Psychological Medicine

cambridge.org/psm

Characterizing the longitudinal course of symptoms and functioning in bipolar disorder

Kristin M. Mignogna[1] and Fernando S. Goes[1,2]

[1]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA and [2]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

Original Article

Cite this article: Mignogna KM, Goes FS (2024). Characterizing the longitudinal course of symptoms and functioning in bipolar disorder. *Psychological Medicine* **54**, 79–89. https://doi.org/10.1017/S0033291722001489

Received: 10 January 2022
Revised: 9 April 2022
Accepted: 5 May 2022
First published online: 14 June 2022

Key words:
Bipolar disorder; clinical course; depression; latent class growth analyses; longitudinal; mania; psychosocial disability; trajectories

Author for correspondence:
Fernando S. Goes,
E-mail: fgoes1@jhmi.edu

Abstract

Background. The course of Bipolar Disorder (BD) is highly variable, with marked inter and intra-individual differences in symptoms and functioning. In this study, we identified illness trajectories across major clinical domains that could have etiological, prognostic, and therapeutic relevance.

Methods. Using the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) study, we performed univariate and multivariate trajectory modeling of depressive symptoms, manic symptoms, and psychosocial functioning. Multinomial regression was performed to identify baseline variables associated with poor outcome trajectories.

Results. Depressive symptoms predominated, with most subjects being found in trajectories characterized by various degrees of depressive symptoms and 13% of subjects being classified in a poor outcome 'persistently depressed' trajectory. Most subjects experienced few manic symptoms, although approximately 10% of subjects followed a trajectory of persistently manic symptoms. Trajectory analysis of psychosocial functioning showed impairment in most of the sample, with little improvement during follow up. Multi-trajectory analyses highlighted significant impairment in subjects with persistently mixed and persistently depressed trajectories of illness. In general, poor outcome trajectories were marked by lower educational attainment, higher unemployment and disability, and a greater likelihood of adverse clinical features (rapid cycling and suicide attempts) and comorbid diagnoses (anxiety disorders, PTSD, and substance abuse/dependence disorders).

Conclusions. Subjects with BD can be classified into several trajectories of clinically relevant domains that are prognostically relevant and show differing degrees of associations with a broad range of negative clinical risk factors. The highest level of psychosocial disability was found in subjects with chronic mixed and depressive symptoms, who show limited improvement despite guideline-based treatment.

Introduction

Bipolar disorder (BD) is a common, heterogeneous disorder marked by complex etiology and highly variable clinical presentation, longitudinal course, and response to treatment. Our ability to predict longitudinal outcome (prognosis) is currently limited, representing a significant challenge for implementing personalized approaches to the treatment of BD. Although subtypes of the disorder have been specified based on lifetime symptoms and severity (bipolar I, bipolar II, schizoaffective disorder, and bipolar not otherwise specified), such categories are often interchangeable across the lifespan and still show prominent prognostic heterogeneity within each subtype. Hence there is a need to identify more clinically useful diagnostic subtypes or domains of illness that better predict longitudinal trajectories of signs, symptoms, and overall wellbeing. However, in part due to the challenges of conducting long-term studies in BD, there have been relatively few longitudinal studies of BD, with most studies relying on a single cross-sectional evaluation.

The importance of clinical course is particularly relevant in studies of adolescents and young adults, when early manifestations of illness are less diagnostic. This has motivated several recent longitudinal studies that have provided important insights of the early manifestation of the disorder. The largest of these studies, the Course and Outcome of Bipolar Youth (COBY) study, followed 413 children and adolescents diagnosed with BD or a BD Spectrum diagnosis for four years, and found a predominantly episodic course for major mood syndromes while also highlighting the presence of a sizeable number of subjects (38% of all participants) with persistent subsyndromal mixed and depressive symptoms (Birmaher et al., 2009). Using structural equation modeling (SEM), the COBY investigators identified four common latent classes marked by subjects with persistent symptoms, a moderate but chronic course, an improving class, and a largely asymptomatic/euthymic class (Birmaher et al., 2009). Similar trajectories were also found in a more recent two-year study

CAMBRIDGE UNIVERSITY PRESS

($N = 146$) of youths with BD spectrum receiving one of two psychosocial interventions in addition to evidence-based psychopharmacological care (Weintraub et al., 2020).

In adult samples, the main longitudinal studies of BD utilizing modern diagnostic methods have been the four-site U.S. Collaborative Depression Study ($N = 139$ with BD I and 75 with BD II), and the Stanley Foundation Study of 935 subjects with BD recruited across four US and three European sites. Both studies found a greater burden of depressive symptoms relative to manic symptoms, emphasizing the frequent symptomatic nature of BD despite naturalistic treatment (Judd et al., 2002, 2003; Post et al., 2021). More recently, Cochran et al., followed a longitudinal sample of 209 subjects with BD-type I and applied machine learning to identify three latent classes characterized by (1) relative euthymia, (2) frequent subsyndromal depression, and (3) mood instability (Cochran, McInnis, & Forger, 2016). However, the most extensive longitudinal study of BD to date has been the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), a long-term multisite outpatient study of over 4000 subjects that included a number of 'nested' trials for BD depression, which provided evidence for the efficacy of adjunctive psychotherapies (Miklowitz et al., 2007) but limited effect of antidepressant medications (Sachs et al., 2007).

While the above studies have found common evidence for worsened course with features such as earlier age at onset, comorbid disorders, and persistent subthreshold symptoms, there has, to our knowledge, not been a systematic evaluation of the major clinical domains of longitudinal course that will ultimately be needed to guide individual or sub-group level prognosis. Hence, in the current study, we sought to leverage the large sample size and longitudinal measurements available in the STEP-BD sample to identify trajectories of illness related to three major domains (depression, mania, and functional capacity) of high clinical and therapeutic relevance to subjects with BD. We have used Latent Class Growth Analysis (LCGA), a longitudinal, class-based extension of structural equation modelling (SEM), to identify potentially more homogenous subtypes of illness with implications for etiological, prognostic, and treatment studies (Cheesman et al., 2018; Kwong et al., 2019). We also apply a multivariate extension of LCGA to analyze multiple outcome variables simultaneously, thereby identifying latent clusters based on similarity in trajectories measured across several domains of illness (Nagin & Odgers, 2010). Such models may be particularly relevant in complex, dynamic disorders such as BD, where symptoms vary within and between traditionally conceptualized 'episodes' of illness.

