RESEARCH

Unlocking treatment success: predicting atypical antipsychotic continuation in youth with mania

Xiangying Yang[1], Wenbo Huang[2], Li Liu[1], Lei Li[1], Song Qing[1], Na Huang[2,3], Jun Zeng[2,3] and Kai Yang[3,4*]

Abstract

Purpose This study aimed to create and validate robust machine-learning-based prediction models for antipsychotic drug (risperidone) continuation in children and teenagers suffering from mania over one year and to discover potential variables for clinical treatment.

Method The study population was collected from the national claims database in China. A total of 4,532 patients aged 4–18 who began risperidone therapy for mania between September 2013 and October 2019 were identified. The data were randomly divided into two datasets: training (80%) and testing (20%). Five regularly used machine learning methods were employed, in addition to the SuperLearner (SL) algorithm, to develop prediction models for the continuation of atypical antipsychotic therapy. The area under the receiver operating characteristic curve (AUC) with a 95% confidence interval (CI) was utilized.

Results In terms of discrimination and robustness in predicting risperidone treatment continuation, the generalized linear model (GLM) performed the best (AUC: 0.823, 95% CI: 0.792–0.854, intercept near 0, slope close to 1.0). The SL model (AUC: 0.823, 95% CI: 0.791–0.853, intercept near 0, slope close to 1.0) also exhibited significant performance. Furthermore, the present findings emphasize the significance of several unique clinical and socioeconomic variables, such as the frequency of emergency room visits for nonmental health disorders.

Conclusions The GLM and SL models provided accurate predictions regarding risperidone treatment continuation in children and adolescents with episodes of mania and hypomania. Consequently, applying prediction models in atypical antipsychotic medicine may aid in evidence-based decision-making.

Keywords Mania, Machine learning, Prediction model, SuperLearner, Atypical antipsychotics

*Correspondence:
Kai Yang
a15828075272@163.com
[1]The First Branch, the First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[2]Department of Clinical Epidemiology and Health Economics, School of Public Health, The University of Tokyo, Tokyo, Japan
[3]The First Affiliated Hospital of Chengdu Medical College, Chengdu, Sichuan, China
[4]Key Laboratory of Geriatric Respiratory Diseases of Sichuan Higher Education Institutes, Chengdu, Sichuan, China

Key points

• The study underscored the significance of predicting continuation of mania treatment, particularly among children and adolescents, underscoring the necessity for further exploration in this domain.

• The findings emphasized the importance of various unique clinical and socioeconomic variables, including the frequency of emergency room visits for nonmental health disorders, as significant predictors of treatment continuation.

• The study have the potential to aid treatment decisions for mania in the young population, providing a valuable clinical reference for physicians in China. Accurate prediction models may assist in evidence-based decision-making for atypical antipsychotic treatment.

• The study successfully developed and validated robust machine-learning-based prediction models for the continuation of risperidone therapy in children and teenagers with mania. The models demonstrated good discrimination and robustness in predicting treatment continuation.

• While the current study yielded promising results, the authors acknowledge the limitations of a limited and geographically concentrated dataset. Future studies involving larger and more varied participant groups are recommended to further enhance the understanding and application of prediction models in this field.

Significance

What is already known on this topic

Prior research has highlighted the need for effective prediction models for antipsychotic drug continuation in pediatric patients with mental health problems, given the complexity of treatment decisions and the lack of comprehensive models in this area.

What this study adds

This study introduces robust machine-learning-based prediction models for antipsychotic drug continuation in children and adolescents with mania. It demonstrates the accuracy of the Generalized Linear Model (GLM) and SuperLearner (SL) algorithm in predicting treatment outcomes. The study also identifies novel clinical and socioeconomic variables.

How this study might affect research, practice or policy

The development of accurate prediction models has the potential to significantly impact clinical decision-making in the treatment of pediatric patients with mania. These models could guide healthcare professionals in identifying those most likely to benefit from antipsychotic therapy continuation.

Background

Bipolar disorder, is a common long-term mental health disorder with a global prevalence of approximately 4% [1, 2]. Individuals suffering from mania or manic episodes experience frequent and irregular hyperactivity and delusions throughout their lives, which can lead to social and occupational issues. The patient may be aggressive and impulsive in social situations, lose self-control, or participate in dangerous behavior such as speeding, risky investments, and other detrimental activities [3]. Additionally, patients with symptoms of mania that developed during their early childhood exhibit more severe symptoms, worse premorbid adjustment, cognitive impairment, and extended periods of continuous rapid cycling compared to those who developed mania in adulthood [4]. As bipolar disorder is a chronic mood disorder, the severity of manic symptoms may continuously grow and persevere into adulthood, denying children the opportunity for appropriate emotional and cognitive growth. Consequently, individuals in manic/hypomanic episodes may suffer significant and long-term physical, social, and economic difficulties [1, 3, 5–7].

