al., 2008). However, such analyses were beyond the scope of the current review and could have exacerbated difficulties in reliably synthesising findings.

A meta-analysis was not conducted due to the heterogeneous nature of studies concerned with bipolar disorder probands, their first-degree relatives and controls groups, which would warrant several sub-group analyses (Borenstein et al., 2009). Whilst a meta-analysis would provide a quantitative investigation of a metric continuum of TEMPS scores (Solmi et al., 2016), high-quality studies, with the level of data required for such analyses, are not currently available.

Despite limitations, the review benefitted from a comprehensive, systematic search of three major databases. However, studies published in a language other than English were not included in the review, which means that the findings may only be generalisable to countries whose academics routinely publish in English. The grey literature was not searched within the current review, which could be deemed as a limitation due to a risk of bias. However, it has been suggested that whilst grey literature searches can be systematic, search results are typically dependent on several factors that change over time (Mahood et al., 2014). As such including grey literature can bias the results of the review and create difficulties with the replication of searches.

The addition of a second-rater to assess the quality of the research improves the reliability of the quality appraisal, reducing the risk of researcher bias. In the absence of published cut-off scores, the author created an arbitrary categorisation system to determine overall study quality. This was helpful in interpreting and contextualising the review's findings. However, deriving a total quality score, which gives equal weight to all items in the tool, is a contested issue and limits the utility of this study's quality appraisal (Moola et al., 2020). Finally, the review constitutes a novel attempt at synthesizing the available evidence on TEMPS scores between those at genetic risk of developing a bipolar disorder and bipolar disorder probands.

Clinical Implications

The findings of the current review broaden our understanding of the relationship between temperament and bipolar disorders. The findings evidence an increase in temperamental dysregulation amongst those at risk of bipolar disorders and bipolar disorder patients. The assessment of temperament in

those at risk or within prodromal phases of bipolar disorder could provide important insights into the clinical evolution of mood disorders including symptomatic expression, long-term course and response and adherence to treatment (Vázquez & Gonda, 2013). Cyclothymic temperament has been associated with poor treatment adherence (Fornaro et al., 2013), hyperthymic temperament with manic episodes as well as depressive temperament and lower hyperthymic scores with number of episodes (Henry et al., 1999). Furthermore, suicide attempts have been associated with significantly higher affective temperament scores (Rihmer et al., 2009), with 90% of non-violent suicide attempters presenting with a dominant affective temperament (depressive, cyclothymic, irritable or anxious) compared to just 21% of controls.

Screening for temperament within clinical settings could help clinicians more accurately assess treatment options (Vöhringer et al., 2012) as well as providing important information to support risk monitoring and management. Psychological formulation could also be enhanced by contemporary models of affective temperament, which inherently consider the evolutionary aspects of temperaments. Focusing on both liabilities and assets can help individuals to understand their predisposition and maximise positive attributes to overcome difficulties, in contrast to diagnostic models that often focus on behavioural patterns in their maladaptive forms (Akiskal, 1998; Akiskal, 2005a). Whilst the TEMPS has been adapted into a short version which could be more suitable for clinical settings (Akiskal et al. 2005b). The TEMPS currently lacks any validated cut-off or categorisation of abnormal temperament which would make it more applicable and clinically meaningful within routine practice (Vöhringer et al. 2012).

Future Research

Future research would benefit from larger samples including individuals from across the bipolar spectrum. Larger sample sizes with clearly defined groups would allow for comparisons between bipolar disorder types which could provide insight into the transition between mood states. Longitudinal follow-up would also allow us to better evaluate the impact of affective temperament on the development of bipolar disorders in those who are at genetic risk. Finally, adoption and twin studies would enable us to draw more reliable conclusions regarding the genetic inheritability of affective temperament and bipolar disorders. Alternatively, the development of alternative assessment tools or methods that are appropriate for use with younger relatives could limit contamination from environmental factors and more accurately assess genetically determined and inborn temperamental traits.

Conclusion

This is the first review to systematically evaluate the evidence reporting differences in affective temperament in first-degree relatives of bipolar disorder patients using the TEMPS. Generally, studies reported more marked or dominant depressive, irritable, cyclothymic, or anxious temperament amongst those with a bipolar disorder, which was less pronounced amongst relatives and to a less extent in healthy controls. The review provides partial support for theories of an affective temperament continuum from extreme temperament to affective illness, which suggests that bipolar disorders could have a genetic basis. This review supports the idea that genes that predispose individuals to bipolar disorders could produce a continuous variation of affective phenotypes that includes normal behaviour. However, further research that considers the

contamination of environmental factors and non-genetic transmission is required.

References

- Akiskal, H. S. (1998). Toward a definition of generalized anxiety disorder as an anxious temperament type. *Acta Psychiatrica Scandinavica*, *98*, 66–73. <https://doi.org/10.1111/j.1600-0447.1998.tb05969.x>
- Akiskal, H. S., Placidi, G. F., Maremmani, I., Signoretta, S., Liguori, A., Gervasi, R., Mallya, G., & Puzantian, V. R. (1998). TEMPS-I: Delineating the most discriminant traits of the cyclothymic, depressive, hyperthymic and irritable temperaments in a nonpatient population. *Journal of Affective Disorders*, *51*(1), 7–19. [https://doi.org/10.1016/S0165-0327\(98\)00152-9](https://doi.org/10.1016/S0165-0327(98)00152-9)
- Akiskal, H. S., Akiskal, K., Allilaire, J. F., Azorin, J. M., Bourgeois, M. L., Sechter, D., Fraud, J. P., Chatenêt-Duchêne, L., Lancrenon, S., Perugi, G., & Hantouche, E. G. (2005a). Validating affective temperaments in their subaffective and socially positive attributes: Psychometric, clinical and familial data from a French national study. *Journal of Affective Disorders*, *85*(1–2), 29–36. <https://doi.org/10.1016/j.jad.2003.12.009>
- Akiskal, H. S., & Akiskal, K. K. (2005b). TEMPS: Temperament Evaluation of Memphis, Pisa, Paris and San Diego. *Journal of Affective Disorders*, *85*(1–2), 1–2. <https://doi.org/10.1016/j.jad.2004.12.003>
- Akiskal, H. S., Akiskal, K. K., Haykal, R. F., Manning, J. S., & Connor, P. D. (2005c). TEMPS-A: progress towards validation of a self-rated clinical version of the Temperament Evaluation of the Memphis, Pisa, Paris, and San Diego Autoquestionnaire. *Journal of Affective Disorders*, *85*, 3–16. <https://doi.org/10.1016/j.jad.2004.12.001>
- Akiskal, H. S., Mendlowicz, M. V, Jean-Louis, G., Rapaport, M. H., Kelsoe, J. R., Gillin, J. C., & Smith, T. L. (2005b). TEMPS-A: Validation of a short version of a self-rated instrument designed to measure variations in temperament.

Journal of Affective Disorders, 85(1–2), 45–52.

<https://doi.org/10.1016/j.jad.2003.10.012>

Akiskal, K. K., & Akiskal, H. S. (2005a). The theoretical underpinnings of affective temperaments: Implications for evolutionary foundations of bipolar disorder and human nature. *Journal of Affective Disorders*, 85(1–2), 231–239.

<https://doi.org/10.1016/j.jad.2004.08.002>

Anderson, I. M., Haddad, P. M., & Scott, J. (2012). Bipolar disorder. In *BMJ (Clinical research ed.)* (Vol. 345). British Medical Journal Publishing Group.

<https://doi.org/10.1136/bmj.e8508>

Aromataris E, Munn Z (2020). *JBI Manual for Evidence Synthesis*. JBI.

<https://doi.org/10.46658/JBIMES-20-01>

Borenstein, M., Hedges, V. H., Higgins, J. P. T., & Rothenstein, H. R. (2009).

Introduction to Meta-Analysis. (J. W. & S. Ltd (Ed.s)). *Subgroup Analyses* (pp. 149-186). <https://doi.org/http://dx.doi.org/10.1002/9780470743386>

Carvalho, A. F., Soares, J. C., Maciel, B. R., Cavalcante, L. M., Vale, A. M. O., & Carvalho, A. F. (2017). Measuring affective temperaments: a systematic review of validation studies of the Temperament Evaluation in Memphis Pisa and San Diego (TEMPS) instruments. In *Journal of Affective Disorders* (Vol. 212, pp. 25–37). Elsevier B.V.

<https://doi.org/10.1016/j.jad.2017.01.023>

Chang, K., Steiner, H., & Ketter, T. (2003). Studies of Offspring of Parents with Bipolar Disorder. *American Journal of Medical Genetics - Seminars in Medical Genetics*, 123 C(1), 26–35. <https://doi.org/10.1002/ajmg.c.20011>

Chiaroni, P., Hantouche, E. G., Gouvernet, J., Azorin, J. M., & Akiskal, H. S. (2005). The cyclothymic temperament in healthy controls and familiarly at risk individuals for mood disorder: Endophenotype for genetic studies?

Journal of Affective Disorders, 85(1–2), 135–145.

<https://doi.org/10.1016/j.jad.2003.12.010>

Cloninger, C. R., Svrakic, D. M., & Przybeck, T. R. (1993). A Psychobiological Model of Temperament and Character. *Archives of General Psychiatry*, 50(12), 975–990. <https://doi.org/10.1001/archpsyc.1993.01820240059008>

De Fruyt, F., De Clercq, B. J., Van De Wiele, L., & Van Heeringen, K. (2006). The Validity of Cloninger's Psychobiological Model Versus the Five-Factor Model to Predict DSM-IV Personality Disorders in a Heterogeneous Psychiatric Sample: Domain Facet and Residualized Facet Descriptions. *Copyright the Authors Journal Compilation R*, 74(2).

<https://doi.org/10.1111/j.1467-6494.2006.00382.x>

De Fruyt, F., Van De Wiele, L., & Van Heeringen, C. (2000). Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Personality and Individual Differences*, 29(3), 441–452. [https://doi.org/10.1016/S0191-8869\(99\)00204-4](https://doi.org/10.1016/S0191-8869(99)00204-4)

Di Florio, A., Hamshere, M., Forty, L., Green, E. K., Grozeva, D., Jones, I., Caesar, S., Fraser, C., Gordon-Smith, K., Jones, L., Craddock, N., & Smith, D. J. (2010). Affective temperaments across the bipolar-unipolar spectrum: Examination of the TEMPS-A in 927 patients and controls. *Journal of Affective Disorders*, 123(1–3), 42–51.

<https://doi.org/10.1016/j.jad.2009.09.020>

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016).

Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open*, 6(12), e011458.

<https://doi.org/10.1136/bmjopen-2016-011458>

- Duffy, A., Vandeleur, C., Heffer, N., & Preisig, M. (2017). The clinical trajectory of emerging bipolar disorder among the high-risk offspring of bipolar parents: current understanding and future considerations. *International Journal of Bipolar Disorders*, 5(1). <https://doi.org/10.1186/s40345-017-0106-4>
- Evans, L., Akiskal, H. S., Keck, P. E., McElroy, S. L., Sadovnick, A. D., Remick, R. A., & Kelsoe, J. R. (2005). Familiality of temperament in bipolar disorder: Support for a genetic spectrum. *Journal of Affective Disorders*, 85(1–2), 153–168. <https://doi.org/10.1016/j.jad.2003.10.015>
- Ferensztajn, E., Skibinska, M., Kaczmarek, M., Losy, J., & Rybakowski, J. K. (2015). Neurobiology and temperament in the offspring of excellent lithium responders. *World Journal of Biological Psychiatry*, 16(4), 272–277. <https://doi.org/10.3109/15622975.2014.913810>
- Ferreira, A. De, Vasconcelos, A. G., Neves, F. S., Laks, J., & Correa, H. (2013). Affective temperaments: Familiality and clinical use in mood disorders. *Journal of Affective Disorders*, 148(1), 53–56. <https://doi.org/10.1016/j.jad.2012.11.047>
- Fornaro, M., De Berardis, D., Iasevoli, F., Pistorio, M. L., D'Angelo, E., Mungo, S., Martino, M., Ventriglio, A., Cattaneo, C. I., Favaretto, E., Del Debbio, A., Romano, A., Ciampa, G., Elassy, M., Perugi, G., & De Pasquale, C. (2013). Treatment adherence towards prescribed medications in bipolar-II acute depressed patients: Relationship with cyclothymic temperament and “therapeutic sensation seeking” in response towards subjective intolerance to pain. *Journal of Affective Disorders*, 151(2), 596–604. <https://doi.org/10.1016/j.jad.2013.07.004>

- Fountoulakis, K. N., & Gonda, X. (2019). Modeling human temperament and character on the basis of combined theoretical approaches. *Annals of General Psychiatry, 18*(1), 21. <https://doi.org/10.1186/s12991-019-0247-1>
- Gandotra, S., Ram, D., Kour, J., & Praharaj, S. K. (2011). Association between affective temperaments and bipolar spectrum disorders: Preliminary perspectives from a controlled family study. *Psychopathology, 44*(4), 216–224. <https://doi.org/10.1159/000322691>
- Gardner, R. (1982). Mechanisms in Manic-Depressive Disorder: An Evolutionary Model. *Archives of General Psychiatry, 39*(12), 1436–1441. <https://doi.org/10.1001/archpsyc.1982.04290120066013>
- Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., Hinde, R. A., & McCall, R. B. (1987). Roundtable: What Is Temperament? Four Approaches. *Child Development, 58*(2), 505. <https://doi.org/10.2307/1130527>
- Gonda, X., Rihmer, Z., Zsombok, T., Bagdy, G., Akiskal, K. K., & Akiskal, H. S. (2006). The 5HTTLPR polymorphism of the serotonin transporter gene is associated with affective temperaments as measured by TEMPS-A. *Journal of Affective Disorders, 91*(2–3), 125–131. <https://doi.org/10.1016/j.jad.2005.12.048>
- Gordovez, F. J. A., & McMahon, F. J. (2020). The genetics of bipolar disorder. *Molecular Psychiatry, 25*(3), 544–559. <https://doi.org/10.1038/s41380-019-0634-7>
- Greenwood, T. A., Akiskal, H. S., Akiskal, K. K., & Kelsoe, J. R. (2012). Genome-wide association study of temperament in bipolar disorder reveals significant associations with three novel loci. *Biological Psychiatry, 72*(4), 303–310. <https://doi.org/10.1016/j.biopsych.2012.01.018>

- Greenwood, T. A., Badner, J. A., Byerley, W., Keck, P. E., McElroy, S. L., Remick, R. A., Sadovnick, A. D., Akiskal, H. S., & Kelsoe, J. R. (2013). Heritability and genome-wide SNP linkage analysis of temperament in bipolar disorder. *Journal of Affective Disorders, 150*(3), 1031–1040.
<https://doi.org/10.1016/j.jad.2013.05.035>
- Grucza, R. A., & Goldberg, L. R. (2007). The comparative validity of 11 modern personality inventories: Predictions of behavioral acts, informant reports, and clinical indicators. *Journal of Personality Assessment, 89*(2), 167–187.
<https://doi.org/10.1080/00223890701468568>
- Henry, C., Lacoste, J., Bellivier, F., Verdoux, H., Bourgeois, M. L., & Leboyer, M. (1999). Temperament in bipolar illness: Impact on prognosis. *Journal of Affective Disorders, 56*(2–3), 103–108. [https://doi.org/10.1016/S0165-0327\(98\)00219-5](https://doi.org/10.1016/S0165-0327(98)00219-5)
- Higier, R. G., Jimenez, A. M., Hultman, C. M., Borg, J., Roman, C., Kizling, I., Larsson, H., & Cannon, T. D. (2014). Enhanced neurocognitive functioning and positive temperament in twins discordant for bipolar disorder. *American Journal of Psychiatry, 171*(11), 1191–1198.
<https://doi.org/10.1176/appi.ajp.2014.13121683>
- Judd, L. L., Akiskal, H. S., Schettler, P. J., Endicott, J., Leon, A. C., Solomon, D. A., Coryell, W., Maser, J. D., & Keller, M. B. (2005). Psychosocial disability in the course of bipolar I and II disorders: A prospective, comparative, longitudinal study. *Archives of General Psychiatry, 62*(12), 1322–1330.
<https://doi.org/10.1001/archpsyc.62.12.1322>
- Judd, L. L., Akiskal, H. S., Schettler, P. J., Endicott, J., Maser, J., Solomon, D. A., Leon, A. C., Rice, J. A., & Keller, M. B. (2002). The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Archives of*

General Psychiatry, 59(6), 530–537.

