DOI: 10.1111/bdi.12752

ORIGINAL ARTICLE

WILEY BIPOLAR DISORDERS
AN INTERNATIONAL JOURNAL OF PSYCHIATRY AND NEUROSCIENCES

Predictive modeling for response to lithium and quetiapine in bipolar disorder

Thomas T. Kim[1] | Steven Dufour[2] | Colin Xu[1] | Zachary D. Cohen[1] | Louisa Sylvia[2,3] | Thilo Deckersbach[2,3] | Robert J. DeRubeis[1] | Andrew A. Nierenberg[2,3]

[1]University of Pennsylvania, Philadelphia, Pennsylvania

[2]The Dauten Family Center for Bipolar Treatment Innovation, Massachusetts General Hospital, Boston, Massachusetts

[3]Harvard Medical School, Boston, Massachusetts

Correspondence
Thomas T. Kim, University of Pennsylvania, Philadelphia, PA.
Email: thomastk@sas.upenn.edu

Abstract

Objectives: Lithium and quetiapine are known to be effective treatments for bipolar disorder. However, little information is available to inform prediction of response to these medications. Machine-learning methods can identify predictors of response by examining variables simultaneously. Further evaluation of models on a test sample can estimate how well these models would generalize to other samples.

Methods: Data (N = 482) were drawn from a randomized clinical trial of outpatients with bipolar I or II disorder who received adjunctive personalized treatment plus either lithium or quetiapine. Elastic net regularization (ENR) was used to generate models for lithium and quetiapine; these models were evaluated on a test set.

Results: Predictions from the lithium model explained 17.4% of the variance in actual observed scores of patients who received lithium in the test set, while predictions from the quetiapine model explained 32.1% of the variance of patients that received quetiapine. Of the baseline variables selected, those with the largest parameter estimates were: severity of mania; attention-deficit/hyperactivity disorder (ADHD) comorbidity; nonsuicidal self-injurious behavior; employment; and comorbidity with each of two anxiety disorders (social phobia/society anxiety and agoraphobia). Predictive accuracy of the ENR model outperformed the simple and basic theoretical models.

Conclusion: ENR is an effective approach for building optimal and generalizable models. Variables identified through this methodology can inform future research on predictors of response to lithium and quetiapine, as well as future modeling efforts of treatment choice in bipolar disorder.

KEYWORDS
bipolar disorder, lithium, machine learning, predictive modeling, quetiapine

1 | INTRODUCTION

Bipolar disorder (BD) is a chronic, severely impairing condition characterized by both depressive and manic/hypomanic episodes. Among the conditions described in DSM-5, BD is among those for which the first line of treatment tends to be pharmacologic, with psychosocial interventions generally serving supplemental purposes.[1] Fortunately, several decades of clinical research has yielded numerous effective medications to treat symptoms of BD; however, despite evidence of their clinical effectiveness, on average, many patients do not experience satisfactory outcomes.[2,3]

Guidelines for BD treatment published over the past decade generally recommend lithium, a gold-standard mood stabilizer, or second-generation antipsychotics (SGAs) as first-line treatments.[1,4] However, these pharmacotherapies have only recently been compared against each other, and evidence has not indicated that any particular medication is superior on average.[5-7] Randomized controlled trials (RCTs) of lithium vs quetiapine specifically have yielded findings suggesting similar average levels of response,[6-8] even in studies that treated BD patients with monotherapy or with adjunctive personalized treatments.

Given its widespread use in BD treatment, lithium has received the most attention in the literature, including efforts to identify features that serve a prognostic function in lithium treatment. Kleindienst et al[9] found that an episodic pattern of mania followed by depression or an older age of onset predicted good response to lithium. Poorer response to lithium was observed in patients with an episodic pattern of depression followed by mania, greater numbers of previous hospitalizations, or a course that included continuous cycling. Other studies conclude that no psychiatric comorbidity, the absence of mixed episodes, and a family history of lithium response predict a good response, whereas a history of childhood physical abuse predicts a poor response in lithium treatment.[10-12] Efforts have also been made to identify genetic predictors of lithium response.[13] Despite research into individual predictors of response to lithium pharmacotherapy, there are currently no published reports examining these predictors simultaneously. By examining predictors of response in aggregate, novel predictive associations may be detected.[14] Furthermore, the estimated impact (ie, coefficients) of individual predictors may change when examined simultaneously.[15] Thus, examining the effect of predictors in aggregate increases accuracy in the identification of predictors and improves replicability.

Quetiapine is another treatment for BD which has been shown to be more efficacious than placebo, and as effective as lithium.[5-7] Quetiapine has further advantages compared to lithium in the treatment of BD such as a relatively safer side-effect profile.[16] Nevertheless, to our knowledge, there have been no previously published attempts to identify predictors of response to quetiapine in the treatment of BD.

It is arguable that there has been no systematic attempt to identify predictors of response to lithium or quetiapine treatment in BD. Previous research on individual predictors of response to lithium therapy may have missed novel predictors by not examining them in aggregate[14]; and there has been no research on predictors of response to quetiapine. Subsequently, the aim of the present study was to systematically examine predictors of response to lithium and quetiapine therapy in the treatment of BD.

