ORIGINAL ARTICLE

Clinical predictors of non-response to lithium treatment in the Pharmacogenomics of Bipolar Disorder (PGBD) study

Yian Lin[1] | Adam X. Maihofer[2] | Emma Stapp[1] | Megan Ritchey[1] |
Ney Alliey-Rodriguez[3] | Amit Anand[4] | Yokesh Balaraman[5] | Wade H. Berrettini[6] |
Holli Bertram[7] | Abesh Bhattacharjee[8] | Cynthia V. Calkin[9] | Carla Conroy[10] |
William Coryell[11] | Nicole D'Arcangelo[10] | Anna DeModena[2] |
Joanna M. Biernacka[12] | Carrie Fisher[5] | Nicole Frazier[7] | Mark Frye[12] |
Keming Gao[10] | Julie Garnham[9] | Elliot Gershon[3] | Kara Glazer[13] |
Fernando S. Goes[13] | Toyomi Goto[10] | Elizabeth Karberg[10] | Gloria Harrington[7] |
Petter Jakobsen[14] | Masoud Kamali[7,15] | Marisa Kelly[7] | Susan G. Leckband[8] |
Falk W. Lohoff[16] | Andrea Stautland[17] | Michael J. McCarthy[2,8] | Melvin G. McInnis[7] |
Francis Mondimore[13] | Gunnar Morken[18] | John I. Nurnberger[5] | Ketil J. Oedegaard[17] |
Vigdis Elin Giever Syrstad[17] | Kelly Ryan[7] | Martha Schinagle[10] | Helle Schoeyen[19] |
Ole A. Andreassen[20,21] | Marth Shaw[11] | Paul D. Shilling[2] | Claire Slaney[9] |
Bruce Tarwater[11] | Joseph R. Calabrese[10] | Martin Alda[9,22] | Caroline M. Nievergelt[2] |
Peter P. Zandi[13] | John R. Kelsoe[2]

[1]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[2]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

[3]University of Chicago, Chicago, IL, USA

[4]Center for Behavioral Health, Cleveland Clinic, Cleveland, OH, USA

[5]Department of Psychiatry, Indiana University, Indianapolis, IN, USA

[6]University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[7]University of Michigan, Ann Arbor, MI, USA

[8]Department of Psychiatry, VA San Diego Healthcare System, La Jolla, CA, USA

[9]Dalhousie University, Halifax, Canada

[10]Department of Psychiatry, Case Western Reserve University, Cleveland, OH, USA

[11]University of Iowa, Iowa City, IA, USA

[12]Mayo Clinic, Rochester, MN, USA

[13]Department of Psychiatry and Behavioral Sciences, The Johns Hopkins School of Medicine, Baltimore, MD, USA

[14]Division of Psychiatry, Haukeland University Hospital, Bergen, Norway

[15]Department of Psychiatry, Massachusetts General Hospital and Harvard University, Boston, MA, USA

[16]National Institute of Alcohol Abuse and Alcoholism, NIH, Bethesda, MD, USA

[17]Department of Clinical Medicine, Faculty of Medicine and Dentistry, University of Bergen and Division of Psychiatry, Haukeland University Hospital, Bergen, Norway

[18]Division of Psychiatry, St. Olav University Hospital of Trondheim and Department of Neuroscience, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

[19]Division of Psychiatry, Faculty of Medicine and Dentistry, Stavanger University Hospital, University of Bergen, Stavanger, Norway

----

[20]Norwegian Centre for Mental Disorders Research (NORMENT), Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[21]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[22]National Institute of Mental Health, Klecany, Czech Republic

Correspondence
Peter P. Zandi, Ph.D., The Arlene and Robert Kogod Professor of Mood Disorders, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, The 550 Building, Suite 201, 550 North Broadway, Baltimore, MD 21205, USA.
Email: pzandi1@jhu.edu

Abstract

Background: Lithium is regarded as a first-line treatment for bipolar disorder (BD), but partial response and non-response commonly occurs. There exists a need to identify lithium non-responders prior to initiating treatment. The Pharmacogenomics of Bipolar Disorder (PGBD) Study was designed to identify predictors of lithium response.

Methods: The PGBD Study was an eleven site prospective trial of lithium treatment in bipolar I disorder. Subjects were stabilized on lithium monotherapy over 4 months and gradually discontinued from all other psychotropic medications. After ensuring a sustained clinical remission (defined by a score of ≤3 on the CGI for 4 weeks) had been achieved, subjects were followed for up to 2 years to monitor clinical response. Cox proportional hazard models were used to examine the relationship between clinical measures and time until failure to remit or relapse.

Results: A total of 345 individuals were enrolled into the study and included in the analysis. Of these, 101 subjects failed to remit or relapsed, 88 achieved remission and continued to study completion, and 156 were terminated from the study for other reasons. Significant clinical predictors of treatment failure ($p < 0.05$) included baseline anxiety symptoms, functional impairments, negative life events and lifetime clinical features such as a history of migraine, suicidal ideation/attempts, and mixed episodes, as well as a chronic course of illness.

Conclusions: In this PGBD Study of lithium response, several clinical features were found to be associated with failure to respond to lithium. Future validation is needed to confirm these clinical predictors of treatment failure and their use clinically to distinguish who will do well on lithium before starting pharmacotherapy.

KEYWORDS
bipolar disorder, clinical predictor, lithium, treatment response

1 | INTRODUCTION

Lithium is regarded as a first-line treatment for bipolar disorder (BD),[1-7] but it does not work for all patients. The modern use of lithium for treatment of BD was first introduced by John Cade in 1949, and it has been widely studied since. Although findings from these studies have been at times controversial, the evidence for the efficacy of lithium in acute mania and maintenance treatment is now well established. In a meta-analysis of five randomized controlled trials of BD comparing prophylactic lithium therapy with placebo, Geddes and colleagues found that lithium is more effective than placebo in preventing recurrence of illness, with 60% in the lithium group remaining well over 1–2 years compared with 40% in the placebo group.[8] In a subsequent meta-analysis of six studies of lithium in the treatment of acute mania, Yildiz and colleagues found that 48% of patients responded to lithium compared to 31% for placebo.[9] While these seminal reviews unequivocally demonstrate the efficacy of lithium for both acute mania and maintenance treatment of BD, they also highlight that anywhere from 40%–50% of patients do not respond adequately over a 2-year period and require either the addition of or a change to another psychotropic drug.[9] These findings are consistent with observational data from longitudinal cohort studies.[10-12]

