Received: 23 December 2023 | Revised: 8 March 2024 | Accepted: 10 April 2024

DOI: 10.1111/acps.13692

Acta Psychiatrica Scandinavica WILEY

ORIGINAL ARTICLE

Phenotypic clustering of bipolar disorder supports stratification by lithium responsiveness over diagnostic subtypes

Katie Scott[1] | **Claire O'Donovan**[1] | **Giulio Emilio Brancati**[2] | **Pablo Cervantes**[3] | **Rafaella Ardau**[4] | **Mirko Manchia**[5,6] | **Giovanni Severino**[7] | **Janusz Rybakowski**[8] | **Leonardo Tondo**[9,10] | **Paul Grof**[11,12] | **Martin Alda**[1] | **Abraham Nunes**[1,13]

[1]Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada

[2]Department of Clinical and Experimental Medicine, University Hospital of Pisa, Pisa, Italy

[3]Department of Psychiatry, McGill University Health Centre, Montreal, Quebec, Canada

[4]Unit of Clinical Pharmacology, University Hospital of Cagliari, Cagliari, Italy

[5]Department of Medical Sciences and Public Health, University of Cagliari, Cagliari, Italy

[6]Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia, Canada

[7]Department of Biomedical Sciences, University of Cagliari, Cagliari, Italy

[8]Department of Adult Psychiatry, Poznan University of Medical Sciences, Poznan, Poland

[9]Department of Psychiatry, Harvard Medical School, Boston, Nova Scotia, USA

[10]Lucio Bini Mood Disorder Center, Cagliari, Italy

[11]Mood Disorders Center of Ottawa, Ottawa, Ontario, Canada

[12]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

[13]Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada

Correspondence
Abraham Nunes and Martin Alda, Department of Psychiatry, Dalhousie University, 5909 Veterans' Memorial Lane, Abbie J. Lane Memorial Building (Room 3088), QEII Health Sciences Centre, Halifax, NS B3H 2E2, Canada.
Email: nunes@dal.ca and malda@dal.ca

Funding information
Canadian Institutes of Health Research, Grant/Award Number: 166098; Dalhousie Medical Research Foundation; Dalhousie Faculty of Medicine Graduate Studentship; Research Nova Scotia; Nova Scotia Health Research Fund

Abstract

Introduction: The aim of this study was to determine whether the clinical profiles of bipolar disorder (BD) patients could be differentiated more clearly using the existing classification by diagnostic subtype or by lithium treatment responsiveness.

Methods: We included adult patients with BD-I or II ($N = 477$ across four sites) who were treated with lithium as their principal mood stabilizer for at least 1 year. Treatment responsiveness was defined using the dichotomized Alda score. We performed hierarchical clustering on phenotypes defined by 40 features, covering demographics, clinical course, family history, suicide behaviour, and comorbid conditions. We then measured the amount of information that inferred clusters carried about (A) BD subtype and (B) lithium responsiveness using adjusted mutual information (AMI) scores. Detailed

phenotypic profiles across clusters were then evaluated with univariate comparisons.

Results: Two clusters were identified ($n = 56$ and $n = 421$), which captured significantly more information about lithium responsiveness (AMI range: 0.033 to 0.133) than BD subtype (AMI: 0.004 to 0.011). The smaller cluster had disproportionately more lithium responders ($n = 47$ [83.8%]) when compared to the larger cluster (103 [24.4%]; $p = 0.006$).

Conclusions: Phenotypes derived from detailed clinical data may carry more information about lithium responsiveness than the current classification of diagnostic subtype. These findings support lithium responsiveness as a valid approach to stratification in clinical samples.

KEYWORDS

bipolar disorder, classification, clustering, diagnosis, lithium

1 | INTRODUCTION

An estimated 83% of adults with bipolar disorder (BD) have severe disability and functional impairment.[1] Indeed, during the average 12.8 years followed in a longitudinal study, participants were symptomatically ill 47% of the time.[2] Improving treatment outcomes is a priority for researchers, but advancements are impeded by the largely unknown etiology of BD. Efforts to understand the underlying basis of BD have been limited by the clinical heterogeneity of the disorder, and the results of biomarker and genetic studies have widely varied.[3,4] In order to reduce the heterogeneity, some suggest identifying subtypes of BD using phenotypic data.[5–7]

Currently, most countries use either the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)[8] or the International Classification of Diseases, 11th Revision (ICD-11)[9] to classify mental disorders. Both manuals divide BD into two groups: Type I and Type II, with the former defined by the occurrence of at least one manic episode, and the latter is defined by both hypomania and at least one depressive episode. The DSM introduced BD-II with its fourth edition in 1994,[10] but the ICD only recently adopted the dichotomy in 2019.[9] Even so, the assumption that BD-II is a valid clinical diagnosis, of which the sole differentiating factor is the severity of mania, remains a point of contention.[11,12]

The BD-I/II diagnoses originated from the anecdotal accounts of clinicians who noticed two distinct groups of hospitalized patients: those experiencing mania, and those experiencing depressive episodes who had a history of elevated states (hypomania) that did not require hospitalization.[13] When the diagnoses were added to the DSM, it implemented a 4-day cut-off for hypomania, which was not decided through evidence-based evaluations, but rather due to concerns of overdiagnosis.[13] This arbitrary cut-off remains a criterion for BD-II diagnosis.[8,9] For mood disorders, the categorical classification method often results in misdiagnoses. The majority of BD cases begin with a depressive mood episode,[14] which can lead to an initial diagnosis of major depressive disorder (MDD).[15] It is only once the individual experiences an elevated mood episode that a diagnosis of BD-I or BD-II may be given. Aside from duration, the severity of the episode is used to differentiate hypomania or mania. Specifically, the presence of psychotic features, degree of

Significant outcomes

- The phenotypic clusters generated from detailed clinical data were significantly more informative about lithium treatment responsiveness than diagnostic subtype of bipolar disorder.
- Furthermore, the cluster comparisons revealed differential response to lithium treatment.

Limitations

- There were varying degrees of missingness across variables, particularly in the comorbid conditions, and the distribution of missingness was unequal between clusters.
- There may be small effects that were not detected in the univariate comparisons of clusters due to the heterogeneous nature of bipolar disorder.
- Replication is needed to support the present study's findings.

functional impairment, and need for hospitalization are considered when assessing severity.[8] Being hospitalized qualifies an elevated mood episode as mania. However, access to psychiatric care varies substantially and often depends on factors other than clinical needs.[16]

Still, there is support for the dichotomization of BD I and II since several phenotypic features segregate across these subtypes.[17–19] Compared to BD-I, individuals with BD-II exhibit more frequent depressive episodes,[20] higher number of overall mood episodes,[21–23] higher prevalence of rapid cycling,[23–25] more anxiety and substance abuse disorder comorbidities,[20,23,26] higher prevalence in women,[20,27] later age of onset,[17,20,28] more frequent family history of mood disorders,[17,28,29] lower incidence of psychosis,[20,21,30] and fewer hospitalizations.[17,21,30] In addition, both BD-I and BD-II are said to "breed true", as individuals with BD-I tend to have more relatives with BD-I, whereas those with BD-II tend to have more relatives with BD-II.[31–33] This fact suggests potential genetic differences between diagnostic subtypes, strengthening the argument for the BD-I and BD-II divide.