One way to systematically identify predictors from a large pool of variables is through machine-learning techniques.[17] These data-mining methods, when used appropriately, not only select variables, but also provide unbiased parameter estimates in the context of a multivariable model.[18] For instance, previous research has used machine-learning methods such as elastic net regularization (ENR) [18] to predict depressed patients' response to antidepressant treatment; this research found that the predictors of treatment response identified can generalize to external samples.[14,19]

The aim of the present study was to examine data obtained from the multisite *Clinical Health Outcomes Initiative in Comparative Effectiveness* RCT of lithium and quetiapine. We applied ENR to pretreatment patient data to identify predictors of response to treatment with lithium or quetiapine as well as create models for both treatment conditions. Since the available biological predictors from this RCT cannot distinguish lithium and quetiapine mechanistically, we decided to focus on building within-treatment models with the goal of identifying individual predictors of response to each treatment rather than differential response to treatment. We further compared the effectiveness of our model generated through ENR against a basic model of symptom severity and a theoretical model informed through past research.[9-12]

2 | MATERIALS AND METHODS

2.1 | Study

This paper draws from data collected from the *Clinical Health Outcomes Initiative in Comparative Effectiveness* (Bipolar CHOICE) study.[6] Bipolar CHOICE was a 6-month RCT that compared the efficacy of lithium and quetiapine for individuals with bipolar I or II disorder. The original study was approved by the institutional review board (IRB) at the Massachusetts General Hospital-Partners HealthCare as well as the IRBs at the other 10 sites. Patients signed approved informed consent forms in the presence of study clinicians prior to any initiation of study procedures. The study is registered on ClinicalTrials.gov (identifier: NCT01331304). Necessary clinical adjustments to the medication regimens of individual participants beyond the primary study medication were permitted as adjunctive personalized treatments (APT). Guidelines for these adjustments followed the Texas Implementation of Medication Algorithms.[20] Adjunctive personalized treatments allowed clinicians to flexibly use the best evidence-based BD treatment(s); however, treatment was restricted in that the lithium + APT group could not receive quetiapine or any other SGA, and the quetiapine + APT group could not receive lithium or any other SGA. Further rationale for study procedures is reported elsewhere.[21]

2.2 | Participants

Patients were screened based on age, DSM-IV-TR criteria for bipolar I or II disorder, and current symptomatic status. Exclusion criteria included contraindication to lithium or quetiapine, risk of harm to one's self or others, and current medical use of lithium or quetiapine. Some contraindications include prior hypersensitivity to lithium or quetiapine and pregnancy. The final sample size comprised 482 patients, with 240 patients receiving lithium + APT (Li + APT) and 242 patients receiving quetiapine + APT (QTP + APT). More sample details (eg, demographics and patient flowchart) are reported in the primary outcomes paper.[6] The dataset was randomly split into a training set for model building and a test set for model evaluation, as described below.

2.3 | Outcome

2.3.1 | Clinical global impressions scale-bipolar version (CGI-BP)

The CGI-BP is a clinician-rated scale that measures severity of mania, depression, and overall BD illness.[22] The CGI-BP for overall BD illness at 6 months was used as the primary outcome measure for analysis. Scores on the CGI-BP overall BD illness severity item range from 1 to 7, with higher scores indicating greater severity. The mean posttreatment CGI-BP scores of patients in both conditions were not significantly different (lithium mean = 3.16; quetiapine mean = 2.92; $t(236) = 1.34$, $P = 0.18$). Because the CGI-BP is a clinician-rated measure of perceived clinical improvement, patients are not assessed at intake. Patients were rated on the CGI-BP at each visit and assessed on symptom severity 7 days prior to each visit. CGI-BP assessment periods were over the 6-month trial period.

2.4 | Predictors

2.4.1 | Quality of life enjoyment and satisfaction questionnaire (Q-LES-Q)

The Q-LES-Q is a well-established self-report measure of quality of life across several domains. Higher scores indicate greater life satisfaction.[23]

2.4.2 | Concise health risk tracking scale (CHRT)

The CHRT is a self-report measure of depression and suicidal risk validated in a bipolar sample. Higher scores indicate greater risk.[24]

2.4.3 | Longitudinal interval follow-up evaluation—Range of Impaired Functioning Tool (LIFE-RIFT)

This clinician-rated scale assesses disorder-related impairment across several domains of individual and interpersonal functioning. Higher scores indicate greater levels of impairment.[25]

2.4.4 | Bipolar inventory of symptoms scale (BISS)

The BISS is a clinician-rated assessment of symptom severity in BD with subscales for mania, depression, anxiety, irritability, and psychosis. Higher scores indicate greater disorder severity.[26]

2.4.5 | Biological predictors

Blood-circulating levels of white blood cell count, hemoglobin, platelet count, blood urea nitrogen, creatinine, sodium, potassium, chloride, bicarbonate, thyroid stimulating hormone, fasting blood glucose, and cholesterol were regularly monitored at each site as standard treatment protocol given that some patients were randomly assigned to lithium therapy. These biological predictors were measured at intake and included as predictors; no reference criteria were used to standardize across sites given that raw scores for each patient were entered into the prediction algorithm.

2.4.6 | Demographics and other predictors

Additional patient characteristics measured at baseline included: demographics, DSM-IV-TR diagnoses at intake, employment status, physical health conditions, and substance abuse (see Table S1).

Since variable selection requires at least 15% of the sample per group for each predictor, adjunctive personalized treatments were grouped together to possibly include as many APTs as possible into variable selection (eg, Prozac and Trazadone were considered part of the antidepressant group). The most frequent adjunctive personalized treatment categories were: antidepressants, mood stabilizers, and stimulants. Unfortunately, concurrent antidepressants and stimulants did not meet the variance threshold and could not be included in variable selection. For example, only 58 patients or 12.0% of the dataset received a concurrent antidepressant and only 15 patients or 3.1% received a concurrent stimulant. However, patients that received other concurrent mood stabilizers (aside from either lithium or quetiapine; this could include antipsychotics) were included in the analysis (146 patients or 30.3% received another mood stabilizer).