Methods

Samples

We obtained phenotypic data from the STEP-BD study, a large NIMH-sponsored, multisite, naturalistic study of BD incorporating measurement-based care that also included several randomized treatment 'pathways', depending on the clinical state of the participant. The available dataset included 4360 subjects aged 15 to 82, followed every three months for the first year, then bi-annually for up to 6 subsequent years. Subjects were initially diagnosed with a modified version of the MINI version 5 (Sheehan et al., 1998) and the Affective Disorder Evaluation (Sachs et al., 2003), a semi-structured interview including relevant sections of the SCID (Structured Clinical Interview for DSM Disorders) (Spitzer, Williams, Gibbon, & First 1992) and systematic questions relating to clinical risk factors and overall clinical course. Subjects were followed for an average of 3.5 years; however, since missing data rates correlated with participation duration, we restricted our analyses to the first 12 months of the study to minimize the effect of subject attrition. In addition, to avoid potential bias from the active treatment arms of STEP-BD, we only used subjects in the naturalistic Standardized Care Pathway (SCP) to reflect the course of illness under standard clinical care.

Phenotypes

As outcome measures, we selected the most frequently used mood rating scales for depression and mania [Montgomery–Äsberg Rating Scale (MADRS), Young Mania Rating Scale (YMRS), as well as the Life-Range Impaired Functioning Tool (LRIFT) to measure functional impairment (Leon et al., 1999; Montgomery & Asberg, 1979; Young, Biggs, Ziegler, & Meyer, 1978). MADRS and YMRS are clinician-administered scales of 10 and 11 individual mood-related domains, respectively, that are widely used to measure mood ratings across time in both naturalistic and therapeutic studies. LRIFT represents a scale of psychosocial function based on four domains reflective of daily functioning (work, interpersonal, recreation, and satisfaction with daily activities) with high inter-rater reliability and robust psychometric properties (Leon et al., 1999). All scores were collected at baseline, every three months for each participant's first year of study participation, and every six months until study exit.

After data from the active treatment arms (Randomized Care Pathway) and duplicate time points within samples were removed, there were 3927 samples in the MADRS and 3923 in the YMRS datasets. Of the 3921 samples that overlapped between these two datasets, 2937 samples had data available for at least two time points within the first year of measurements within the SCP. These samples were used for our Latent Class Growth Analyses (LCGA) of MADRS and YMRS. Of these samples, 2483 also had LRIFT measures in at least two-time points and were used for the univariate LRIFT model. Most of these samples, ($N = 2442$) had sufficient ratings for all three analyses (MADRS, YMRS and LRIFT) and were used in the multi-trajectory analysis. The average age of the participants at study entry was 40.5 years (± 12.8 S.D.), while the average age of onset (of depressive or manic symptoms) was 17.3 years (± 8.8 S.D.). Subjects not included in the LCGA showed few differences in covariate associated clinical characteristics (see below) except for baseline history of suicidal ideation (40.9% *v.* 35.7%), a history of rapid cycling (27.7% *v.* 24.2%), and earlier age of onset (16.4 *v.* 17.4 years). There were, however, more prominent differences in baseline demographics, including: higher rates of Black ancestry (7.6% *v.* 4.7%), older interview age (37.3 *v.* 41 years of age), lower likelihood of being married (30.2% *v.* 38.4%) and lower rates of completion of college (36% *v.* 46.0%).

Latent Class Growth Analyses (LCGA)

Longitudinal trajectories were modeled in Mplus version 8 (Muthén & Muthén), using an intercept, a linear slope, and a quadratic slope (in 2-slope models only). We initially attempted to use Growth Mixture Models but found that some, but not all, of the models converged successfully across our main outcome phenotypes. To maintain comparability across all phenotypes, we

https://doi.org/10.1017/S0033291722001489 Published online by Cambridge University Press

elected to perform LCGA analyses, acknowledging that the fixed-variance model was potentially an oversimplification, while often producing more easily interpretable models (Jung & Wickrama, 2008). Fit statistics used to compare models were the Lo-Mendell-Rubin statistic, entropy, loglikelihood, Akaike information criterion, parsimony, and proportion of samples assigned to the smallest class (minimum allowable was 5%). Fit statistics of competing models are shown in online Supplementary Table S1. Models were initially run first with only a linear slope, followed by a linear and quadratic slope. Fit statistics were subsequently compared, and models were visually examined for marked differences in slope showing a substantial change in curvature or class composition with the inclusion of a quadrative slope parameter. For multi-trajectory LCGA, each phenotype was assigned its own latent growth variables (intercept, linear slope, and quadratic slope), with only one class variable regressed on all growth variables. Covariates were not included in the LCGA models in order to be used for the characterization of the derived latent classes.

Covariate analyses

We selected 17 common clinical covariates to test across trajectory classes. For each of the best fitting LCGA models, we performed multinomial regression modeling in STATA version 15 to examine the effects of each covariate on latent class membership, using as a reference the baseline asymptomatic or minimally symptomatic class. Effect sizes are reported as relative risk ratios, and association p values are reported before and after Bonferroni correction for the 17 individual covariates (online Supplementary Tables S2–S5). Self-reported age, sex, race, educational history, and employment status were collected at study entry. Given the small number of subjects in non-white racial categories, we restricted the analyses to these two most prevalent non-white racial groups [Black (5.3%) and Asian (2.1%)]. Clinical features [history of psychosis ('Psychos2' variable), age of onset, and the number of prior depressive and manic episodes] were extracted from the Affective Disorder Evaluation form. Rapid cycling was defined as four or more episodes in the past year, based on a prior STEP-BD publication (Schneck et al., 2008). Lifetime history of comorbid diagnoses (probable or definite) was extracted from the Mini International Psychiatric Interview (Sheehan et al., 1998) performed at study entry. History of alcohol and substance dependence and abuse were reduced to a single binary variable representing a lifetime diagnosis. History of anxiety was defined as having or not having (binary) a history of probable or definite panic disorder, social phobia, or generalized anxiety disorder. As a diagnosis of PTSD may also be a proxy for childhood trauma, this diagnosis was considered a separate category from the other anxiety disorders.

Results

The final sample size after quality control (see *Methods*) was 2938 for the analyses of MADRS and YMRS scores, and 2483 for the analysis of the LRIFT scores, with 2442 samples available for the analyses of all three phenotypes. Of these samples, 65.3% were diagnosed with Bipolar Disorder type I (BD-I), 26.8% with Bipolar Disorder type II (BD-II), 6.3% with Bipolar Disorder-Not Otherwise Specified (BD-NOS), and 1.5% with Schizoaffective Disorder, Bipolar Type. The average age at entry was 40.48 years old, and the average age of onset was 17.3 years of age. Overall MADRS, YMRS, and LRIFT scores (mean entry values of 16.1, 6.8, and 11.6, respectively), reflected initial moderate levels of symptoms and showed a clear trend towards symptom improvement (YMRS: $\beta = -0.44$, $p = 1.51 \times 10^{-17}$; MADRS: $\beta = -0.90$, $p = 2.87 \times 10^{-29}$; LRIFT: $\beta = -0.25$, $p = 3.07 \times 10^{-26}$) between study entry and the final 12-month assessment.