There is considerable continued interest in identifying predictors of response to lithium before starting treatment in order to avoid the typical trial and error process of finding the right medication for a particular patient during which time he or she may continue to experience devastating symptoms and be at risk for suicide. This is the goal of precision medicine (also referred to as individualized or personalized medicine). Although the promise of precision medicine has garnered a great deal of attention recently,[13] the search for predictors of lithium response dates back to the very first studies of its prophylactic effect in mood disorders.[14]

Indeed, there is a long history of searching for clinical predictors of response to lithium treatment that can help guide treatment decisions. In 2005, Kleindienst and colleagues[15,16] carried out two comprehensive systematic reviews of predictors of lithium response in which they identified nearly 2,000 studies published between 1966 and 2003 on this topic. In one review, they focused on studies that examined psychosocial and demographic predictors and identified nine that emerged as consistently associated with lithium response. Four were associated with good response (high social status, social support, good compliance, and "dominance" personality trait), while five were associated with poorer response (stress, high expressed emotion, neurotic personality trait, unemployment, and high number of life events). In the other review, they focused on studies that examined clinical predictors of lithium response and identified five that were consistently associated with lithium response across studies. These included a pattern of mania-depression-interval in bi-phasic episodes (so-called MDI polarity sequence) and older age at onset associated with better response, and high number of hospitalizations, a pattern of depression-mania-interval (i.e., DMI polarity sequence), and continuous cycling associated with poorer response. Both reviews concluded that the effect sizes of these factors on treatment response were relatively small.

In 2019, Hui and colleagues[17] carried out a subsequent meta-analysis of clinical predictors of lithium response that included more recent data from 71 studies with over 12,000 patients. They identified six predictors of good lithium response, some of which overlapped the earlier review by Kleindienst and colleagues,[15,16] and included manic-depression-interval pattern, absence of rapid cycling, absence of psychotic symptoms, family history of bipolar disorder, shorter pretreatment illness duration, and later age at onset. They noted, however, that the included studies tended to have small sample sizes and there was considerable heterogeneity in results.

The Pharmacogenomics of Bipolar Disorder (PGBD) Study (www.clinicaltrials.gov, NCT01272531) was a large multi-center study designed to prospectively identify clinical and molecular predictors of lithium response. We report here the results of an analysis of clinical data from this study to examine clinical predictors of lithium response. The advantage of this study over previous ones is that patients were prospectively followed on lithium monotherapy for up to 2 years to better identify predictors of long-term treatment response specifically to lithium.

2 | METHODS

2.1 | Study overview

The PGBD was one of 14 research projects in the Pharmacogenetic Research Network funded by the National Institute of Health to support multi-disciplinary, collaborative research on how genetic factors contribute to inter-individual differences in responses to medications. The PGBD set out to conduct a multi-site prospective study of lithium monotherapy in the treatment of BD.

The details of the trial have been described elsewhere.[18] Briefly, the goal of the study was prevention of illness recurrence by lithium monotherapy. All patients were observed in an observation phase lasting 4 weeks to confirm they were in remission defined by having a Clinical Global Impression of Severity Scale (CGI-S) score of ≤3 (mildly ill) for at least 4 weeks. After the observation phase, the patients entered a 2-year maintenance phase, during which they were assessed every 2 months to monitor their on-going clinical response. Patients who came into the trial clinically unstable and/or not on lithium monotherapy were first transitioned to lithium monotherapy in a stabilization phase that lasted a maximum of 16 weeks which included visits every other week for the first 8 weeks and one visit per month for the next 2 months. The treatment dosage of lithium was not fixed by study protocol but instead was titrated by the treating clinicians as clinically indicated. Throughout the follow-up, patients were allowed to take a benzodiazepine for anxiety and/or zolpidem for sleep. A range of clinical measures (described below) was collected at the screening and subsequent visits to monitor clinical progress and enable investigation of clinical predictors of response.

2.2 | Participants

Patients were enrolled into the study from outpatient psychiatry clinics in academic medical centers at nine sites within the United States and two international sites. The nine domestic sites included: University of California, San Diego; Indiana University; University of Chicago; University of Pennsylvania; University of Iowa; Johns Hopkins University; Case Western Reserve University; University of Michigan; and the Mayo Clinic. The two international sites were University of Bergen, Norway, and Dalhousie University in Halifax, Canada.

Patients were included in the study if they: (1) had bipolar I disorder in any phase of illness; (2) were naïve to or not presently on lithium and had at least one affective episode meeting DSM-IV criteria in the last 12 months or were currently on lithium and did not have any history of mood episodes meeting DSM-IV criteria in the last 6 months; (3) were able to give informed consent; (4) were 18 years or older; and (5) were currently symptomatic, as defined as a CGI-S score of at least 3 (mild severity), unless the patient entered the study already stable on lithium monotherapy. Women of child-bearing potential were included if they agreed to use adequate contraception and inform their doctor at the earliest possible time of their plans to conceive.

Patients were excluded if they: (1) were unwilling or unable to comply with study requirements; (2) had renal impairment (serum creatinine >1.5 mg/dL); (3) had thyroid stimulating hormone (TSH) level over >20% above the upper normal limit or, if on thyroid medication, had not been euthyroid for at least 3 months before the first visit; (4) were currently in crisis such that inpatient hospitalization or other crisis management should take priority; (5) met criteria for physical dependence requiring acute detoxification from alcohol, opiates or barbiturates; (6) were pregnant or breastfeeding; (7) had

participated in a clinical trial of an investigational drug within the past 1 month; or (8) had a history of lithium toxicity, not due to mismanagement or overdose, that required treatment.

All study procedures were approved by local Institutional Review Boards (IRBs), and all patients provided written informed consent. This analysis included data on the first 345 BD patients who enrolled into the study and had sufficient follow-up of at least 4 weeks as of the date of data freeze on June 26, 2017. There were four patients who were still active in the study but had not yet reached the maintenance phase by the time of this data freeze and were not included in these analyses.