<https://doi.org/10.1001/archpsyc.59.6.530>

Kelsoe, J. R. (2003). Arguments for the genetic basis of the bipolar spectrum. In

Journal of Affective Disorders (Vol. 73, Issues 1–2, pp. 183–197).

[https://doi.org/10.1016/S0165-0327\(02\)00323-3](https://doi.org/10.1016/S0165-0327(02)00323-3)

Kesebir, S., Vahip, S., Akdeniz, F., Yüncü, Z., Alkan, M., & Akiskal, H. (2005).

Affective temperaments as measured by TEMPS-A in patients with bipolar I disorder and their first-degree relatives: A controlled study. *Journal of Affective Disorders*, 85(1–2), 127–133.

<https://doi.org/10.1016/j.jad.2003.10.013>

Kesebir, S., & Yosmaoğlu, A. (2020). QEEG - spectral power density of brain regions in predicting risk, resistance and resilience for bipolar disorder: A comparison of first degree relatives and unrelated healthy subjects.

Heliyon, 6(6). <https://doi.org/10.1016/j.heliyon.2020.e04100>

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass

correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155-163. <http://dx.doi.org/10.1016/j.jcm.2016.02.012>

Koukopoulos, A., Sani, G., Koukopoulos, A. E., Albert, M. J., Girardi, P., &

Tatarelli, R. (2006). Endogenous and exogenous cyclicity and temperament in bipolar disorder: Review, new data and hypotheses. *Journal of Affective Disorders*, 96(3), 165–175. <https://doi.org/10.1016/j.jad.2006.08.031>

Liu, T., Xu, G., Lu, W., Zhang, R., Chen, K., McIntyre, R. S., Teopiz, K. M., So, K. F., & Lin, K. (2021). Affective Temperament Traits Measured by TEMPS-A and Their Associations with Cognitive Functions among Offspring of Parents with Bipolar Disorder with and without Subthreshold Symptoms.

Journal of Affective Disorders, 283(January), 377–383.

<https://doi.org/10.1016/j.jad.2021.01.061>

Loftus, J., Etain, B., & Scott, J. (2016). What can we learn from offspring studies in bipolar disorder? *BJPsych Advances*, 22(3), 176–185.

<https://doi.org/10.1192/apt.bp.114.013086>

Loftus, J., Etain, B., & Scott, J. (2016). What can we learn from offspring studies in bipolar disorder? *BJPsych Advances*, 22(3), 176–185.

<https://doi.org/10.1192/apt.bp.114.013086>

Mahood, Q., Van Eerd, D., & Irvin, E. (2014). Searching for grey literature for systematic reviews: Challenges and benefits. *Research Synthesis Methods*, 5(3), 221–234. <https://doi.org/10.1002/jrsm.1106>

<https://doi.org/10.1002/jrsm.1106>

Mahon, K., Perez-Rodriguez, M. M., Gunawardane, N., & Burdick, K. E. (2013). Dimensional endophenotypes in bipolar disorder: Affective dysregulation and psychosis proneness. *Journal of Affective Disorders*, 151(2), 695–701.

<https://doi.org/10.1016/j.jad.2013.08.003>

Mendlowicz, M. V., Jean-Louis, G., Kelsoe, J. R., & Akiskal, H. S. (2005). A comparison of recovered bipolar patients, healthy relatives of bipolar probands, and normal controls using the short TEMPS-A. *Journal of Affective Disorders*, 85(1–2), 147–151.

<https://doi.org/10.1016/j.jad.2004.01.012>

Misiak, B., Stramecki, F., Gawęda, Ł., Prochwicz, K., Sąsiadek, M. M., Moustafa, A. A., & Frydecka, D. (2035). *Interactions Between Variation in Candidate Genes and Environmental Factors in the Etiology of Schizophrenia and Bipolar Disorder: a Systematic Review*. <https://doi.org/10.1007/s12035-017-0708-y>

- Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, Currie M, Qureshi R, Mattis P, Lisy K, Mu P-F (2020). In: Aromataris E, Munn Z (Ed.). *JBI Manual for Evidence Synthesis: Chapter 7: Systematic reviews of etiology and risk*. (pp.267-269). Available from: <https://synthesismanual.jbi.global>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), n71. <https://doi.org/10.1186/s13643-021-01626-4>
- Perugi, G., Toni, C., Traverso, M. C., & Akiskal, H. S. (2003). The role of cyclothymia in atypical depression: Toward a data-based reconceptualization of the borderline-bipolar II connection. *Journal of Affective Disorders*, 73(1–2), 87–98. [https://doi.org/10.1016/S0165-0327\(02\)00329-4](https://doi.org/10.1016/S0165-0327(02)00329-4)
- Pössl, J., & von Zerssen, D. (1990). A case history analysis of the “manic type” and the “melancholic type” of premorbid personality in affectively ill patients. *European Archives of Psychiatry and Neurological Sciences*, 239(6), 347–355. <https://doi.org/10.1007/BF01734540>
- Rasic, D., Hajek, T., Alda, M., & Uher, R. (2013). Risk of mental illness in offspring of parents with schizophrenia, bipolar disorder and major depressive disorder: a meta-analysis of family high-risk studies. *Comprehensive Psychiatry*, 54(8), e34. <https://doi.org/10.1016/j.comppsy.2013.07.057>

- Rózsa, S., Rihmer, Z., Gonda, X., Szili, I., Rihmer, A., Kő, N., Németh, A., Pestality, P., Bagdy, G., Alhassoon, O., Akiskal, K. K., & Akiskal, H. S. (2008). A study of affective temperaments in Hungary: Internal consistency and concurrent validity of the TEMPS-A against the TCI and NEO-PI-R. *Journal of Affective Disorders, 106*, 45–53.
<https://doi.org/10.1016/j.jad.2007.03.016>
- Rihmer, A., Rozsa, S., Rihmer, Z., Gonda, X., Akiskal, K. K., & Akiskal, H. S. (2009). Affective temperaments, as measured by TEMPS-A, among nonviolent suicide attempters. *Journal of Affective Disorders, 116*(1–2), 18–22. <https://doi.org/10.1016/j.jad.2008.10.024>
- Saguem, B. N., Mtiraoui, A., Nakhli, J., Mannai, J., Ben Salah, N., El Kissi, Y., & Ben Nasr, S. (2019). Affective temperaments and their relationships with life events in bipolar patients and siblings: a controlled study. *Journal of Mental Health, 30*(1), 36–42.
<https://doi.org/10.1080/09638237.2019.1608924>
- Savitz, J., van der Merwe, L., & Ramesar, R. (2008). Hypomanic, cyclothymic and hostile personality traits in bipolar spectrum illness: A family-based study. *Journal of Psychiatric Research, 42*(11), 920–929.
<https://doi.org/10.1016/j.jpsychires.2007.10.011>
- Solmi, M., Zaninotto, L., Toffanin, T., Veronese, N., Lin, K., Stubbs, B., Fornaro, M., & Correll, C. U. (2016). A comparative meta-analysis of TEMPS scores across mood disorder patients, their first-degree relatives, healthy controls, and other psychiatric disorders. *Journal of Affective Disorders, 196*, 32–46.
<https://doi.org/10.1016/j.jad.2016.02.013>
- Solmi, M., Zaninotto, L., Toffanin, T., Veronese, N., Lin, K., Stubbs, B., Fornaro, M., & Correll, C. U. (2016). A comparative meta-analysis of TEMPS scores

across mood disorder patients, their first-degree relatives, healthy controls, and other psychiatric disorders. *Journal of Affective Disorders*, 196, 32–46.

<https://doi.org/10.1016/j.jad.2016.02.013>

Van Os, J., Rutten, B. P. F., & Poulton, R. (2008). Gene-environment interactions in schizophrenia: Review of epidemiological findings and future directions. *Schizophrenia Bulletin*, 34(6), 1066–1082.
<https://doi.org/10.1093/SCHBUL/SBN117>

Vázquez, G. H., & Gonda, X. (2013). Affective Temperaments and Mood Disorders: A Review of Current Knowledge. *Current Psychiatry Reviews*, 9(1), 21–32. <https://doi.org/10.2174/1573400511309010004>

Vázquez, G. H., Kahn, C., Schiavo, C. E., Goldchluk, A., Herbst, L., Piccione, M., Saidman, N., Ruggeri, H., Silva, A., Leal, J., Bonetto, G. G., Zaratiegui, R., Padilla, E., Vilapriño, J. J., Calvo, M., Guerrero, G., Strejilevich, S. A., Cetkovich-Bakmas, M. G., Akiskal, K. K., & Akiskal, H. S. (2008). Bipolar disorders and affective temperaments: A national family study testing the “endophenotype” and “subaffective” theses using the TEMPS-A Buenos Aires. *Journal of Affective Disorders*, 108(1–2), 25–32.
<https://doi.org/10.1016/j.jad.2007.09.011>

Vöhringer, P. A., Whitham, E. A., Thommi, S. B., Holtzman, N. S., Khrad, H., & Ghaemi, S. N. (2012). Affective temperaments in clinical practice: A validation study in mood disorders. *Journal of Affective Disorders*, 136(3), 577–580. <https://doi.org/10.1016/j.jad.2011.10.028>

Whalley, H. C., Sussmann, J. E., Chakirova, G., Mukerjee, P., Peel, A., McKirdy, J., Hall, J., Johnstone, E. C., Lawrie, S. M., & McIntosh, A. M. (2011). The neural basis of familial risk and temperamental variation in individuals at

high risk of bipolar disorder. *Biological Psychiatry*, 70(4), 343–349.

<https://doi.org/10.1016/j.biopsych.2011.04.007>

Youngstrom, E. A., Van Meter, A., & Freeman, A. J. (2011). Bipolar Disorder. In

Encyclopedia of Adolescence (Vol. 3, pp. 76–84). Elsevier Inc.

<https://doi.org/10.1016/B978-0-12-373951-3.00103-4>

Appendix A – The TEMPS

REMOVED DUE TO COPYRIGHT RESTRICTIONS

Appendix B – Search terms

mania OR hypomania OR "bipolar disorder" OR "bipolar spectrum disorder*" OR bipolar OR "manic-depress*"

AND

"genetic high-risk" OR "famil* risk" OR "famil* high-risk" OR relative OR "1ST degree relative*" OR "first degree relative*" OR "first-degree relative*" OR parent* OR mother OR father OR child* OR toddler OR infant OR offspring OR off-spring

AND

"affective temperament*" OR temperament* OR "TEMPS*" or "The Temperament Evaluation of Memphis Pisa and San Diego"

Appendix C – JBI Critical Appraisal checklist for Cross-sectional Studies

REMOVED DUE TO COPYRIGHT RESTRICTIONS

Appendix D – Quality appraisal table

Table 1a (Appendix)

Outcomes for the JBI Critical Appraisal Checklist for Cross-Sectional Studies

Study	Item number								Quality rating (%)	Comments
	1	2	3	4	5	6	7	8		
Chiaroni et al., (2005)	1	1	1	2	1	2	0	2	62%	<p>Strengths: Interview by qualified and experienced psychiatrist to determine lifetime diagnoses of controls and BD diagnoses for BD group. Blind assessment of participants. Relatives and controls were matched.</p> <p>Weaknesses: Inclusion and exclusion criteria limited in detail. Unclear what the familial relationship was. Limited information provided on proband diagnoses and no TEMPS data collected for probands. Used self-report TEMPS rather than interview.</p>
Evans et al., (2005)	0	1	1	1	0	1	1	2	43%	<p>Strengths: Participants assessed by direct standardised interview by trained professionals as well as review of medical records.</p> <p>Weaknesses: No description of inclusion or exclusion criteria. No clear consideration of confounding factors. Controls recruited through advertisement. No discussion of stage of course of illness for BD group. No reporting on familial relationship. Despite different scores BD types were grouped due to limited number of participants.</p>
Ferensztajn et al., (2015)	1	1	2	2	1	0	1	2	62%	<p>Strengths: All participants were assessed using standardised clinical interviews by a psychiatrist and checked by a second psychiatrist when there was evidence of any disorder.</p>

										Weaknesses: No information was provided about exclusion criteria, limited description of BD group including stage or course of disorder. Sample limited in size and unrepresentative given categorisation by medication (lithium response).
Ferreira et al., (2012)	1	1	1	1	0	2	1	2	56%	Strengths: All participants interviewed by trained psychiatrist using standardised interview procedures as well as interviews with family members and medical record review. Weaknesses: Limited information provided regarding inclusion and exclusion criteria as well as limited concern for confounding variables. Did not describe familial relationship. Unclear status for BD group. Controls recruited via advertisement and were not blind to the topics of the research i.e. psychiatric disorders. BDI and BDII group although differences in scores were observed.
Gandotra et al., (2011)	2	1	2	1	1	2	1	2	75%	Strengths: Clear inclusion criteria and predetermined criteria for the selection of relatives. Groups were well matched and duration of contact with probands considered. Stage of illness for BD group reported. Weaknesses: Small sample size. Unclear how diagnoses were verified. Controls only screened by general health with no assessment of psychiatric history.
Greenwood et al., (2013)	1	1	2	0	0	1	1	2	50%	Strengths: All participants interview and diagnosed using standardised procedures by extensively trained clinicians. Reliability of diagnoses was tested, and medical records were consulted. Full version of TEMPS employed.

											Weaknesses: Control group not matched. No description of exclusion criteria or consideration of confounding factors. No description or assessment of stage of BD. Self-reported TEMPS.
Higier et al., (2014)	2	1	2	2	1	2	0	2	75%		Strengths: Clear inclusion and exclusion criteria. Reporting of possible confounding factors and strategies to limit bias. Inclusion of twin pairs. Weaknesses: TEMPS adapted which could limit validity and reliability. Unclear how the new measure was administered.
Kesebir et al., (2005)	2	1	1	2	1	2	1	2	68%		Strengths: Stage of BD group reported. Clear inclusion criteria. Trained psychiatrist verified diagnoses using standardised clinical interview. Controls well matched. Included those with BD that were in remission to control for effects of BD on temperament scores. Weaknesses: Familial relationship not reported. Some first-degree relatives not contacted as requested by BD participants. Unclear how the TEMPS was administered.
Kesebir et al., (2020)	1	1	1	1	1	2	1	2	62%		Strengths: BD group all in euthymic period. Controls matched by age and gender. Weaknesses: Unclear how diagnoses of participants were assessed. Unclear exclusion and inclusion criteria. No indication of how the TEMPS was administered. Relatively small sample size.
Liu et al., (2021)	2	2	2	1	1	1	1	2	75%		Strengths: BD assessed by psychiatrists using standardised clinical assessment, interrater reliability assessed. Clear inclusion and exclusion criteria. Describes relationship of first-degree relatives. Categorises relatives by affected and unaffected.

											Weaknesses: Small sample size. Alternative version of the TEMPS employed. Measures not completed by BD group.
Mahon et al., (2013)	1	1	2	1	2	1	1	1	62%		Strengths: Diagnoses verified by psychologist and psychiatrists and reviewed using standardised assessment methods. Control groups assessed for presence of psychiatric disorders. Familial relationship pre-determined and described. Weaknesses: Self-report TEMPS. Small sample size and risk of type II error. Study would benefit from more clarity around inclusion and exclusion. BDI heavy sample although paper refers to BD spectrum.
Mendlowicz et al., (2005)	1	1	0	1	1	2	1	2	56%		Strengths: Clinically well BD group which was operationalised and measured. Attempt to control for state effect. All participants interviewed by trained professionals and interrater reliability assessed. Weaknesses: Small sample which mean all BD types grouped. Self-report TEMPS employed. Unclear relation of the first-degree relatives.
Saguem et al., (2021)	2	1	2	2	1	2	2	2	87%		Strengths: Clear inclusion criteria and remission well operationalised. Groups well matched. Participants assessed by psychiatrist to confirm diagnoses. Full TEMPS employed and administered via interview. Attempts to control for mood state – participants not included until remission was achieved. Weaknesses: Small sample which mean all BD types grouped for analyses.