Pharmacotherapy, along with electroconvulsive therapy and psychosocial therapy, is one of the principal treatments of manic/hypomanic episodes [8, 9]. Most professional guidelines advocate atypical antipsychotics and mood stabilizers as first-line treatments for individuals of any age who suffer from mania or hypomania [10, 11]. For instance, second-generation antipsychotics (SGAs), lithium, and valproate, have been found to be more effective than placebos. However, due to inconsistencies in the pharmacological mechanisms of antipsychotics and the clinical variability of mania, deciding on the course of treatment is challenging. Furthermore, there is a scarcity of information concerning the selection of the most appropriate atypical antipsychotics for the continued treatment of patients during the maintenance phase, particularly in pediatric patients. Assessing treatment effects and responses becomes challenging due to treatment-emergent affective switch (TEAS) and subsyndromal mood fluctuations throughout remission [12]. Conventional predictive modeling methodologies such as logistic regression and Cox regression may be less successful at capturing higher-order interactions or nonlinear effects, which can be attributed to the abundance and intricacy of longitudinal individual-level data in clinical and claims data. Although most machine learning (ML) methods do not outperform traditional logistic regression in terms of discriminating and validation performance [13], there are no relevant studies in which treatment durability in manic/hypomanic episodes clinical pharmacotherapy is evaluated using ML. In the present study, a variety of conventional ML techniques were employed in addition to SuperLearner, which is an integrated ML technique that

aimed to forecast teenage bipolar disorder medication adherence. This forecast was achieved by merging multiple pre-established prediction algorithms into a unified optimal algorithm. The objective of the present study was to develop and validate robust clinical prediction models to predict the persistence of atypical psychotropic medicines (risperidone) in adolescents and children in manic/hypomanic episodes, as well as to explore prospective variables that may be beneficial for clinical treatment.

Methods

Clinical and administrative claims data were utilized from 4,552 patients aged 4–18 who began atypical psychiatric therapy (risperidone, 0.5 to 2 milligrams once a day) for manic episodes (mania/hypomania/depression/mixed episodes) between September 2013 and October 2019 in Chengdu Secondary People's Hospital, Chengdu, Sichuan Province, China. The data currently cover patients with mental health disorders from five southwestern provinces and one autonomous region of China, including reports on the status and treatment of approximately 500,000 patients, with a significant proportion of patients over the age of 35. These patients were mostly covered by the Urban and Rural Resident Basic Medical Insurance (URRBMI), which is required for Chinese urban residents. The database contains demographic factors such as age, gender, and health insurance information (e.g., frequency of outpatient visits for mental health disorders). The clinical data comprised information provided by medical professionals, such as diagnoses, prescription medications, therapeutic methods, and comorbidities. Psychiatric professionals conducted examinations to identify code groups in accordance with the 10th edition of the International Classification of Diseases (ICD-10) until an agreement was reached on which codes may be grouped as part of the same condition, the diagnoses were grouped based on similarities in symptoms, etiology, or clinical presentation by psychiatric professionals, this strategy was employed to streamline the analysis and facilitate comparisons between different conditions, ensuring a coherent and manageable framework for the study. In the psychiatric department, patients were first evaluated by observation and psychometric testing. Serum corticosteroid, thyroid, and cerebrospinal fluid examinations were also conducted, and using the diagnostic classification criteria for mental illness (ICD-10), the diagnosis was determined based on the symptoms and examination results. Regarding loss to follow-up in the clinical setting, the record would be discontinued if a patient did not return or reply to any contact. Only the dataset with no missing values was included in the present study, and the missing data were assessed via a k-nearest neighbor (KNN) algorithm. First, data on individuals who were administered risperidone and with mania, hypomania, or other manic episodes (ICD-10: Manic Episodes (F30): Including codes for hypomania (F30.0), mania without psychotic symptoms (F30.1), and mania with psychotic symptoms (F30.2). Bipolar Affective Disorder (F31): Including codes for bipolar disorder with current hypomanic episode (F31.0), bipolar disorder with current manic episode without psychotic symptoms (F31.1), and bipolar disorder with current manic episode with psychotic symptoms (F31.2)) between 2013 and 2019 were gathered. To identify the first episode of mania in new users of antipsychotics, every patient needed to have a retrospective period (refers to the recorded or monitored period) of a minimum of two years. Patients with a history of bipolar disorder diagnosis or any other antipsychotic/mood stabilizer prescription in the first two years were disregarded due to the absence of a suitable look-back time. The index date for each patient was the first date of antipsychotic medication treatment. Finally, those using risperidone as their first-line antipsychotic treatment were chosen because risperidone and olanzapine were the most frequently utilized drugs in the present complete dataset. Despite not being the first option to control manic or hypomanic episodes, risperidone is effective for acute episodes of mental illnesses [14–16]. This study was performed in accordance with the good research practices and Ethics and the protocol was approved by the research ethics committee of Chengdu Secondary People's Hospital (2022CYFYIRB-BA). All participants among health professionals, case managers, community health workers and patients were asked consent to participate, having signed a written declaration of informed consent. Informed consent was also obtained from all subjects and/or their legal guardian(s).