Alternatively, lithium treatment responsiveness has been proposed as a subtype of BD.[34] It is argued that valid subtypes of BD should have implications for patient outcomes, particularly treatment response.[12] Although early research regarding response to lithium treatment and subtype suggested that individuals with BD-II responded more favourably than BD-I,[35] subsequent studies have reported similar benefits for both subtypes from common pharmacological treatments including lithium,[36,37] lamotrigine,[38] and quetiapine.[39,40] In the absence of differential treatment efficacy, the currently defined diagnostic subtypes appear to simply describe two general presentations of BD. Like the diagnostic subtypes, lithium treatment response is associated with specific clinical characteristics, wherein individuals who respond well to lithium treatment tend to have a completely episodic clinical course and no history of rapid cycling, compared to those who are poor lithium responders.[41] In addition, lithium response is familial in nature,[42] and ongoing research continues to identify contributing genetic variants.[43] There is also evidence of cellular differences related to lithium response, as demonstrated by the hyperexcitability in hippocampal neurons derived from induced pluripotent stem-cells of subjects with BD, where lithium treatment could reverse this characteristic in the cells of only the lithium responders.[44] Given that these biological and clinical phenotypes are associated with contrasting treatment efficacy, lithium responsiveness may serve a more clinically useful subtype than the BD-I/II classification. On the other hand, excellent lithium responders are approximately 30% of the BD population,[45] making the majority of BD patients non-responders. Treatment of lithium non-responders can be difficult because they often have more complex clinical profiles, such as having comorbid conditions, serious functional impairment, and a chronic course of illness.[46] There is evidence suggesting that positive therapeutic response to lithium and other pharmacological interventions is more likely in the early stages of BD, before the illness becomes chronic.[36,47] Although, there are conflicting results reporting no change in response in later stages of illness.[48,49] If the effectiveness of lithium treatment throughout the progression of BD varies, its utility as a clinical descriptor is less certain.

Both methods of categorizing BD have their own merits and support, but stratifying by diagnostic subtype versus lithium treatment responsiveness has not been empirically evaluated in clinical data. In practice, patients present with individual phenotypes that clinicians assess to determine the appropriate diagnosis or classification. While each classification is associated with specific characteristics, it remains unclear whether phenotypic profiles of BD are more informative about the diagnostic subtype or lithium response. In the proposed study, clinical profiles were generated from detailed participant data using clustering algorithms in order to determine whether the profiles contained more information about diagnostic subtype or lithium treatment responsiveness.

2 | MATERIALS AND METHODS

2.1 | Data

With patients' informed consent and institutional ethical approval, we collected data from patients with BD involved in longitudinal cohorts across four centres. These included the Mood Disorders Program at Nova Scotia Health (Maritime Bipolar Registry), the Mood Disorders Program at McGill University Health Centre, the Mood Disorders Centre of Ottawa, and the former Hamilton Psychiatric Hospital. We have described the clinical assessment procedure and samples elsewhere.[41] Patients with BD included in the present study were treated with lithium as their principal prophylactic mood stabilizer for at least 1 year; the use of additional mood stabilizer medications was not permitted for the minimum 1-year period to qualify, but antidepressants and/or antipsychotics were allowed as adjunct medication, if needed. We chose to measure prophylactic treatment response due to the episodic nature of BD and the detrimental impact that recurrences have on those affected.[50]

Our clinical phenotypic data included 40 features, covering demographics, clinical course, family history,

suicide behaviour, and commonly comorbid conditions of physical and mental health. Treatment responsiveness was assessed using the Retrospective Assessment of Lithium Response Phenotype Scale ("Alda scale").[51] This scale is broken down into two subscales. The "A score" provides a unidimensional evaluation of overall improvement, and ranges from 0 (no improvement) to 10 (complete response, no recurrences during adequate treatment, no residual symptoms, and full functional recovery). The "B score" functions as a five-item penalty for factors that confound lithium's relationship with response on the A score. The B items can be scored 0, 1, or 2 points each. B1 scores the number of episodes before/off the treatment, where "0" = 4 or more episodes, "1" = 2 or 3 episodes, and "2" = 1 episode. B2 scores the frequency of episodes before/off the treatment, where "0" = average to high, including rapid cycling, "1" = low, spontaneous remissions of 3 or more years on average, and "2" = 1 episode only, risk of recurrence cannot be established. B3 scores the duration of the treatment, where "0" = 2 or more years, "1" = 1 to 2 years, and "2" = less than 1 year. B4 scores the compliance during period(s) of stability, where "0" = Excellent, for example, documented by drug levels in the therapeutic range, "1" = good, more than 80% levels in the therapeutic range, and "2" = poor, repeatedly off treatment, less than 80% levels in the therapeutic range. B5 scores the use of additional medication during the period of stability, where "0" = none except infrequent sleep medication (1 per week or less); no other mood stabilizers, antidepressants, or antipsychotics for control of mood symptoms, "1" = low-dose antidepressants or antipsychotics for control of mood symptoms, and "2" = prolonged or systematic use of an antidepressant or antipsychotic. The total Alda score is computed as $max(AScore - BScore, 0)$. In the present study, we define lithium responsiveness categorically as an Alda Score of ≥7. This is necessary to facilitate comparison of empirically derived phenotypic cluster classes with lithium responsiveness and bipolar I/II diagnoses. This threshold was previously shown to have good interrater reliability,[51] and using different thresholds did not change the results in sensitivity analysis. Variable definitions are available in the Supporting Information.

2.2 | Demographics and sample characteristics

Feature distributions were summarized across all subjects, as well as stratified separately by (A) bipolar I versus II diagnosis and (B) lithium responsiveness. Continuous variables were summarized using means and standard deviations, and categorical variables as counts and percentages. Comparison of feature distributions across strata used *K*-sample permutation tests for continuous variables, and randomization chi-square tests for categorical variables. Monte Carlo simulations (B = 2000) were used with both tests; therefore, the smallest possible *p*-value was 0.0005 before correction by the Benjamini-Hochberg method.

2.3 | Clustering analysis

All analyses were conducted in R,[52] using the "tableone",[53] "mice",[54] "cluster",[55] and "aricode"[56] packages. Before imputation, several variables were adjusted. Episode frequency for each of the lifetime episode variables (i.e., depressive, (hypo)manic, mixed, multiphasic, total) was determined as the number of episodes divided by illness duration (age minus age of onset). Variables that were excluded from being used as imputers were age, and any variables associated with lithium treatment other than the categorical lithium response variable (ie. all Alda scores, episodes before and while treated with lithium), to avoid biasing the imputed information towards lithium response.

Prior to conducting the clustering analysis, the variables of categorical lithium response and diagnosis (BD-I or BD-II) were removed, since these variables are the "ground truth" classes against which cluster assignments were compared. To ensure that information about lithium response scores would not be included in the feature set being clustered, the Alda scores, and number of episodes before and on lithium treatment were excluded. Due to the mixed nature of our data, Gower's distance was utilized to calculate dissimilarity,[57] and hierarchical clustering was performed using Ward's linkage method.[58] The optimal number of clusters was determined by maximization of silhouette scores. To identify the degree to which assigned clusters captured information about lithium responsiveness or bipolar subtype, the assigned clusters were compared to the categorically defined lithium response and BD-I versus BD-II classification, respectively. This was done using adjusted mutual information (AMI) scores[59] which were calculated to quantify the amount of information contained in the cluster assignments about diagnosis and lithium response. The AMI scores were bootstrapped (B = 1000) to generate 95% confidence intervals (CIs) for all five imputed datasets.

2.4 | Evaluation of phenotypes across clusters

To evaluate the distribution of phenotypic profiles across clusters, we performed univariate comparisons of the

cluster assignments across each feature in our dataset using the same statistical tests described for the demographics and sample characteristics. The descriptive statistics of comparisons across imputed datasets were pooled according to Rubin's Rules,[60] and the Median-P-Rule[61] was used for the reported *p*-values.

3 | RESULTS

3.1 | Demographics and sample characteristics

A detailed description of the sample characteristics is shown in Table 1. Overall, the sample included 477 adults diagnosed with BD-I ($n = 346$) or BD-II ($n = 131$) and classified as a lithium responder (LiR; $n = 150$) or non-responder (LiNR; $n = 327$). The average duration of lithium treatment for LiRs was 12.5 ± 9.4 years (ranging from 1 to 38 years), and the average duration for LiNRs was 1.9 ± 0.4 years (ranging from 1 to 2 years).