Savitz et al., (2008)	1	1	0	1	1	2	1	2	56%	<p>Strengths: Assessment of euthymia which was well operationalised. Analysis attempted to control for confounding factors.</p> <p>Weaknesses: No healthy unrelated controls. High rate study refusal which could indicate a bias in selected sample. Self-report TEMPS employed. Unclear inclusion and exclusion criteria. Not all participants included were euthymic which is unclear in the method and results section.</p>
Vázquez et al., (2008)	1	1	2	1	1	1	1	2	62%	<p>Strengths: Reported on relationship to proband. Full version of TEMPS utilised. Groups were matched.</p> <p>Weaknesses: Self-report measures employed. Inclusion criteria somewhat unclear. Did not include BD group in study.</p>
Whalley et al., (2011)	2	1	1	1	1	2	1	2	68%	<p>Strengths: Clear operationalised definition of unaffected relative. Clear inclusion and exclusion criteria. Groups well matched. All participants interviewed by trained psychiatrist to determine diagnoses.</p> <p>Weaknesses: Self-report TEMPS measure employed. BD group not included in study. Family members identified by BD group which could lead to a risk of bias. Second degree relatives included as well as first degree relatives, but groups not separated.</p>

Note. 2: Yes
1: Partially/unclear
0: No

Section II: Empirical Study

A Network Analysis of Bipolar Disorders and Common Psychiatric Diagnoses

Abstract

Objectives

The research aimed to explore mania as a network of its symptoms to gain insight into the structure of bipolar disorders within a nationally representative sample. A secondary aim was to construct a network of manic symptoms and common psychiatric diagnoses, to explore comorbidity and identify possible pathways to bipolar disorders from other psychopathological states.

Methods

The study utilised a retrospective cross-sectional analysis of data drawn from the Adult Psychiatric Morbidity Survey 2014. Life-time experience of mood symptoms assessed via questionnaire in 7076 participants was used to estimate networks. An additional network consisting of mood disorder symptoms and other psychiatric diagnoses was estimated and community detection was employed. Bootstrapping analyses were used to determine the reliability of network parameters.

Results

Increased activity and increased energy were the most central symptoms within the network (most highly correlated with other symptoms). The introduction of other psychiatric diagnoses caused a shift in the network, in which racing thoughts became the most central node in the network. Four communities were revealed 1) irritability, distractibility and racing thoughts with depression and anxiety disorders; 2) risky behaviour and troublesome money spending with alcoholism; 3) a mania only community including increased energy, activity, sociability, speech and hyperactivity; and 4) an 'other diagnoses' community. Racing thoughts was highlighted as a possible bridge between communities.

Conclusions

Symptoms of activity, energy and racing thoughts appear to be core features of bipolar disorders. Racing thoughts could be a risk factor for convergence to bipolar disorders or a comorbidity pathway.

Practitioner Points:

- Symptoms of energy and activity should be a key consideration in the assessment of bipolar disorders.
- The identification of racing thoughts may indicate the onset of a hypomanic/manic episode and should be a target for assessment and intervention.

Keywords: Bipolar disorders, mania, comorbidity, network analysis, centrality

Introduction

Bipolar disorders have been characterised by two apparently contrasting phases: recurring episodes of depressive symptoms (feeling intensely low) and prominent manic (feeling extremely happy and becoming overactive) or hypomanic (not as extreme as mania) symptoms. Despite the traditional view that these episodes are discrete and opposite presentations of extreme mood, research has highlighted that depressive and manic symptoms can co-occur as 'mixed states' (Marneros, 2001, Swann et al., 2007) and, that, longitudinally the two types of mood symptoms fluctuate relatively independently with a small, positive correlation between the two (Johnson et al., 2011). It has been suggested that the presence of mixed states may be indicative of a more severe presentation of illness, with implications for poor treatment outcomes and a severely recurrent and complicated course (Swann et al., 2013). However, accounts of mixed states continue to be debated, with ongoing disagreements regarding diagnostic definitions (Berk, 2005; Koukopoulos, 2013; Parker, 2019).

Current classification systems separate bipolar disorders into four major subtypes based on the pattern and frequency of depression and mania. According to the Diagnostic and Statistical Manual of Mental Disorders [DSM-5], (2013) bipolar disorders include bipolar I, bipolar II, cyclothymia and bipolar disorder not otherwise specified (American Psychiatry Association, 2013). Bipolar I is characterised by episodes of depression and at least one episode of mania or a mixed episode. Bipolar II is classified as several protracted depressive episodes and at least one hypomanic episode. Cyclothymic disorder consists of several periods of hypomanic and depressive symptoms, in which depressive symptoms do not meet the criteria for depressive episodes. Finally, for a diagnosis of bipolar disorder not otherwise specified, depressive and

hypomanic-like symptoms and episodes must be present and may alternate rapidly but cannot meet the full diagnostic criteria for any of the other classifications.

Clinically, bipolar disorders have proven difficult to diagnose taking an average of 8.8 years for bipolar disorder to be correctly identified (Baldessarini et al., 2007). Clinical studies have highlighted difficulties in differentiating bipolar I and bipolar II subtypes from each other (Judd, 2003; Mantere, 2008) and diagnostic manuals have historically failed to discriminate between the two (Dios, 2014; World Health Organisation, 1993). Diagnosis of bipolar disorders is further complicated by difficulties in differentiating bipolar disorders from unipolar depression. Studies show that clinicians generally have a greater awareness of symptoms of major depression than mania, although clinicians' diagnostic skills may also be hindered by patients' lack of insight which is more impaired in mania than in depression (Ghaemi et al., 2002). In addition, patients with bipolar disorders spend predominantly more time experiencing depressive symptoms which are subjectively more distressing than symptoms of mania (Judd et al., 2002). Patients are therefore more likely to seek treatment when experiencing symptoms of depression.

Bipolar disorder is thought to be more severe than unipolar depression given the rates of comorbidity with other psychiatric disorders, particularly anxiety disorders and substance use disorders (Merikangas et al., 2011). Several studies have highlighted the importance of early detection and intervention in the hope of enhancing the likelihood of positive long-term outcomes (Kessing et al., 2021). However, misdiagnosis does not only delay access to appropriate treatment but leads to unsuitable treatments such as antidepressants which can worsen the course of illness in the long term

(Ghaemi et al., 2002) and trigger mixed states (Parker & Ricciardi, 2019). Some also argue the risk of misdiagnosis has been exacerbated by the availability and accessibility of antidepressants (Ghaemi et al., 2002) which means that symptoms of depression are more likely to be recognised, diagnosed and treated as a unipolar presentation. Misdiagnosis also has significant economic consequences (Stensland et al., 2010), which can be considerably reduced by early detection and treatment (Mccombs et al., 2006).

The difficulties associated with detecting and diagnosing bipolar disorders can partly be attributed to the lack of agreement about how they are defined (Angst, 2007; Ghaemi 2002) and the lack of better validated diagnostic criteria (Kessing et al., 2021). This problem applies particularly to mania, and whilst traditional methods of analysing psychopathology such as factor analyses have had some success in demonstrating its multidimensional nature (Martino et al., 2020) conventional epidemiological and statistical methods may be incapable of capturing the heterogeneous nature of manic symptoms or their relationship with other symptoms of psychopathology. Network models provide an alternative approach to understanding the structure of psychopathology and identifying putative causal processes that lead to illness (Borsboom & Cramer, 2013). According to this approach, “mental disorders” (clusters of symptoms) occur, not because the symptoms have common underlying causes, but because symptoms are connected in a network of causal relationships.

To model these interactions, network analyses examine associations between symptoms (termed “nodes”) connected by “edges”, which are visually depicted in a “network”. It is proposed that an episode of a disorder occurs whenever the necessary number of symptoms become activated for a sufficient duration (McNally, 2016), with the triggering of one symptom leading to a

cascade of others. Thus, if a symptom (e.g., low mood) has many connections within a system, it may trigger the onset of its connected symptoms; the number of connections of a symptom is known as degree centrality (Fried & Cramer, 2017). Consequently, high centrality may be a risk factor for developing further symptoms (Borsboom & Cramer, 2013) and may play an important role in maintaining disorders (Borsboom, 2017). Although mania is a network of symptoms, almost by definition, network analysis has seldom been applied to bipolar disorders.

Koenders and colleagues (2015) applied a network approach to a sample of 125 patients with diagnoses of bipolar disorder. Participants were separated into three longitudinal courses (minimally impaired, depressed, and cycling). The results showed that symptom networks significantly differed between the three groups, with the most strongly interconnected symptoms associated with the more severe courses of bipolar disorder. Importantly, in this study 'manic' and 'depressive' symptoms did not form isolated clusters; rather both poles were interconnected. However, networks were not presented for the overall sample, the groups were pre-determined and the sample size was small (see methods section for a discussion of this issue). Voigt et al., (2018) also applied an individual network approach to longitudinal data from a single person diagnosed with bipolar disorder. Despite limitations the case study illustrated that mood fluctuated strongly over time and that symptoms of depression were present and central in both the depressive and hypomanic period.

In a recent study by Weintraub and colleagues (2020) network analyses were used to explore the development of bipolar disorders in 272 adolescents who had diagnoses of bipolar I and bipolar II or were considered genetically high risk for developing a bipolar disorder due to family history. As expected,

symptoms were most interrelated with the symptoms of the same mood pole, with mood lability and irritability found to be “bridge” symptoms that connected the two mood poles. Community analysis (which groups symptoms in a manner analogous to factor analysis) revealed four communities of symptoms within the network; a solely depressive symptom community and three communities of mania symptoms. Symptoms of activity and depressed mood were found to be most central within bipolar networks, which supports the growing literature highlighting increased energy and activity as a key construct within mania (Martino, 2020; Scott, 2017).

As highlighted by recent research (Koender, 2015; Voigt, 2018; Weintraub, 2020), network models provide a novel approach to analysing mood shifts in bipolar disorder. Initial findings support research that challenges the current categorical approach to psychopathology by demonstrating the interconnectedness of symptoms on both mood poles. Network models also show promise in being able to explore pathways through the bipolar spectrum as well as mechanisms that may predict bipolar course (Koender, 2015; Weintraub, 2020). There is also some indication that networks may be able to help in identifying warning signs of emergent episodes (Weintraub et al., 2020). With an increased understanding of the interaction between symptoms of bipolar disorder, we will likely be able to develop a better understanding of the development and maintenance of bipolar disorders.

A limitation of the existent studies is that they were with clinical samples, which creates two risks. First, given that a substantial proportion of the population experiences subclinical bipolar symptoms (Akiskal, 2002), network studies based on clinical samples may underestimate the range and variability of symptoms. Second, statistical studies show that selecting samples based on

clinical criteria (symptom severity) creates a risk of Berkson's bias and the misestimation of symptom-symptom associations, sometimes leading to spurious negative associations and the failure to recover the true network structure (De Ron et al. 2019). The purpose of the current study is, therefore, to investigate the structure of bipolar disorders within a nationally representative sample using data from the 2014 Adult Psychiatric Morbidity Survey (APMS; McManus et al., 2016) which, in contrast to previous waves of the survey, used the Mood Disorders Questionnaire (MDQ; Hirschfeld et al., 2000) to measure lifetime manic symptoms. Additionally, due to the significant rates of comorbidity with other psychiatric disorders in bipolar disorder and the significant overlap of symptoms (Goekoop & Goekoop, 2014), the study aims to explore links and pathways to other common psychiatric diagnoses. Given that network approaches are exploratory and data-driven, any proposed hypotheses are tentative.

Aims and Hypothesis

The primary aim of this study is to construct a network of manic symptoms in a community sample, to explore the structure of bipolar disorder. Based on previous research (Scott et al., 2017) we hypothesize that energy symptoms will be most central within the bipolar disorder network. A further aim is to construct a network that includes both manic symptoms and also common psychiatric diagnoses, to explore comorbidity between manic and other symptoms and hence identify possible pathways to bipolar disorders from other psychopathological states. Based on previous research we hypothesize that major depression will be most strongly connected to the manic network (reflecting the co-occurrence of depression and mania in historic and current

definitions of bipolar disorders) and that irritability will be a bridge or pathway between major depression and mania.

Method

Design

The study was a retrospective cross-sectional analysis of data drawn from the APMS (McManus et al., 2016).

Sample

The APMS is a stratified, multi-stage probability sample of the general population in England. The APMS includes assessment of common mental disorders, psychosis, autism, substance misuse, suicidal thoughts, attempts and self-harm, as well as screening for bipolar disorders and personality disorders. The survey is designed to be representative of the whole population and covers adults aged over 16 years old living in private households. The 2014 APMS sample contains 7546 individuals who had been randomly selected from postal areas stratified according to socioeconomic variables. The response rate was 57%. Data are weighted to account for non-response and selection probability.

Due to missing item-level data, 470 participants were removed prior to analysis resulting in a final sample size of $N = 7076$ participants with complete data on the measure of interest. Overall, 2% of the sample screened positive for bipolar disorder ($n = 130$). This is comparable to the World Mental Health Survey which identified a rate of 2.4% across 11 other countries (Marwaha et al, 2016). A summary of demographic variables are presented in Table 1.

There is currently no comprehensive guidance on the issue of sample size and statistical power within the network analysis field, specifically, how

many participants are required to reliably model the association between nodes in a cross-sectional network model (Fried and Cramer, 2017). It has been suggested that due to the complex structure of network models, “sample size” is not directly applicable within network studies (Stadtfeld et al., 2018). As such it remains that there is no clear consensus regarding the minimum number of participants per parameter needed to generate stable networks. Despite this, recent network analyses of bipolar disorder mood symptoms have reported that a network can be considered adequately powered when there are at least five participants per node (Weintraub et al., 2020), which would suggest the current sample size is satisfactory.

Nonetheless, a network estimation methodology ‘*eLasso*’, which includes regularisation and model selection criterion will be applied when estimating networks and is discussed in the analysis section below (van Borkulo et al., 2014). Validation studies show that the ‘*eLasso*’ method provides adequate network recovery in samples as small as 100 observations in which the most important connections were correctly identified and there was a low false-positive rate (van Borkulo et al., 2014).

Ethics

Ethical approval was granted by the University of Sheffield’s Department of Psychology research ethics committee (see Appendix A). For the original study, ethical approval was obtained from the West London National Research Ethics Committee (Reference 14/LO/0411) and all participants gave informed consent. A copy of the anonymised 2014 APMS dataset was made available for this research project. The dataset was password protected and stored on a university computer which was only accessible to the lead researcher. Anonymity was maintained throughout the study, and individual cases were not

identifiable to the researchers.

Measures

The MDQ (Hirschfeld et al., 2000; appendix B), is a 15-item, self-report scale based on DSM-IV criteria and clinical experience designed to screen for bipolar spectrum disorders (i.e. bipolar I, bipolar II, cyclothymia and bipolar not otherwise specified). The scale assesses lifetime experience of hypomanic and manic symptoms across 13 yes/no response items. Positive screening for bipolar disorder is dependent on endorsement of at least seven items of hypomanic/manic symptoms, together with affirmative answers regarding symptoms occurring at the same time and moderate to severe functional impairment queried on a 4-point scale (*no problem, minor problem, moderate problem or serious problem*).

The MDQ was developed and validated using a psychiatric outpatient sample. The MDQ has a sensitivity in the general population of 0.28 and high specificity of 0.97 meaning it is likely to screen out nearly all true negatives (Hirschfeld et al., 2003). The MDQ has been found to correctly identify seven out of 10 people with bipolar disorder, and successfully screen out nine out of 10 people that did not (APA, 2013 as cited in Marwaha Sal & Bebbington, 2016).

The lifetime experience of common mental disorders was assessed as part of the APMS 2014 self-report survey. Participants were presented with a show card listing different psychiatric diagnoses and were asked which they had experienced at some point in their life and whether this had been diagnosed by a professional. The self-diagnosis of depression, post-natal depression, generalised anxiety, panic attacks, phobia, post-traumatic stress disorder

(PTSD), alcohol or drug dependence, psychosis, obsessive-compulsive disorder (OCD) and attention deficit hyperactivity disorder (ADHD), were included in the current research.

Analysis

Data cleaning and preparation

The R Project for Statistical Computing was used for all analyses using RStudio. Initial data cleaning and preparation involved importing relevant data and creating a data matrix. To ensure the data matrix was positive definite and that variables were not linear combinations of each other the eigenvalue computation within the R package '*corpcor*' was applied. Subsequently, we examined the variables for any highly inter-correlated ($r > 0.50$) nodes using the '*goldbricker*' function of the R package '*networktools*' (Jones, 2018). The '*goldbricker*' function is a data-driven method used to determine whether any of the nodes in the psychometric network may be measuring the same underlying construct (i.e. are colinear). No redundant variables were identified.