Outcomes

Treatment persistence is a powerful determinant of antipsychotic medication success since it reflects pharmacological effectiveness, safety, and tolerability for patients and clinicians. Discontinuation is widely perceived as a concern during remission from manic/hypomanic episodes. As a result, treatment persistence over one year was an outcome of interest for the present study [2]. The continuation of therapy was characterized as continuing treatment with a primary antipsychotic medication for a duration of at least one year without stopping more than 60 days of prescriptions or experiencing any event of discontinuation from treatment for any cause. Any of the following occurrences during follow-up constituted all-cause treatment discontinuation: (a) Immediate adverse effects; (b) Long-standing/extended adverse effects; (c) Patient preference (typically upon symptom relief); (d) Guided by healthcare professionals (e.g., regimen simplification, therapeutic education and support); (e) Unsatisfactory response; (f) Development of new physical

health conditions or interactions with other medications: (i) Chronic or progressive diseases (e.g., heart or kidney conditions); (ii) Transient/self-limiting conditions (e.g., pregnancy, breastfeeding, drug overdose/self-poisoning). Patients who satisfied all the treatment criteria were assigned a 1 (yes), while others were assigned a 0 (no).

Predictors

With clinical and administrative claims data collected from the psychiatry department and insurance company, potential risk indicators were identified based on several feature selection methods. Twenty-eight predictors were also included: patient demographic information (gender and age) and clinical characteristics, including number of inpatient visits for mental health disorders, number of route visits for mental health disorders, number of emergency room visits for mental health disorders, number of outpatient visits for mental health disorders, number of routine visits for nonmental health disorders, number of inpatient visits for nonmental health disorders, number of outpatient visits for nonmental health disorders, number of emergency room visits for nonmental health disorders, depression, anxiety disorder, ADHD (attention deficit hyperactivity disorder), intellectual disability, personality disorders, substance use disorder, somatic symptom disorder, other mental health disorders, neurological and neuromuscular problems, cardiovascular disease, renal and urologic disorders, gastrointestinal disease, metabolic disorders, asthma, diabetes mellitus, anticholinergic drugs, SSRI/SNRI (selective serotonin reuptake inhibitors and serotonin-norepinephrine reuptake inhibitors), and MAO (monoamine oxidase) inhibitors. Such predictors were measured during the year before the index date. Moreover, each variable was formulated as a binary or continuous predictor. Continuous predictors were normalized by utilizing min-max ratios, and the significance of these predictors on the outcomes was validated using two random forest-based feature selection methods (recursive feature elimination (RFE) [17] and Boruta [18]) in addition to the traditional regularization technique LASSO (least absolute shrinkage and selection operator) regression [19]. Ultimately, the most significant predictors were determined.

Statistical analysis

The total dataset was randomly split into training (80%) and testing (20%) datasets. Subsequently, seven distinct ML prediction models were formulated for risperidone therapy continuation over one year for patients aged 4–18. The seven distinct prediction algorithms were (1) generalized linear model (GLM); (2) stochastic gradient boosting machine (GBM); (3) generalized additive model (GAM); (4) random forest (RF) algorithm; (5) support vector machine (SVM); and (6) SuperLearner (SL). GLM

serves as a statistical model extending linear regression, GBM operates as an ensemble machine learning technique constructing a sequence of weak predictive models, subsequently amalgamating them to form a more robust predictive model. GAM, an extension of GLM, offers increased flexibility for modeling intricate relationships. RF, an ensemble learning approach, creates numerous decision trees during training, yielding the class mode for classification problems or mean prediction for regression issues. SVM, a supervised machine learning algorithm, is applied for classification or regression tasks. Finally, SuperLearner functions as a meta-algorithm that amalgamates diverse models to enhance predictive performance.