Psychopharmaceutical medications were sorted for past drug history through a similar method.

2.5 | Data preprocessing

For each treatment (lithium and quetiapine), we randomized the full sample into a training and test set. An 80/20 split was implemented, whereby a random 80% of the full sample (192 lithium, 194 quetiapine) constituted the training set, and the test set comprised the remaining 20% (48 lithium, 48 quetiapine). Previous researchers have recommended an 80/20 split between the training and testing set, as this achieves a balanced compromise between bias and variance for moderate sample sizes.[27] Because of the moderate-large sample size (N = 482), by using an 80/20 split (as opposed to a 90/10 or 50/50 split), we can prioritize having a sufficiently large training set for the purposes of model building, while maintain an adequately large sample size for model evaluation.[28]

After forming a training and test set, imputation of missing data was performed separately for each training and testing set with the *R* package missForest.[28] Categorical variables were made binary where appropriate.[29] Variables were removed if more than 20% of patients had missing values and if variables did not have enough variance (binary variables required at least 15% of the sample in each category). Previous research recommends at least 10% of the sample in each category (for our dataset, this recommendation asks for 48.2 patients per group for each predictor); however, given the possibility of strong, but rare predictors (eg, suicidal behavior), we decided to use 15% of the sample

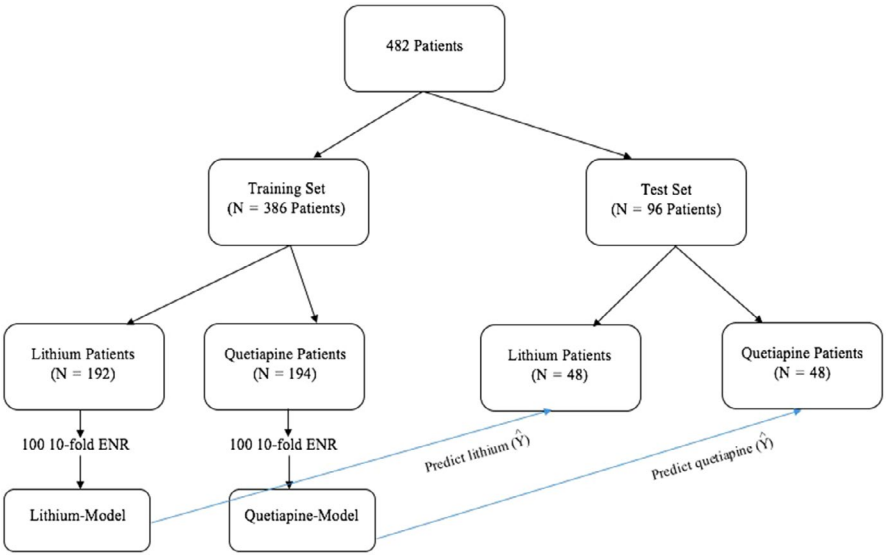

FIGURE 1 Analysis scheme [Colour figure can be viewed at wileyonlinelibrary.com]

in each category (ie, 72.3 patients per group for each predictor) as an appropriate cutoff.[27] Variables that showed a high correlation ($r \geq 0.65$) with another variable were removed depending on interpretability or significance in meaning. For example, a correlation of 0.89 was obtained between cholesterol and LDL. LDL was removed because total cholesterol was more interpretable. Outliers for continuous variables—defined as values ranging above and below 1.5% of the interquartile range—were winsorized, and variables with skewed distributions were log-transformed, where appropriate. All baseline predictors were then mean-centered and standardized using the *R* package caret (the function is preProcess[30]) to protect against potential errors in statistical inference.[31]

2.6 | Model building

All analyses were conducted in R 3.4.2. Figure 1 shows the analysis scheme. For each training set, two separate models were constructed using ENR with the glmnet package[32]: one model for patients who received lithium and one for patients who received quetiapine. Models were built to predict patient CGI-BP end-of-treatment scores at 6 months. We entered 100 predictors into each model building procedure (see Table S1 for full list of variables). ENR has been shown to robustly maintain predictive accuracy even with a large number of predictors relative to the number of observations.[18] Interactions between predictors were not modeled to maintain parsimony.

ENR uses features of least absolute shrinkage and selection operator (LASSO) and ridge regression to regularize parameter estimates.[18] While an ordinary least squares regression may produce an overfitted model when including many predictors, regularization protects from overfitting by incorporating a penalty term to the regression model that limits the size of coefficients.[33] Thus, regularization allows for more parsimonious and replicable models.[33]

The LASSO uses the L1 penalty that shrinks regression terms by the absolute value of the magnitude of coefficients. LASSO can perform feature selection by shrinking coefficients to 0, but it is inefficient when selecting variables from groups of highly correlated predictors.[18] Ridge regression uses the L2 penalty that shrinks regression terms equal to the square of the magnitude of coefficients. Ridge regression is able to shrink parameter estimates, but it cannot perform feature selection. ENR combines the L1 penalty of LASSO regression with the L2 penalty of ridge regression to achieve a compromise for variable selection while providing generalizable parameter estimates, even if variables are highly correlated.[18]

This compromise of L1 and L2 is achieved by optimizing the two tuning parameters of ENR: alpha and lambda. The alpha parameter determines the ratio of L1/L2 penalization, where alpha = 1 corresponds with pure LASSO regularization, and alpha = 0 corresponds with pure ridge regression. By tuning the alpha parameter, an optimal balance between L1 and L2 penalization can be determined to build the final model.[18] The lambda parameter determines the degree of penalization, where larger lambda values correspond with heavier shrinkage of regression coefficients.[33]

In order to determine the optimal alpha values for the lithium model and the quetiapine model, we ran 100 iterations of 10-fold cross validation with randomly drawn splits within each of the training samples. Within each iteration of 10-fold cross validation, we tested alpha values from 0.05 to 0.95, in 0.05 increments. Values of 0 and 1 were excluded because they correspond to pure ridge regression and LASSO, respectively. The optimal alpha values were those that produced the lowest cross-validation error (mean-squared error of prediction) over the 100 iterations.