Network estimation

Networks were estimated using the R Package '*IsingFit*' (van Borkulo et al., 2014) which is based on the Ising model. In this model, variables interact with their nearest neighbours and can be in either of two states (i.e. -1 or +1). The model assumes that the 'activation' of one variable is dependent on the activation of its neighbouring variables. The model contains specific parameters and within the *IsingFit* procedure, the model estimates these parameters with logistic regressions (one variable is regressed on all others in turn). The *IsingFit* package implements the Ising model alongside a regularised estimation method

(EBIC: Extended Bayesian Information Criterion) which is referred to as the ‘*eLasso*’ methodology.

The package produces undirected edges between nodes that are estimated by fitting a regularised logistic regression model to each node. Spurious effects are then weighted down to zero, removing potential false positives. This results in conservative networks of only the strongest edges, which in turn are more interpretable. The EBIC model subsequently selects the best neighbourhood sets according to the pre-selected gamma which controls the trade-off between including false-positive edges and removing true edges. The gamma coefficient is typically set between 0 and 0.5, with larger values applying more stringent penalization. Thus, a gamma of 0 typically produces a model with more edges, whilst a gamma of 0.5 produces a model with fewer edges. For the current research a gamma of 0.5 was set, to minimize the likelihood of false edges (van Borkulo et al., 2014).

As logistic regression models are applied to each node and are undirected, it is possible that an edge can be estimated in a neighbourhood set (i.e. node A predicting node B) whilst the reciprocal edge (node B to node A) remains absent. However, by applying the ‘*AND*’ rule which specifies that an edge should only be included if it is present in both neighbour sets, the model produces undirected edges that are the mean values of the two logistic regression coefficients (i.e. node A predicting node B, and node B predicting node A); this rule was applied in the current study. The Ising networks were graphically illustrated using the ‘*qgraph*’ R package (Epskamp et al., 2012) which provides information about the relative importance of each node in the network and automatically implements the Fruchterman–Reingold (“spring”) algorithm which computes the optimal layout of the network to place highly

connected nodes closer to each other. Green edges indicate positive interactions, whilst red edges indicate negative relationships. The colour saturation and the thickness of the edges correspond to the absolute weight and scale relative to the strongest weight in the graph (Epskamp et al., 2012).

Centrality Analysis

The '*qgraph*' package was also used to calculate centrality indices which provide information on the relative importance of each node in a network (Epskamp et al, 2017). Several types of centrality indices exist; strength, betweenness and closeness were estimated in the current study. Strength is determined by summing the standardised weights of all significant edges in the network; a node that is high in strength can directly influence other nodes when activated. Betweenness sums the number of times each node lies on the shortest path between two other nodes; nodes that are high in betweenness are important for transmitting effects between other nodes in the network. Closeness is calculated by taking the inverse of the sum of distances of the node of interest from all other nodes in the network; high closeness means a node is likely to be quickly affected by changes in other nodes in the network.

Network accuracy

To assess the accuracy of the centrality outputs and determine whether the ordering of values is reliable, bootstrapping analyses were computed using the '*Bootnet*' package in R following methods outlined by Epskamp and colleagues (2017). The stability of centrality indices is computed using the correlation stability coefficient (CS-coefficient). The CS-coefficient measures the maximum number of cases (participants in the data set) that can be dropped whilst maintaining a correlation of 0.7 with the original centrality indices with

95% certainty. There is some evidence to suggest that centrality is stable if the CS-coefficient is at least 0.25, but preferably above 0.5. The default 'bootnet' bootstrapping technique was applied in the current study, which uses a non-parametric approach and a total of 1000 bootstrapped samples.

Community detection

To examine the possibility of subnetworks or "communities" within the bipolar disorder and common psychiatric diagnoses network the spin-glass algorithm was performed using the 'igraph' (Csardi & Nepusz, 2006) R package. A community is described as "groups of densely interconnected nodes that are only sparsely connected with the rest of the network" (Reichardt & Bornholdt, 2006, p. 2). When applying the spin-glass algorithm, a community is established when the number of edges and weighted strength of edges within a cluster exceeds that of the nodes in another cluster, meaning that nodes can only belong to one community. As the spin-glass algorithm can produce different results every time the analysis is run, a seed was set to run the algorithm 1000 times. The median number of communities was then taken as the result (Fried, 2016).

Community centrality analysis

Similarly, to the network centrality analyses, bridge centrality, which includes bridge strength, bridge closeness, and bridge betweenness, was analysed for each node. The bridge centrality indices provide information regarding the relative importance of each node in the network. However, unlike network centrality, the associated nodes within bridge centrality, are from different communities. Thus, bridge centrality measures the importance of a symptom in linking the communities. Bridge strength measures the sum of the

absolute values of all edges between a node and all nodes of the other communities i.e. how connected a symptom is with other disorders. Bridge betweenness measures the number of times a node lies on the shortest path between any two nodes from two different and distinct communities. Bridge closeness reflects the average distance of a node to all nodes that are not in the same community i.e. the shortest path from one community to another (Jones et al, 2019).

Results

Sample characteristics

A summary of demographic variables are presented in Table 1. Overall, 7546 participants completed the 2014 APMS survey, $N = 7076$ of which had complete data on the MDQ and were included in the analyses. In total $n = 130$ participants screened positive for bipolar disorders. There were no significant differences in rates of positive screens for men (2.1%) and women (1.8%). However, scores indicating the likely presence of bipolar disorders varied by age. Bipolar disorders were more common in the younger groups with 3.4% of 16-24 year olds and 3.1% of 25-35 year olds screening positive. In contrast, 0.4% of participants aged 65-74 years and no participants over the age of 75 years old screened positive for bipolar disorders. Those aged 16-64 years old were more likely to screen positive if they were unemployed or economically inactive (3.9% and 4.3% respectively) compared to those in employment (1.9%). There were no significant differences in positive screening rates by ethnic group. Most of the participants screening positive for bipolar disorder were not in receipt of psychotropic medication or psychological therapy at the time of the survey (59.2%).

Table 1

Sample demographics for the overall sample with complete MDQ data and for participants that screened positive for a bipolar disorder

Characteristic	Overall sample (<i>N</i> = 7039) <i>N</i> (%)	Positive screen for a bipolar disorder† (<i>n</i> = 130) <i>n</i> (%)
Sex		
Male	2871 (40.8)	58 (44.6)
Female	4168 (59.2)	74 (55.4)
Age in years		
16-24	548 (7.8)	24 (18.5)
25-34	995 (14.1)	32 (24.6)
35-44	1124 (16.0)	23 (17.7)
45-54	1236 (17.6)	28 (21.5)
55-64	1151 (16.4)	18 (13.8)
65-74	1114 (15.8)	5 (3.8)
Over 75	871 (12.4)	0 (0)
Ethnic origin		
White British/other	6388 (90.8)	114 (87.7)
Black	179 (2.5)	8 (6.2)
British/African/Caribbean		
Asian/Asian British	317 (4.5)	5 (3.8)
Mixed/multiple ethnic groups	133 (1.9)	3 (2.3)
Did not respond	44 (0.3)	-

Note. †A positive screening for bipolar disorder is dependent on endorsement of at least seven items of hypomanic/manic symptoms, together with affirmative answers regarding symptoms occurring at the same time and moderate to severe functional impairment queried on a 4-point scale ("no problem" to "serious problem").

Network Analysis

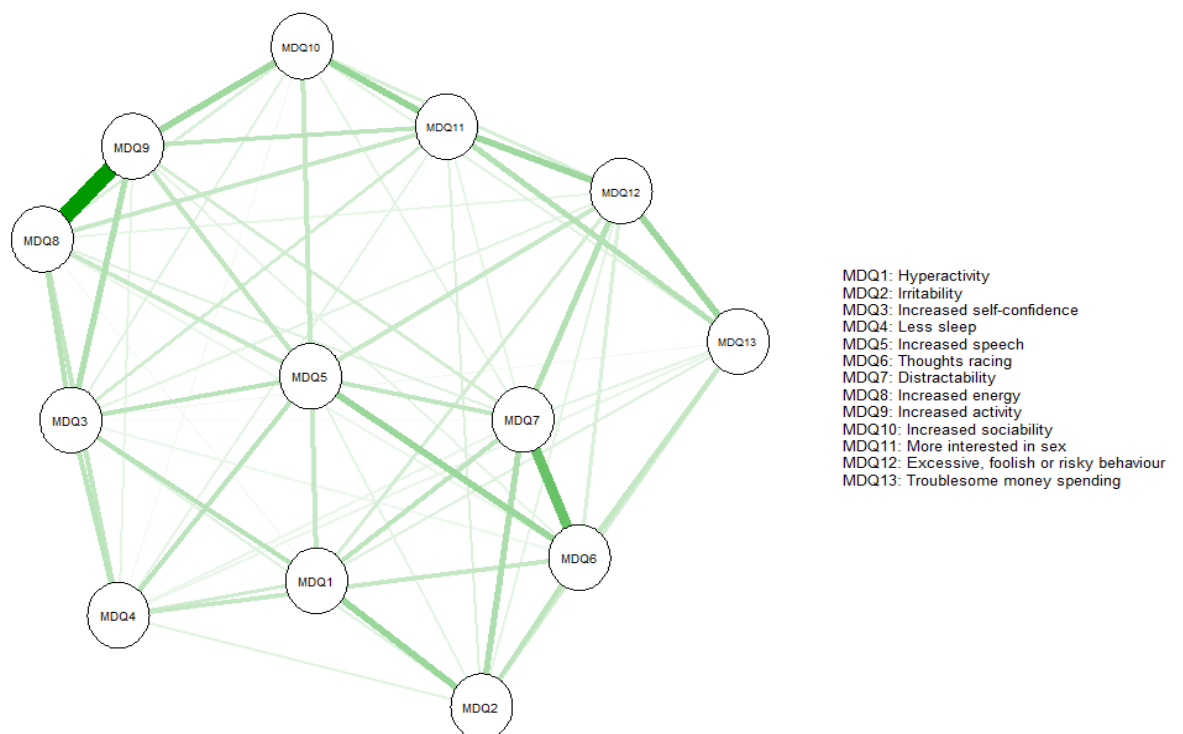
Aim 1: Bipolar disorder network estimation

The *eLasso* regularised network is presented in Figure 1. Each node represents an item of the MDQ whilst the edges represent regularised partial correlations between the items. The network was highly interconnected; out of a potential 78 edges, 71 (91%) were estimated to be above zero. The bootstrapped difference test suggests that the strongest edges are significantly different from weaker edges and bootstrapped CI's show that the edges appear

stable with relatively small CI's. (see Figure C1 and C2, appendix C). The bootstrap analyses indicate that the rank ordering of edge weights (i.e. the thickness of edges) could be interpreted with confidence. All the associations between the variables were positive, with no negative associations between items. Overall, the strongest edges were between MDQ8¹ "Increased energy" and MDQ9 "Increased activity" followed by MDQ6 "Thoughts racing" and MDQ7 "Increased distractibility". Moderate edges are shown between MDQ5 "Increased speech" and MDQ6 "Thoughts racing" similarly to MDQ1 "Hyperactivity" and MDQ2 "Irritability" (see Figure C3, appendix C).

Figure 1

Estimated network structure of 13 bipolar disorder mood symptoms

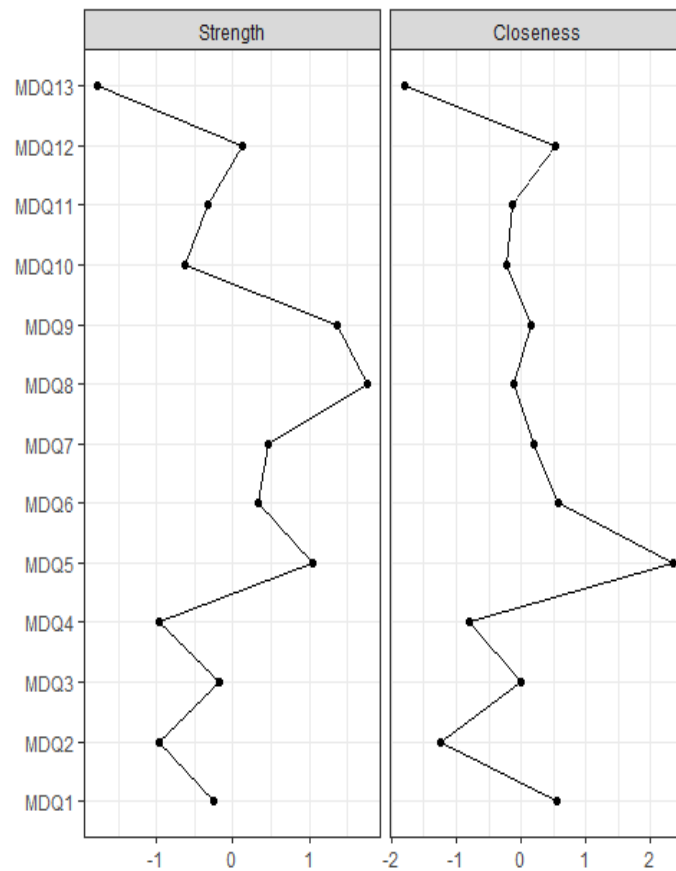


Note. Edges (lines) can be interpreted as partial correlations, with thickness denoting strength of effect. Green lines indicate a positive association.

¹Shortened summaries of MDQ items are utilised. For full MDQ item names, see Appendix B

The case-dropping bootstrap method (Figure C4, Appendix C) indicated that the strength values ($CS[\text{cor} = 0.7] = 0.75$) and the closeness values ($CS[\text{cor} = 0.7] = 0.67$) of the centrality indices were reliable at the preferable threshold. The results indicate that the node strength and closeness values are stable, and the order of node strength and closeness is reliable across 1000 bootstrapping models. However, the stability of node betweenness dropped below the necessary cut off ($CS[\text{cor} = 0.7] = 0.21$), suggesting that the rank order should be interpreted with caution. As such only the strength and closeness metrics will be discussed.

The most central item in the network in terms of node strength was MDQ9 “Increased activity”, indicating that it has the strongest associations with other nodes in the network. MDQ8 “Increased energy”, MDQ7 “Distractibility” and MDQ5 “Increased speech” were also high in strength. The weakest node in the network was MDQ13 “troublesome money spending”. MDQ4 “less sleep” and MDQ2 “irritability” were also relatively weaker nodes within the network. In terms of the centrality metric closeness, MDQ5 “increased speech” had the highest z-score indicating it is most likely to be affected by changes in other nodes in the network (see Figure 2).

Figure 2*Centrality indices of strength and closeness*

Note. Indices presented as standardised z-scores. Higher values indicate greater overall importance to network.

Aim 2: Bipolar disorder and common psychiatric diagnoses network estimation

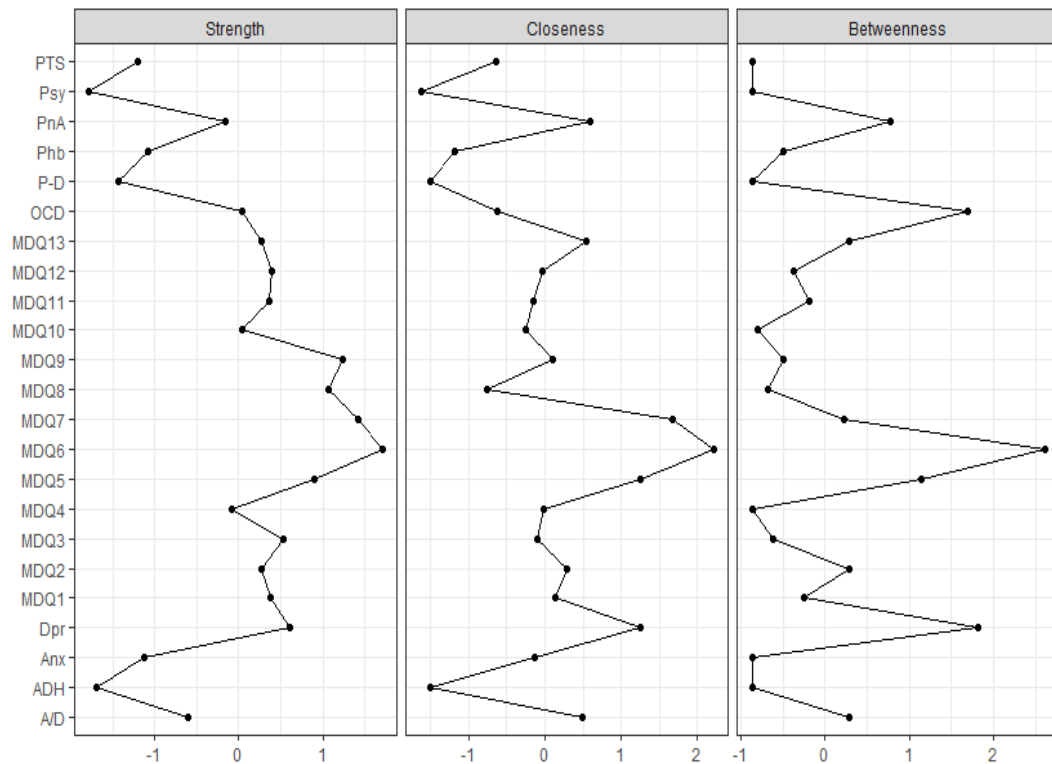
An *eLasso* regularised network comprising mood symptoms of bipolar disorder and common psychiatric diagnoses was then constructed. The network was less interconnected; out of a potential 253 edges, 118 (46%) were estimated to be above zero. The bootstrapped difference test suggests that the strongest edges are significantly different from weaker edges and bootstrapped CI's show that the edges appear stable with relatively small CI's. (see Figures D1 and D2, appendix D). The bootstrap analyses indicate that the rank ordering

of edge weights could be interpreted with confidence. All the associations between the variables were positive, with no negative associations between items. Similarly, to the network consisting only of mood symptoms of bipolar disorder, the strongest edges in the network were between MDQ8 “Increased energy” and MDQ9 “Increased activity” followed by MDQ6 “Thoughts racing” and MDQ7 “Increased distractibility”. Other strong connections were revealed between MDQ13 “troublesome money spending” and “A/D” alcohol/drug dependence, similarly to “Pna” panic attacks and “Dpr” depression (see Figure D3, appendix D).