First, a seed and "tune Length" were selected for complexity and hyperparameter tuning using the "caret" package. Furthermore, although GLM is a machine learning approach, it was employed as a traditional prediction model in the present study. GLM was used in this manner because, unlike other methods, a complicated procedure for adjusting hyperparameters is not needed. The SL algorithm was also used to forecast treatment continuance over one year. The SL algorithm is an integrated machine learning technique that uses cross-validation to identify the optimum weighted combination of several candidate learners [20]. Additionally, the selection of an algorithm in advance is not necessary. The SL algorithm is theoretically capable of selecting the optimal set of algorithms from a variety of alternatives and combining the results of the relevant options. Potential algorithms can fall into either parametric or nonparametric categories, and the algorithms all perform k-fold cross-validation on the datasets. Tenfold cross-validation was applied to the dataset to divide it into k=10 almost equal-sized sets of mutually exclusive and exhaustive enumerations, this method was utilized to adjust the hyperparameters if applicable [21]. For each k-fold, one of the k-sets served as the validation set, while the remaining sets served as the training set. Each technique was applied to the training set to create an estimator, while its performance (known as risk or squared error) was evaluated using the validation set. The performance of the validation dataset had to be evaluated, as overly flexible algorithms strive to exploit random fluctuations in training data to improve accuracy. The process was repeated until each set had functioned as both a training and validation dataset, and predictions for all observations had been obtained. Subsequently, predictors and algorithms that were insignificant were discarded. To accommodate varying data complexity levels, the statistical properties of the algorithms had to be diverse. Moreover, bootstrap resampling was utilized with 25 repetitions to ensure that the results were generalized [22]. The area under the receiver operating characteristic curve (AUC) was the fundamental

metric for assessing model prediction performance. As a result, this approach was employed to select and preserve the best-performing final model. DeLong's test was employed to compare the AUCs between two models, with the results being deemed statistically significant if the p value was less than 0.05. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1-score, recall, and precision were all

Table 1 Baseline demographic parameters and indicators included in the training and testing datasets

Variables	Training (3,621)	Testing (911)
Age (year)	15.2 ± 4.7[a]	14.8 ± 4.5
Male	2,021 (55.5%)[b]	520 (57.1%)
Anxiety disorder	899 (24.7%)	227 (24.9%)
Depression	1,058 (29.8%)	285 (31.3%)
Intellectual disability	353 (9.7%)	81 (8.9%)
Personality disorders	262 (7.2%)	70 (7.7%)
Somatic Symptom disorder	171 (4.7%)	31 (3.4%)
Substance use disorder	455 (12.5%)	110 (12.1%)
Other mental health disorders	870 (23.9%)	204 (22.5%)
SSRI/SNRIs	13,187 (21.2%)	3,325 (22.3%)
Anticholinergic drugs	16,333 (1.2%)	4,147 (0.9%)
MAO inhibitors	10,646 (5.5%)	2,669 (4.9%)
Neurological and neuromuscular diseases	138 [3.8%]	40 [4.4%]
Cardiovascular diseases	62 (1.7%)	17 (1.9%)
Renal and urological disorders	51 [1.4%]	11 [1.2%]
Gastrointestinal diseases	36 [1.0%]	12 [1.3%]
Metabolic disorders	145 [3.0%]	30 [3.3%]
Asthma	477 [13.1%]	120 [13.2%]
Diabetes mellitus	98 [2.7%]	26 [2.9%]
Number of any route visits for mental health disorders	2.32 ± 2.41	2.70 ± 2.45
Number of inpatient visits for mental health disorders	0.50 ± 2.43	0.73 ± 3.19
Number of outpatient visits for mental health disorders	3.44 ± 3.62	3.70 ± 3.49
Number of emergency room visits for mental health disorders	1.21 ± 2.81	1.40 ± 2.33
Number of any route visits for nonmental health disorders	17.52 ± 15.26	17.71 ± 13.93
Number of inpatient visits for nonmental health disorders	1.34 ± 2.41	1.30 ± 3.02
Number of outpatient visits for nonmental health disorders	18.70 ± 16.73	16.97 ± 17.60
Number of emergency room visits for nonmental health disorders	1.91 ± 1.76	1.87 ± 1.49
Continuous treatment	543 (14.9%)	129 (14.2%)

[a] Values represent mean (standard deviation)

[b] Values represent n (%)

Abbreviations: MAO, monoamine oxidase inhibitor; ADHD, attention deficit hyperactivity disorder; SSRI, selective serotonin reuptake inhibitor; SNRI, selective norepinephrine reuptake inhibitor

utilized to evaluate prediction performance. Internal validation measures such as intercept and slope values, which should gradually approach 0.0 and 1.0, respectively, were also used to assess model generalizability in large, clustered datasets. Validated prediction models were constructed using data over two years (2013–2015), and findings from 2016 as longitudinal data may yield a higher predictive value. The model was then evaluated using the anticipated variables from 2013 to 2015 and the outcomes from 2016. R 4.0.5 was used to perform all analyses (R foundation, Vienna, Austria).

TRIPOD statement
The present study adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD) initiative for designing algorithms that can be tested using other datasets with wholly different circumstances [23].