Trajectories of depressive, manic and psychosocial functional domains

As shown in Table 1, the best fitting models for most longitudinal latent classes were represented by linear slopes, except for manic symptoms, which showed a significantly improved fit with a two-slope, quadratic model. In general, model fit significantly increased with increasing numbers of classes (max = 4) for all models except the univariate LGCA of manic symptoms, where the 3-class model had the best fit. Figures 1 to 3 show the results of the univariate latent class trajectory analyses.

Longitudinal trends in depressive symptoms (Fig. 1a, online Supplementary Table S2) were best characterized by four classes, described as 'persistently depressed' ($N = 391$; 13.3%), 'worsening depression' ($N = 153$; 5.2%), 'improving but moderate depression' ($N = 1030$; 35.1%), and 'non-depressed' ($N = 1364$; 46.4%). The large 'improving but moderate depression' class showed changes in symptoms across the 12 months but was still marked by depressive symptoms at the end of the follow-up period (end mean MADRS score of 12.2 ± 6.8). Compared to the baseline 'non-depressed' class (Fig. 1b), trajectory-based classes with more persistent symptoms were marked by lower educational attainment, a greater likelihood of being unemployed and disabled, and had a greater prevalence of more morbid clinical features (history of rapid cycling and suicide attempts) and comorbid diagnoses (anxiety disorder, PTSD, and substance abuse/dependence disorders). However, there was no association with earlier age at onset or a lifetime history of psychotic features.

Manic symptoms (Fig. 2a, online Supplementary Table S3) were less prevalent, less severe, and best captured by a three-class model with both linear and quadratic slopes. We characterized these classes as a 'persistent manic symptoms' ($N = 284$; 9.7%); an 'improving manic symptoms' class ($N = 349$; 11.9%); and a baseline, 'non-manic' ($N = 2305$; 78.5%) class. The overall prevalence of manic symptoms was lower compared to depressive symptoms, with most subjects (78%) having no or very minimal manic symptoms throughout the follow-up period. Baseline predictors of class membership are shown in Fig. 2b. Compared to the baseline class, subjects with a persistent manic trajectory had lower educational status but were more likely to have been diagnosed with schizoaffective disorder and to have experienced greater antecedent clinical morbidity and comorbid anxiety and substance abuse disorders.

In the univariate LCGA of psychosocial functioning using the Life-Range of Impaired Functioning Tool (LRIFT), the best fitting model converged on four linear classes (Fig. 3a, online Supplementary Table S4). In contrast to the depressive and manic symptom-based classes, psychosocial functioning showed relatively consistent levels of functional impairment across time, with a trend towards small improvement in all classes. We have characterized these functional classes as 'severely impaired' ($N = 287$, 11.6%), 'moderately impaired' ($N = 715$; 28.8%), 'mildly impaired' ($N = 1092$; 44.0%), and 'unimpaired' ($N = 390$; 15.7%). Baseline predictors showed a clear dose-response relationship (Fig. 3b), with increasing associations of less favorable

https://doi.org/10.1017/S0033291722001489 Published online by Cambridge University Press

Table 1. Latent Class Growth Analyses (LCGA) best-fitting model statistics

Model	Slopes (*N*)	Classes (*N*)	Parameters (*N*)	Loglikelihood (H_0)	AIC	BIC	Adjusted BIC	Entropy	LMR *p* value	Max Class Prop	Min Class Prop
Manic Symptoms	linear & quadratic	3	16	−32 699.3	65 430.6	65 526.3	65 475.5	0.84	$<1 \times 10^{-3}$	0.78	0.10
Depressive Symptoms	linear	4	16	−37 771.4	75 574.8	75 670.6	75 619.7	0.66	5.2×10^{-3}	0.46	0.05
Psychosocial Functioning	linear	4	16	−20 729.1	41 490.2	41 583.3	41 532.4	0.53	7.1×10^{-2}	0.44	0.12
Multi-variate trajectory	linear	4	42	−81 092.6	162 269.3	162 512.9	162 379.5	0.80	1.9×10^{-3}	0.38	0.10

AIC, Akaike information criterion; BIC, Bayesian information criterion; *N*-adjusted BIC, sample size-adjusted BIC; LMRp Value, Lo-Mendell-Rubin *p* value; Max and Min Class Prop, proportion of the sample assigned (according to maximum class membership statistic) to the largest and the smallest class. (Smaller-magnitude values for Loglikelihood, AIC, and BIC indicate better model fit; Higher entropy values indicate higher distinguishability between classes; and the proportion of total sample assigned to the largest and smallest classes, recommended to be above 0.05, helps differentiate true class distinction from statistical artifact).

demographic variables (low educational attainment and unemployment/disability), negative clinical features, and comorbid disorders. Notably, there were no association between the functioning-based trajectories and DSM-based BD subtypes.

Multi-trajectory analyses of co-occurring symptoms and psychosocial functioning

In complex disorders such as BD, various domains of symptoms and functioning levels may be co-occurring and influence the trajectory of an individual. To better model the combined effects of clinical (manic and depressive) symptoms and their relationship to psychosocial functioning, we performed multi-trajectory modeling of the three domains described above in univariate analyses. This joint analysis identifies groups of individuals that follow a distinct combination of longitudinal courses across symptoms domains, as measured by the depression (MADRS), mania (YMRS), and psychosocial functioning rating scales (LRIFT). As shown in Fig. 4a and online Supplementary Table S5, the multi-trajectory analysis identified four classes labelled as: (1) 'minimally symptomatic' ($N = 894$; 36.6%); (2) 'persistent moderate depression and moderate impairment' ($N = 965$; 39.5%); (3) 'persistently depressed and impaired' ($N = 358$, 14.7%); and (4) 'persistently mixed' ($N = 226$; 9.3%) with moderate impairment. Notably, the large class of subjects with minimal depressive or manic symptoms (Fig. 4 'minimal depression' group) still showed evidence of moderate but persistent psychosocial impairment. The second major group ('persistent moderate depression') was characterized by more prominent but improving depressive symptoms and a more significant degree of psychosocial impairment. The class with consistent, mild to moderate manic symptoms ('persistent mixed symptoms') was similarly marked by persistent, moderate depressive symptoms and disability, with limited improvement in any symptoms domains. Finally, 15% of subjects experienced a more severe and enduring depression with minimal manic symptoms; this class of subjects ('persistently depressed and impaired') were also found to have the greatest burden of psychosocial disability.