The case-dropping bootstrap method (Figure D4, appendix D) indicated that the strength values ($CS[\text{cor} = 0.7] = 0.75$) and the closeness values ($CS[\text{cor} = 0.7] = 0.75$) of the centrality indices were reliable at the preferable threshold. The stability of node betweenness met the minimum threshold for reliability but dropped below the preferable cut-off ($CS[\text{cor} = 0.7] = 0.43$). The results indicate that the node strength and closeness values are stable, and the order of node and closeness strength is reliable across 1000 bootstrapping models, however, the betweenness values should be interpreted with caution. The most central item in the network in terms of node strength, closeness and betweenness were MDQ6 “thoughts racing”. Similarly, to the bipolar disorder mood symptom network, MDQ7 “distractibility”. MDQ9 “increased activity” and MDQ8 “increased energy” were also high in strength. In terms of the centrality metric closeness, MDQ7 “distractibility” and MDQ5 “increased speech” had the highest z-scores after MDQ6 “thoughts racing”. Finally, for the centrality metric betweenness, after MDQ6, “drp” depression had the highest z-score indicating its importance in connecting other nodes in the network (see Figure 3).

Figure 3

Centrality indices of strength, closeness and betweenness

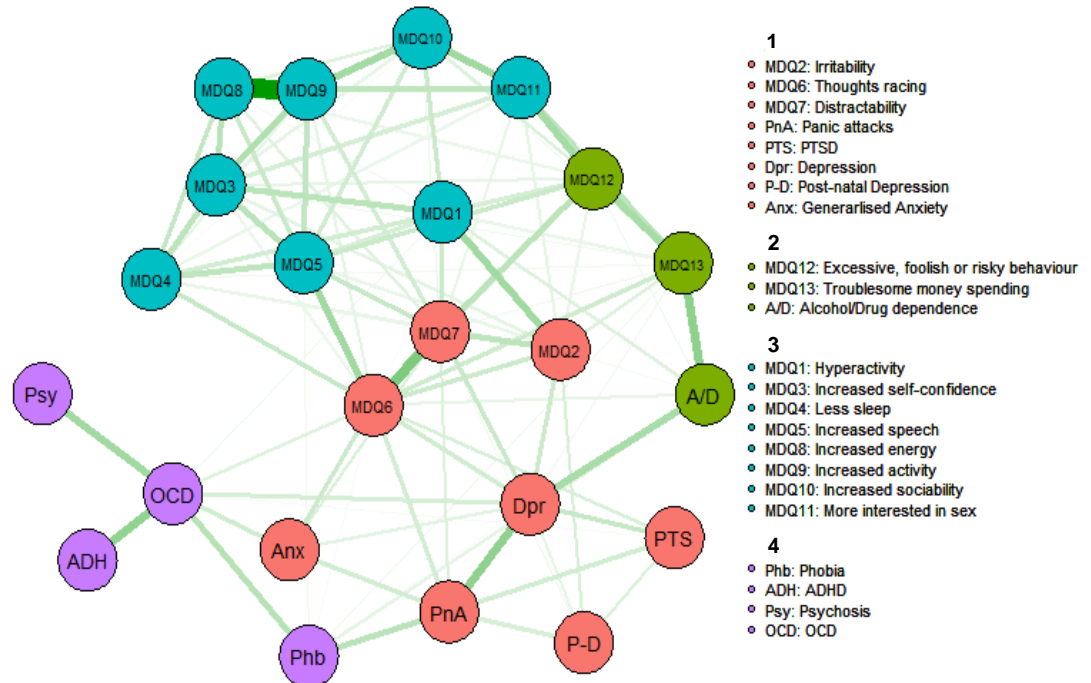


Note. Indices presented as standardised z-scores. Higher values indicate greater overall importance to network.

The spin-glass algorithm detected four communities of nodes (see Figure 4). The first community included three bipolar disorder mood symptoms (irritability, thoughts racing and distractibility) and five psychiatric diagnoses (depression, post-natal depression, generalised anxiety, panic attacks and PTSD). A second community included two bipolar disorder mood symptoms (risky behaviour and troublesome money spending) and alcohol/drug dependence. A third community consisting of the remaining bipolar disorder mood symptoms was detected (hyperactivity, increased self-confidence, less sleep, increased speech, increased energy, increased activity, increased sociability and more interest in sex). The fourth community contained the remaining psychiatric diagnoses (phobia, ADHD, psychosis and OCD).

Figure 4

Estimated network structure of 13 bipolar disorder mood symptoms and common psychiatric diagnoses



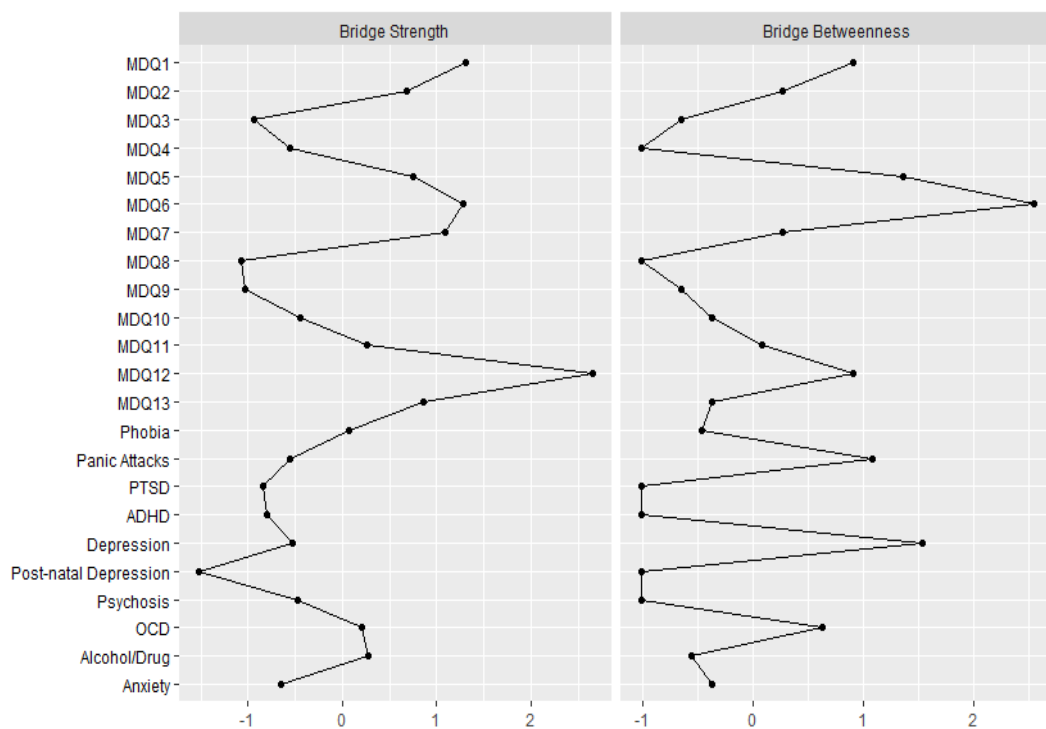
Note. Edges (lines) can be interpreted as partial correlations, with thickness denoting strength of effect. Green lines indicate a positive association. Nodes of the same colour illustrate membership to a shared 'community'.

The case-dropping bootstrap method indicated that the bridge strength values ($CS[\text{cor} = 0.7] = 0.75$) and the bridge betweenness values ($CS[\text{cor} = 0.7] = 0.75$) of the bridge centrality indices were reliable at the preferable threshold. The results indicate that the bridge strength and bridge betweenness values are stable, and the order of bridge strength and bridge is reliable across 1000 bootstrapping models. However, the stability of bridge closeness dropped below the necessary cut off ($CS[\text{cor} = 0.7] = 0.12$), suggesting that the rank order should be interpreted with caution. As such only the bridge strength and bridge betweenness metrics will be discussed.

Bridge centrality indices (Figure. 5) show that MDQ12 “excessive, foolish or risky behaviour” has especially high bridge strength suggesting that it is most connected with all other communities. MDQ1 “hyperactivity” MDQ6 “thoughts racing” and MDQ7 “distractibility” also have higher bridge strength values than other nodes in the network. Regarding bridge betweenness, MDQ6 “thoughts racing” has the highest bridge betweenness suggesting that it is a potential bridge between communities. Other nodes that have relatively higher bridge betweenness scores are “dpr” depression and MDQ5 “increased speech”.

Figure 5

Centrality indices of bridge strength and bridge betweenness



Note. Indices presented as standardised z-scores. Higher values indicate greater overall importance to network

Summary of results

The network estimation identified that the most central nodes in the bipolar disorder mood symptom network were MDQ9 “increased activity” and MDQ8 “increased energy”. The introduction of other psychiatric diagnoses caused a slight shift in the network, in which MDQ6 “thoughts racing” became the most central node in the network amongst MDQ7 “distractibility”, MDQ9 “increased activity” and MDQ8 “increased energy”, which remained as central bipolar symptoms. Depression also showed importance with high betweenness centrality, suggesting significance in connecting nodes in the overall network.

Community analyses revealed four communities. A ‘mixed community’ consisting of the bipolar mood symptoms irritability, distractibility and thoughts racing alongside psychiatric diagnoses of depression, post-natal depression, generalised anxiety, PTSD and panic attacks. A ‘poor judgement’ community containing bipolar mood symptoms associated with risky behaviour and troublesome money spending as well as alcohol dependence. A ‘pure mania’ community consisting solely of bipolar mood symptoms such as increased energy, activity, sociability, speech and hyperactivity. A final ‘other diagnoses’ community was revealed that consisted of the remaining psychiatric conditions such as ADHD, psychosis and OCD. Bridge centrality revealed MDQ6 “thoughts racing” and depression as possible bridges between communities. MDQ12 “risky behaviour” was shown to have the strongest links to all other communities.

Discussion

The present study used a cross-sectional probability sample of the general population to investigate the network structure of bipolar disorder mood

symptoms. Our main finding was that increased energy and increased activity were the most central symptoms in terms of the centrality values node strength, suggesting that the activation of increased energy and increased activity, drives the activation of the rest of the network. This finding is consistent with a wide range of research that highlights what has been termed 'activation symptoms' as core symptoms of bipolar disorders (Scott et al., 2017). The DSM-5 criteria for mania and hypomania were also amended several years ago to include the requirement for increased energy or increased activity alongside mood disturbances as part of criterion A, to meet diagnostic criteria (APA, 2013).

Unsurprisingly increased energy and increased activity were the most strongly connected symptoms in the network, however, the network as a whole was highly interconnected (91% of possible edges estimated to be above zero) suggesting that individuals who experience a symptom within the network are likely to also experience other symptoms within the network. The re-estimation of the network to include other psychiatric diagnoses caused a shift in the network in which 'thoughts racing' became the most central node within the overall network. Racing thoughts had strong associations with 'distractibility', which was consistently central within both networks. Whilst this appears reasonable theoretically, the possible significance of 'distractibility' as a core feature of bipolar disorder is seldom reported in the literature and has rarely been included or investigated (Martino et al., 2020).

Community analyses of the bipolar disorder mood symptoms and common psychiatric diagnoses showed an amalgamation of bipolar disorder symptoms and other psychiatric conditions. Depression formed a community with a select number of symptoms of bipolar disorder ('irritability', 'distractibility' and 'thoughts racing'), as well as other psychiatric diagnoses, including anxiety

disorders. Racing thoughts, irritability and distractibility are all symptoms of bipolar disorder which are thought to occur in both depression and mania. Racing thoughts is considered a core feature of mixed depression (Koukopoloulos, 2018; Weiner, 2019) and has been acknowledged in the DSM-5 (APA, 2013) as part of the 'mixed features specifier'. Although irritability and distractibility appear to have been overlooked in the mixed states classification, Koukopoloulos and colleagues (2013) submit that irritable mood and symptoms of psychomotor agitation are characteristic of mixed depression, with research highlighting irritability and anxiety as common experiences for patients diagnosed with bipolar mania with depressive symptoms (Suppes et al., 2017). Thus, the findings from the community analyses provide support for research highlighting the association between depression, irritability and distractibility and the limitations of the DSM-5 mixed features specifier.

Analyses revealed one community that consisted solely of bipolar disorder mood symptoms. This community containing only MDQ items appeared to consist of what could be considered the principal features of mania. This 'mania' community contained the most central symptoms from the bipolar disorder network (increased energy and increased activity) alongside other psychomotor symptoms (increased speech, less sleep) which are thought to be the most consistent features of bipolar disorders rather than changes in mood (Rossi 2001; Scott, 2017; Swann 2001). Other symptoms within this community such as hyperactivity, increased sociability and increased interest in sex have also been found amongst the core features of mania within several factor analytic studies (Martino et al., 2020). These findings may support theories of a 'manic disorder' in which individuals experience typical mania in the absence of any depressive episodes (Schweitzer et al., 2005). Schweitzer and colleagues

(2005) highlight that whilst depression is often seen to coexist with mania (i.e. in mixed states) or present before or after episodes of mania, there is no strong evidence to suggest that depression is any more linked to mania than other psychiatric diagnoses, which is reflected in the networks from the current study.

Depression formed a strong edge with alcohol/drug dependence consistent with previous research highlighting the association between mania with depressive symptoms and higher rates of comorbid conditions particularly anxiety disorders and substance abuse (Swann et al., 2013). Alcohol and drug dependence formed a community together with mood symptoms associated with poor judgement (foolish or risky behaviour and troublesome money spending). As individuals with diagnoses of bipolar disorder show higher rates of alcohol dependence than that of any other mental health condition (Merikangas et al., 2011), with an estimated lifetime prevalence of 46% (Regier et al., 1990), it is unsurprising that alcohol dependence formed a community amongst symptoms of bipolar disorder. There is evidence to show that substance abuse has been associated with a more severe and persistent bipolar disorder course (Koenders et al., 2015), with some research attributing this to shared risk factors for bipolar disorders and substance abuse (Biseul et al., 2017). For example, the onset of manic symptoms increases the likelihood of risk-taking behaviours which could lead to the use of substances. Likewise, the use of substances may lead to an increase in poor judgement and risky behaviours.

The bridge centrality findings provide us with important insight into shared risk factors and how symptoms within different communities may interact. There is considerable overlap between the symptoms of psychiatric diagnoses (e.g., changes in sleep is common in several psychiatric conditions

(Appavoo & Chirwa, 2019) and the presence of a symptom within one condition could be a risk factor for another, leading to co-morbidity. According to network theories, the symptoms that increase the risk of activating a node that links to another community, or in this case psychiatric disorder, are bridge symptoms (Jones et al., 2019). Within the current study, thoughts racing and depression were identified as having high bridge betweenness, meaning they could be risk factors for moving between possible dimensions of bipolar disorder or even comorbid conditions.