Results
Table 1 exhibits the demographics of the participants and the percentage of individuals continuing 1-year risperidone treatment in children and adolescents. There were 3,621 patients in the training dataset (80% of total dataset) and 911 patients (20% of total dataset) in the testing dataset, with approximately 14.7% receiving risperidone medicine continuously over one year. The mean ages of patients in the training dataset and test dataset were 15.2 (SD 4.7) years and 14.8 (SD 4.5) years, respectively. Male patients outnumbered females (55.5% in the training dataset and 57.1% in the test dataset), and most patients (96.8%) were covered by national health insurance. Depression was the most common disorder among mental health problem comorbidities (training 29.8%, testing 31.3%). Relatively equal distributions for other pharmaceutical prescriptions and healthcare system consumption were observed from both datasets. The most identified discontinued events were Unsatisfactory Responses (138 in the training dataset, 32 in the testing dataset). Additionally, there was no significant difference between the training and test datasets in the distribution of baseline variables or outcome measures of interest. Figure 1 indicated the methodology flow gram employed for prediction model development.

Table 2; Fig. 2 denote the performance of the machine learning techniques, including the SL model, for predicting risperidone treatment continuance. The GLM and SL models outperformed the other machine learning algorithms regarding discrimination power and calibration capabilities. GLM was the best and most robust modeling approach for predicting risperidone treatment continuation, producing the highest AUC value (0.823, 95% CI: 0.792–0.854) and the lowest slope (1.004) and intercept (-1.248) values, which were closely followed by SL
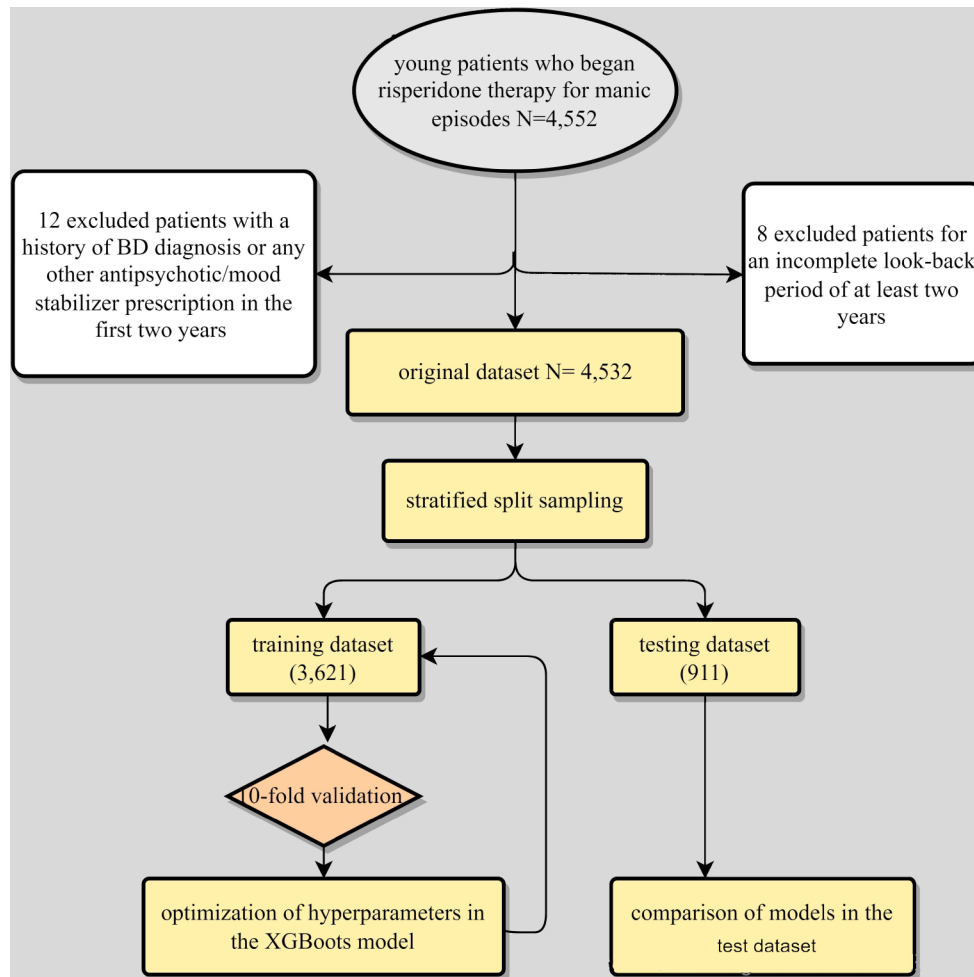
Fig. 1 The methodology flowgram employed for prediction model development and validation

Table 2 Performance measures for assessing the predictive accuracy of mania treatment continuation prediction models