When stratified by BD diagnosis type, significant differences were observed between groups, after correction. The age of onset for (hypo)manic episodes was younger in the BD-I group (M = 28.20 [SD = 10.52]) compared to BD-II (33.08 [12.64], $p = 0.018$). The groups significantly differed regarding clinical course ($p = 0.018$) with a completely episodic clinical course being more prevalent for BD-I (36.2%; BD-II: 22.9%) and chronic clinical course more prevalent for BD-II (49.5%, BD-I: 28.9%). The majority of both groups had a depressive polarity of onset (BD-I: 52.7%, BD-II: 81.7%); however, a (hypo)manic episode at onset was more common in BD-I (36.9%, BD-II: 15.1%; $p = 0.006$). A history of rapid cycling was more prevalent for BD-II (45.4%) compared to BD-I (22.1%, $p = 0.006$), while psychosis was more often observed in the BD-I group (73.6%) than in the BD-II group (17.4%, $p = 0.006$). A higher B1 score (0.37 [0.64]) on the Alda scale was present for BD-I, compared to BD-II (0.19 [0.47], $p = 0.022$), which corresponds to fewer episodes experienced before/off lithium treatment for BD-I. Lastly, the prevalence of migraines was increased for BD-II (29.9%) than for BD-I (15.6%, $p = 0.023$).

Significant group effects were evident for lithium responsiveness as well. Individuals in the LiR group were older (50.27 [13.65]) in comparison to the LiNRs (45.19 [13.18], $p = 0.002$). Furthermore, the age of illness onset was significantly later for LiRs (26.20 [9.05]) than for LiNRs (23.45 [9.43], $p = 0.020$). Similarly, the age of the first depressive episode was older for LiRs (27.87 [9.71]) compared to LiNRs (24.64 [9.75], $p = 0.004$). LiRs were more likely to have a completely episodic clinical course (75.7%), whereas LiNRs generally fit into one of three categories of clinical course: chronic (39.6%), completely episodic (22.6%), or episodic with residual symptoms (34.9%, $p = 0.002$). In addition to experiencing a later onset, the frequency of depressive episodes was decreased in LiRs (0.18 [0.19]) compared to LiNR (0.32 [0.39], $p = 0.002$), and a history of rapid cycling was less prevalent in LiRs (7.5%, LiNR: 35.2%; $p = 0.002$). LiRs were more likely to have a first-degree relative with BD (36.7%) when compared to LiNRs (25.4%, $p = 0.035$). Expectedly, LiRs had significantly fewer episodes when receiving lithium treatment (0.52 [1.03]) in comparison with LiNRs (5.32 [7.75], $p = 0.002$). Finally, generalized anxiety disorder affected a smaller proportion of LiRs (14.7%) than LiNRs (33.9%, $p = 0.008$).

3.2 | Concordance between phenotypic clusters, bipolar diagnostic subtype, and treatment responsiveness

The hierarchical clustering analysis determined that the data fit two clusters (silhouette score range: 0.22 to 0.24). The AMI scores for lithium responsiveness and diagnostic subtype were computed for all five imputed datasets (Figure 1). The results consistently demonstrated that the cluster assignments contained significantly more information regarding lithium treatment response compared to BD type.

3.3 | Distribution of phenotypic profiles across clusters

The pooled results of univariate comparisons across all imputed datasets are shown in Table 2. The results of each individual dataset can be found in Tables S1-S5.

Regarding demographic characteristics, both Cluster 1 ($n = 56$) and Cluster 2 ($n = 421$) had similar ages and sex distributions. There were no significant differences in diagnosis between clusters. However, individuals in Cluster 1 were more likely to be lithium responders (83.8%) compared to those in Cluster 2 (24.4%, $p = 0.006$). Participants in Cluster 1 had greater clinical improvement while treated with lithium, indicated by the higher A score (8.70 [SD = 1.82]) and total lithium score (7.92 [2.56]) compared to Cluster 2 (A score: 5.72 [3.18], $p = 0.006$; total score: 3.79 [3.19], $p = 0.006$). Additionally, Cluster 1 showed lower B1 (0.06 [0.27]), B2 (0.08 [0.30]), and B5 (0.56 [0.66]) scores than Cluster 2 (B1: 0.36 [0.62], $p = 0.007$; B2: 0.34 [0.61], $p = 0.013$; B5: 1.35 [0.84], $p = 0.006$), conveying that individuals in Cluster 1 experienced more mood episodes before/off lithium treatment, had a higher frequency of episodes before/off

TABLE 1 Sample demographics and characteristics.

	Overall ($N = 477$)	BD-I ($N = 346$)	BD-II ($N = 131$)	*p*	Corrected-*p*	LiNR ($N = 327$)	LiR ($N = 150$)	*p*	Corrected-*p*
Age	46.79 (13.53)	45.88 (13.88)	49.19 (12.28)	0.021	0.072	45.19 (13.18)	50.27 (13.65)	< 0.001	**0.002***
Sex = Female	276 (57.9)	193 (55.8)	83 (63.4)	0.144	0.266	200 (61.2)	76 (50.7)	0.036	0.079
Diagnosis = BD-I	346 (72.5)	346 (100.0)	0 (0.0)	—	—	237 (72.5)	109 (72.7)	1.000	1.000
SES				0.167	0.284			0.025	0.068
Disabled	90 (25.6)	70 (27.3)	20 (20.8)			81 (28.5)	9 (13.2)		
Employment insurance	32 (9.1)	25 (9.8)	7 (7.3)			27 (9.5)	5 (7.4)		
Employed	100 (28.4)	74 (28.9)	26 (27.1)			75 (26.4)	25 (36.8)		
Other	27 (7.7)	15 (5.9)	12 (12.5)			20 (7.0)	7 (10.3)		
Retired	38 (10.8)	23 (9.0)	15 (15.6)			26 (9.2)	12 (17.6)		
Social assistance	47 (13.4)	35 (13.7)	12 (12.5)			38 (13.4)	9 (13.2)		
Student	18 (5.1)	14 (5.5)	4 (4.2)			17 (6.0)	1 (1.5)		
Age of onset	24.32 (9.39)	23.95 (8.99)	25.29 (10.34)	0.157	0.280	23.45 (9.43)	26.20 (9.05)	0.005	**0.020***
Onset depression	25.58 (9.84)	25.29 (9.47)	26.29 (10.70)	0.351	0.501	24.64 (9.75)	27.87 (9.71)	0.001	**0.004***
Onset (hypo)mania	29.47 (11.30)	28.20 (10.52)	33.08 (12.64)	< 0.001	**0.006***	28.90 (11.22)	30.79 (11.43)	0.091	0.153
Illness duration	22.45 (12.73)	21.94 (13.13)	23.78 (11.53)	0.169	0.284	21.70 (12.50)	24.07 (13.11)	0.053	0.108
Clinical course				0.002	**0.018***			< 0.001	**0.002***
Chronic	135 (34.4)	83 (28.9)	52 (49.5)			126 (39.6)	9 (12.2)		
Completely episodic	128 (32.7)	104 (36.2)	24 (22.9)			72 (22.6)	56 (75.7)		
Continuous cycling	1 (0.3)	1 (0.3)	0 (0.0)			1 (0.3)	0 (0.0)		
Episodic with residual symptoms	119 (30.4)	90 (31.4)	29 (27.6)			111 (34.9)	8 (10.8)		
Single episode	9 (2.3)	9 (3.1)	0 (0.0)			8 (2.5)	1 (1.4)		
Frequency of depressions	0.27 (0.34)	0.25 (0.32)	0.35 (0.40)	0.010	0.052	0.32 (0.39)	0.18 (0.19)	0.001	**0.002***
Frequency of manias	0.23 (0.49)	0.26 (0.53)	0.14 (0.34)	0.030	0.095	0.22 (0.27)	0.24 (0.77)	0.795	0.872
Frequency of mixed	0.01 (0.04)	0.01 (0.05)	0.00 (0.00)	0.058	0.144	0.01 (0.04)	0.01 (0.05)	0.668	0.764
Frequency of multiphasic	0.03 (0.07)	0.03 (0.07)	0.03 (0.08)	0.980	1.000	0.02 (0.06)	0.03 (0.08)	0.219	0.298
Frequency of total episodes	0.57 (0.71)	0.57 (0.70)	0.60 (0.73)	0.705	0.837	0.62 (0.65)	0.48 (0.80)	0.078	0.139
Polarity of onset				< 0.001	**0.006***			0.044	0.093
Biphasic (D-M)	11 (2.4)	7 (2.1)	4 (3.2)			6 (1.9)	5 (3.5)		
Biphasic (M-D)	24 (5.2)	24 (7.1)	0 (0.0)			17 (5.3)	7 (4.9)		
Depressive	280 (60.6)	177 (52.7)	103 (81.7)			208 (65.0)	72 (50.7)		