A major finding of the current study was the shift in the network following the addition of common psychiatric diagnoses, in which 'thoughts racing' became the most central node within the overall network. Previous research has highlighted significant associations between racing thoughts and major depression with some research suggesting that depression with racing thoughts could lie on a continuum between major depression without racing thoughts and bipolar disorders (Benazzi, 2005). Similarly, racing thoughts has been found amongst the most common prodromal symptoms before the first hypomanic/manic episode (Zeschel et al., 2013). Thus, the identification of racing thoughts as a possible bridge between communities in the current study, supports research highlighting racing thoughts as a possible risk factor for conversion to bipolar disorder (Weiner et al., 2019). However, interpretations are tentative, as bridge centrality results are contingent on how communities are defined within a network. Research suggests that the most reliable method of exploring co-morbidity relies on communities being determined theoretically rather than through network estimation methods, particularly in cross-sectional psychometric networks (Jones et al., 2019).

Clinical Implications

The findings of this study have several clinical implications. Firstly, the networks provide insight into the complex and multidimensional structure of bipolar disorders that are routinely observed in clinical practice. The communities observed in the current research somewhat reflect course types that have been identified in practice and within the literature (Koenders et al., 2015) and therefore could be clinically meaningful. However, there is some longstanding scepticism regarding specific course groups (Kraepelin 1921 cited in Akiskal, 2003; Koukopoulos, 2013). Insights into the possible interactions between symptoms and pathways through bipolar disorder and associated comorbidities may enable us to develop more targeted interventions (Borsboom, 2017). The research highlighted increased energy and activity as central features of bipolar disorders suggesting that these should be a key consideration in the assessment of bipolar disorders. Further, centrality indices within the current research suggest that thoughts racing may be a risk factor for activating symptoms associated with a more complicated and difficult to treat course of bipolar disorder, thus this may be an important focus of assessment and target for intervention.

Whilst the relationship between mania and depression remains somewhat unclear and the validity of 'mixed state' diagnoses continues to be questionable (Berk, 2005; Koukopoulos, 2013; Parker, 2019). The current study highlights depression as a hub of connectivity and a possible bridge to other psychiatric diagnoses, supporting research that highlights depression as a complicating factor of bipolar disorders. Whilst network approaches may lead us to believe that targeting depression directly may be an appropriate course of action, there is a significant amount of research highlighting the detrimental

impacts of antidepressants in worsening the course of bipolar disorder (Ghaemi, 2003; Koukopoulos, 2018; Parker, 2019). Thus, the clinical implications of the findings are more complicated than targeting the most central symptoms in the network.

Whilst there is good evidence to suggest that reducing the activation of highly central nodes may reduce activation in the overall network, it remains unclear how best to 'deactivate' these nodes (Robinaugh et al., 2016). For instance, it has been suggested that interventions should focus on 'decoupling' strongly connected nodes, meaning interventions should target significant edge weights to reduce these interactions. Therefore, whilst the findings of the current research provide some insight into possible risk factors for comorbidities, any intervention should be closely monitored, and the onset of symptoms associated with depression should warrant further assessment and considerable thought around treatment in clinical practice.

Limitations

The findings of this study should be considered in the context of its limitations. Firstly, the study is based on a community sample, which is a strength in terms of capturing those with sub-threshold symptoms such as hypomania which often missed within the literature. However, the analysis of data from a general population sample limits our conclusions about symptom relationships in individuals with a more severe or chronic presentation of bipolar disorders, such as those in purely clinical samples. In addition, due to the typical characteristics of a community sample, the networks in the current study may represent the lower range of symptom severity which could have impacted the overall connectivity of networks, particularly the comorbidity network.

Furthermore, the symptoms were assessed cross-sectionally meaning they provide limited insight into how the networks may change over time. In addition, the research included depression and other psychiatric diagnoses as global scores rather than individual symptoms, which limits insight into the relationship between mania, depression and other diagnoses. However, the relatively small sample size restricted the number of variables that could be included reliably. Finally, the current research lacks some capacity to provide new clinical insights, due to the reliance on diagnostic rating scales. Some have suggested that if research continues to be solely determined by measures of symptoms (Kessing et al., 2021) or current diagnostic criterion alone (Angst, 2007), we will struggle to make progress in developing more valid descriptions of bipolar disorders, which could lead to novel treatment developments (Phillips & Kupfer, 2013).

Future Research

Due to the limitations associated with cross-sectional networks, future research would benefit from longitudinal or time-series data to examine how networks change and evolve. Time-series analyses could also be strengthened using alternative symptom measures. Advances in technology have seen the emergence of electronic mood monitoring software which uses apps, text messages and web interfaces to capture data about patients daily or weekly experience of psychiatric symptoms, quality of life and medication (Goodyday et al., 2020). A recent study by Gordon-Smith et al. (2021) compared the responses of individuals with a bipolar disorder on standardised self-rated questionnaires and personalised reporting using an electronic mood monitoring tool. The results showed differences in scores on the standardised rating scales compared to patients' views of their mood, highlighting some of the limitations

associated with standardised assessments. Incorporating innovative data collection methods may enable investigators to gather clinically relevant, time-series data whilst minimising the burden on patients and provide new insights into the structure of bipolar disorders and psychopathology more broadly. Furthermore, such methods may also enable clinicians to monitor the effectiveness of interventions in targeting central symptoms.

Conclusions

The current research examined the structure of bipolar disorders and relationships with comorbid conditions using network analysis, which has rarely been applied within this population. In addition, the study acknowledged critical issues around under-diagnosis by utilising a general population sample and measures designed to detect early stages of bipolar disorders. The results provide important insights into the structure of bipolar disorder including how core features of bipolar disorders may interact and reinforce each other, as well as risk factors for increased complexity and comorbidity. However, future applications of network approaches should focus on larger, longitudinal samples to gain greater insight into the reliability of bipolar networks over time and thus their utility in predicting mood course and links to other conditions. In addition, network models could be used to gain insight into interventions for bipolar disorders and how they impact the shape and connectedness of symptom networks. More generally, it appears that research into bipolar disorders continues to highlight the heterogeneous nature of the disorder with limited consensus around accurate descriptions of the condition and its subtypes. As such, future research will undoubtedly benefit from more objective measures or patient-led monitoring, to enhance clinical applicability.

References

- Akiskal, H. S., Azorin, J. M., & Hantouche, E. G. (2003). Proposed multidimensional structure of mania: Beyond the euphoric-dysphoric dichotomy. *Journal of Affective Disorders*, 73(1–2), 7–18. [https://doi.org/10.1016/S0165-0327\(02\)00318-X](https://doi.org/10.1016/S0165-0327(02)00318-X)
- Angst, J. (2007). The bipolar spectrum. In *British Journal of Psychiatry* (Vol. 190, Issue MAR., pp. 189–191). Cambridge University Press.
<https://doi.org/10.1192/bjp.bp.106.030957>
- Appavoo, S., & Chirwa, S. (2019). The role of sleep in assessing and treating psychiatric conditions. *Sleep Medicine*, 64, 16.
<https://doi.org/10.1016/j.sleep.2019.11.047>
- Baldessarini, R. J., Tondo, L., Baethge, C. J., Lepri, B., & Bratti, I. M. (2007). Effects of treatment latency on response to maintenance treatment in manic-depressive disorders. *Bipolar Disorders*, 9(4), 386–393. <https://doi.org/10.1111/j.1399-5618.2007.00385.x>
- Bebbington, P., & Ramana, R. (1995). The epidemiology of bipolar affective disorder. *Social Psychiatry and Psychiatric Epidemiology*, 30(6), 279–292.
<https://doi.org/10.1007/BF00805795>
- Benazzi, F. (2005). Unipolar depression with racing thoughts: A bipolar spectrum disorder? *Psychiatry and Clinical Neurosciences*, 59(5), 570–575.
<https://doi.org/10.1111/j.1440-1819.2005.01416.x>
- Berk, M., Dodd, S., & Malhi, G. S. (2005). “Bipolar missed states”: The diagnosis and clinical salience of bipolar mixed states. In *Australian and New Zealand Journal of Psychiatry* (Vol. 39, Issue 4, pp. 215–221). <https://doi.org/10.1111/j.1440-1614.2004.01557.x>

- Birmaher, B., Gill, M. K., Axelson, D. A., Goldstein, B. I., Goldstein, T. R., Yu, H., Liao, F., Iyengar, S., Diler, R. S., Strober, M., Hower, H., Yen, S., Hunt, J., Merranko, J. A., Ryan, N. D., & Keller, M. B. (2014). Longitudinal trajectories and associated baseline predictors in youths with bipolar spectrum disorders. *American Journal of Psychiatry*, *171*(9), 990–999. <https://doi.org/10.1176/appi.ajp.2014.13121577>
- Biseul, I., Icick, R., Seguin, P., Bellivier, F., & Scott, J. (2016). *Feasibility and Acceptability of the “HABIT” Group Programme for Comorbid Bipolar and Alcohol and Substance use Disorders*. <https://doi.org/10.1002/cpp.2053>
- Cassano, G. B., Rucci, P., Frank, E., Fagiolini, A., Dell’Osso, L., Shear, M. K., & Kupfer, D. J. (2004). The mood spectrum in unipolar and bipolar disorder: Arguments for a unitary approach. *American Journal of Psychiatry*, *161*(7), 1264–1269. <https://doi.org/10.1176/appi.ajp.161.7.1264>
- Cogle, J. R., Hakes, J. K., Macatee, R. J., Chavarria, J., & Zvolensky, M. J. (2015). Quality of life and risk of psychiatric disorders among regular users of alcohol, nicotine, and cannabis: An analysis of the National Epidemiological Survey on Alcohol and Related Conditions (NESARC). *Journal of Psychiatric Research*, *66–67*, 135–141. <https://doi.org/10.1016/j.jpsychires.2015.05.004>
- Cramer, A. O. J., Waldorp, L. J., Van Der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2–3), 137–150. <https://doi.org/10.1017/S0140525X09991567>
- De Ron, J., Fried, E. I., & Epskamp, S. (2021). Psychological networks in clinical populations: Investigating the consequences of Berkson’s bias. *Psychological Medicine*, *51*(1), 168–176. <https://doi.org/10.1017/S0033291719003209>

- Double, D. B. (1990). The factor structure of manic rating scales. *Journal of Affective Disorders, 18*(2), 113–119. [https://doi.org/10.1016/0165-0327\(90\)90067-I](https://doi.org/10.1016/0165-0327(90)90067-I)
- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212. <https://doi.org/10.3758/s13428-017-0862-1>
- Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*. <https://doi.org/10.18637/jss.v048.i04>
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617–634. <https://doi.org/10.1037/met0000167>
- Fried, E. (2016). *R tutorial: how to identify communities of items in networks*. Psych Networks - Organized Incoherence. <https://psych-networks.com/r-tutorial-identify-communities-items-networks/>
- Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: a review of recent insights. *Social Psychiatry and Psychiatric Epidemiology, 52*(1), 1–10. <https://doi.org/10.1007/s00127-016-1319-z>
- Ghaemi, S. N., Ko, J. Y., & Goodwin, F. K. (2002). “Cade’s disease” and beyond: Misdiagnosis, antidepressant use, and a proposed definition for bipolar spectrum disorder. *Canadian Journal of Psychiatry, 47*(2), 125–134. <https://doi.org/10.1177/070674370204700202>
- Goekoop, R., & Goekoop, J. G. (2014). A network view on psychiatric disorders: Network clusters of symptoms as elementary syndromes of psychopathology. *PLoS ONE, 9*(11). <https://doi.org/10.1371/journal.pone.0112734>

- Goodday, S. M., Atkinson, L., Goodwin, G., Saunders, K., South, M., Mackay, C., Denis, M., Hinds, C., Attenburrow, M. J., Davies, J., Welch, J., Stevens, W., Mansfield, K., Suvilehto, J., & Geddes, J. (2020). The true colours remote symptom monitoring system: A decade of evolution. *Journal of Medical Internet Research*, 22(1), 15188. <https://doi.org/10.2196/15188>
- Gordon-Smith, K., Saunders, K. E., Savage, J., Craddock, N., Jones, I., & Jones, L. (2021). Have I argued with my family this week?": What questions do those with lived experience choose to monitor their bipolar disorder? *Journal of Affective Disorders*, 281, 918–925. <https://doi.org/10.1016/j.jad.2020.11.034>
- Haro, J. M., Reed, C., Gonzalez-Pinto, A., Novick, D., Bertsch, J., & Vieta, E. (2010). 2-year course of bipolar disorder type I patients in outpatient care: Factors associated with remission and functional recovery. <https://doi.org/10.1016/j.euroneuro.2010.08.001>
- Hirschfeld, R. M. A., Williams, J. B. W., Spitzer, R. L., Calabrese, J. R., Flynn, L., Keck, J., Lewis, L., McElroy, S. L., Post, R. M., Rapport, D. J., Russell, J. M., Sachs, G. S., & Zajecka, J. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The mood disorder questionnaire. *American Journal of Psychiatry*, 157(11), 1873–1875. <https://doi.org/10.1176/appi.ajp.157.11.1873>
- Hirschfeld, R. M. A., Holzer, C., Calabrese, J. R., Weissman, M., Reed, M., Davies, M., Frye, M. A., Keck, P., McElroy, S., Lewis, L., Tierce, J., Wagner, K. D., & Hazard, E. (2003). Validity of the mood disorder questionnaire: A general population study. *American Journal of Psychiatry*, 160(1), 178–180. <https://doi.org/10.1176/appi.ajp.160.1.178>

- Humpston, C. S., Bebbington, P., & Marwaha, S. (2021). Bipolar disorder: Prevalence, help-seeking and use of mental health care in England. Findings from the 2014 Adult Psychiatric Morbidity Survey. *Journal of Affective Disorders*, *282*, 426–433. <https://doi.org/10.1016/j.jad.2020.12.151>
- Johnson, S. L., Morriss, R., Scott, J., Paykel, E., Kinderman, P., Kolamunnage-Dona, R., & Bentall, R. P. (2011). Depressive and manic symptoms are not opposite poles in bipolar disorder. *Acta Psychiatrica Scandinavica*, *123*(3), 206–210. <https://doi.org/10.1111/j.1600-0447.2010.01602.x>
- Jones, P. J., Ma, R., & McNally, R. J. (2019). Bridge Centrality: A Network Approach to Understanding Comorbidity. *Multivariate Behavioral Research*, 1–15. <https://doi.org/10.1080/00273171.2019.1614898>
- Judd, L. L., Akiskal, H. S., Schettler, P. J., Endicott, J., Maser, J., Solomon, D. A., Leon, A. C., Rice, J. A., & Keller, M. B. (2002). The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Archives of General Psychiatry*, *59*(6), 530–537. <https://doi.org/10.1001/archpsyc.59.6.530>
- Judd, L. L., Schettler, P. J., Akiskal, H. S., Maser, J., Coryell, W., Solomon, D., Endicott, J., & Keller, M. (2003). Long-term symptomatic status of bipolar I vs. bipolar II disorders. *International Journal of Neuropsychopharmacology*, *6*(2), 127–137. <https://doi.org/10.1017/S1461145703003341>
- Kessing, L. V., González-Pinto, A., Fagiolini, A., Bechdolf, A., Reif, A., Yildiz, A., Etain, B., Henry, C., Severus, E., Reininghaus, E. Z., Morken, G., Goodwin, G. M., Scott, J., Geddes, J. R., Rietschel, M., Landén, M., Manchia, M., Bauer, M., Martínez-Cengotitabengoa, M., ... Vieta, E. (2021). DSM-5 and ICD-11 criteria for bipolar disorder: Implications for the prevalence of bipolar disorder and validity of the diagnosis – A narrative review from the ECNP bipolar disorders

network. *European Neuropsychopharmacology*, 52, 2451-2496.

<https://doi.org/10.1016/j.euroneuro.2021.01.097>

Kodesh, A., Goldshtein, I., Gelkopf, M., Goren, I., Chodick, G., & Shalev, V. (2012).

Epidemiology and comorbidity of severe mental illnesses in the community:

Findings from a computerized mental health registry in a large Israeli health

organization. *Social Psychiatry and Psychiatric Epidemiology*, 47(11), 1775–

1782. <https://doi.org/10.1007/s00127-012-0478-9>

Koenders, M. A., De Kleijn, R., Giltay, E. J., Elzinga, B. M., Spinhoven, P., & Spijker, A.

T. (2015). *A Network Approach to Bipolar Symptomatology in Patients with*

Different Course Types. <https://doi.org/10.1371/journal.pone.0141420>

Koenders, M. A., De Kleijn, R., Giltay, E. J., Elzinga, B. M., Spinhoven, P., & Spijker, A.