As expected (Fig. 4b), subjects with persistent, severe depression and persistent manic symptoms were associated with decreased educational attainment, a higher likelihood of being disabled and unemployed, and a greater lifetime prevalence of negative illness features (earlier age at onset, more frequent histories of rapid cycling and suicide attempts, and higher rates of comorbid anxiety and substance abuse disorders). History of psychotic symptoms was generally associated with a better illness course, as was self-reported Asian ancestry. In the only significant association with Black race in our study, there was a nominally significant increase in membership in the 'persistently mixed symptoms' class (relative risk ratio = 2.5, $p = 0.0042$).

Discussion

Bipolar Disorder is a syndrome with marked inter and intra-individual clinical heterogeneity. Characterizing the longitudinal course of the illness is necessary for predictive models and potentially for informing the type (or degree) of intervention for an individual patient. Toward this end, we sought to identify data-driven subtypes of illness using major classes of symptom domains and functioning in the STEP-BD sample, a large effectiveness study that still represents the largest phenotyped cohorts of BD subjects available. Depressive symptoms were more

https://doi.org/10.1017/S0033291722001489 Published online by Cambridge University Press

https://doi.org/10.1017/S0033291722001489 Published online by Cambridge University Press

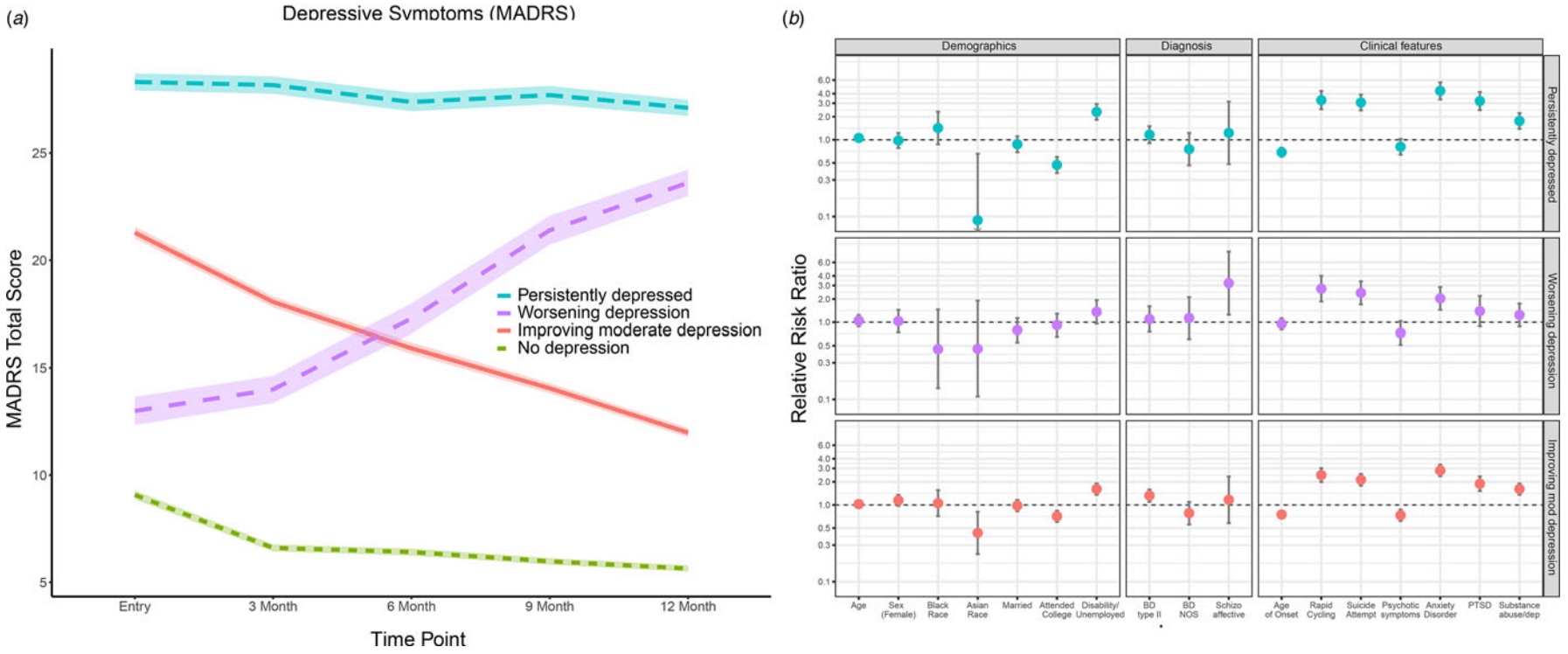

Fig. 1. (a) Course Trajectories based on Longitudinal Class Grow Analyses (LCGA) of Depressive Symptoms using scores from the Montgomery–Äsberg Rating Scale (MADRS); (b) Results from multinomial regression analysis using the 'non-depressed' class as the reference group.

https://doi.org/10.1017/S0033291722001489 Published online by Cambridge University Press