TABLE 1 (Continued)

	Overall ($N = 477$)	BD-I ($N = 346$)	BD-II ($N = 131$)	p	Corrected-p	LiNR ($N = 327$)	LiR ($N = 150$)	p	Corrected-p
(Hypo)manic	143 (31.0)	124 (36.9)	19 (15.1)			86 (26.9)	57 (40.1)		
Mixed	4 (0.9)	4 (1.2)	0 (0.0)			3 (0.9)	1 (0.7)		
Rapid cycling	117 (28.1)	68 (22.1)	49 (45.4)	< 0.001	**0.006***	109 (35.2)	8 (7.5)	< 0.001	**0.002***
Psychosis	236 (49.5)	217 (73.6)	19 (17.4)	< 0.001	**0.006***	183 (59.0)	53 (56.4)	0.705	0.788
FDR with mood disorder	235 (55.8)	159 (52.5)	76 (64.4)	0.034	0.103	160 (56.5)	75 (54.3)	0.670	0.764
FDR with BD	138 (28.9)	98 (28.3)	40 (30.5)	0.663	0.821	83 (25.4)	55 (36.7)	0.010	**0.035***
Total FDR	6.08 (3.46)	5.86 (3.42)	6.66 (3.49)	0.020	0.072	5.89 (3.38)	6.50 (3.60)	0.075	0.139
Episodes pre-Li	6.11 (7.96)	5.46 (7.69)	8.06 (8.47)	0.020	0.072	5.72 (7.36)	7.54 (9.79)	0.127	0.196
Episodes on Li	4.28 (7.16)	4.11 (6.84)	4.79 (8.05)	0.517	0.702	5.32 (7.75)	0.52 (1.03)	< 0.001	**0.002***
Li response = Responder	150 (31.4)	109 (31.5)	41 (31.3)	1.000	1.000	0 (0.0)	150 (100.0)	—	—
Suicide attempts	0.54 (1.12)	0.52 (1.09)	0.60 (1.20)	0.559	0.741	0.61 (1.23)	0.30 (0.63)	0.019	0.057
Suicide ideation	171 (46.1)	119 (44.1)	52 (51.5)	0.246	0.390	137 (48.4)	34 (38.6)	0.113	0.180
Social anxiety	80 (20.7)	58 (20.6)	22 (21.0)	1.000	1.000	72 (22.9)	8 (11.0)	0.027	0.069
Panic disorder	85 (20.5)	59 (19.3)	26 (23.9)	0.339	0.501	73 (23.0)	12 (12.4)	0.028	0.069
Generalized anxiety	117 (30.2)	83 (29.3)	34 (32.4)	0.625	0.803	106 (33.9)	11 (14.7)	0.002	**0.008***
OCD	38 (9.2)	24 (7.8)	14 (13.0)	0.134	0.258	34 (10.7)	4 (4.1)	0.077	0.139
Substance abuse	130 (30.9)	101 (32.8)	29 (25.7)	0.183	0.299	106 (33.3)	24 (23.3)	0.065	0.129
ADHD	20 (5.2)	15 (5.3)	5 (4.8)	1.000	1.000	19 (6.0)	1 (1.4)	0.158	0.226
Learning disability	19 (4.9)	14 (5.0)	5 (4.8)	1.000	1.000	18 (5.7)	1 (1.4)	0.146	0.214
Primary insomnia	48 (12.3)	28 (9.9)	20 (19.0)	0.018	0.072	42 (13.3)	6 (8.1)	0.246	0.327
Personality disorder	52 (13.4)	37 (13.1)	15 (14.3)	0.877	1.000	47 (15.0)	5 (6.8)	0.090	0.153
Diabetes	38 (10.2)	29 (10.7)	9 (8.7)	0.693	0.837	32 (10.5)	6 (8.8)	0.824	0.886
Hypertension	58 (15.7)	39 (14.6)	19 (18.6)	0.431	0.600	45 (14.9)	13 (19.7)	0.363	0.450
Menstrual abnormality	62 (31.6)	44 (32.8)	18 (29.0)	0.634	0.803	53 (31.7)	9 (31.0)	1.000	1.000
Thyroid disease	114 (31.0)	73 (27.2)	41 (41.0)	0.012	0.057	93 (30.9)	21 (31.3)	1.000	1.000
Head injury	85 (26.3)	70 (30.0)	15 (16.7)	0.013	0.059	73 (27.1)	12 (22.2)	0.504	0.598
Migraine	69 (19.5)	40 (15.6)	29 (29.9)	0.004	**0.023***	63 (21.9)	6 (9.1)	0.032	0.074

Note: Categorical variables reported as N (%) and continuous variables reported as Mean (SD). Corrected for multiple comparisons using the Benjamini-Hochberg method. Smallest possible p-value before correction was 0.0005. Frequency of episode variables are the number of specified episodes divided by years of illness.

Abbreviations: BD-I, bipolar I disorder; BD-II, bipolar II disorder; Li, lithium; LiNR, lithium non-responder (Li total score <7); LiR, lithium responder (Li total score); (hypo)mania, includes both hypomania and mania; SES, socioeconomic status; Biphasic (D-M) or (M-D), biphasic mood episode depressive to (hypo)manic or (hypo)manic to depressive, respectively; FDR, first-degree relative; OCD, obsessive-compulsive disorder; ADHD, attention deficit hyperactivity disorder.

*Corrected-$p < 0.05$.

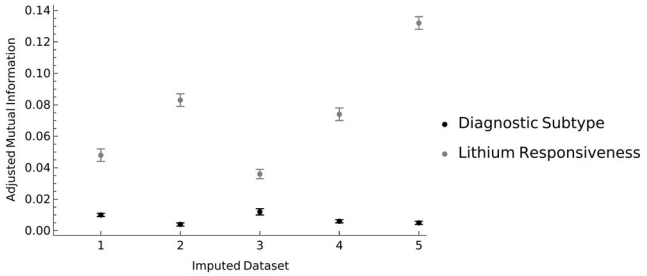

FIGURE 1 Adjusted mutual information of each imputed dataset for bipolar disorder diagnostic subtype and lithium responsiveness from clusters.

treatment, and did not require the use of additional medication to manage their symptoms as often as those in Cluster 2. On the other hand, a history of rapid cycling was less common in Cluster 1 (7.6%) than in Cluster 2 (29.6%, $p = 0.041$). Cluster 1 had a higher frequency of multiphasic episodes (0.05 [0.10]) compared to Cluster 2 (0.02 [0.06], $p = 0.021$).