T. (2015). A network approach to bipolar symptomatology in patients with

different course types. *PLoS ONE*, 10(10), 1–16.

<https://doi.org/10.1371/journal.pone.0141420>

Koukopoulos, A., Sani, G., & Ghaemi, S. N. (2013). Mixed features of depression: Why

DSM-5 is wrong (and so was DSM-IV). In *British Journal of Psychiatry* (Vol. 203,

Issue 1, pp. 3–5). Cambridge University Press.

<https://doi.org/10.1192/bjp.bp.112.124404>

Mantere, O., Suominen, K., Valtonen, H. M., Arvilommi, P., Leppämäki, S., Melartin, T.,

& Isometsä, E. (2008). Differences in outcome of DSM-IV bipolar I and II

disorders. *Bipolar Disorders*, 10(3), 413–425. <https://doi.org/10.1111/j.1399->

5618.2007.00502.x

Marneros, A. (2001). Origin and development of concepts of bipolar mixed states.

Journal of Affective Disorders, 67(1–3), 229–240. <https://doi.org/10.1016/S0165->

0327(01)00437-2

- Martino, D. J., Valerio, M. P., & Parker, G. (2020). The structure of mania: An overview of factorial analysis studies. *European Psychiatry, 63*(1).
<https://doi.org/10.1192/j.eurpsy.2020.18>
- McCombs, J. S., Ahn, J., Tencer, T., & Shi, L. (2007). The impact of unrecognized bipolar disorders among patients treated for depression with antidepressants in the fee-for-services California Medicaid (Medi-Cal) program: A 6-year retrospective analysis. *Journal of Affective Disorders, 97*(1–3), 171–179.
<https://doi.org/10.1016/j.jad.2006.06.018>
- McElroy, E., McIntyre, J. C., Brent, R. P., Wilson, T., Holt, K., Kullu, C., Nathan, R., Kerr, A., Panagaki, K., McKeown, M., Saini, P., Gabbay, M., & Corcoran, R. (2019). Mental Health, Deprivation, and the Neighborhood Social Environment: A Network Analysis. *Clinical Psychological Science, 7*(4), 719–734.
<https://doi.org/10.1177/2167702619830640>
- McManus, S., Bebbington, P., Jenkins, R., Brugha, T., S, M., P, B., R, J., Brugha, T., Stansfeld, S., Clark, C., Bebbington, P., King, M., Jenkins, R., & Hinchliffe, S. (2016). Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. In *Adult Psychiatric Morbidity Survey 2014*.
<https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-survey-survey-of-mental-health-and-wellbeing-england-2014>
- McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy, 86*, 95–104. <https://doi.org/10.1016/j.brat.2016.06.006>
- Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M. A., Petukhova, M., & Kessler, R. C. (2007). Lifetime and 12-month prevalence of bipolar spectrum disorder in the national comorbidity survey replication.

Archives of General Psychiatry, 64(5), 543–552.

<https://doi.org/10.1001/archpsyc.64.5.543>

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>

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). *American Journal of Psychiatry*, 163(2), 217–224.

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

Phillips, M. L., & Kupfer, D. J. (2013). Bipolar Disorder 2 - Bipolar disorder diagnosis: Challenges and future directions. *The Lancet*, 381(9878), 1663–1671.

[https://doi.org/10.1016/S0140-6736\(13\)60989-7](https://doi.org/10.1016/S0140-6736(13)60989-7)

Regier, D. A., Farmer, M. E., Rae, D. S., Locke, B. Z., Keith, S. J., Judd, L. L., & Goodwin, F. K. (1990). Comorbidity of Mental Disorders With Alcohol and Other Drug Abuse: Results From the Epidemiologic Catchment Area (ECA) Study. *JAMA: The Journal of the American Medical Association*, 264(19), 2511–2518.

<https://doi.org/10.1001/jama.1990.03450190043026>

Reichardt, J., & Bornholdt, S. (2006). *Statistical mechanics of community detection*.

<https://doi.org/10.1103/PhysRevE.74.016110>

- Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology, 125*(6), 747–757. <https://doi.org/10.1037/abn0000181>
- Rossi, A., Daneluzzo, E., Arduini, L., Di Domenico, M., Pollice, R., & Petruzzi, C. (2001). A factor analysis of signs and symptoms of the manic episode with Bech-Rafaelsen Mania and Melancholia Scales. *Journal of Affective Disorders, 64*(2–3), 267–270. [https://doi.org/10.1016/S0165-0327\(00\)00228-7](https://doi.org/10.1016/S0165-0327(00)00228-7)
- Schweitzer, I., Maguire, K., & Ng, C. H. (2005). Should bipolar disorder be viewed as manic disorder? Implications for bipolar depression. *Bipolar Disorders, 7*(5), 418–423. <https://doi.org/10.1111/j.1399-5618.2005.00246.x>
- Scott, J., Murray, G., Henry, C., Morken, G., Scott, E., Angst, J., Merikangas, K. R., & Hickie, I. B. (2017). Activation in bipolar disorders a systematic review. In *JAMA Psychiatry* (Vol. 74, Issue 2, pp. 189–196). American Medical Association. <https://doi.org/10.1001/jamapsychiatry.2016.3459>
- Shoham, N., Lewis, G., Hayes, J., McManus, S., Kiani, R., Brugha, T., Bebbington, P., & Cooper, C. (2020). Psychotic symptoms and sensory impairment: Findings from the 2014 adult psychiatric morbidity survey. *Schizophrenia Research, 215*, 357–364. <https://doi.org/10.1016/j.schres.2019.08.028>
- Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria: Rationale and Reliability. *Archives of General Psychiatry, 35*(6), 773–782. <https://doi.org/10.1001/archpsyc.1978.01770300115013>
- Stadtfeld, C., Snijders, T. A. B., Steglich, C., Van Duijn, M., & Zürich, E. (n.d.). *Statistical Power in Longitudinal Network Studies*.

- Stensland, M. D., Schultz, J. F., & Frytak, J. R. (2010). Depression diagnoses following the identification of bipolar disorder: Costly incongruent diagnoses. *BMC Psychiatry, 10*. <https://doi.org/10.1186/1471-244X-10-39>
- Suppes, T., Eberhard, J., Lemming, O., Young, A. H., & McIntyre, R. S. (2017). Anxiety, irritability, and agitation as indicators of bipolar mania with depressive symptoms: a post hoc analysis of two clinical trials. *International Journal of Bipolar Disorders, 5*(1), 36. <https://doi.org/10.1186/s40345-017-0103-7>
- Swann, A. C., Janicak, P. L., Calabrese, J. R., Bowden, C. L., Dilsaver, S. C., Morris, D. D., Petty, F., & Davis, J. M. (2001). Structure of mania: Depressive, irritable, and psychotic clusters with different retrospectively-assessed course patterns of illness in randomized clinical trial participants. *Journal of Affective Disorders, 67*(1–3), 123–132. [https://doi.org/10.1016/S0165-0327\(01\)00447-5](https://doi.org/10.1016/S0165-0327(01)00447-5)
- Swann, A. C., Lafer, B., Perugi, G., Frye, M. A., Bauer, M., Bahk, W.-M., Scott, J., Ha, K., & Suppes, T. (2013). Bipolar mixed states: an international society for bipolar disorders task force report of symptom structure, course of illness, and diagnosis. *The American Journal of Psychiatry, 170*(1), 31–42. <https://doi.org/10.1176/appi.ajp.2012.12030301>
- Van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4*(September). <https://doi.org/10.1038/srep05918>
- Voigt, A., Kreiter, D., Jacobs, C., Revenich, E., Serafras, N., Wiersma, M., Os, J. van, Bak, M., & Drukker, M. (2018). Clinical Network Analysis in a Bipolar Patient Using an Experience Sampling Mobile Health Tool: An n=1 Study. *Bipolar*

Disorder: Open Access, 04(01), 1–7. <https://doi.org/10.4172/2472-1077.1000121>

Waleeprakhon, P., Ittasakul, P., Lotrakul, M., Wisajun, P., Jullagate, S., & Ketter, T. A. (2014). Development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire Thai version.

Neuropsychiatric Disease and Treatment, 10, 1497–1502.

<https://doi.org/10.2147/NDT.S67842>

Weintraub, M. J., Schneck, C. D., & Miklowitz, D. J. (2020). Network analysis of mood symptoms in adolescents with or at high risk for bipolar disorder. *Bipolar Disorders*, 22(2), 128–138. <https://doi.org/10.1111/bdi.12870>

<https://doi.org/10.1111/bdi.12870>

Zeschel, E., Correll, C. U., Hausleiter, I. S., Krüger-Özgürdal, S., Leopold, K., Pfennig, A., Bechdorf, A., Bauer, M., & Juckel, G. (2013). The bipolar disorder prodrome revisited: Is there a symptomatic pattern? *Journal of Affective Disorders*, 151(2), 551–560. <https://doi.org/10.1016/j.jad.2013.06.043>

Appendices

Appendix A – Ethical Approval



Downloaded: 09/05/2019
Approved: 09/05/2019

Adele Ghandte
Registration number: 170149422
Psychology
Programme: Doctorate of Clinical Psychology

Dear Adele

PROJECT TITLE: A network analysis of Bipolar Disorder
APPLICATION: Reference Number 024198

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 09/05/2019 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 024198 (dated 08/05/2019).
- Participant information sheet 1059761 version 1 (08/04/2019).
- Participant consent form 1059762 version 1 (08/04/2019).

If during the course of the project you need to [deviate significantly from the above-approved documentation](#) please inform me since written approval will be required.

Yours sincerely

Jilly Gibson-Miller
Ethics Administrator
Psychology

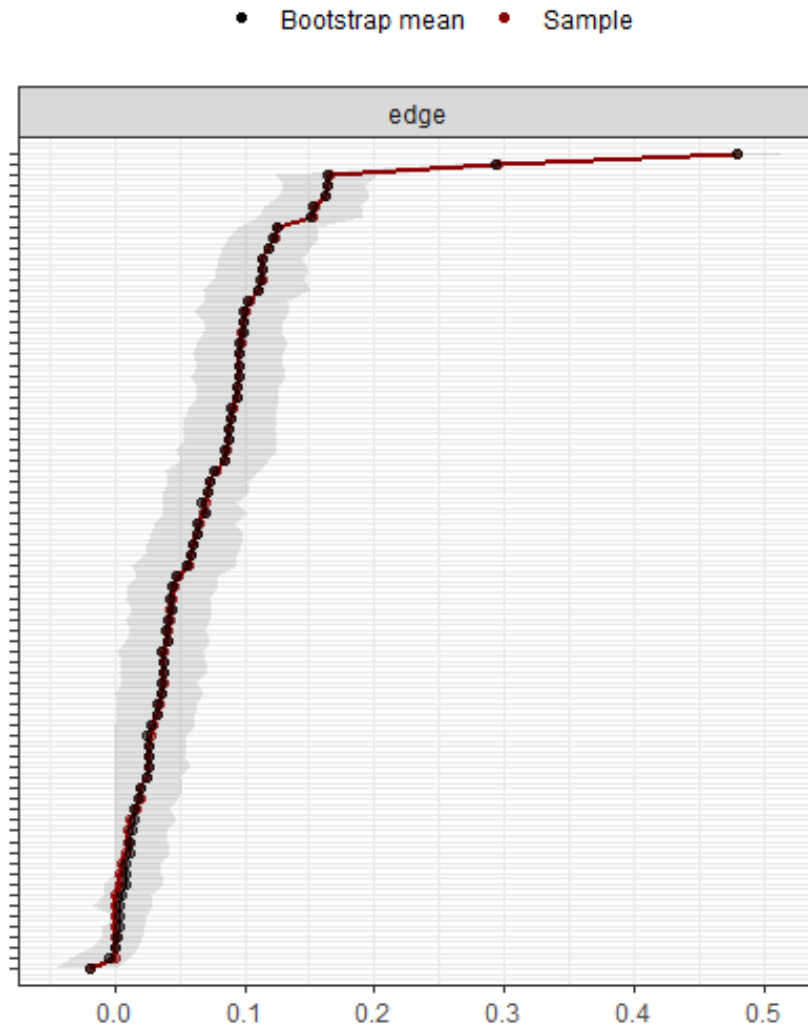
Appendix B – Mood Disorder Questionnaire

REMOVED DUE TO COPYRIGHT RESTRICTIONS

Appendix C – Network accuracy analyses for the 13 mood symptoms of bipolar disorder network (Aim 1)

Figure C1

Graph depicting bootstrapped confidence intervals of estimated edge-weights for the estimated network of 13 bipolar disorder mood symptoms



Note. The red line indicates the sample values, the dark line indicates the mean bootstrapped values and the grey area indicates the 95% bootstrapped confidence intervals. Each horizontal line represents one edge of the network, ordered from the edge with the highest edge-weight to the edge with the lowest edge-weight. In the case of ties (for instance, multiple edge-weights were estimated to be exactly 0), the mean of the bootstrap samples was used in ordering the edges. The y-axis labels have been removed to avoid cluttering.

Figure C2

Graph depicting bootstrapped difference tests ($\alpha = 0.05$) between edge weights that were non-zero in the estimated network

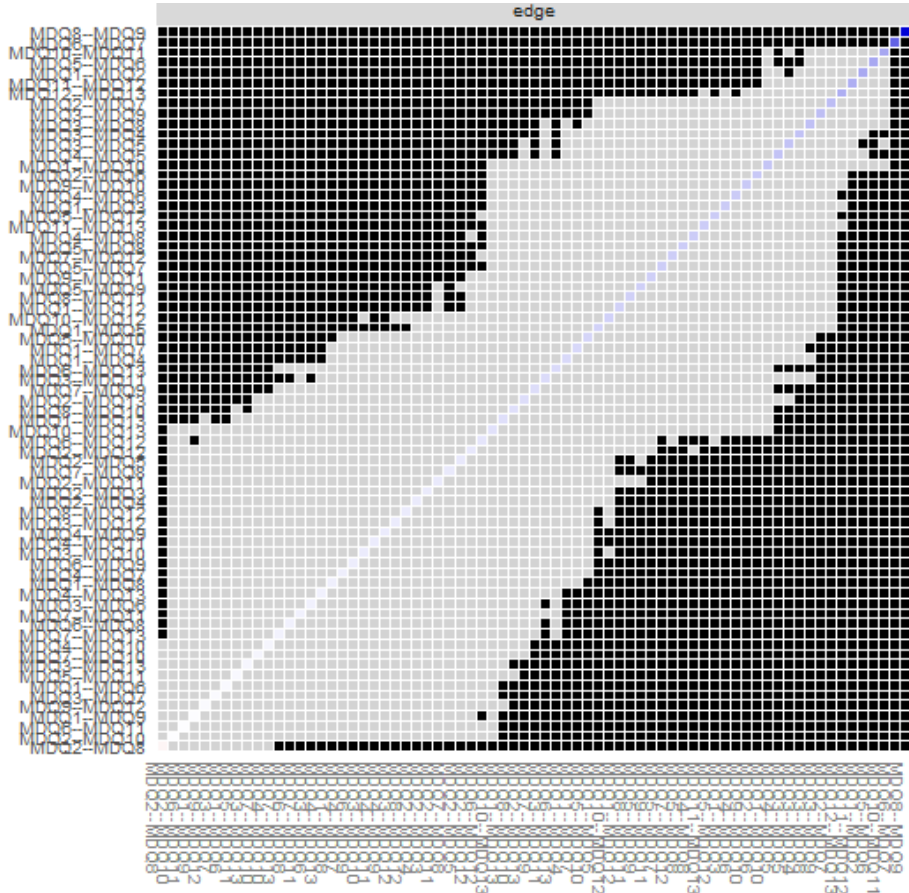
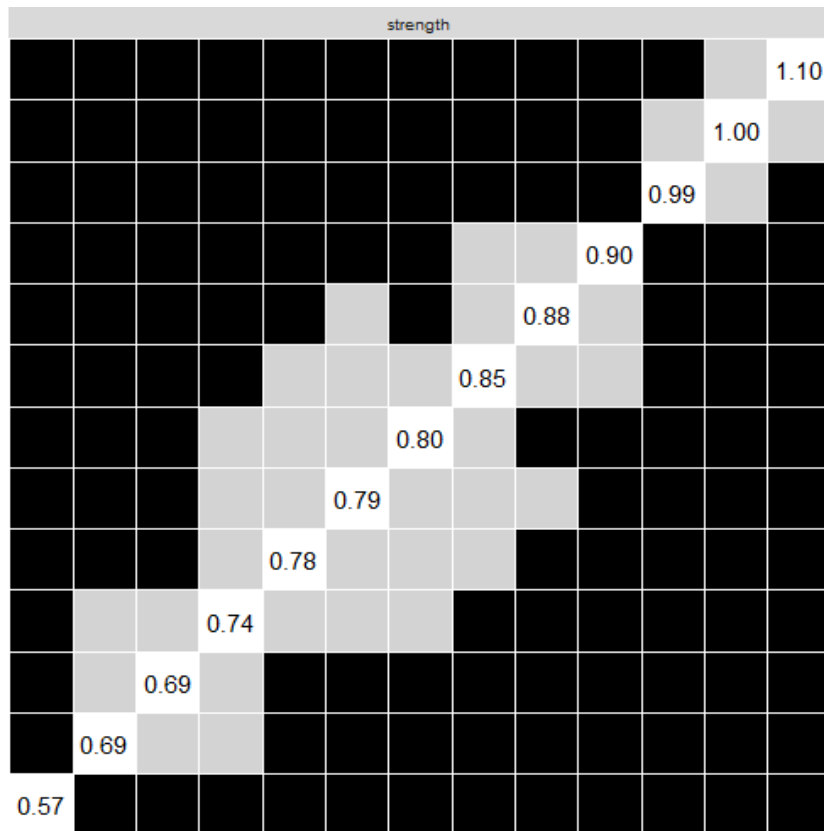