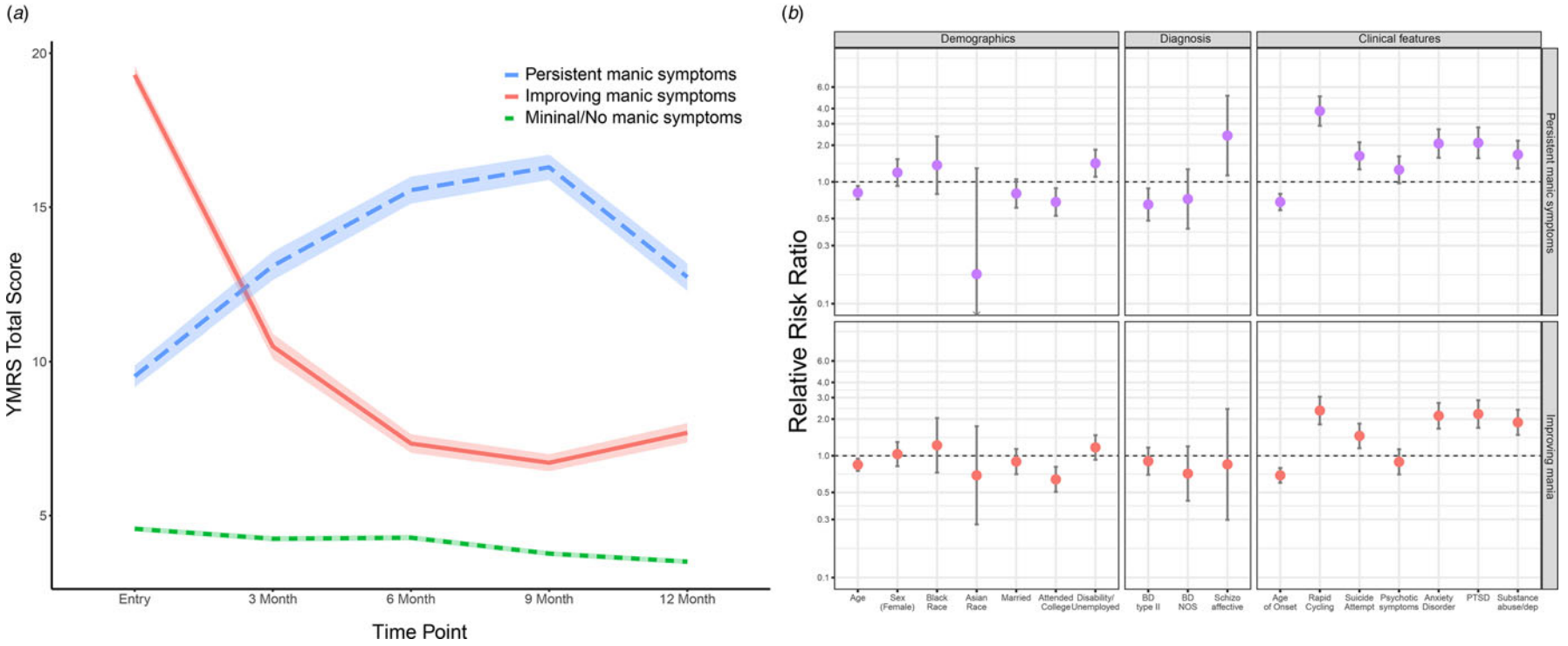

Fig. 2. (a) Course Trajectories based on Longitudinal Class Grow Analyses (LCGA) of Manic Symptoms using scores from the Young Mania Rating Scale (YMRS); (b) Results from multinomial regression analysis using the 'non-manic' class as the reference group.

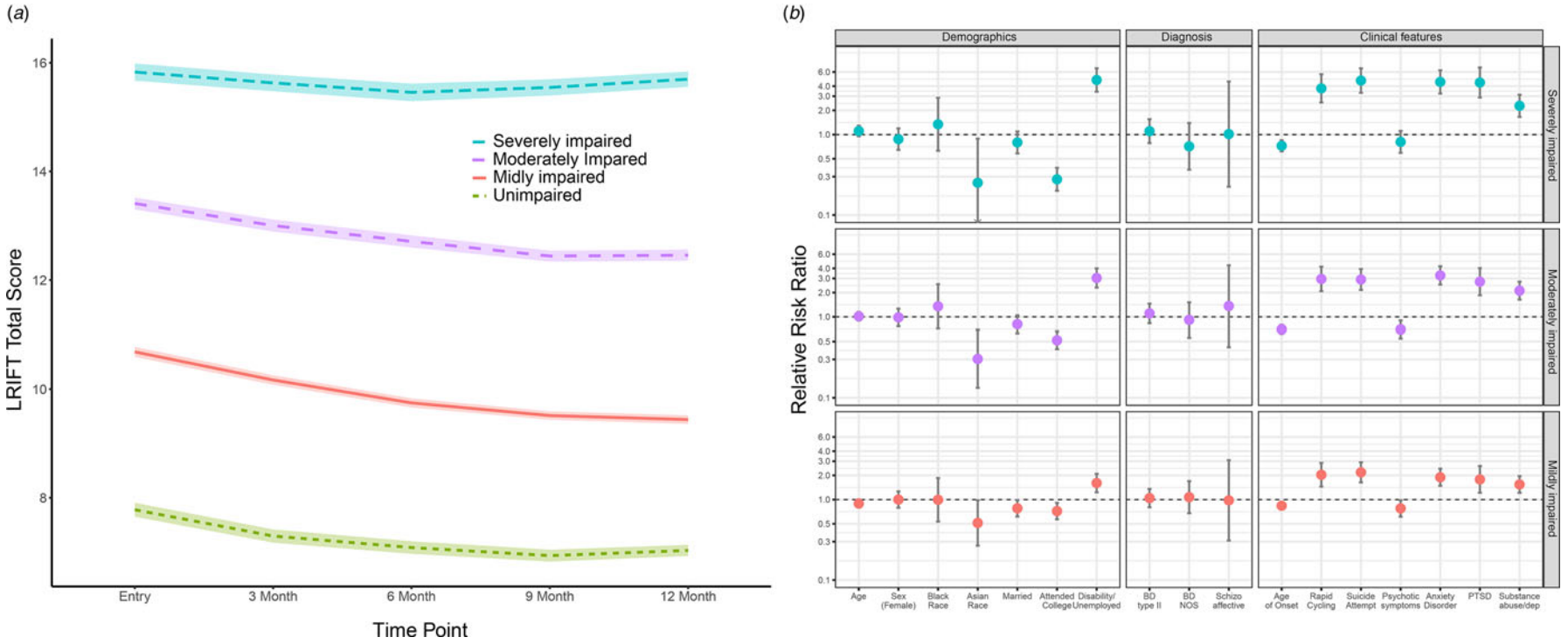

Fig. 3. (a) Course Trajectories based on Longitudinal Class Grow Analyses (LCGA) of Psychosocial Functioning using the Life-Range Impaired Functioning Tool (LRIFT) scale; (b) Results from multinomial regression analysis using the 'unimpaired' class as the reference group.