Furthermore, Cluster 1 exhibited a higher prevalence of attention-deficit hyperactivity disorder (ADHD, 84.8%), learning disabilities (50.0%), and head injury (85.3%) compared to Cluster 2 (ADHD: 4.0%, $p = 0.006$; learning disability: 4.2%, $p = 0.013$; head injury: 25.5%, $p = 0.029$). It is important to note, once clustered, Cluster 1 had significantly more missingness in these three variables than Cluster 2 had. When considering missing values, Cluster 1 had a higher prevalence of ADHD and learning disabilities (ADHD: Yes = 8.2%, No = 1.9%, Missing = 90.0%; learning disability: Yes = 5.0%, No = 5.0%, Missing = 90.0%) compared to Cluster 2 (ADHD: Yes = 3.7%, No = 87.2%, Missing = 9.1%; learning disability: Yes = 3.9%, No = 86.8%, Missing = 9.4%). However, Cluster 1 showed a lower prevalence of head injury (Yes = 7.1%, No = 1.5%, Missing = 91.4%) than Cluster 2 (Yes = 19.3%, No = 56.4%, Missing = 24.4%). Detailed results of variables for which the clusters largely differed in missingness are available in Table S6.

4 | DISCUSSION

The present study has shown that a data-driven induction of phenotypic clusters, based on detailed and prospectively verified clinical profiles of patients with BD, identifies more information about treatment responsiveness than the BD-I/II subtypes. Furthermore, univariate comparisons found differential lithium responsiveness, not diagnostic subtype, between phenotypic clusters. In the following, we discuss the differences observed during preliminary group comparisons, implications of the clustering results, how the findings relate to the current literature, study limitations, and recommendations for subsequent research.

Our initial sample stratification by diagnostic subtype and lithium response revealed several group differences consistent with prior literature. Participants with BD-I were more likely to experience psychotic features, a (hypo)manic polarity of onset, and a younger age at the first (hypo)manic episode.[17] Individuals with BD-II more often experienced comorbid migraine,[62] and were more likely to have a history of rapid cycling.[63] Differences in clinical course included BD-I being associated with a completely episodic course or episodic course with residual symptoms, whereas the BD-II group most often had a chronic course of illness.[64]

When stratified by lithium treatment response, LiRs primarily experienced a completely episodic clinical course with no history of rapid cycling and a later age of illness onset.[36,65] Responders were also older during their first depressive episode, which could be explained by the later age of onset. However, LiRs had less frequent depressive episodes than LiNRs, aligning with prior reports.[66] Additionally, LiRs were more likely to have a first-degree relative diagnosed with BD.[36] In our sample, the LiNRs were younger in age at the time of interview, but neither group differed regarding the duration of illness. Moreover, LiNRs had a higher prevalence of generalized anxiety disorder when compared to LiRs.[67] Finally, the initial group stratification did not reveal varied lithium treatment response by diagnostic subtype, suggesting lithium therapy is similarly effective for BD-II as for BD-I.[36,37] These consistencies indicate our sample was comparable to the populations described in past studies.

The clustering analysis found the phenotypic profiles contained differential lithium responsiveness independent of diagnostic subtype, suggesting that the subtypes do not accurately capture the clinical variation in BD. This is likely a consequence of the diagnostic subtypes being defined by a single dimension. Considering the substantial heterogeneity observed with BD, the dissimilarities between the clusters are likely due to subtle differences across many variables rather than large effects from a select few. Instead, the results indicate that the clinical phenotypes were more informative of lithium treatment response, meaning lithium response and its associated features were driving the phenotypic variation of BD in this case. The effects observed in the B subscale items suggest that in addition to the greater clinical improvement from lithium treatment in the predominantly responder profile, there were fewer confounding factors related to the number and frequency of mood episodes before/off treatment and the use of additional

TABLE 2 Comparisons across phenotypic clusters.

	Cluster 1 ($N = 56$)	Cluster 2 ($N = 421$)	Median p-value
Age	46.91 (14.14)	46.75 (13.44)	0.769
Sex = Female	28.6 (50.9)	247.4 (58.8)	0.930
Diagnosis = BD-I	39.6 (70.5)	306.4 (72.8)	0.861
SES			0.556
Disabled	0.6 (7.0)	89.4 (25.9)	
EI	0.2 (2.0)	31.8 (9.2)	
Employed	4.6 (67.5)	95.4 (27.7)	
Other	0.2 (2.5)	26.8 (7.8)	
Retired	1.0 (11.5)	37.0 (10.7)	
Social assistance	0.8 (9.5)	46.2 (13.4)	
Student	0.0 (0.0)	18.0 (5.2)	
Age of onset	25.00 (8.14)	24.21 (9.53)	0.916
Onset depression	27.06 (8.70)	25.41 (9.95)	0.584
Onset (hypo)mania	29.88 (10.48)	29.42 (11.38)	0.742
Illness duration	21.94 (12.40)	22.50 (12.78)	0.869
Clinical course			0.511
Chronic	1.8 (19.5)	133.2 (34.8)	
Completely episodic	5.2 (57.1)	122.8 (32.1)	
Continuous cycling	0.0 (0.0)	1.0 (0.3)	
Episodic with residual symptoms	1.8 (19.6)	117.2 (30.6)	
Single episode	0.4 (3.8)	8.6 (2.2)	
Frequency of depressions	0.22 (0.30)	0.28 (0.34)	0.564
Frequency of (hypo)manias	0.15 (0.16)	0.24 (0.52)	0.452
Frequency of mixed	0.00 (0.00)	0.01 (0.04)	0.391
Frequency of multiphasic	0.05 (0.10)	0.02 (0.06)	**0.021***
Frequency of any episode	0.49 (0.48)	0.59 (0.73)	0.555
Polarity of onset			0.444
Biphasic (D-M)	2.6 (5.1)	8.4 (2.0)	
Biphasic (M-D)	3.4 (6.4)	20.6 (5.0)	
Depressive	30.6 (58.9)	249.4 (60.8)	
Manic	15.4 (29.6)	127.6 (31.1)	
Mixed	0.0 (0.0)	4.0 (1.0)	
Rapid cycling	2.2 (7.6)	114.8 (29.6)	**0.041***
Psychosis	6.4 (32.0)	229.6 (59.8)	0.081
FDR with mood disorder	21.6 (40.7)	213.4 (58.0)	0.181
FDR with BD	18.8 (33.4)	119.2 (28.3)	0.679
Total FDR	6.18 (3.54)	6.07 (3.45)	0.951
Episodes pre-lithium	4.05 (NA)	6.15 (8.02)	0.554
Episodes on lithium	4.22 (3.51)	4.28 (7.19)	0.851
Lithium response = Responder	47.2 (83.8)	102.8 (24.4)	**0.006***
Lithium A score	8.70 (1.82)	5.72 (3.18)	**0.006***
Lithium B1 score	0.06 (0.27)	0.36 (0.62)	**0.007***
Lithium B2 score	0.08 (0.30)	0.34 (0.61)	**0.013***

(Continues)

TABLE 2 (Continued)

	Cluster 1 ($N = 56$)	Cluster 2 ($N = 421$)	Median p-value
Lithium B3 score	0.06 (0.31)	0.26 (0.59)	0.075
Lithium B4 score	0.04 (0.23)	0.19 (0.50)	0.097
Lithium B5 score	0.56 (0.66)	1.35 (0.84)	**0.006***
Lithium total score	7.92 (2.56)	3.79 (3.19)	**0.006***
Suicide attempts	0.34 (0.67)	0.55 (1.14)	0.606
Suicide ideation	12.0 (52.1)	159.0 (45.7)	0.511
Social anxiety	3.4 (61.1)	76.6 (20.1)	0.190
Panic disorder	1.6 (8.4)	83.4 (21.1)	0.433
Generalized anxiety	4.2 (58.7)	112.8 (29.6)	0.457
OCD	0.6 (3.4)	37.4 (9.5)	0.631
Substance abuse	6.2 (26.6)	123.8 (31.1)	0.708
ADHD	4.6 (84.8)	15.4 (4.0)	**0.006***
Learning disability	2.8 (50.0)	16.2 (4.2)	**0.013***
Primary insomnia	2.6 (36.7)	45.4 (11.9)	0.264
Personality disorder	1.4 (28.6)	50.6 (13.3)	0.490
Diabetes	1.4 (29.9)	36.6 (10.0)	0.430
Hypertension	0.6 (9.0)	57.4 (15.8)	0.861
Menstrual abnormality	1.9 (6.0)	60.2 (16.6)	0.266
Thyroid disease	4.2 (73.0)	109.8 (30.3)	0.118
Head injury	4.0 (85.3)	81.0 (25.5)	**0.029***
Migraine	3.6 (57.3)	65.4 (18.8)	0.263

Note: Pooled results from all datasets. Categorical variables reported as N (%) and continuous variables reported as Mean (SD). Counts may be decimal numbers as the result of pooling. Corrected for multiple comparisons using the Benjamini-Hochberg method. Smallest possible p-value before correction was 0.0005. Frequency of episode variables are the number of specified episodes divided by years of illness.