Once the alpha parameters for the lithium model and quetiapine model were determined, the lambda parameters for each respective model were then tuned to determine the degree of penalization to further improve the predictive accuracy of the final model. With these alpha parameters, 100 iterations of 10-fold cross-validations were again run within each training sample; the optimal lambda value was determined by averaging the lambda value for the iterations corresponding with the lowest mean

squared error. The resulting alpha and lambda parameters were then used to create the final models for lithium and quetiapine, respectively.

2.7 | Basic and theoretical models

In order to evaluate the relative predictive accuracy of the model generated through ENR, we constructed two other models for each of lithium and quetiapine: a basic symptom-severity model and a theoretical model. For the basic models, we fitted a linear regression of baseline BISS mania and baseline BISS depression predicting 6-month CGI-BP end-of-treatment scores for patients within each respective training set. For the theoretical models, we fitted a linear regression in each respective training set predicting 6-month CGI-BP scores using baseline BISS mania and baseline BISS depression in combination with predictors identified by previous research, specifically: age of disorder onset, psychiatric comorbidity, history of childhood abuse, previous hospitalization, and family history of BD.[9-12]

2.8 | Model evaluation

We then applied the lithium and quetiapine models to their respective held-out test samples. Each model generated a prediction of 6-month CGI-BP scores for patients in the respective held-out group. We then regressed actual observed 6-month CGI-BP on these predicted 6-month CGI-BP scores to evaluate the predictive accuracy of our models. By regressing the observed values in the held-out test sample against the predicted values for the held-out test sample, we gain an estimate (ie, the R^2) of how much variance these models would explain in a new sample.

The basic symptom-severity and theoretical models were also applied to their respective held-out test samples to generate predictions of CGI-BP scores for patients in each respective held-out group. Predictive accuracy was similarly evaluated by regressing actual observed 6-month CGI-BP scores against predicted 6-month CGI-BP scores, generating R^2's to reflect the predictive accuracy of each model.

Importantly, since our final models were derived entirely from the training data, any estimates of variance explained in the test set (in the form of an R^2) is an unbiased estimate of the generalizable predictive accuracy of our models.[34,35] Thus, a model producing a larger R^2 reflects the predictive accuracy of a model to prospectively predict outcomes in a held-out sample.

3 | RESULTS

Tables 1 and 2 present the ENR models, with beta weights, for lithium and quetiapine, respectively (the largest 20 parameters are listed; the full model with coefficients can be found in Tables S2 and S3). The lithium model featured non-suicidal self-injurious behavior (NSSI), attention-deficit/hyperactivity disorder (ADHD), BISS mania, social phobia/social anxiety disorder, and suicide risk. The quetiapine model featured employment, agoraphobia, BISS irritability, substance dependence marijuana, and BISS depression.

3.1 | Predictive accuracy

The predictive accuracy of actual observed 6-month CGI-BP scores regressed against predicted CGI-BP scores for the ENR models produced an R^2 of 0.174 for patients receiving lithium, and an R^2 of 0.321 for patients receiving quetiapine (see Figure 2). Thus, our lithium model explained 17.4% of the variance in the actual observed CGI-BP scores at 6 months, while our quetiapine model explained 32.1% of the variance in the actual observed CGI-BP scores at 6 months.

Conversely, the basic symptom-severity model for lithium produced an R^2 of 0.142 and the basic symptom-severity model for quetiapine produced an R^2 of 0.235. The theoretical model for lithium produced an R^2 of 0.111 and the theoretical model for quetiapine produced an R^2 of 0.207.

4 | DISCUSSION

This study presents a data-driven method to systematically examine predictors of response to lithium and quetiapine in BD. Previous research has only looked at individual predictors of response to

TABLE 1 Elastic net regularization lithium model (top 20 largest parameters)

Predictor	Coefficient
Nonsuicidal self-injurious behavior	0.207
Attention-deficit/hyperactivity disorder	0.191
BISS mania	0.150
Social phobia/social anxiety disorder	0.102
Suicide risk	0.097
BISS psychosis	0.096
BISS Anxiety	0.087
Agoraphobia	0.081
Past antidepressant	0.076
Depression	−0.074
CHRT	0.068
Alcohol family history	−0.067
Height	−0.066
Bipolar family history	−0.064
Major depressive disorder	0.057
Bipolar disorder II	−0.053
Blood pressure	0.050
Depression family history	−0.047
BISS irritability	0.047
QLES SUM health	−0.046

BISS, bipolar inventory of symptoms scale; CHRT, concise health risk tracking scale; QLES, quality of life enjoyment and satisfaction questionnaire.