2.3 | Clinical outcomes

Patients were followed until they: (1) completed all study visits over 2 years of the maintenance phase (or had achieved the maintenance phase and were still active in the on-going study by the date of the data freeze), (2) were terminated from the study before completion of all visits because of failure to achieve (i.e., failure to remit) or maintain (i.e., relapse) stabilization on lithium, or (3) were terminated from the study for other reasons. Failure to remit was defined by the inability to achieve clinically sustained remission (where remission was documented as described above) by the end of the observation phase or based on clinical judgment that the patient was unable to adequately stabilize on lithium monotherapy. Relapse was evaluated using the Mood Episode Checklist which summarizes DSM-IV criteria for mania and depression and was collected at each visit during the maintenance phase. Relapse was defined by the following: (1) meets criteria for mania and has a CGI-S of 5 (markedly ill) or greater; (2) meets criteria for a major depressive episode with 4-week duration; (3) meets criteria for a mixed episode with CGI-S of 5 or greater; (4) psychiatric hospitalization for a mood episode is required; or (5) in the physician's judgment the patient cannot be managed on monotherapy and a change in medication is required. Episodes of hypomania without impairment of function were not considered relapses. These criteria were designed to be stringent so as to detect clear failures of prophylaxis, rather than brief episodes that might not require a medication change in clinical practice. Serum lithium levels were routinely monitored as clinically recommended over the course of follow-up. On average, lithium levels were maintained at appropriate therapeutic levels[19] and were, in fact, slightly higher for those who failed to remit or relapse compared to others (0.68 vs 0.63 mEq/L, $p = 0.05$).

2.4 | Clinical predictors

Patients were evaluated with the Diagnostic Interview for Genetic Studies (DIGS) in order to establish a diagnosis of bipolar I disorder by DSM-IV criteria and collect detailed historical clinical information about current and lifetime mental illnesses. Patients also completed a range of self and clinician rated scales at the screening and subsequent visits to document the clinical course of illness and factors that may relate to the course. Self-rated scales included the Childhood Life Events Scale; the Lifetime History of Aggression Scale; the Columbia Suicide Symptom Severity Scale; the Basic Language Morningness Scale (BALM); the Temperament Evaluation of Memphis, Pisa, Paris and San Diego – auto-questionnaire version (TEMPS-A); the 16 item Quick Inventory of Depression Symptomatology Self-Report (QIDS-SR-16); the Sheehan Disability Scale; the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q); and the Life Events Questionnaire (LEQ). Clinician rated scales included the following: the Clinical Global Impressions of Severity Scale (CGI-S); the Hamilton Rating Scale for Anxiety (HAM-A); the Montgomery Asberg Depression Rating Scale (MADRS); the Clinician Administered Rating for Mania (CARS-M); and the Modified Scale for Suicidal Ideation (MSSI). From the assessments collected either at the screening or baseline visits, we derived 43 clinical variables for analysis that were selected based on clinical experience and an expert review of the literature involving three of us (JK, MA, and JC). These included variables on socio-demographic factors, baseline symptoms, clinical history and course, co-morbid illnesses, family history of mental illness, childhood and current life events, and level of functioning. See Table 1 for a full list of variables that were examined.

2.5 | Statistical analyses

Differences in socio-demographic factors between patients who completed all study procedures, those who failed to remit or experienced a relapse, and those who were terminated from the study for other reasons were compared using chi-square tests for categorical variables and one-way ANOVA for continuous variables. We then used survival analysis with Cox Proportional Hazard models to examine the relationship between clinical predictors measured at baseline and the time from study entry to treatment failure, which was defined as the time of the last visit at which the patient was determined to have failed to remit or to have relapsed. All other patients were censored at the time of their last visit in the on-going study. We examined each clinical predictor individually in models that additionally controlled for potential confounders including age at study entry, sex, race, and lithium status upon entry into the study. These variables were selected from the available data because they are important socio-demographic factors that experience indicated may be relevant and/or they were found to differ with treatment outcome. Race was captured as a categorical variable for Whites, Blacks, Asians, or other. Lithium status upon entry into the study was captured as a categorical variable to distinguish those who entered the study stable on lithium monotherapy, on lithium plus other psychotropic medications, or not on lithium. We used two-tailed $p < 0.05$ to declare associations statistically significant. We did not correct for multiple testing because the clinical predictors were carefully selected based on prior hypotheses that they may be relevant to treatment response.

To determine if the associations with treatment response of the clinical predictors identified through the above procedures differed in the initial versus later phases of follow-up, we stratified the survival analyses and looked first at survival over the stabilization/

TABLE 1 Clinical predictors examined for association with treatment response

Baseline symptoms	Comorbidity
Anxiety symptoms[a]	Comorbid alcohol abuse/dependence[c]
Hypermotor activity[b]	Comorbid substance abuse/dependence[c]
Irritability and aggressiveness[b]	Comorbid anxiety disorder[c]
Clinical history	Comorbid personality disorder[c]
Age of onset[c]	**Functioning**
Chronicity of affective disorder[c]	Disability at baseline: impairment[f]
Chronicity of substance abuse[c]	Disability at baseline: family/home life[f]
History of delusions[e]	Disability at baseline: social life[f]
History of auditory hallucinations[e]	Disability at baseline: total[f]
History of visual hallucinations[e]	Disability at baseline: work/school[f]
History of any hallucinations[e]	Functioning during most severe depression[c]
History of headaches lasting 4 to 72 hours[d]	Functioning during most severe mania[c]
History of migraines[d]	Functioning overall[c]
History of suicidal thought/behavior[c]	Years of education[e]
History of suicide attempt[e]	Marital status[e]
Affective psychosis[c]	**Life events**
Independence of psychosis episodes[c]	Childhood life events[g]
Mania type: irritable vs. elated[e]	Childhood physical abuse[g]
Number hospitalizations: inpatient[e]	Life events at last visit: total[h]
Number hospitalizations: inpatient + day[e]	Life events at last visit: negative[h]
Presence of mixed episodes[c]	Life events at last visit: positive[h]
Presence of rapid cycling[c]	**Family history**
	First degree history completed suicide[i]
	First degree history bipolar disorder[i]
	First degree history depression[i]

[a]Hamilton Anxiety Rating Scale (HAM-A) administered by clinician at the baseline visit.

[b]Clinician-Administered Rating Scale for Mania (CARS-M) administered by clinician at the baseline visit.

[c]Final Best Estimate form of the DIGS evaluation administered by clinician at the screening visit.

[d]Migraine Questionnaire self-rated at the screening visit.

[e]Diagnostic Interview for Genetic Studies (DIGS) administered by clinician at the screening visit.