Abbreviations: BD-I, bipolar I disorder; BD-II, bipolar II disorder; Li, lithium; LiNR, lithium non-responder (Li total score <7); LiR, lithium responder (Li total score); (hypo)mania, includes both hypomania and mania; SES, socioeconomic status; Biphasic (D-M) or (M-D), biphasic mood episode depressive to (hypo) manic or (hypo)manic to depressive, respectively; FDR, first-degree relative; OCD, obsessive-compulsive disorder; ADHD, attention deficit hyperactivity disorder; Lithium A score, measure of clinical improvement; Lithium B1 score, number of episodes before/off the treatment; Lithium B2 score, frequency of episodes before/off the treatment; Lithium B3 score, duration of the treatment; Lithium B4 score, compliance during period(s) of stability; Lithium B5 score, use of additional medication during the period of stability.

*Corrected-$p < 0.05$.

medications during treatment. While an overall difference of lithium treatment response was found, it is important to stress that both clusters contained a mixture of responders and non-responders. The distribution of LiRs and LiNRs per cluster likely contributed to the lack of effect found in certain variables such as the number of episodes during lithium treatment, which was present in the sample stratification by lithium responsiveness but absent in the phenotypic cluster comparisons. Other significant differences between clusters included the history of rapid cycling and the frequency of multiphasic episodes. The univariate comparisons of clusters revealed disproportionate prevalence of comorbid ADHD, learning disabilities, and head injury. Specifically, the cluster containing significantly more LiRs had increased prevalence of these conditions, which contradicts the typical clinical profiles observed in individuals who respond well to lithium treatment.[68] When accounting for the missingness in the head injury variable, the prevalence was lower for Cluster 1, while comorbid ADHD and learning disability remained more prevalent. Further investigation revealed that the occurrence of ADHD and learning disability in Cluster 1 was entirely attributed to the small number of LiNRs. In fact, only two responders of the whole sample had either condition, and those individuals were part of Cluster 2.

Overall, our results demonstrate that lithium responsiveness has promise as a method for reducing sample heterogeneity in BD. This is supported by evidence of a biological basis,[44] as well as the intrinsic validity of treatment response as a subtype.[12] Even so, it is important to recognize that the diagnostic subtypes are not entirely

without value. In practice, the BD subtype can serve as a heuristic for clinicians such as when determining the appropriate treatment, since individuals with BD-II are less likely to experience antidepressant-induced (hypo) mania, compared to persons diagnosed with BD-I.[69] Moreover, the BD-II diagnosis may act as a reminder to clinicians to assess for possible history of hypomania when an individual presents with depressive symptoms, therein reducing the population with BD who are misdiagnosed and incorrectly treated for unipolar depression.[70]

Our study's primary limitation was the degree of missingness in the clinical data. In particular, the comorbid conditions exhibited a substantial amount of missing information, with up to 32% missing. We employed strategies to minimize the impact of missing data, including performing multiple imputation by chained equations which has been shown to reliably manage even greater amounts of missingness.[71,72] Also, we repeated the imputation and clustering analysis at several thresholds of the percent of missing values per variable to demonstrate the stability of the AMI results (Figure S1). Nonetheless, there remains the possibility of bias. Furthermore, we were unable to utilize additional datasets to increase our sample size due to between-site heterogeneity that has previously been shown to complicate analyses of phenotypic predictors of lithium responsiveness.[41,73] The objective of the study was to compare the validity of stratification by diagnostic subtype or lithium treatment responsiveness within our sample. Hierarchical clustering emerged as the most suitable method for our analysis, based on factors such as the ability to analyze mixed data,[74,75] robustness to noise and outliers,[76] and performance in smaller samples.[75] However, because the results from hierarchical clustering cannot be extrapolated to external samples, it is crucial to replicate these findings in a well-characterized dataset for validation.

Traditionally, researchers have focused on single traits and predictors rather than a more comprehensive examination of BD, partly due to limited data collection. This narrow approach to BD research neglects the reality that an individual's phenotype is composed of many interconnected traits within a complex network, and that a comprehensive understanding of the phenotypic spectrum of BD may require modelling small differences across many features. Fortunately, there is growing recognition of the need for a more thorough understanding of the disorder. To facilitate in-depth research, more extensive data collection is required, along with the development of standardized procedures for the clinical information to reduce information bias caused by between-site heterogeneity.[77] With larger detailed datasets, investigators may consider characterizing the phenotypic landscape in a continuous framework, instead of relying solely on categorical classification. An emphasis on elucidating the correlation structure across all features of BD and unravelling the network of clinical traits could offer more realistic and powerful representations, which empower us to map out the phenotypic landscape of BD. Additionally, it is essential to explore treatment response beyond lithium therapy and to collect more extensive, longitudinal data to illustrate any change in clinical course and outcomes as the result of various treatments. Subsequent studies could leverage family data, wherein relatedness can serve a tool to discern subgroups with a potential genetic basis.

Thorough phenotypic characterization is crucial for enhancing our comprehension of BD. Identifying subtypes will allow researchers to reduce heterogeneity in samples, making it easier to examine genetic and biological factors associated with particular phenotypes, and the underlying mechanisms and etiology of BD overall. These subtypes could provide predictive value for clinical practice regarding the course of illness and efficacy of specific treatments, leading to improved illness management and prognosis. In summary, a broader strategy is needed to account for and capture the heterogeneity of BD, and to facilitate impactful results with implications for researchers and clinicians.

ACKNOWLEDGEMENTS

This work was supported by the Canadian Institutes of Health Research (Martin Alda, #166098), Dalhousie Medical Research Foundation (Martin Alda, Abraham Nunes), the Dalhousie Faculty of Medicine Graduate Studentship (Martin Alda), Research Nova Scotia (Martin Alda, Abraham Nunes), the Nova Scotia Health Research Fund (Martin Alda), and the Dalhousie Faculty of Medicine Graduate Studentship (Katie Scott).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Katie Scott https://orcid.org/0000-0001-5395-5036
Giulio Emilio Brancati https://orcid.org/0000-0003-4460-0406
Mirko Manchia https://orcid.org/0000-0003-4175-6413
Martin Alda https://orcid.org/0000-0001-9544-3944
Abraham Nunes https://orcid.org/0000-0002-4955-9150