Reference model[c]	AUC	*P* value[d]	Sensitivity	Specificity	PPV	NPV	F1 Score	Intercept	Slope	Recall	Precision
Generalized Linear Model (Reference)	0.82 (0.79–0.85)	NA	0.71	0.73	0.72	0.82	0.73	-1.24	1.00	0.73	0.75
Support Vector Machines with Linear Kernel	0.82 (0.79–0.85)	0.73	0.72	0.73	0.72	0.76	0.78	-1.23	1.14	0.77	0.76
Generalized Additive Model using LOESS	0.82 (0.78–0.86)	0.58	0.73	0.74	0.75	0.88	0.74	-1.25	1.12	0.73	0.75
Stochastic Gradient Boosting	0.81 (0.78–0.84)	0.49	0.70	0.76	0.73	0.89	0.74	-1.26	1.19	0.72	0.75
Random Forest	0.78 (0.75–0.81)	**0.02**	0.70	0.88	0.74	0.83	0.71	-0.29	1.78	0.74	0.75
Super Learner	0.82 (0.79–0.85)	0.84	0.74	0.72	0.74	0.89	0.74	-1.24	1.01	0.74	0.74

[c] As the reference model, we employed a generalized linear model

[d] Using the DeLong's test, we compared the area under the curve of each machine-learning-based prediction model with the reference model

P values less than 0.05 are bold
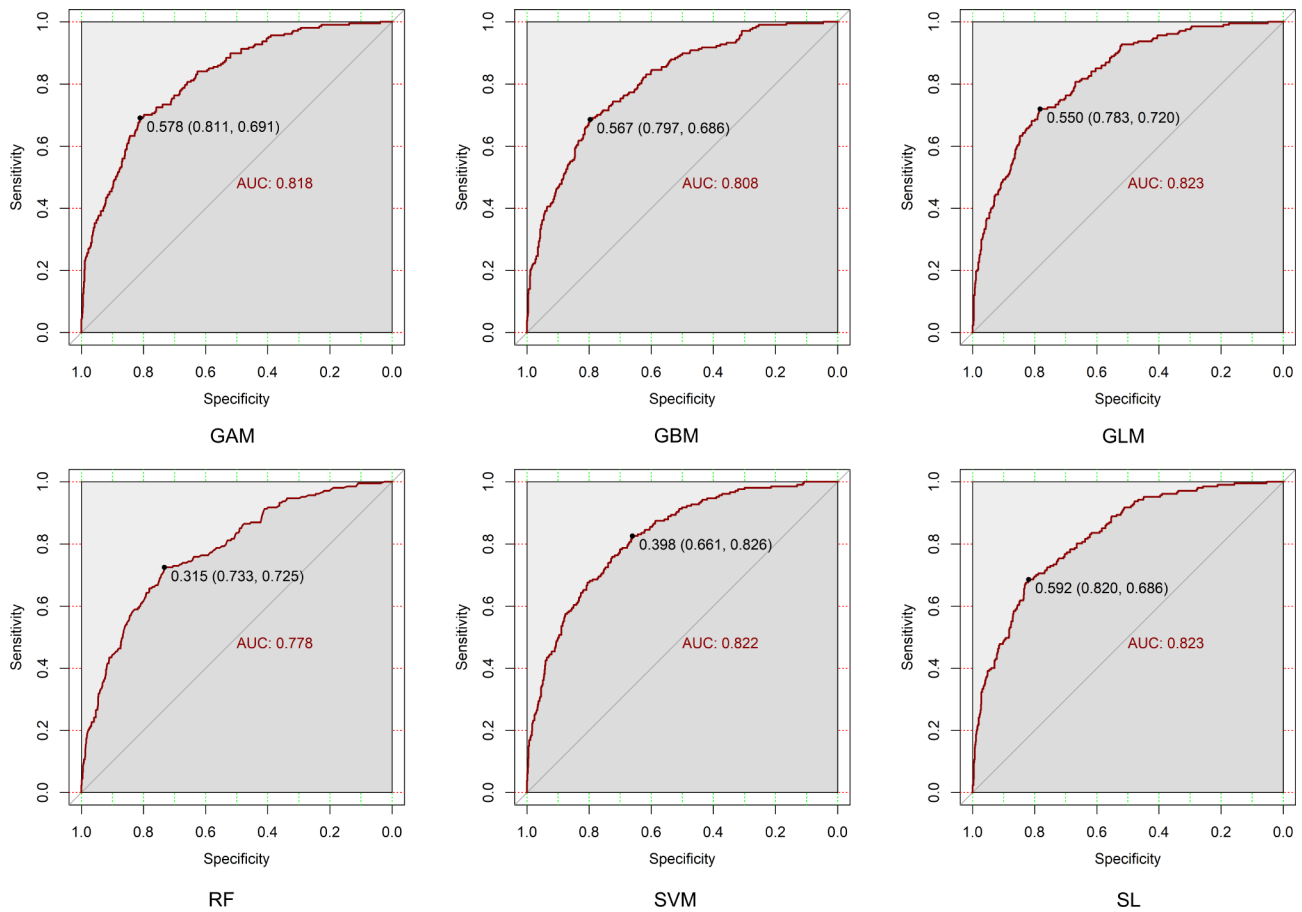
Fig. 2 The area under the receiver operating characteristic curve (AUC) plot of six prediction models. Abbreviations: GLM, generalized linear models. RF, random forests. GBM, stochastic gradient boosting machine. GAM, generalized additive models. SVM, support vector machine. SL, Superlearner