[f]Sheehan Disability Scale self-rated at the baseline visit.

[g]Childhood Life Events scale self-rated at the screening visit.

[h]Life Events Questionnaire from last study visit.

[i]Family History scale administered by clinician at screening visit.

observation phases among all patients who entered the study, and then separately over the maintenance phase among patients who entered the maintenance phase. To formally test for differences in association, we combined the stratified survival data and included in the Cox Proportional Hazard models an interaction term between the specific predictor and an indicator variable for the stabilization/observation versus maintenance phases.

To assess the robustness of observed associations to the assumptions of the survival analysis, we carried out two additional analyses. We defined two alternative but related response variables for analysis: (1) an acute response variable based on whether patients proceeded to the maintenance phase or not; and (2) a prophylactic response variable which contrasted patients who completed all study visits or who had reached the maintenance phase and were still active on study as of the data freeze on June 26, 2017 versus those who failed to remit or who relapsed on lithium monotherapy before completing all study visits. We then used logistic regression to examine the association between the clinical predictors and the two different dichotomous response variables in models that controlled for the same potential confounders as in the survival analysis. The inferences drawn from these two alternative logistic regression analyses were nearly identical to those from the survival analysis, so we report here the results from the survival analysis because it uses more of the available information provided by the prospective data and it provides a unified framework for analyzing the data over the entire time course of the study.

Finally, to evaluate the predictive ability of a model that included all clinical predictors individually found to be significantly associated with treatment failure, we carried out a receiver-operating curve (ROC) analysis specifically for survival data. We first carried out multiple imputation to fill in missing covariate data and maximize the available data for the ROC analysis. We note that we only used the multiple imputation procedure for this and not the primary analyses described above, and we used it only after confirming that analyses with the imputed dataset yielded results that were consistent with those reported from the primary analyses described above. Multiple imputation was performed on the predictor dataset with the mi command in STATA to generate 35 imputed datasets. A consensus imputed dataset was generated by taking the median (for continuous covariates) or modal (for categorical covariates) values across the 35 imputed datasets. We note that this procedure does not take into account the uncertainty in the consensus imputed estimates, but we reasoned it would be sufficient for obtaining reasonable estimates from the ROC analysis. We then proceeded to compare the ROC curves of nested models, including a base model that included the base variables controlled for in all analyses (age at study entry, sex, race, and lithium status upon entry into the study) and a full model that included the base variables plus all clinical predictors that were individually associated with treatment failure (see Table 3). The consensus imputed dataset was randomly split into ten non-overlapping subsets of approximately equal size, with approximately the same proportion of censored and event observations across all subsets. Cox models for the nested models were then fit using nine out of ten subsets, leaving the tenth

subset as a hold-out set. Using the results of the fitted models, linear predictor scores were obtained for observations in the hold-out set. Model fitting and prediction were repeated ten times, where a different subset of data was held out each time. Predicted survival ROC curves over 2 years were estimated for the linear predictions using the CoxWeights function from the risksetROC R package.[20,21] The area under the curve (AUC) for the ROC of the nested models were generated, and the differences in AUC were recorded. This process was repeated across 10,000 permutations of survival status and time of censoring pairings. The *p*-value for AUC difference between models was derived as the proportion of permuted AUC differences that were greater than the unpermuted AUC difference.

3 | RESULTS

Figure 1 shows a CONSORT-like flow diagram of the study. A total of 345 individuals were enrolled into the study and included in the analysis. Of these, a total of 194 patients successfully advanced to the maintenance phase, while 60 patients failed to remit on lithium monotherapy during stabilization and/or observation phases. Another 91 patients were terminated from the study for other reasons prior to the maintenance phase. Of the 194 patients who entered the maintenance phase, 41 experienced a relapse, 65 were terminated for other reasons, and 88 completed the study or were still in active treatment as of the date of data freeze.

Table 2 shows basic socio-demographic characteristics of the study sample broken down by the final outcome status of the patients, whether they completed the study (or were stabilized in maintenance and still active on the study), experienced a treatment failure, or were terminated for other reasons. There were no significant differences in age, sex or race between these three broad outcomes. Patients who entered the study stable on lithium monotherapy were significantly more likely to complete the study compared with those who either were on lithium and other psychotropic medications or were not on lithium on study entry. There were also significant differences between the sites in the outcomes achieved by the patients. These differences were largely explained by the proportion of patients at each site that entered the study stable on lithium, highlighting the importance of controlling for this potential confounder in subsequent analyses.

We then examined the association between hypothesized clinical predictors of lithium response and treatment response. Table 1 shows the list of clinical predictors that were selected a priori for investigation and the self and clinician rated scales from which they were derived. We examined each predictor individually in survival models controlling for factors that we reasoned may confound the relationship with treatment response because they are important socio-demographic factors or were found to differ with outcome status, including age at study entry, sex, race, and lithium status upon entry into the study. Table 3 shows the results for those clinical predictors that were significantly associated with treatment response at nominal significance of *p* < 0.05.
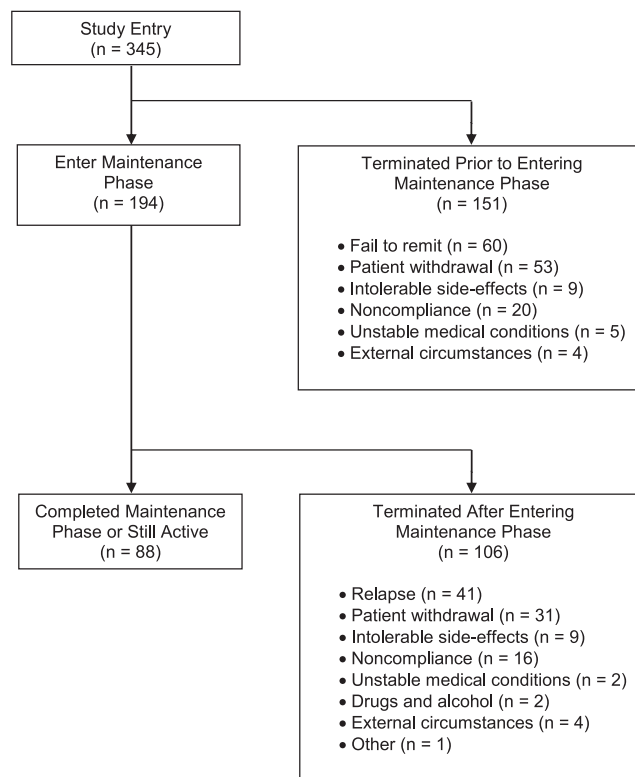


```
Study Entry
(n = 345)
    │
    ├──────────────────────────────────────┐
    ▼                                        ▼
Enter Maintenance              Terminated Prior to Entering
    Phase                          Maintenance Phase
  (n = 194)                            (n = 151)

                               • Fail to remit (n = 60)
                               • Patient withdrawal (n = 53)
                               • Intolerable side-effects (n = 9)
                               • Noncompliance (n = 20)
                               • Unstable medical conditions (n = 5)
                               • External circumstances (n = 4)
    │
    ├──────────────────────────────────────┐
    ▼                                        ▼
Completed Maintenance          Terminated After Entering
Phase or Still Active              Maintenance Phase
    (n = 88)                           (n = 106)

                               • Relapse (n = 41)
                               • Patient withdrawal (n = 31)
                               • Intolerable side-effects (n = 9)
                               • Noncompliance (n = 16)
                               • Unstable medical conditions (n = 2)
                               • Drugs and alcohol (n = 2)
                               • External circumstances (n = 4)
                               • Other (n = 1)
```