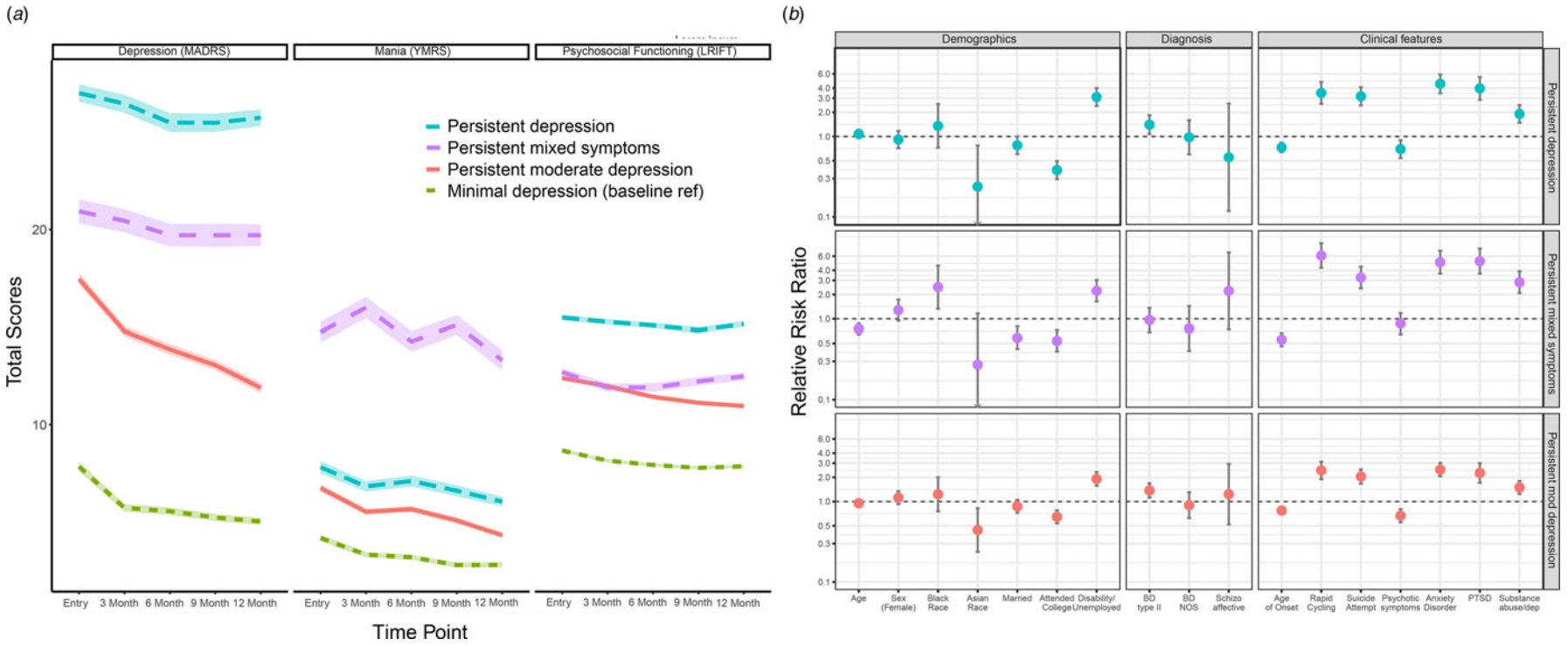

Fig. 4. (a) Course Trajectories based on Multi-Trajectory Longitudinal Class Grow Analyses (LCGA) of Depressive Symptoms (MADRS), Manic Symptoms (YMRS), and psychosocial functioning (LRIFT); (b) Results from multinomial regression analysis using the 'minimally symptomatic' class as the reference group.

prevalent than manic symptoms, and while many subjects had a relatively benign course, a small but significant number of subjects were found to have persistent manic ($N = 284$; 9.7%) and depressive ($N = 391$; 13.3%) symptoms despite receiving guideline-based treatment. While depressive or manic symptoms generally improved during the follow-up period, impairments in psychosocial functioning remained relatively unchanged. In the multivariate trajectory analysis, we found that subjects with persistent mild to moderate hypomanic symptoms also experienced persistent depressive symptoms, consistent with a subclass of subjects with bipolar disorder who may suffer from a trait-like occurrence of mixed states (Akiskal, 1996; Sato et al., 2004). Nevertheless, it is notable that the most impaired class was marked solely by chronic depressive symptoms, consistent with a previously shown stronger association between depressive symptoms and psychosocial impairment (Judd et al., 2005).

Overall, the analyses of baseline clinical covariates showed similar but increasing associations between indicators of poorer health and clinical severity, and trajectories marked by more persistent symptoms and greater functional impairment. In terms of demographic risk factors, there was little effect of current age, female sex, or marital status. Compared to subjects of white race, subjects of Asian descent were generally associated with class membership in the baseline (minimally symptomatic) classes, whereas those of self-reported Black race were more likely to be found in the chronic mixed category. However, due to the small number of subjects of Asian ($N = 51$, 2.1%) or Black race ($N = 104$, 4.2%), no association met significance after correction for multiple testing. On the other hand, low educational attainment and current unemployment and disability were consistently associated with poorer outcome classes across all tested symptom and functional domains. These broad but ubiquitous indices of poorer outcomes may represent risk factors that are consequences of having bipolar disorder, although there is also moderate evidence from Mendelian Randomization studies consistent with a potential causal role (Mullins et al., 2020; Wendt et al., 2021). Interestingly, in epidemiological studies, BD is generally associated with *increased* educational attainment (e.g. MacCabe et al., 2010), highlighting the heuristic value of utilizing longitudinal courses to identify subgroups of patients that differ from the average and may require alternative or more intensive clinical care.

Like the demographic risk factors, baseline clinical features indicative of more active illness (earlier age at onset, rapid cycling, and history of suicide attempt) and comorbid disorders (particularly anxiety and substance abuse disorders) were almost ubiquitously associated with the more severe illness trajectories. Stronger associations were seen in poorer outcome trajectories across depressive, manic, and functional domains. These clinical factors are well known to be associated with greater overall morbidity (Bauer et al., 2018; Goes et al., 2012b; Post et al., 2021), decreased responsiveness to first-line mood stabilizers (Calabrese et al., 2005; Tohen et al., 2007), and represent, with the exception of rapid cycling, transdiagnostic risk factors increasingly the focus of attention in clinical care in general (Hickie et al., 2019). Perhaps surprisingly, the lifetime presence of psychotic symptoms was primarily associated with membership in the baseline reference classes with few active symptoms. One hypothesis to explain this unexpected association could be that the designation of psychosis in the STEP-BD was limited to only a few relative broad questions from the MINI diagnostic interview, whereas a more fine-grained assessment of psychotic symptoms and their duration may have facilitated the identification of a subset of subjects with more persistent or problematic psychotic features (Goes et al., 2012a). Alternatively, it is also conceivable that psychotic symptoms are more linked to episodic, 'classic' forms of bipolar disorder, which may manifest a less symptomatic and higher functioning baseline when outside of well-demarcated episodes (Burton et al., 2018).

In contrast to the above clinical features, DSM-based subtypes of illness (BD-type II, BD-NOS, and schizoaffective disorder) showed fewer clear associations and none surviving correction for multiple testing (online Supplementary Tables S2–S5). Schizoaffective disorder was overrepresented in the persistently manic class; there were no clear associations between DSM-based subtypes of illness and trajectories of depressive symptoms or psychosocial functioning. This lack of association across the DSM-based subtypes highlights the known limitations of current diagnostic approaches and reiterates the need for alternative approaches that more directly address residual symptoms experienced in BD.