(AUC: 0.823, 95% CI: 0.791–0.853, slope: 1.016, intercept, -1.248). When compared with GLM and SL, SVM, GAM, GBM, and RF performed insignificantly in terms of AUC (SVM: 0.822, GAM: 0.818, GBM: 0.808, and RF: 0.778). Moreover, concerning the accuracy, sensitivity, specificity, PPV, NPV, precision, F1-score, intercept, and slope, such models performed worse regarding internal validation compared to GLM and SL. Although traditional algorithms (GLM) and ensemble algorithms (SL) outperformed machine learning approaches in terms of AUC, the differences were generally not statistically significant. Findings were also made that models applying consequent two-year data did not significantly outperform those using single-year data (Table 3).

Our supplemental figures respectively displayed the essential predictors in the RFE, LASSO, and Boruta models that contributed to the present outcomes (Supplemental material-001 Figs. 1, 2 and 3). All contributions of the predictors for each model can be observed in the supplemental material. Both the RFE and Boruta prediction models identified the count of emergency room visits for nonmental health disorders and the count of outpatient visits for nonmental health disorders as the

main predictors. Furthermore, the number of emergency room visits for nonmental health disorders was recognized as the most significant predictor in the LASSO regression prediction model due to a positive significance value of 40.58. Based on these findings, emergency room visits may serve as a significant prospective predictor of the continuation of manic therapy in children and adolescents.

Discussion

To the best of our knowledge, the present study represents the first instance in which a prediction model for the continuation of atypical antipsychotic treatment in children and adolescents experiencing manic/hypomanic episodes has been constructed and validated. The persistence of treatment is crucial for the successful management of manic/hypomanic episodes. Sustained treatment helps maintain stability in patients over the long term, reducing the risk of symptom recurrence. For individuals with mania, timely and continuous treatment effectively manages mood fluctuations, decreasing the frequency and severity of depressive and manic episodes. Furthermore, good treatment persistence improves patients'

Table 3 Performance measures for assessing the predictive accuracy of mania treatment continuation prediction models. Single-year data are from 2016, and two years of data are from 2013–2015

Model	AUC		Intercept		Slope	
	(single year)	(consecutive 2-year)	(single year)	(consecutive 2-year)	(single year)	(consecutive 2-year)
Generalized Linear Model (Reference)	0.82 (0.79–0.85)	0.82 (0.79–0.85)	-1.24	-1.24	1.00	1.01
Support Vector Machines with Linear Kernel	0.82 (0.79–0.85)	0.82 (0.79–0.85)	-1.23	-1.24	1.14	1.14
Generalized Additive Model using LOESS	0.82 (0·78–0.86)	0.81 (0·78–0.85)	-1.25	-1.25	1.12	1.12
Stochastic Gradient Boosting	0.81 (0·75–0·81)	0.81 (0·75–0·81)	-1.26	-1.26	1.19	1.17
Random Forest	0.78 (0·75–0·81)	0.77 (0·75–0·80)	-0.29	-0.28	1.78	1.76
Super Learner	0.82 (0·79–0·85)	0.82 (0·79–0·85)	-1.24	-1.24	1.01	1.00

quality of life, enhances their social functioning, and reduces reliance on emergency medical services. Therefore, emphasizing and promoting the continuity of treatment for mania is paramount for the overall health and well-being of patients [24, 25]. Prediction models offer valuable insights into the likelihood of treatment continuation for individuals with mania. By incorporating factors such as medication adherence patterns, comorbidities, and patient demographics, these models assist clinicians in making more informed decisions about treatment strategies. Understanding patients' adherence behaviors allows clinicians to tailor treatment plans to maximize therapeutic effectiveness and minimize the risk of discontinuation. The application of prediction models aims to improve health outcomes and enhance the quality of life for individuals with bipolar disorder by optimizing treatment adherence and continuity. These models contribute to better symptom management, reduced relapse rates, and improved overall well-being for patients. Through targeted interventions and personalized care plans, clinicians can support patients in achieving optimal treatment outcomes and maintaining long-term stability. The model can help clinicians tailor treatment plans to each patient's specific needs, thereby improving medication adherence and reducing the risk of relapse. For example, a high-risk score might prompt the clinician to schedule more frequent follow-up appointments or consider adjunct therapies. In addition, the prediction model can be integrated with existing electronic health record (EHR) systems to streamline its use, when a clinician reviews a patient's chart, the EHR system can provide real-time alerts if the model predicts a high risk of treatment discontinuation. This integration ensures that clinicians have immediate access to actionable insights without additional administrative burden.