TABLE 2 Elastic net regularization quetiapine model (top 20 largest parameters)

Predictor	Coefficient
Employment	0.155
Agoraphobia	0.105
BISS irritability	0.090
Substance dependence marijuana	0.082
BISS depression	0.080
Marital status	0.080
BISS anxiety	0.069
Generalized anxiety disorder	0.068
LIFE RIFT	0.067
Manic episode	0.064
Race—white	−0.061
Major depressive disorder	0.060
QLES SUM work	−0.052
Asthma	−0.051
Previous psychiatric hospitalization	0.048
Hyperlipidemia	0.046
QLES SUM leisure	−0.045
HDL	0.044
QLES SUM health	−0.043
Insomnia	0.041

BISS, bipolar inventory of symptoms scale; CHRT, concise health risk tracking scale; LIFE-RIFT, Longitudinal interval follow-up evaluation-range of impaired functioning tool; QLES, quality of life enjoyment and satisfaction questionnaire.

lithium; furthermore, to our knowledge, there has been no research of individual predictors of response to quetiapine. Examining multiple predictors simultaneously can identify novel predictive associations that could have been overlooked.

We used ENR with internal cross-validation to determine optimal alpha and lambda tuning parameters; these parameters were used to generate two final predictive models for lithium and quetiapine. Internal cross-validation minimized the risk of overfitting to the training data.[27]

By using a training set for model building and testing these models on a separate test set, we were able to obtain unbiased estimates of model generalizability.[27] Furthermore, by regressing observed outcomes against the model-predicted outcomes, we were able to calculate an estimate for how much variance our models would explain in a novel test sample. Our lithium model explained 17.4% of the variance in actual observed CGI-BP scores in the test sample, and our quetiapine model explained 32.1% of the variance in the test sample. When compared to a similar analysis using ENR to predict response to antidepressant medication, their model explained 17.5% of the variance in response.[14] To our knowledge, there is not much research to compare the R^2 utility of a model in the field of mental disorders; however, using Cohen's suggestions for behavioral sciences, he would consider the R^2 for the lithium model to be a large effect and the R^2 for the quetiapine model to be a very large effect.[36]

Thus, using only variables measured at intake, our lithium model was able to explain a medium proportion of the variance (equivalent to an r = 0.41) of CGI-BP scores at 6 months, while our quetiapine model was able to explain a large proportion of the variance (equivalent to an r = 0.57).[37] Furthermore, the models generated by ENR explained a much larger proportion of variance in the test sample compared to the basic symptom-severity models (lithium model: 14.2% of variance; quetiapine model: 23.5% of variance) and the theoretical models (lithium mode: 11.1% of variance; quetiapine model: 20.7% of variance).

ENR was able to produce generalizable models that predict a significant portion of the variance in a completely held-out test sample using only variables measured at intake. External validity was preserved by dividing our data into training and testing sets prior to any data preprocessing, data cleaning, missing data imputation, or model building.

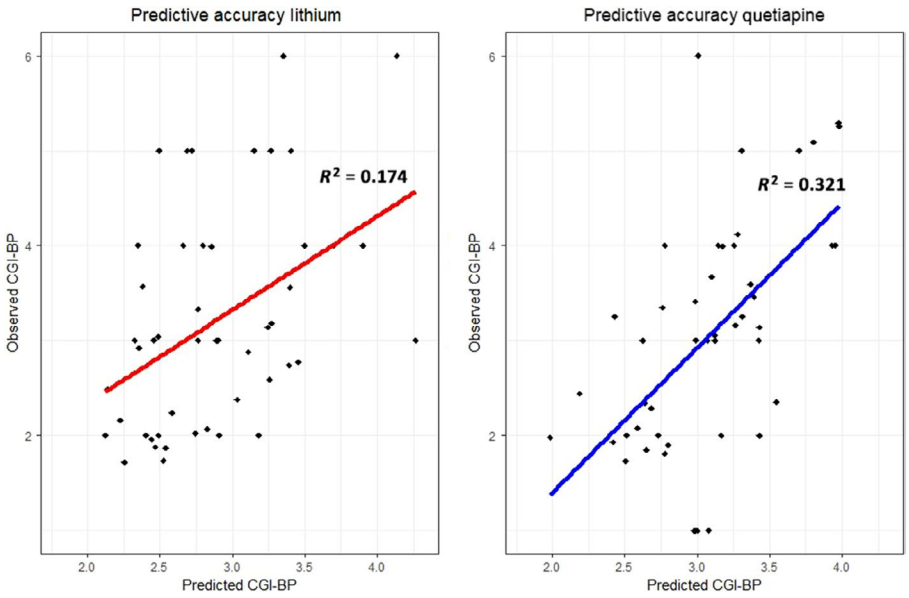

FIGURE 2 Predictive accuracy of lithium and quetiapine models [Colour figure can be viewed at wileyonlinelibrary.com]

Furthermore, when comparing the predictive accuracy of models generated by ENR against a basic symptom-severity or theoretical model, the ENR models demonstrated greater predictive accuracies for both patients that received lithium and quetiapine. One reason ENR outperformed both the basic symptom-severity and theoretical model is by ENR's protections against overfitting to a sample.[16] Similarly, we suspect the reason the theoretical model performed poorly compared to the basic symptom-severity model is in part due to its complexity; the theoretical model overfitted the training sample, which decreased its ability to generalize to the held-out test sample. Interestingly, the quetiapine theoretical model outperformed the lithium theoretical model even though the theoretical models were informed by lithium predictor research. This suggests that previous research may have identified common predictors of response to both lithium and quetiapine treatment of BD.

It is also worth noting that the basic symptom-severity models outperformed the theoretical models, despite the theoretical models including more predictors. This illustrates one of the shortcomings of aggregating variables identified from multiple different studies on individual predictors of response into a single model. Specifically, the estimated impact of variables (coefficients) may change when modeled in aggregate; furthermore, the lack of penalization of coefficients means that although a more complex model may produce a better fit within the training sample, the model could overfit and suffer in predictive accuracy on an external test sample.[14] Thus, even though the models generated using ENR included a greater number of variables than the theoretical models, they still maintained external validity due to the ability of ENR to shrink coefficients to minimize overfitting.