REFERENCES

1. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617-627. doi:10.1001/archpsyc.62.6.617
2. Judd LL, Akiskal HS, Schettler PJ, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry*. 2002;59(6):530-537. doi:10.1001/archpsyc.59.6.530
3. O'Connell KS, Coombes BJ. Genetic contributions to bipolar disorder: current status and future directions. *Psychol Med*. 2021;51(13):2156-2167. doi:10.1017/S0033291721001252
4. Wartchow KM, Cordeiro RC, Scaini G. Advances in the pathophysiology of bipolar disorder. *Curr Opin Psychiatry*. 2023;36(1):20-27. doi:10.1097/YCO.0000000000000836
5. Houenou J, Perlini C, Brambilla P. Epidemiological and clinical aspects will guide the neuroimaging research in bipolar disorder. *Epidemiol Psychiatr Sci*. 2015;24(2):117-120. doi:10.1017/S2045796014000766
6. de Almeida JRC, Phillips ML. Distinguishing between unipolar depression and bipolar depression: current and future clinical and neuroimaging perspectives. *Biol Psychiatry*. 2013;73(2):111-118. doi:10.1016/j.biopsych.2012.06.010
7. Chen YL, Tu PC, Huang TH, et al. Identifying subtypes of bipolar disorder based on clinical and neurobiological characteristics. *Sci Rep Sci Rep*. 2021;11(1):17082. doi:10.1038/s41598-021-96645-5
8. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Fifth ed. American Psychiatric Association; 2013.
9. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 11th ed. 2019 https://icd.who.int/
10. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Fourth ed. American Psychiatric Association; 1994.
11. Gitlin M, Malhi GS. The existential crisis of bipolar II disorder. *Int J Bipolar Disord*. 2020;8(1):5. doi:10.1186/s40345-019-0175-7
12. Ostacher MJ. Bipolar II should only exist if we can actually study treatments of it. Otherwise, what purpose does it serve? *Bipolar Disord*. 2018;20(4):395-396. doi:10.1111/bdi.12628
13. Dunner DL. Bipolar II disorder. *Bipolar Disord*. 2017;19(7):520-521. doi:10.1111/bdi.12567
14. Daban C, Colom F, Sanchez-Moreno J, García-Amador M, Vieta E. Clinical correlates of first-episode polarity in bipolar disorder. *Compr Psychiatry*. 2006;47(6):433-437. doi:10.1016/j.comppsych.2006.03.009
15. Chen H, Zhang L, Xu C, Zhu J, Chen M, Fang Y. Analysis of misdiagnosis of bipolar disorder in an outpatient setting. *Shanghai Arch Psychiatry*. 2018;30(2):93-101. doi:10.11919/j.issn.1002-0829.217080
16. Xu J, Wang J, King M, et al. Rural–urban disparities in the utilization of mental health inpatient services in China: the role of health insurance. *Int J Health Econ Manag*. 2018;18(4):377-393. doi:10.1007/s10754-018-9238-z
17. Tondo L, Miola A, Pinna M, Contu M, Baldessarini RJ. Differences between bipolar disorder types 1 and 2 support the DSM two-syndrome concept. *Int J Bipolar Disord*. 2022;10(1):21. doi:10.1186/s40345-022-00268-2
18. Ha K, Ha TH, Hong KS. Bipolar I and bipolar II: it's time for something new for a better understanding and classification of bipolar disorders. *Can J Psychiatr Rev Can Psychiatr*. 2019;64(8):548-549. doi:10.1177/0706743719861279
19. Nierenberg AA, Bipolar II. Disorder is NOT a myth. *Can J Psychiatr*. 2019;64(8):537-540. doi:10.1177/0706743719852096
20. Serafini G, Gonda X, Aguglia A, et al. Bipolar subtypes and their clinical correlates in a sample of 391 bipolar individuals. *Psychiatry Res*. 2019;281:112528. doi:10.1016/j.psychres.2019.112528
21. Vieta E, Gast C, Otero A, Nieto E, Vallejo J. Differential features between bipolar I and bipolar II disorder. *Compr Psychiatry*. 1997;38(2):98-101. doi:10.1016/S0010-440X(97)90088-2
22. Goodwin FK, Jamison KR. *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression*. Oxford University Press; 2007.
23. Dell'Osso B, Holtzman JN, Goffin KC, et al. American tertiary clinic-referred bipolar II disorder compared to bipolar I disorder: more severe in multiple ways, but less severe in a few other ways. *J Affect Disord*. 2015;188:257-262. doi:10.1016/j.jad.2015.09.001
24. Maj M, Pirozzi R, Formicola AMR, Tortorella A. Reliability and validity of four alternative definitions of rapid-cycling bipolar disorder. *Am J Psychiatry*. 1999;156(9):1421-1424. doi:10.1176/ajp.156.9.1421
25. Baldessarini RJ, Tondo L, Floris G, Hennen J. Effects of rapid cycling on response to lithium maintenance treatment in 360 bipolar I and II disorder patients. *J Affect Disord*. 2000;61(1–2):13-22. doi:10.1016/S0165-0327(99)00196-2
26. Rihmer Z, Szádóczky E, Füredi J, Kiss K, Papp Z. Anxiety disorders comorbidity in bipolar I, bipolar II and unipolar major depression: results from a population-based study in Hungary. *J Affect Disord*. 2001;67(1–3):175-179. doi:10.1016/S0165-0327(01)00309-3
27. Merikangas KR, Jin R, He JP, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry*. 2011;68(3):241-251. doi:10.1001/archgenpsychiatry.2011.12
28. Benazzi F. Bipolar II disorder family history using the family history screen: findings and clinical implications. *Compr Psychiatry*. 2004;45(2):77-82. doi:10.1016/j.comppsych.2003.12.003
29. Endicott J, Nee J, Andreasen N, Clayton P, Keller M, Coryell W. Bipolar II: combine or keep separate? *J Affect Disord*. 1985;8(1):17-28. doi:10.1016/0165-0327(85)90068-0
30. Bega S, Schaffer A, Goldstein B, Levitt A. Differentiating between bipolar disorder types I and II: results from the National Epidemiologic Survey on alcohol and related conditions (NESARC). *J Affect Disord*. 2012;138(1–2):46-53. doi:10.1016/j.jad.2011.12.032
31. Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet*. 2003;123C(1):48-58. doi:10.1002/ajmg.c.20013
32. Cheng CM, Chen MH, Chang WH, et al. Risks of coaggregation of major psychiatric disorders among first-degree relatives of patients with bipolar I and bipolar II disorder: evidence from a nationwide population-based study. *J Clin Psychiatry*. 2021;82(5):e1-e8. doi:10.4088/JCP.20m13810