FIGURE 1 Consort-like flow diagram of patients in the Pharmacogenetics of Bipolar Disorder Prospective trial

The significant clinical predictors fell into four main categories: baseline anxiety symptoms, lifetime clinical features, daily functioning, and life events. The severity of anxiety symptoms at baseline as measured by total score on the HAM-A was significantly associated with increased risk of treatment failure when examined as a continuous covariate (hazard ratio [HR] 1.05, 95% confidence interval [CI] 1.03 to 1.08) and categorically as none, mild, moderate and severe (results not shown). Interestingly, a pre-existing diagnosis of co-morbid anxiety disorder meeting DSM-IV criteria was not associated with treatment response, suggesting that baseline symptoms rather than lifetime diagnosis are more relevant. The lifetime clinical features assessed at baseline that positively associated with increased risk of treatment failure included a history of migraine (HR 1.62, 95% CI 1.03 to 2.55), suicidal behavior (with an apparent dose-response relationship of HR 1.65, 95% CI 0.95 to 2.86 for ideation and HR 2.03, 95% CI 1.16 to 3.53 for more serious attempts) and history of mixed episodes (HR 1.60, 95% CI 1.01 to 2.53), as well as a chronic (non-episodic) pre-treatment course of illness (HR 2.92, 95% CI 1.76 to 4.83). Overall, functional disability related to illness was also an important predictor of treatment failure as assessed by the clinician with regard to lifetime disability (HR 1.80, 95% CI 1.13 to 2.86) and self-rated current disability on the Sheehan Disability Scale completed at baseline (HR 1.06, 95% CI 1.03 to 1.08). The self-rated current disability measure encompassed functional impairment in work, family and social life, all of which were found to be significantly associated with treatment failure, but for simplicity only results of total impairment are shown. Finally, life adversity in the form of past childhood physical abuse (HR

TABLE 2 Socio-demographic characteristics of the study sample by final outcome status

	Completed study[a] (n = 88)	Treatment failure[b] (n = 101)	Terminated other[c] (n = 156)	p-value
Age, mean years ±SD	43.84 ± 15.48	42.20 ± 13.32	41.66 ± 14.60	0.526
Sex, n (%)				0.492
Male	41 (46.59)	51 (50.50)	67 (42.95)	
Female	47 (53.41)	50 (49.50)	89 (57.05)	
Race, n (%)				0.055
Asian	2 (2.27)	1 (0.99)	4 (2.56)	
Black	7 (7.95)	7 (6.93)	28 (17.95)	
White	77 (87.50)	89 (88.12)	115 (73.72)	
More than one race	2 (2.27)	4 (3.96)	9 (5.77)	
Ethnicity, n (%)				0.936
Hispanic	3 (3.41)	3 (3.00)[d]	6 (3.85)	
Non-Hispanic	85 (96.59)	97 (97.00)	150 (96.15)	
Li status, n (%)[e]				<0.001
Li monotherapy	56 (63.64)	16 (15.84)	25 (16.03)	
Li plus other meds	19 (21.59)	47 (46.53)	58 (37.18)	
Not on Li	13 (14.77)	38 (37.62)	73 (46.79)	
Site, n (%)				0.001
UCSD	7 (7.95)	11 (10.89)	11 (7.05)	
Case Western	10 (11.36)	21 (20.79)	40 (25.64)	
Indiana	8 (9.09)	12 (11.88)	6 (3.85)	
Johns Hopkins	5 (5.68)	8 (7.92)	28 (17.95)	
Bergen	10 (11.36)	8 (7.92)	21 (13.46)	
Chicago	2 (2.27)	4 (3.96)	11 (7.05)	
Iowa	12 (13.64)	9 (8.91)	13 (8.33)	
Michigan	17 (19.32)	14 (13.86)	10 (6.41)	
Penn	4 (4.55)	4 (3.96)	8 (5.13)	
Dalhousie	12 (13.64)	9 (8.91)	5 (3.21)	
Mayo Clinic	1 (1.14)	1 (0.99)	3 (1.92)	

[a]Includes subjects who completed all study visits or achieved maintenance and still active on study.
[b]Includes subjects who failed to remit or who relapsed on lithium monotherapy.
[c]Includes subjects who withdrew from the study or were terminated for other reasons.
[d]The ethnicity of one patient was unknown.
[e]Lithium status at study entry; other medications refer to psychotropic medications except benzodiazepines or zolpidem.

1.97, 95% CI 1.28 to 3.03) or recent negative life events (HR 1.02, 95% CI 1.00 to 1.03) as captured by the Childhood Life Events scale or the Lifetime Events Questionnaire, respectively, were also associated with increased risk of treatment failure. Stratification of the survival analyses by study phase (Table S1) showed that the association of these clinical predictors with treatment failure did not significantly differ between the stabilization/observation versus maintenance phases (all interactions, $p > 0.05$).