To our knowledge, this analysis is among the first and the largest longitudinal trajectory analyses of BD. If replicated and sufficiently generalizable, our results have several clinical implications. First, they highlight the importance of early identification and focused treatment for patients within the class of *persistent* symptoms that appear minimally responsive to conventional treatments. High comorbidity and occupational dysfunction associated with these classes point to the need for multi-dimensional treatment approaches earlier in the illness course to mitigate the extent of 'multi-morbidity' often seen in poor-outcome BD. Second, the very modest improvements in psychosocial functioning obtained with conventional care underscore the need for novel approaches that more specifically target functional impairment (Torrent et al., 2013). Third, our multivariate analyses reveal that a sizeable minority of subjects ($N = 226$, 9.3%) experienced persistently mixed symptoms throughout the course of the study, suggesting that mixed symptoms may also have trait-like manifestations in addition to their traditional conceptualization as transient or brief episodic states.

Limitations

There are several limitations of the sample that may limit the external validity of our analyses. First, as is common in the follow-up of subjects with active psychiatric illness (Etain et al., 2021; Findling et al., 2013), there was attrition throughout the study, leading us to restrict our analyses to the first year of follow-up. Our results may therefore be less applicable to the sizeable number of subjects who entered the study but dropped out within a year of follow-up – subjects who, as described in the Methods, were notable for higher rates of unfavorable baseline demographic features. Second, although the STEP-BD sample is more clinically representative than most clinical trials (Bowden et al., 2012), racial and ethnic diversity was still limited, and there was likely to have been selected for less acutely severe patients (who may be more likely to agree to participate in a longitudinal outpatient study). Third, the composition of subjects also did not generally include children or adolescents, limiting our ability to study early childhood risk factors relevant for the course of illness. Fourth, all subjects in STEP-BD received guideline base but "uncontrolled" clinical care, during the follow-up period. Hence, the trajectories described in our analyses cannot be said to represent a natural history of the illness, representing

https://doi.org/10.1017/S0033291722001489 Published online by Cambridge University Press

instead an 'observed' history under usual clinical care. Fifth, the degree of granularity in the dataset was also a likely limitation since measures of clinical state were performed every three months, representing a potentially crude, averaged recollection of symptoms that may fluctuate and change on a much shorter time scale.

Moreover, we note two important limitations of our modeling strategies. First is the acknowledgment that the key assumption of LGCA, fixed intra-class variance, is an oversimple assumption made for modeling efficiency. Second, we note that due to relatively low levels of entropy (<0.8), in the psychosocial functioning model, class membership was not as certain as it was for the other models. Therefore, class size and covariate effects may have been misestimated and these results should be taken as preliminary until replication is performed in an independent sample.

Conclusions

Improved understanding and prediction of the longitudinal course of bipolar disorder are among the most important advancements needed to help usher in an era of personalized treatment. Our results show that adults with BD can be characterized in partially distinct classes, which are differentially associated with common risk factors that may merit earlier and more personalized interventions. Future research will be needed to generalize our findings to more diverse and clinically representative samples, and to link these common, often transdiagnostic, risk factors with therapeutically actionable interventions.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291722001489

References

Akiskal, H. S. (1996). The prevalent clinical spectrum of bipolar disorders: Beyond DSM-h;lkIV. *Journal of Clinical Psychopharmacology*, *16*(2), 4S. Retrieved from https://journals.lww.com/psychopharmacology/Fulltext/1996/04001/The_Prevalent_Clinical_Spectrum_of_Bipolar.2.aspx.

Bauer, M., Andreassen, O. A., Geddes, J. R., Vedel Kessing, L., Lewitzka, U., Schulze, T. G., & Vieta, E. (2018). Areas of uncertainties and unmet needs in bipolar disorders: Clinical and research perspectives. *The Lancet Psychiatry*, *5*(11), 930–939. https://doi.org/10.1016/S2215-0366(18)30253-0.

Birmaher, B., Axelson, D., Goldstein, B., Strober, M., Gill, M. K., Hunt, J., … Keller, M. (2009). Four-year longitudinal course of children and adolescents with bipolar spectrum disorders: The Course and Outcome of Bipolar Youth (COBY) study. *The American Journal of Psychiatry*, *166*(7), 795–804. https://doi.org/10.1176/appi.ajp.2009.08101569.

Bowden, C. L., Perlis, R. H., Thase, M. E., Ketter, T. A., Ostacher, M. M., Calabrese, J. R., … Sachs, G. S. (2012). Aims and results of the NIMH systematic treatment enhancement program for bipolar disorder (STEP-BD). *CNS Neuroscience & Therapeutics*, *18*(3), 243–249. https://doi.org/10.1111/j.1755-5949.2011.00257.x.

Burton, C. Z., Ryan, K. A., Kamali, M., Marshall, D. F., Harrington, G., McInnis, M. G., & Tso, I. F. (2018). Psychosis in bipolar disorder: Does it represent a more "severe" illness? *Bipolar Disorders*, *20*(1), 18–26. https://doi.org/10.1111/bdi.12527.

Calabrese, J. R., Shelton, M. D., Rapport, D. J., Youngstrom, E. A., Jackson, K., Bilali, S., … Findling, R. L. (2005). A 20-month, double-blind, maintenance trial of lithium versus divalproex in rapid-cycling bipolar disorder. *The American Journal of Psychiatry*, *162*(11), 2152–2161. https://doi.org/10.1176/appi.ajp.162.11.2152.

Cheesman, R., Purves, K. L., Pingault, J.-B., Breen, G., Rijsdij, K., Plomin, R. & Eley, T. C. (2018). Extracting stability increases the SNP heritability of emotional problems in young people. *Translational Psychiatry*, *8*(1), 1–9. https://doi.org/10.1038/s41398-018-0269-5.

Cochran, A. L., McInnis, M. G., & Forger, D. B. (2016). Data-driven classification of bipolar I disorder from longitudinal course of mood. *Translational Psychiatry*, *6*(10), e912. https://doi.org/10.1038/tp.2016.166.

Etain, B., Bellivier, F., Olié, E., Aouizerate, B., Aubin, V., Belzeaux, R., … Bing, A. (2021). Clinical predictors of recurrences in bipolar disorders type 1 and 2: A FACE-BD longitudinal study. *Journal of Psychiatric Research*, *134*, 129–137. https://doi.org/10.1016/j.jpsychires.2020.12.041.

Findling, R. L., Jo, B., Frazier, T. W., Youngstrom, E. A., Demeter, C. A., Fristad, M. A., … Horwitz, S. M. (2013). The 24-month course of manic symptoms in children. *Bipolar Disorders*, *15*(6), 669–679. https://doi.org/10.1111/bdi.12100.

Goes, F. S., Hamshere, M. L., Seifuddin, F., Pirooznia, M., Belmonte-Mahon, P., Breuer, R., … Potash, J. B. (2012a). Genome-wide association of mood-incongruent psychotic bipolar disorder. *Translational Psychiatry*, *2*(10), e180. doi: 10.1038/tp.2012.106.