Since every model was generated using only the training data, and then prospectively evaluated on a held-out test sample, the estimates of predictive accuracy obtained reflect the model's ability to generalize to a completely novel sample from the same population.[25] Consequently, future efforts to obtain accurate estimates of patient response to lithium or quetiapine should consider generating predictions of patient response using a machine-learning method rather than relying on basic symptom-severity models or even more complex theoretical models.

The top five variables selected in the lithium model were: NSSI, ADHD, BISS mania, social phobia/social anxiety disorder, and suicide risk.

The top five variables selected in the quetiapine model were employment, agoraphobia, BISS irritability, substance dependence marijuana, and BISS depression.

Predictors with smaller parameter estimates should be taken with caution; nevertheless, the fact they were retained suggests that these other predictors may reliably improve prediction accuracy when used in aggregate. Some of the variables selected by our data-driven approach concurred with past findings. For example, ADHD comorbidity has been associated with early age-of-onset, higher number of depressive episodes, worse outcomes, and poorer response to treatment in patients diagnosed with BD.[38-41]

Literature on anxiety comorbidity with BD shows a trend toward lithium nonresponsiveness.[42] Additionally, a recent analysis in the Bipolar CHOICE dataset found that benzodiazepine users experienced significantly less improvement in CGI-BP than benzodiazepine nonusers[43]; while this study did not examine comorbid anxiety disorder as a predictor of outcome in BD patients, a majority of benzodiazepine users in this analysis did have a comorbid anxiety disorder diagnosis. SAD, agoraphobia, and panic disorder comorbidity with BD is associated with poor symptomatic outcome and reduced overall function. Our analysis also suggested that BD patients with SAD or agoraphobia had poorer prognoses than those without these comorbidities in both the quetiapine and lithium conditions.

Past studies have found that severity of mania predicts poorer outcome for BD patients.[44,45] The current study supports these findings, as our machine-learning methods revealed that severity of mania predicted poorer response in patients receiving lithium, and also revealed that severity of mania also predicted poorer response in quetiapine.

ENR selected NSSI as the top predictor in the lithium model and did not select NSSI in the quetiapine model. Patients who received lithium with NSSI were predicted to have worse outcomes. Literature has not examined NSSI as a predictor of outcome in lithium and quetiapine; however, Hayes et al[46] did examine NSSI as the primary outcome amongst patients during treatment of lithium and quetiapine. They found that patients taking lithium had reduced self-harm and unintentional injuries compared to patients prescribed quetiapine. Future studies should examine NSSI as a predictor of outcome amongst BD patients either receiving lithium or quetiapine.

BISS irritability was selected as an important predictor in the quetiapine model, and not for the lithium model. This finding does align with a systematic review that did not find a significant association between irritability and lithium response.[9] Another study found that patients with irritability had better outcomes in risperidone, an antipsychotic, compared to lithium.[47] While our analysis was prognostic, irritability was not highly selected in the lithium model compared to the quetiapine model. More studies should examine whether irritability is an important predictor in BD patients receiving lithium or quetiapine, and whether there are differential effects.

Literature has found that pretreatment depression is a robust predictor of lithium nonresponse[48]; however, literature has not examined depression as a predictor of quetiapine response. Similarly, not much research has explored marijuana dependence as a predictor of outcome amongst BD patients, even though substance abuse may be more common for BD patients.[49] Quetiapine has been associated with improvement in psychiatric symptoms and cocaine cravings in BD patients with cocaine dependence[50]; however, it would be interesting to examine whether marijuana dependency predicts response for patients receiving quetiapine.

The available biological predictors in this dataset could not distinguish lithium and quetiapine, making it difficult to build models that can be used to differentiate patients who should be offered one treatment in preference to the other. Future model building and model testing efforts in precision mental health would do well to use this approach with a dataset that comes from a comparison of mechanistically distinct treatments. In that context, after separate ENR models are generated, predicted outcomes for each patient in each treatment can be compared, in the manner of a personalized

advantage index (PAI).[51] Furthermore, following the recommendations of Kessler,[52] a possible next step would be the implementation of the method in large observational samples.

One of the advantages of using ENR to identify predictors of response to treatment in a secondary analysis of a pre-existing dataset is that ENR can narrow down the pool of predictors of treatment response and identify the most promising predictors. Although 100 predictors were entered into the model procedure—which ENR can adequately handle[18]—this does not necessarily mean that future efforts in identifying predictors of response to lithium or quetiapine should include 100 predictors. Notably, many of the biological predictors entered into the modeling procedure were not retained in the final models, suggesting that future research efforts into predictors of response to lithium or quetiapine for BD may be able to adequately predict treatment response without the arduous task of collecting these data for each patient. If anything, using machine-learning methods to identify putative predictors of response is a far more cost-effective approach than trying to individually identify predictors of treatment response, especially considering that the models constructed through machine-learning outperformed the theoretical model constructed through research on individual predictors of treatment response.

4.1 | Limitations

We did not have access to a prospective validation sample to further test the model's validity. While a completely held-out test sample is statistically equivalent to a prospective sample from the same population,[27] separate validation from a truly prospective sample can further examine the model's validity across different treatment settings for BD. [It is important to note that the predictions from this model are in the context of lithium or quetiapine therapy in combination with adjunctive personalized treatment; which, while may reflect real-life clinical practice,[6] may not necessarily align with past research using only lithium or quetiapine monotherapy.]