To evaluate how well a model that included the significant clinical predictors identified above could predict lithium treatment failure over a 2-year period, we carried out an additional ROC analysis (Figure 2). The ROC figure shows the true positive rate (i.e., sensitivity) versus false positive rate (i.e., 1-specificity) of a model in predicting treatment failure in patients at different probability thresholds

of the model to declare outcome status. A key performance metric of the prediction model is the AUC. In our case, the AUC can be interpreted as the probability that the prediction model is able to correctly identify the patient who experienced a treatment failure when randomly presented with two patients, one who had a treatment failure and the other who did not. An AUC of 0.5 suggests random chance in choosing the correct patient, while an AUC of 1.0 suggests a 100% probability. Although the clinical predictors identified in our analysis above were highly correlated, we retained them all in the prediction model because multi-collinearity does not affect predictive performance (see Table S2 for parameter estimates of the multivariate prediction model). The ROC curve for a full model with all the identified clinical predictors plus the base variables had an AUC of 0.74, which was significantly different from the null ($p = 0.0001$).

TABLE 3 Hazard ratio (HR)[a] associations between clinical predictors and treatment response

	# treatment failures/total person-days[b]	HR (95% CI); p-value
Baseline anxiety symptoms[c]	98/105053	1.05 (1.03–1.08); p < 0.001
Chronicity of affective disorder		
Non-chronic	32/65683	1.00
Chronic	58/28693	2.92 (1.76–4.83); p < 0.001
History of migraine		
No	71/86811	1.00
Yes	28/17094	1.62 (1.03–2.55); p = 0.037
History of suicidal behavior		
None	22/42134	1.00
Suicidal ideation	32/30103	1.65 (0.95–2.86); p = 0.077
Suicide attempt	36/22915	2.03 (1.16–3.53); p = 0.012
History of mixed episodes		
No	46/65213	1.00
Yes	44/28388	1.60 (1.01–2.53); p = 0.046
Overall functioning		
Not disabled	46/67890	1.00
Disabled	39/24587	1.80 (1.13–2.86); p = 0.013
Disability at baseline: total[c]	97/102885	1.06 (1.03–1.08); p < 0.001
Disability at baseline: impairment		
No	36/68392	1.00
Yes	61/35389	1.85 (1.15–2.97); p = 0.011
Childhood physical abuse		
No	61/82077	1.00
Yes	37/21477	1.97 (1.28–3.03); p = 0.002
Negative life events[c]	91/94193	1.02 (1.00–1.03); p = 0.020

[a]Hazard ratios estimated from models that control for age, sex, race, and lithium status upon study entry.

[b]This is the number of treatment failures defined as failure to remit or a relapse on lithium monotherapy over the total number of days of follow-up from study entry to the last visit for all patients in the specific category; sums of treatment failures and person-days of follow-up may differ across covariates due to missing data.

[c]Covariate is continuous and for continuous covariates the number of treatment failures per total person-days of follow-up is shown for all subjects with non-missing data for that covariate.

This was better than the base model that included only the base variables, which had an AUC of 0.68, although the improvement was not significant (p = 0.13).

4 | DISCUSSION

We report here results from the PGBD Study, in which we examine clinical predictors of response to lithium treatment for bipolar disorder
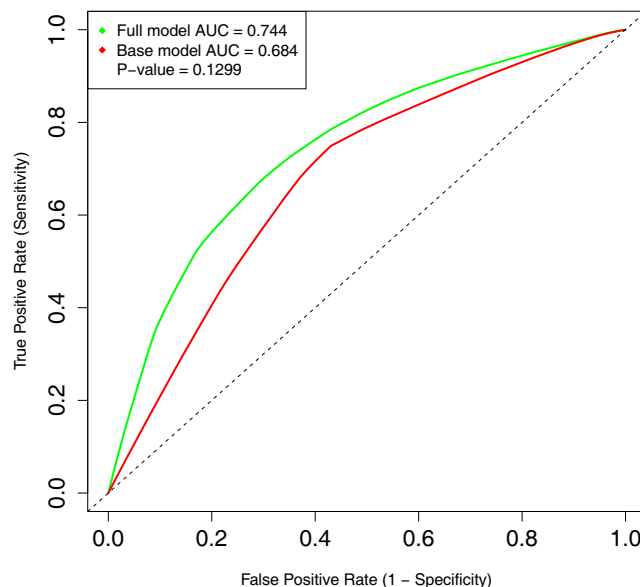

FIGURE 2 ROC curves for the prediction of lithium treatment failure over a 2-year period for a base model that included the base factors controlled for in all analyses (age at study entry, sex, race, and lithium status upon entry into the study) and a full model that included these base factors and all clinical predictors that were individually associated with treatment failure as reported in Table 3

(BD). Lithium is a first line treatment for BD and can be remarkably effective in controlling the devastating symptoms of BD. However, it is not effective in everyone and anywhere between 40%–50% of patients, or even more depending upon the length of follow-up, may require alternative therapeutic regimens. We identified several clinical markers that are associated with failure to respond to lithium treatment. These include current anxiety symptoms, functional impairments, negative life events and certain lifetime clinical features such as a history of migraine, suicidal ideation/attempts, and mixed episodes, as well as co-morbid personality disorder and a chronic course of illness. Future validation will be required to confirm whether these clinical markers are associated with treatment failure and whether they can be used clinically to effectively distinguish who will and will not do well on lithium before starting therapy. The particular significance of the present study is that it represents a prospective clinical evaluation of lithium response, in contrast to the extensive literature evaluating lithium response retrospectively.

The findings from our study agree with some, but not all, of the conclusions from previous studies in the literature. Similar to these studies, we found that poor functioning prior to treatment (as captured by unemployment in previous studies), negative life events, and personality disturbances were associated with poor treatment response. On the other hand, we found no evidence for an association of treatment response with age at onset, number of hospitalizations or rapid cycling. We did not have a direct measure of social status; however, we did have the number of years of education, which is a reasonable proxy for social status, but was not associated with treatment response. We also did not have sufficient data

to examine associations with episode pattern, which is a compelling observation that has been implicated by previous studies.