Goes, F. S., McCusker, M. G., Bienvenu, O. J., MacKinnon, D. F., Mondimore, F. M., Schweizer, B., … Potash, J. B. (2012b). Co-morbid anxiety disorders in bipolar disorder and major depression: Familial aggregation and clinical characteristics of co-morbid panic disorder, social phobia, specific phobia and obsessive-compulsive disorder. *Psychological Medicine*, *42*(7), 1449–59. doi: 10.1017/S0033291711002637.

Hickie, I. B., Scott, E. M., Cross, S. P., Iorfino, F., Davenport, T. A., Guastella, A. J., … Hermens, D. F. (2019). Right care, first time: A highly personalised and measurement-based care model to manage youth mental health. *Medical Journal of Australia*, *211*(59), S3–S46. doi: 10.5694/mja2.50383. Retrieved from https://www.mja.com.au/journal/2019/211/9/right-care-first-time-highly-personalised-and-measurement-based-care-model.

Judd, L. L., Akiskal, H. S., Schettler, P. J., Coryell, W., Endicott, J., Maser, J. D., … Keller, M. B. (2003). A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. *Archives of General Psychiatry*, *60*, 261–269. https://doi.org/10.1001/archpsyc.60.3.261.

Judd, L. L., Akiskal, H. S., Schettler, P. J., Endicott, J., Leon, A. C., Solomon, D. A., … 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., … 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.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling: Latent trajectory classes. *Social and Personality Psychology Compass*, *2*(1), 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x.

Kwong, A. S. F., López-López, J. A., Hammerton, G., Manley, D., Timpson, N. J., Leckie, G., & Pearson, R. M. (2019). Genetic and Environmental Risk Factors Associated With Trajectories of Depression Symptoms From Adolescence to Young Adulthood. *JAMA Network Open*, *2*(6), e196587. https://doi.org/10.1001/jamanetworkopen.2019.6587.

Leon, A. C., Solomon, D. A., Mueller, T. I., Turvey, C. L., Endicott, J., & Keller, M. B. (1999). The Range of Impaired Functioning Tool (LIFE-RIFT): A brief measure of functional impairment. *Psychological Medicine*, *29*(4), 869–878. https://doi.org/10.1017/s0033291799008570.

MacCabe, J. H., Lambe, M. P., Cnattingius, S., Sham, P. C., David, A. S., Reichenberg, A., … Hultman, C. M. (2010). Excellent school performance at age 16 and risk of adult bipolar disorder: National cohort study. *The British Journal of Psychiatry: The Journal of Mental Science*, *196*(2), 109–115. https://doi.org/10.1192/bjp.bp.108.060368.

Miklowitz, D. J., Otto, M. W., Frank, E., Reilly-Harrington, N. A., Wisniewski, S. R., Kogan, J. N., … Sachs, G. S. (2007). Psychosocial treatments for bipolar depression: A 1-year randomized trial from the systematic treatment enhancement program. *Archives of General Psychiatry*, *64*(4), 419–426. https://doi.org/10.1001/archpsyc.64.4.419.

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry : The Journal of Mental Science*, *134*, 382–389. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/444788.

Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., I. Coleman, J. R., Qiao, Z., … Andreassen, O. A. (2020). *Nature Genetics*, *53*(6), 817–829.

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*, 109–138. https://doi.org/10.1146/annurev.clinpsy.121208.131413.

Post, R. M., Altshuler, L. L., Kupka, R., McElroy, S. L., Frye, M. A., Grunze, H., … Nolen, W. A. (2021). 25 Years of the International Bipolar Collaborative Network (BCN). *International Journal of Bipolar Disorders*, *9*(1), 13. https://doi.org/10.1186/s40345-020-00218-w.

Sachs, G. S., Nierenberg, A. A., Calabrese, J. R., Marangell, L. B., Wisniewski, S. R., Gyulai, L., … Thase, M. E. (2007). Effectiveness of adjunctive antidepressant treatment for bipolar depression. *New England Journal of Medicine*, *356*(17), 1711–1722. https://doi.org/10.1056/NEJMoa064135.

Sachs, G. S., Thase, M. E., Otto, M. W., Bauer, M., Miklowitz, D., Wisniewski, S. R., … Rosenbaum, J. F. (2003). Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biological Psychiatry*, *53*(11), 1028–1042. https://doi.org/10.1016/S0006-3223(03)00165-3.

Sato, T., Bottlender, R., Sievers, M., Schröter, A., Kleindienst, N., & Möller, H.-J. (2004). Evaluating the inter-episode stability of depressive mixed states. *Journal of Affective Disorders*, *81*(2), 103–113. https://doi.org/10.1016/S0165-0327(03)00159-9.

Schneck, C. D., Miklowitz, D. J., Miyahara, S., Araga, M., Wisniewski, S., Gyulai, L., … Sachs, G. S. (2008). The prospective course of rapid-cycling bipolar disorder: Findings from the STEP-BD. *The American Journal of Psychiatry*, *165*(3), 370–377, quiz 410. https://doi.org/10.1176/appi.ajp.2007.05081484.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59*(Suppl. 20), 22–33, quiz 34–57.

Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Archives of General Psychiatry*, *49*(8), 624–629. https://doi.org/10.1001/archpsyc.1992.01820080032005.

Tohen, M., Calabrese, J., Vieta, E., Bowden, C., Gonzalez-Pinto, A., Lin, D., … Corya, S. (2007). Effect of comorbid anxiety on treatment response in bipolar depression. *Journal of Affective Disorders*, *104*(1–3), 137–146.

Torrent, C., Bonnin, C. D. M., Martínez-Arán, A., Valle, J., Amann, B. L., González-Pinto, A., … Vieta, E. (2013). Efficacy of functional remediation in bipolar disorder: A multicenter randomized controlled study. *The American Journal of Psychiatry*, *170*(8), 852–859. https://doi.org/10.1176/appi.ajp.2012.12070971.

Weintraub, M. J., Schneck, C. D., Axelson, D. A., Birmaher, B., Kowatch, R. A., & Miklowitz, D. J. (2020). Classifying mood symptom trajectories in adolescents with bipolar disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *59*(3), 381–390. https://doi.org/10.1016/j.jaac.2019.04.028.

Wendt, F. R., Pathak, G. A., Lencz, T., Krystal, J. H., Gelernter, J., & Polimanti, R. (2021). Multivariate genome-wide analysis of education, socioeconomic status and brain phenome. *Nature Human Behaviour*, *5*(4), 482–496. https://doi.org/10.1038/s41562-020-00980-y.

Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for mania: Reliability, validity and sensitivity. *The British Journal of Psychiatry : The Journal of Mental Science*, *133*, 429–435. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/728692.

https://doi.org/10.1017/S0033291722001489 Published online by Cambridge University Press