Figure C3

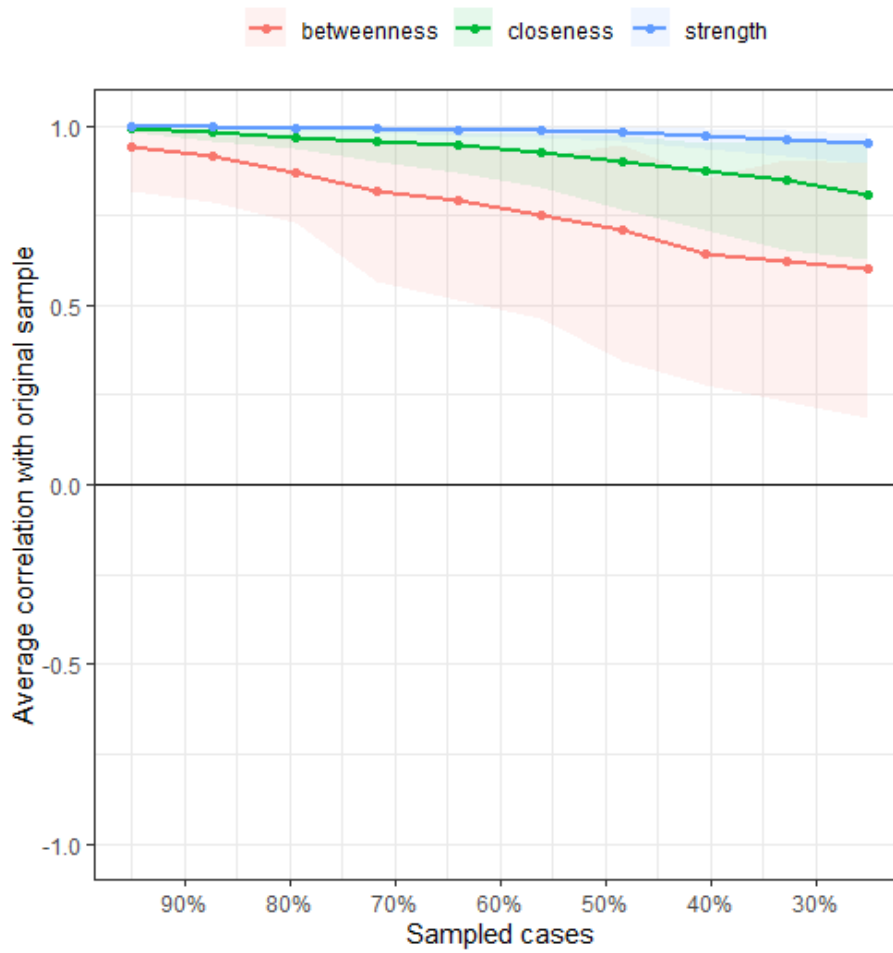
Graph depicting node strength of the 13 bipolar disorder symptoms



Note. Grey boxes indicate nodes that do not differ significantly from one-another and black boxes represent nodes that do differ significantly from one another white boxes show the value of node strength.

Figure C4

Graph depicting average correlations between centrality indices of networks sampled with persons dropped and the original sample

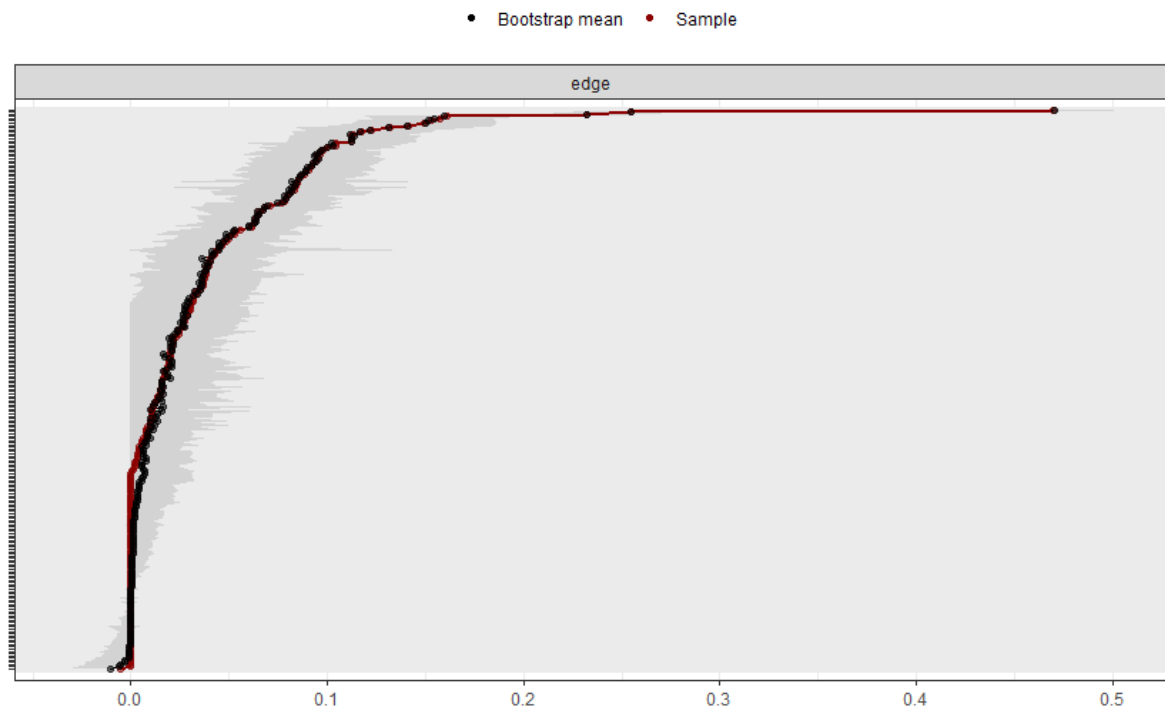


Note. Lines indicate the means and areas indicate the range from the 2.5th quantile to the 97.5th quantile.

Appendix D – Network accuracy analyses for the 13 mood symptoms of bipolar disorder network and 10 psychiatric diagnoses (Aim 2)

Figure D1

Graph depicting bootstrapped confidence intervals of estimated edge-weights for the estimated network of 13 bipolar disorder mood symptoms and 10 psychiatric diagnoses



Note. The red line indicates the sample values, the dark line indicates the mean bootstrapped values and the grey area indicates the 95% bootstrapped confidence intervals. Each horizontal line represents one edge of the network, ordered from the edge with the highest edge-weight to the edge with the lowest edge-weight. In the case of ties (for instance, multiple edge-weights were estimated to be exactly 0), the mean of the bootstrap samples was used in ordering the edges. The y-axis labels have been removed to avoid cluttering.

Figure D2

Graph depicting bootstrapped difference tests ($\alpha = 0.05$) between edge weights that were non-zero in the estimated network

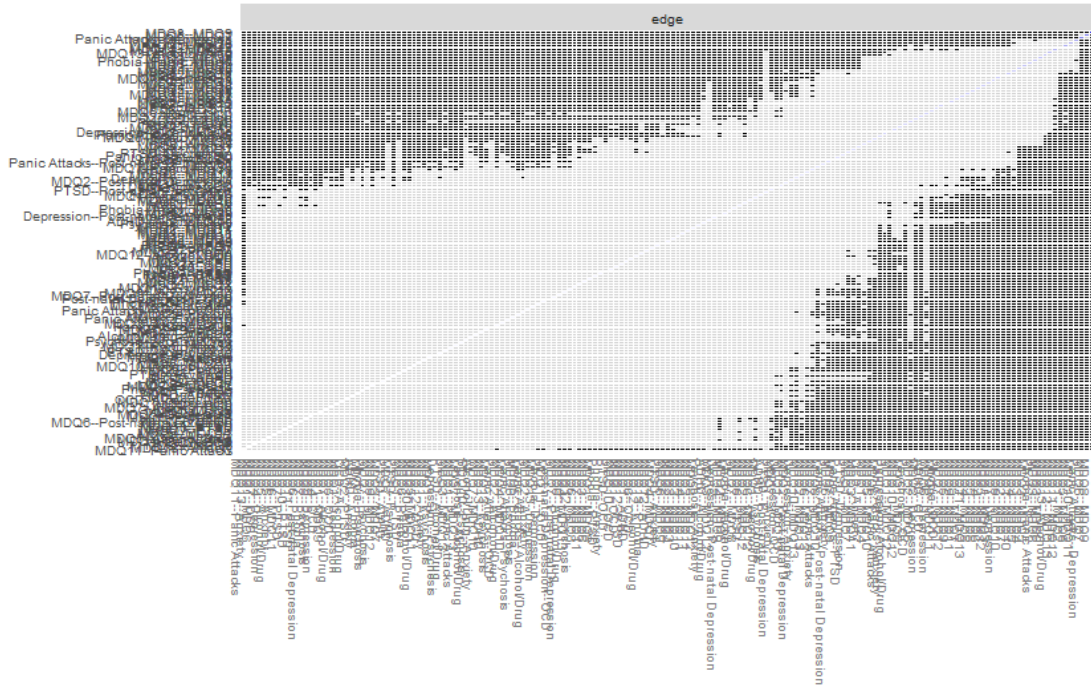
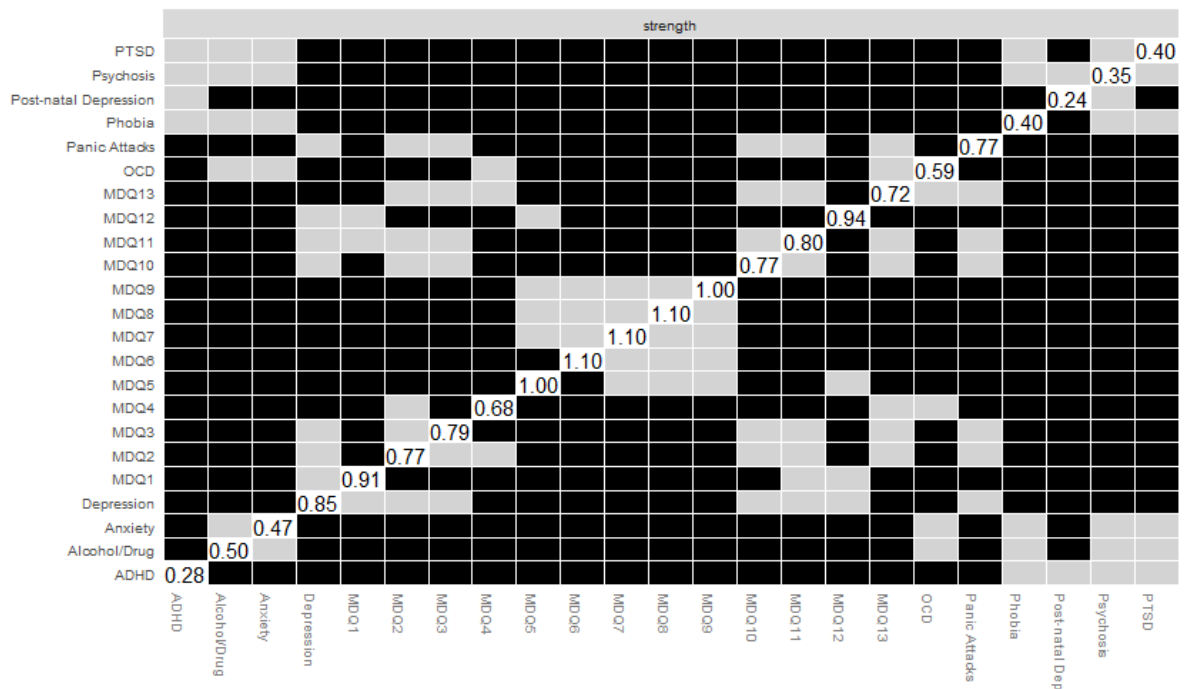


Figure D3

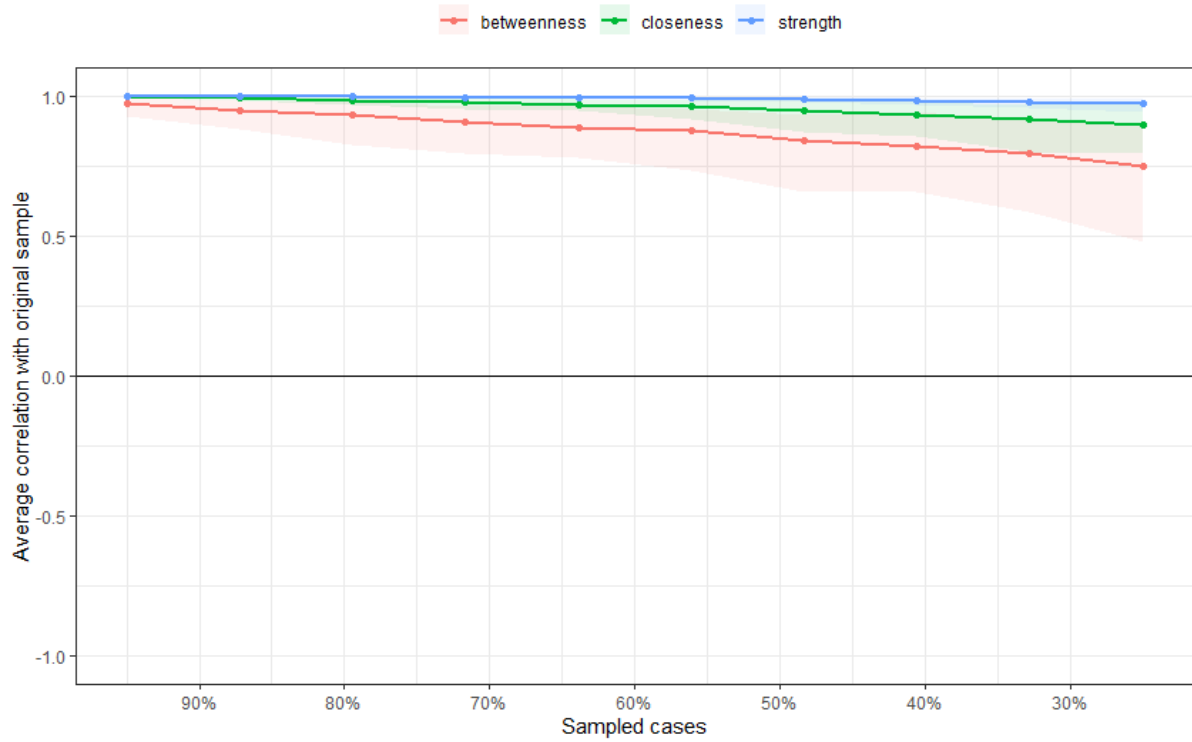
Graph depicting node strength of the 13 bipolar disorder symptoms and 10 psychiatric diagnoses



Note. Grey boxes indicate nodes that do not differ significantly from one-another and black boxes represent nodes that do differ significantly from one another white boxes show the value of node strength.

Figure D4

Graph depicting average correlations between centrality indices of networks sampled with persons dropped and the original sample



Note. Lines indicate the means and areas indicate the range from the 2.5th quantile to the 97.5th quantile.