Unique to our study, we found intriguing associations of treatment response with current anxiety symptoms as well a history of migraine, suicidal ideation/attempts, and mixed episodes. The observation that symptom level, but not lifetime diagnosis, of anxiety was associated with treatment response echoes findings from the NIMH Collaborative Depression Study that the severity of anxiety is predictive of long-term morbidity in BD.[22] With regard to migraines, it has been shown that the prevalence of migraines in patients with BD is 2–3 times higher than in the overall population. Moreover, antiepileptic drugs, such as valproate, are used to treat migraines whereas lithium has no indication in their prophylactic treatment. Thus, comorbid migraine could mark an etiologically distinct sub-type of BD that is less responsive to lithium treatment.[23] With regard to suicidal behavior, we observed a "dose-response" relationship between the severity of suicidal behavior and increased risk for treatment failure, which lends further credence to the finding. However, this finding should not be taken as a reason for not prescribing lithium to suicidal patients, because lithium has been shown to be effective in reducing the risk of suicide,[24,25] even in people who do not experience full mood stabilization on lithium.[26] Interestingly, we found that the association with suicidal behavior was noticeably stronger in the maintenance phase of the study, although a formal interaction test of a difference by study phase was not significant. The interpretation of this finding is unclear and warrants further investigation.

If our findings are validated, they may help complete a clinical picture for the types of patients that do not respond well to lithium treatment and lead to clues about the underlying mechanisms that explain poor response. However, these findings should be interpreted in the light of certain limitations of the study. At least four such limitations merit further consideration. First, approximately 45% of enrolled patients were withdrawn or terminated from the study for a variety of reasons before the pre-specified endpoints of treatment failure or completion of all visits. Although relatively high, this attrition rate is not inconsistent with previous long-term maintenance treatment trials of bipolar disorder. A systematic review of such trials observed attrition rates ranging from 19% to 98% (median = 68%), with higher study withdrawals for lithium compared to other mood stabilizers.[27,28] The survival analysis we carried out assumed the risk of treatment failure for patients lost to follow-up was the same as for those who stayed on the study per protocol. It is possible this assumption was not true, and patients who did not complete the study per protocol did so because they were different somehow and possibly experiencing complications that were a precursor to treatment failure. Consistent with this, we did observe differences in certain baseline characteristics for those who did not complete the study per protocol. These individuals tended to be younger ($p = 0.091$), non-white ($p = 0.027$), and not stable on lithium monotherapy upon study entry ($p < 0.001$). However, we carried out two alternative analyses of the data and the findings were remarkably similar, suggesting the findings were robust to assumptions made by the survival analysis.

Second, in order to broaden the available population for study, we included patients who were naïve to lithium as well as those who may have taken lithium in the past or were currently on it. It is likely the response trajectories while on study would be different for these patients. Indeed, over one-quarter of the patients entered the study stable on lithium monotherapy and their treatment outcomes were notably better. To account for these differences, we tightly controlled for lithium status in the analysis, so that inferences about the associations with treatment response would not be confounded by these differences.

Third, the sample size may not have provided sufficient power to detect significant associations with important clinical predictors with smaller effect sizes. However, we emphasize this is one of the largest prospective studies specifically designed from inception to investigate predictors of lithium response. Indeed, it is the only such study that sought to treat patients with monotherapy in order to more firmly link treatment predictors with lithium response unclouded by the use of other psychotropic medications that are frequently taken by patients with BD. This is a unique and noteworthy strength of this study.

Finally, given the limitations of the current study, we could not distinguish whether the identified clinical predictors are specific for non-response to lithium versus other medications or reflect the natural course of illness. This is a limitation that is common to many previous studies of lithium treatment in BD. The ultimate goal of precision medicine is to identify predictors that can predict response to one treatment versus another so that they can be used to make clinical decisions about starting one over the other. We are unable to make such recommendations from the current findings. However, they do offer hypotheses that can be tested in other samples to determine if they can predict response to other treatments.

Given the devastating burden of BD, there is considerable motivation to develop more effective strategies for treating the disorder. Lithium is an inexpensive and effective treatment, but it does not work for everyone. It would be of tremendous clinical benefit if we could identify predictors of who will and will not respond to lithium before starting treatment. This study provides new evidence that certain clinical factors could be used to help with such predictions and help inform decisions about whether patients presenting with BD should be started on lithium or alternative medications. Interestingly, we found that a model which included these clinical factors could predict lithium treatment failure with an AUC of 0.74 that was significantly better than the null. The AUC observed in our study is similar, albeit less, to the one achieved in a recent report on predicting lithium response using a machine learning method against 180 clinical predictors.[29] Another recent study reported on a clinical prediction model that explained 17.4% of the variance in observed outcome scores in response to treatment with lithium over a 6-month period in a randomized comparative effectiveness trial.[30] The hope is that eventually we will be able to improve these prediction models by incorporating both clinical and biological (e.g., neuroimaging, neurophysiology, and molecular) factors. This is the goal of the PGBD, and this report is a first step toward this goal.

AUTHOR CONTRIBUTIONS

Wade Berrettini, William Coryell, Mark Frye, Elliot Gershon, Susan Leckband, Melvin McInnis, John Nurnberger, Ketil Oedegaard, Paul Shilling, Joseph Calabrese, Martin Alda, Caroline Nievergelt, Peter Zandi, and John Kelsoe contributed to conception and design. Yian Lin, Adam Maihofer, Emma Stapp, Megan Ritchey, Ney Alliey-Rodriguez, Amit Anand, Yokesh Balaraman, Wade Berrettini, Holli Bertram, Abesh Bhattacharjee, Cynthia Calkin, Carla Conroy, William Coryell, Nicole D'Arcangelo, Anna DeModena, Joanna Biernacka, Carrie Fisher, Nicole Frazier, Mark Frye, Keming Gao, Julie Garnham, Elliot Gershon, Kara Glazer, Fernando Goes, Toyomi Goto, Elizabeth Karberg, Gloria Harrington, Petter Jakobsen, Masoud Kamali, Marisa Kelly, Susan Leckband, Falk Lohoff, Andrea Stautland, Michael McCarthy, Melvin McInnis, Francis Mondimore, Gunnar Morken, John Nurnberger, Ketil Oedegaard, Kelly Ryan, Martha Schinagle, Helle Schoeyen, Ole Andreassen, Martha Shaw, Paul Shilling, Claire Slaney, Bruce Tarwater, Joseph Calabrese, Martin Alda, Caroline Nievergelt, Peter Zandi, and John Kelsoe contributed to data acquisition. Yian Lin, Adam Maihofer, Emma Stapp, Megan Ritchey, Wade Berrettini, William Coryell, Mark Frye, Elliot Gershon, Andrea Stautland, Melvin McInnis, John Nurnberger, Ketil Oedegaard, Paul Shilling, Joseph Calabrese, Martin Alda, Caroline Nievergelt, Peter Zandi, and John Kelsoe contributed to data analysis and interpretation. Yian Lin, Adam Maihofer, Emma Stapp, Megan Ritchey, Amit Anand, Wade Berrettini, William Coryell, Mark Frye, Elliot Gershon, Fernando Goes, Michael McCarthy, Melvin McInnis, Francis Mondimore, John Nurnberger, Ketil Oedegaard, Joseph Calabrese, Martin Alda, Caroline Nievergelt, Peter Zandi, and John Kelsoe contributed to manuscript writing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study will be made openly available on dbGaP (https://www.ncbi.nlm.nih.gov/gap/).