This is the first effort to systematically examine predictors of response to lithium or quetiapine in the treatment of BD. Previous studies have looked at predictors individually, but the unbiased, data-driven approach outlined in this work could inform future directions for research and clinical practice. For instance, accurate prediction of patient outcome can be used to inform treatment planning decisions, where modeling the likelihood of patient recovery is a critical outcome of interest to health care systems, providers, and stakeholders.[53] Future development of these tools with larger training samples can improve prediction of patient outcome, even to the point where differential predictions of outcome for the personalization of treatment of BD may be possible.[54] Hopefully, this research could provide efficient treatment allocation; with these models, clinicians could collect data on the most informative predictors (not all predictors would have to be used but rather the ones with the largest parameter estimates). Clinicians could then input this data to receive their patient's expected response to different medications (in this case, lithium and quetiapine) to provide the most effective treatment to their patients.

In conclusion, this work, at the minimum, serves as a demonstration that data-driven machine-learning approaches can handle prediction of treatment outcomes in BD better than theoretically driven single variable research. The further development of the tools and methods exemplified in this work should provide information that, ultimately, can be used by clinicians to make better-informed treatment decisions for their patients.[54]

ORCID


Thomas T. Kim https://orcid.org/0000-0003-2412-6647

REFERENCES

1. Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. *The Lancet*. 2013;381(9878):1672-1682.
2. Baca-Garcia E, Gronzalez de Rivera JL, Saiz-Ruiz J, et al. Diagnostic stability and evolution of bipolar disorder in clinical practice: a prospective cohort study. *Acta Psychiatr Scand*. 2007;115:473-480.
3. Treuer T, Tohen M. Predicting the course and outcome of bipolar disorder: a review. *Eur Psychiatry*. 2010;25(6):328-333.
4. Connolly KR, Thase ME. The clinical management of bipolar disorder: a review of evidence-based guidelines. *Prim Care Companion CNS Disord*. 2011;13(4).
5. Hayes JF, Marston L, Walters K, Geddes JR, King M, Osborn DP. Lithium vs. valproate vs. olanzapine vs. quetiapine as maintenance monotherapy for bipolar disorder: a population-based UK cohort study using electronic health records. *World Psychiatry*. 2016;15(1):53-58.
6. Nierenberg AA, McElroy SL, Friedman ES, et al. Bipolar CHOICE (Clinical Health Outcomes Initiative in Comparative Effectiveness): a pragmatic 6-month trial of lithium versus quetiapine for bipolar disorder. *J Clin Psychiatry*. 2016;77(1):90-99.
7. Bowden CL, Grunze H, Mullen J, et al. A randomized, double-blind, placebo-controlled efficacy and safety study of quetiapine or lithium as monotherapy for mania in bipolar disorder. *J Clin Psychiatry*. 2005;66(1):111-121.
8. Young AH, McElroy SL, Bauer M, et al. A double-blind, placebo-controlled study of quetiapine and lithium monotherapy in adults in the acute phase of bipolar depression (EMBOLDEN I). *J Clin Psychiatry*. 2010;71(2):150-162.
9. Kleindienst N, Engel R, Greil W. Which clinical factors predict response to prophylactic lithium? A systematic review for bipolar disorders. *Bipolar Disord*. 2005;7(5):404-417.
10. Grof P. Sixty years of lithium responders. *Neuropsychobiology*. 2010;62(1):8-16.
11. Backlund L, Ehnvall A, Hetta J, Isacsson G, Ågren H. Identifying predictors for good lithium response–a retrospective analysis of 100 patients with bipolar disorder using a life-charting method. *Eur Psychiatry*. 2009;24(3):171-177.
12. Tighe SK, Mahon PB, Potash JB. Predictors of lithium response in bipolar disorder. *Ther Adv Chronic Dis*. 2011;2(3):209-226.
13. Hou L, Heilbronner U, Degenhardt F, et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. *Lancet*. 2016;387(10023):1085-1093.
14. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry*. 2016;3(3):243-250.
15. Ludlow L, Klein K. Suppressor variables: the difference between 'is' versus 'acting as'. *J Stat Educ*. 2014;22(2): 1-28.
16. Vieta E, Parramon G, Padrell E, et al. Quetiapine in the treatment of rapid cycling bipolar disorder. *Bipolar Disord*. 2002;4(5):335-340.