33. Coryell W, Endicott J, Reich T, Andreasen N, Keller M. A family study of bipolar II disorder. *Br J Psychiatry J Ment Sci*. 1984; 145:49-54. doi:10.1192/bjp.145.1.49
34. Alda M. The phenotypic spectra of bipolar disorder. *Eur Neuropsychopharmacol*. 2004;14(SUPPL. 2):S94-S99. doi:10.1016/j.euroneuro.2004.03.006
35. Tondo L, Baldessarini RJ, Hennen J, Floris G. Lithium maintenance treatment of depression and mania in bipolar I and bipolar II disorders. *Am J Psychiatry*. 1998;155(5):638-645. doi:10.1176/ajp.155.5.638
36. Hui TP, Kandola A, Shen L, et al. A systematic review and meta-analysis of clinical predictors of lithium response in bipolar disorder. *Acta Psychiatr Scand*. 2019;140(2):94-115. doi:10.1111/acps.13062
37. Baldessarini RJ, Tondo L, Hennen J, Floris G. Latency and episodes before treatment: response to lithium maintenance in bipolar I and II disorders 1. *Bipolar Disord*. 1999;1(2):91-97. doi:10.1034/j.1399-5618.1999.010206.x
38. Ginsberg LD. Efficacy and safety of lamotrigine for adults with bipolar disorder in a private practice setting. *CNS Spectr*. 2006; 11(5):376-382. doi:10.1017/s1092852900014504
39. Thase ME, Macfadden W, Weisler RH, et al. Efficacy of quetiapine monotherapy in bipolar I and II depression: a double-blind, placebo-controlled study (the BOLDER II study). *J Clin Psychopharmacol*. 2006;26(6):600-609. doi:10.1097/01.jcp.0000248603.76231.b7
40. Datto C, Pottorf WJ, Feeley L, LaPorte S, Liss C. Bipolar II compared with bipolar I disorder: baseline characteristics and treatment response to quetiapine in a pooled analysis of five placebo-controlled clinical trials of acute bipolar depression. *Ann General Psychiatry*. 2016;15(1):9. doi:10.1186/s12991-016-0096-0
41. Nunes A, Ardau R, Berghöfer A, et al. Prediction of lithium response using clinical data. *Acta Psychiatr Scand*. 2020;141(2): 131-141. doi:10.1111/acps.13122
42. Grof P, Duffy A, Cavazzoni P, et al. Is response to prophylactic lithium a familial trait? *J Clin Psychiatry*. 2002;63(10):942-947. doi:10.4088/JCP.v63n1013
43. Hou L, Heilbronner U, Degenhardt F, et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. *Lancet*. 2016;387(10023): 1085-1093. doi:10.1016/S0140-6736(16)00143-4
44. Mertens J, Wang QW, Kim Y, et al. Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature*. 2015;527(7576):95-99. doi:10.1038/nature15526
45. Garnham J, Munro A, Slaney C, et al. Prophylactic treatment response in bipolar disorder: results of a naturalistic observation study. *J Affect Disord*. 2007;104:185-190. doi:10.1016/j.jad.2007.03.003
46. Lin Y, Maihofer AX, Stapp E, et al. Clinical predictors of non-response to lithium treatment in the pharmacogenomics of Bipolar disorder (PGBD) study. *Bipolar Disord*. 2021;23(8):821-831. doi:10.1111/bdi.13078
47. Berk M, Hallam KT, McGorry PD. The potential utility of a staging model as a course specifier: a bipolar disorder perspective. *J Affect Disord*. 2007;100(1):279-281. doi:10.1016/j.jad.2007.03.007
48. Baldessarini RJ, Tondo L, Hennen J. Treatment-latency and previous episodes: relationships to pretreatment morbidity and response to maintenance treatment in bipolar I and II disorders. *Bipolar Disord*. 2003;5(3):169-179. doi:10.1034/j.1399-5618.2003.00030.x
49. Baethge C, Smolka MN, Gruschka P, et al. Does prophylaxis-delay in bipolar disorder influence outcome? Results from a long-term study of 147 patients. *Acta Psychiatr Scand*. 2003; 107(4):260-267. doi:10.1034/j.1600-0447.2003.02057.x
50. Di Marzo S, Giordano A, Pacchiarotti I, Colom F, Sánchez-Moreno J, Vieta E. The impact of the number of episodes on the outcome of Bipolar disorder. *Eur J Psychiatr*. 2006;20(1):21-28.
51. Manchia M, Adli M, Akula N, et al. Assessment of response to lithium maintenance treatment in bipolar disorder: a consortium on lithium genetics (ConLiGen) report. *PLoS One*. 2013; 8(6):e65636. doi:10.1371/journal.pone.0065636
52. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021 https://www.R-project.org/
53. Yoshida K, Bartel A. Tableone: Create "Table 1" to Describe Baseline Characteristics with or without Propensity Score Weights. 2022 https://CRAN.R-project.org/package=tableone
54. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1-67. doi:10.18637/jss.v045.i03
55. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: Cluster Analysis Basics and Extensions. 2022 https://CRAN.R-project.org/package=cluster
56. Chiquet J, Rigaill G, Sundqvist M. Aricode: Efficient Computations of Standard Clustering Comparison Measures. 2022 https://CRAN.R-project.org/package=aricode
57. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics*. 1971;27(4):857-871. doi:10.2307/2528823
58. Ward J. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*. 1963;58(301):236-244. doi:10.1080/01621459.1963.10500845
59. Vinh NX, Epps J, Bailey J. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J Mach Learn Res*. 2010;11(95):2837-2854.
60. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. 1st ed. Wiley; 1987. doi:10.1002/9780470316696
61. Eekhout I, van de Wiel MA, Heymans MW. Methods for significance testing of categorical covariates in logistic regression models after multiple imputation: power and applicability analysis. *BMC Med Res Methodol*. 2017;17(1):129. doi:10.1186/s12874-017-0404-7
62. Leo RJ, Singh J. Migraine headache and bipolar disorder comorbidity: a systematic review of the literature and clinical implications. *Scand J Pain*. 2016;11(1):136-145. doi:10.1016/j.sjpain.2015.12.002
63. Miola A, Tondo L, Pinna M, Contu M, Baldessarini RJ. Characteristics of rapid cycling in 1261 bipolar disorder patients. *Int J Bipolar Disord*. 2023;11(1):21. doi:10.1186/s40345-023-00300-z
64. Serra G, Koukopoulos A, De Chiara L, et al. Early clinical predictors and correlates of long-term morbidity in bipolar disorder. *Eur Psychiatry J Assoc Eur Psychiatr*. 2017;43:35-43. doi:10.1016/j.eurpsy.2017.02.480
65. Grof P. Selecting effective long-term treatment for Bipolar patients: monotherapy and combinations. *J Clin Psychiatry*. 2003;64(suppl 5):19133.

66. Díaz-Zuluaga AM, Vélez JI, Cuartas M, et al. Ancestry component as a major predictor of lithium response in the treatment of bipolar disorder. *J Affect Disord*. 2023;332:203-209. doi:10.1016/j.jad.2023.03.058
67. Gaudiano BA, Miller IW. Anxiety disorder comobidity in Bipolar I disorder: relationship to depression severity and treatment outcome. *Depress Anxiety*. 2005;21(2):71-77. doi:10.1002/da.20053
68. Bauer M, Grof P, Muller-Oerlinghausen B, eds. *Lithium in Neuropsychiatry: the Comprehensive Guide*. CRC Press; 2013. doi:10.3109/9780203007051
69. Altshuler LL, Suppes T, Black DO, et al. Lower switch rate in depressed patients with bipolar II than bipolar I disorder treated adjunctively with second-generation antidepressants. *Am J Psychiatry*. 2006;163(2):313-315. doi:10.1176/appi.ajp.163.2.313
70. Vieta E, Bipolar II. Disorder: frequent, valid, and reliable. *Can J Psychiatr*. 2019;64(8):541-543. doi:10.1177/0706743719855040
71. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. doi:10.1002/sim.4067
72. Austin PC, van Buuren S. The effect of high prevalence of missing data on estimation of the coefficients of a logistic regression model when using multiple imputation. *BMC Med Res Methodol*. 2022;22(1):196. doi:10.1186/s12874-022-01671-0
73. Nunes A, Stone W, Ardau R, et al. Exemplar scoring identifies genetically separable phenotypes of lithium responsive bipolar disorder. *Transl Psychiatry*. 2021;11(1):1-13. doi:10.1038/s41398-020-01148-y
74. dos Santos TRL, Zárate LE. Categorical data clustering. *Expert Syst Appl Int J*. 2015;42(3):1247-1260. doi:10.1016/j.eswa.2014.09.012
75. Abbas O. Comparisons between data clustering algorithms. *Int Arab J Inf Technol*. 2008;5(3):320-325.
76. Erman N, Korosec A, Suklan J. Performance of selected agglomerative hierarchical clustering methods. *Innov Issues Approach Soc Sci*. 2015;8(1):180-204. doi:10.12959/issn.1855-0541.IIASS-2015-no1-art11
77. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet Lond Engl*. 2002;359(9302):248-252. doi:10.1016/S0140-6736(02)07451-2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Scott K, O'Donovan C, Brancati GE, et al. Phenotypic clustering of bipolar disorder supports stratification by lithium responsiveness over diagnostic subtypes. *Acta Psychiatr Scand*. 2024;1-14. doi:10.1111/acps.13692