ORCID

Yian Lin https://orcid.org/0000-0002-6343-7942
Amit Anand https://orcid.org/0000-0003-1016-6904
William Coryell https://orcid.org/0000-0003-3989-7276
Joanna M. Biernacka https://orcid.org/0000-0001-9350-4440
Keming Gao https://orcid.org/0000-0003-1130-2461
John I. Nurnberger https://orcid.org/0000-0002-7674-1767
Martin Alda https://orcid.org/0000-0001-9544-3944
Peter P. Zandi https://orcid.org/0000-0001-8423-2623

REFERENCES

1. Berghofer A. Lithium and suicide. *BMJ*. 2013;347:f4449.
2. Coryell W. Maintenance treatment in bipolar disorder: a reassessment of lithium as the first choice. *Bipolar Disord*. 2009;11:77-83.
3. Grof P, Müller-Oerlinghausen B. A critical appraisal of lithium's efficacy and effectiveness: the last 60 years. *Bipolar Disord*. 2009;11:10-19.
4. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian network for mood and anxiety treatments (CANMAT) and international society for bipolar disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord*. 2018;20:97-170.
5. Miura T, Noma H, Furukawa TA, et al. Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2014;1:351-359.
6. Malhi GS, Bell E, Boyce P, et al. The 2020 royal australian and new zealand college of psychiatrists clinical practice guidelines for mood disorders: Bipolar disorder summary. *Bipolar Disord*. 2020;22:805-821.
7. Sakurai H, Kato M, Yasui-Furukori N, et al. Medical Education Panel of the Japanese Society of Clinical Neuropsychopharmacology: pharmacological management of bipolar disorder: Japanese expert consensus. *Bipolar Disord*. 2020;22:822-830.
8. Geddes JR, Burgess S, Hawton K, Jamison K, Goodwin GM. Long-term lithium therapy for bipolar disorder: Systematic review and meta-analysis of randomized controlled trials. *Am J Psychiatry*. 2004;161:217-222.
9. Yildiz A, Vieta E, Leucht S, Baldessarini RJ. Efficacy of antimanic treatments: meta-analysis of randomized, controlled trials. *Neuropsychopharmacology*. 2011;36:375-389.
10. Baldessarini RJ, Tondo L. Does lithium treatment still work? Evidence of stable responses over three decades. *Arch Gen Psychiatry*. 2000;57:187-190.
11. 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.
12. Rybakowski JK, Chlopocka-Wozniak M, Suwalska A. The prophylactic effect of long-term lithium administration in bipolar patients entering treatment in the 1970s and 1980s. *Bipolar Disord*. 2001;3:63-67.
13. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372:793-795.
14. Baastrup PC, Mogens S. Lithium as a prophylactic agent: its effect against recurrent depressions and manic-depressive psychosis. *Arch Gen Psychiatry*. 1967;16:162-172.
15. Kleindienst N, Engel RR, Greil W. Psychosocial and demographic factors associated with response to prophylactic lithium. A systematic review for bipolar disorders. *Psychol Med*. 2005;35:1685-1694.
16. Kleindienst N, Engel RR, Greil W. Which clinical factors predict response to prophylactic lithium? A systematic review for bipolar disorders. *Bipolar Disord*. 2005;7:404-417.
17. 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;140:94-115.
18. Oedegaard KJ, Alda M, Anand A, et al. The pharmacogenomics of bipolar disorder study (PGBD): identification of genes for lithium response in a prospective sample. *BMC Psychiatry*. 2016;16:129.
19. Nolen WA, Licht RW, Young AH, et al. ISBD/IGSLI Task Force on the treatment with lithium: what is the optimal serum level for lithium in the maintenance treatment of bipolar disorder? A systematic review and recommendations from the ISBD/IGSLI task force on treatment with lithium. *Bipolar Disord*. 2019;21:394-409.
20. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005;61:92-105.
21. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56:337-344.

22. Coryell W, Solomon DA, Fiedorowicz JG, Endicott J, Schettler PJ, Judd LL. Anxiety and outcome in bipolar disorder. *Am J Psychiatry.* 2009;166:1238-1243.

23. Jacobsen KK, Nievergelt CM, Zayats T, et al. Genome wide association study identifies variants in NBEA associated with migraine in bipolar disorder. *J Affect Disord.* 2015;172:453-461.

24. Baldessarini RJ, Tondo L, Davis P, Pompili M, Goodwin FK, Hennen J. Decreased risk of suicides and attempts during long-term lithium treatment: A meta-analytic review. *Bipolar Disord.* 2006;8:625-639.

25. Cipriani A, Hawton K, Stockton S, Geddes JR. Lithium in the prevention of suicide in mood disorders: updated systematic review and meta-analysis. *BMJ.* 2013;346:f3646.

26. Ahrens B, Müller-Oerlinghausen B. Does lithium exert an independent antisuicidal effect? *Pharmacopsychiatry.* 2001;34:132-136.

27. Smith LA, Cornelius V, Warnock A, Bell A, Young AH. Effectiveness of mood stabilizers and antipsychotics in the maintenance phase of bipolar disorder: a systematic review of randomized controlled trials. *Bipolar Disord.* 2007;9:394-412.

28. Sylvia LG, Reilly-Harrington NA, Leon AC, et al. Methods to limit attrition in longitudinal comparative effectiveness trials: lessons from the lithium treatment - moderate dose use study (LiTMUS) for bipolar disorder. *Clin Trials.* 2012;9:94-101.

29. Nunes A, Ardau R, Berghofer A, et al. Prediction of lithium response using clinical data. *Acta Psychiatr Scand.* 2020;141:131-141.

30. Kim TT, Dufour S, Xu C, et al. Predictive modeling for response to lithium and quetiapine in bipolar disorder. *Bipolar Disord.* 2019;21:428-436.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.