17. Cohen ZD, Kim TT, Van HL, Dekker J, Driessen E. A demonstration of a multi-method variable selection approach for treatment selection: Recommending cognitive–behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychother Res*. 2019;1-14. https://doi.org/10.1080/10503307.2018.1563312.
18. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(2):301-320.
19. Webb CA, Trivedi MH, Cohen ZD, et al. Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. *Psychol Med*. 2018;1-10.
20. Suppes T, Dennehy EB, Hirschfeld RM, et al. The Texas implementation of medication algorithms: update to the algorithms for treatment of bipolar I disorder. *J Clin Psychiatry*. 2005;66(7):870-886.
21. Nierenberg AA, Sylvia LG, Leon AC, et al. Clinical and Health Outcomes Initiative in Comparative Effectiveness for Bipolar Disorder (Bipolar CHOICE): a pragmatic trial of complex treatment for a complex disorder. *Clin Trials*. 2014;11(1):114-127.
22. Spearing MK, Post RM, Leverich GS, Brandt D, Nolen W. Modification of the Clinical Global Impressions (CGI) Scale for use in bipolar illness (BP): the CGI-BP. *Psychiatry Res*. 1997;73(3):159-171.
23. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. *Psychopharmacol Bull*. 1993;29(2):321-326.
24. Trivedi MH, Wisniewski SR, Morris DW, et al. Concise Health Risk Tracking scale: a brief self-report and clinician rating of suicidal risk. *J Clin Psychiatry*. 2011;72(6):757.
25. Leon AC, Solomon DA, Mueller TI, et al. A brief assessment of psychosocial functioning of subjects with bipolar I disorder: the LIFE-RIFT. *J Nerv Ment Dis*. 2000;188(12):805-812.
26. Gonzalez JM, Bowden CL, Katz MM, et al. Development of the Bipolar Inventory of Symptoms Scale: concurrent validity, discriminant validity and retest reliability. *International journal of methods in psychiatric research*. 2008;17(4):198-209.
27. Kuhn M, Johnson K. *Applied predictive modeling*, Vol. 26. New York, NY:Springer; 2013.
28. Stekhoven DJ, Buhlmann P. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118.
29. Kuhn M, Johnson K. *Applied predictive modeling*. New York, NY:Springer Science & Business Media; 2013.
30. Williams CK, Engelhardt A, Cooper T, et al. Package 'caret'. 2018. https://cran.r-project.org/web/packages/caret/caret.pdf. Accessed August 20, 2018.
31. Kraemer HC, Blasey CM. Centring in regression analyses: a strategy to prevent errors in statistical inference. *Int J Methods Psychiatr Res*. 2004;13(3):141-151.
32. Friedman J, Hastie T, TibshiraniR. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.
33. TibshiraniRJ. Regression shrinkage and selection via the. *J R Stat Soc Series B Stat Methodol*. 1996;267-288.
34. Piñeiro G, Perelman S, Guerschman JP, Paruelo JM. How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecol Modell*. 2008;216(3–4):316-322.
35. Smith EP, RoseKA. Model goodness-of-fit analysis using regression and related techniques. *Ecol Modell*. 1995;77(1):49-64.
36. Cohen J. Statistical power analysis for the behavioral sciences (2nd ed). In Hillsdale NJ. L. Erlbaum Associates; 1988.
37. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155-159.
38. Karaahmet E, Konuk N, Dalkilic A, et al. The comorbidity of adult attention-deficit/hyperactivity disorder in bipolar disorder patients. *Compr Psychiatry*. 2013;54(5):549-555.
39. Nierenberg AA, Miyahara S, Spencer T, et al. Clinical and diagnostic implications of lifetime attention-deficit/hyperactivity disorder comorbidity in adults with bipolar disorder: data from the first 1000 STEP-BD participants. *Biol Psychiat*. 2005;57(11):1467-1473.
40. Tamam L, Karakus G, Ozpoyraz N. Comorbidity of adult attention-deficit hyperactivity disorder and bipolar disorder: prevalence and clinical correlates. *Eur Arch Psychiatry Clin Neurosci*. 2008;258(7):385-393.
41. Perroud N, Cordera P, Zimmermann J, et al. Comorbidity between attention deficit hyperactivity disorder (ADHD) and bipolar disorder in a specialized mood disorders outpatient clinic. *J Affect Disord*. 2014;168:161-166.
42. Young LT, Cooke RG, Robb JC, Levitt AJ, Joffe RT. Anxious and non-anxious bipolar disorder. *J Affect Disord*. 1993;29:49-52.
43. Bobo WV, Reilly-Harrington NA, Ketter TA, et al. Effect of adjunctive benzodiazepines on clinical outcomes in lithium- or quetiapine-treated outpatients with bipolar I or II disorder: results from the Bipolar CHOICE trial. *J Affect Disord*. 2014;161:30-35.
44. Post RM, Uhde TW, Roy-Byrne PP, Joffe RT. Antidepressant effects of carbamazepine. *Am J Psychiatry*. 1986;143:29-34.
45. Jann MW. Diagnosis and treatment of bipolar disorders in adults: a review of the evidence on pharmacologic treatments. *Am Health Drug Benefits*. 2014;7(9):489.
46. Hayes JF, Pitman A, Marston L, et al. Self-harm, unintentional injury, and suicide in bipolar disorder during maintenance mood stabilizer treatment: A UK Population-Based Electronic Health Records Study. *JAMA Psychiatry*. 2016;73(6):630-637.
47. Vitiello B, Riddle MA, Yenokyan G, et al. Treatment moderators and predictors of outcome in the Treatment of Early Age Mania (TEAM) study. *J Am Acad Child Adolesc Psychiatry*. 2012;51(9):867-878.
48. Swann AC, Bowden CL, Morris D, et al. Depression during mania. Treatment response to lithium or divalproex. *Arch Gen Psychiatry*. 1997;54(1):37-42.
49. Strakowski SM, DelBello MP, Fleck DE, Arndt S. The impact of substance abuse on the course of bipolar disorder. *Biol Psychiatry*. 2000;48(6):477-485.
50. Brown ES, Nejtek VA, Perantie DC, Bobadilla L. Quetiapine in bipolar disorder and cocaine dependence. *Bipolar Disord*. 2002;4(6):406-411.
51. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS ONE*. 2014;9(1):e83875.
52. Kessler RC. The potential of predictive analytics to provide clinical decision support in depression treatment planning. *Curr Opin Psychiatry*. 2018;31(1):32-39.
53. Wrathall J, Belnap T. Reducing health care costs through patient targeting: risk adjustment modeling to predict patients remaining high cost. *EGEMS (Wash DC)*. 2017;5(2):4.
54. Cohen ZD, DeRubeis RJ. Treatment selection in depression. *Annu Rev Clin Psychol*. 2018;14.

SUPPORTING INFORMATION

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

How to cite this article: 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. https://doi.org/10.1111/bdi.12752