While previous researchers have reported various predictive models for antipsychotic treatment continuation [26], their predictive power was relatively low. Additionally, there is a paucity of studies referring to bipolar disorder treatment persistence, especially in pediatric and young adult populations. However, the present type of study was essential given the diagnostic criteria of mixed-state mania and the identification of the manic symptoms of mild and severe mania in juveniles [27]. In addition, various modeling techniques, including traditional linear regression and modern machine learning as well as a novel SuperLearner, were evaluated in the present study. The predictors in this study incorporated readily available clinical and claim data rather than laboratory data, enhancing the generalizability and stability of the findings. Two ML-based prediction models (GLM and SL) utilizing both clinical and claims data demonstrated strong prognostic and predictive performance (AUC > 0.82) and validation power (intercept near 0, slope close to 1.0) for medication continuation in mania patients over one year. However, prediction models that utilized the 2-year data failed to significantly outperform models that used single-year data ($p > 0.05$). Furthermore, numerous key predictive factors were revealed, such as the number of emergency room visits for nonmental health disorders and the number of outpatient visits for nonmental health disorders. The clinical record data utilized in the present study may provide significant information regarding risk prediction that medical claims data do not (for example, comorbidities and medication history). The findings also highlight the significance of health insurance coverage. Consequently, compared with using models that rely on only one of the datasets, more information was obtained by using models that rely on clinical data from screening procedures as well as individual patient characteristics from medical claims data.

Although risperidone is not the first choice for the treatment of manic or hypomanic episodes, it is still acknowledged in China as yielding significant therapeutic efficacy and relatively low side effects. The present findings highlight the significance of combining clinical and claims data for creating ML prediction models to predict risperidone treatment persistence in young patients with mania and mixed episodes.

In the present study, a selection of standard machine learning algorithms failed to outperform the conventional GLM when predicting antipsychotic medication treatment persistence. However, the superior discriminating performance of the GLM was insufficient in revealing a clear advantage over other machine learning techniques. Previous studies have suggested that SuperLearner algorithms may enhance competence when compared with traditional linear regression and machine learning algorithms [28]. However, the effectiveness of the SuperLearner algorithm was comparable to that of the other algorithms utilized in the present study. The SuperLearner did not outperform the GLM algorithm in the training set, and GLM produced equivalent or higher AUCs in the validation set. Such results support earlier research findings, including those of Christodoulou et al. [13], who compared the performance of linear regression and other machine learning models for clinical predictive modeling algorithms. Christodoulou et al.'s findings revealed little indication that machine learning outperformed linear regression models in predicting clinical risk. By adjusting for linear and nonlinear interaction predictors between outcomes, machine learning enhances discriminating performance. Despite such findings, the techniques suffer the drawback of requiring sophisticated hyperparameter adjustments, involving vast data and expenses, and possessing interpretation complexity. In contrast, traditional regression approaches are easier to comprehend and require less time to process clinical data.

Since the selection of input variables, potential algorithms, and their specifications is advisable before executing an analysis, the present authors evaluated the risk factors that are frequently included in mania assessment tools, which may also be utilized for claims data. Furthermore, to gain a better understanding regarding critical risk variables, two random forest-based feature selection approaches and a classic LASSO regression method for predictor importance maps were assessed. However, care was taken when using random forest approaches to evaluate the outcomes of machine learning predictions due to the occurrence of random variations in numerous decision logs. Nonetheless, risk variables such as administered drugs, sex, age, and comorbidities, as well as socioeconomic factors such as health insurance-related data, appeared to yield a significant influence on

risk prediction. Such results indicate that model performance may depend primarily on a small number of necessary predictor variables. The variable with the greatest influence on the prediction models in the present study was the number of nonmental health visits to emergency rooms. According to a previous retrospective cohort analysis [29], adherence and persistence with atypical antipsychotic medication were both expected to be associated with higher drug costs and lower medical service costs. From the present findings, poor adherence or perseverance may result in diseases or occurrences that are more expensive than drugs, possibly leading to more visits to the nonmental health emergency department. A recent prospective observational study [30] in the United States also evidenced that a large proportion of emergency patients evaluated for nonpsychiatric complaints (for example, chest pain and abdominal pain) suffered comorbid mental health problems, including but not limited to depression and anxiety, indicating a significant association between mental health problems and emergency department visits for nonmental health-related issues.

There are several limitations to the present study. First, because mania was recognized in accordance with ICD-10 codes, the chance of misdiagnosis cannot be ruled out, particularly for depression. Vieta et al. [27] discovered that the diagnosis of depression in claims data for pediatric populations generally possessed excellent validity. However, the number of false positives present during diagnosis was more significant in the outpatient context. In addition, patients may cease or modify their antipsychotic regimen due to deteriorating symptoms or additional episodes of mental illness [31] rather than poor reactions to first-line antipsychotic medication. Several discontinuation cases suggested that deteriorating mental symptoms were included to identify unsatisfactory treatment results. Furthermore, including all the predictive characteristics that may increase the accuracy of prediction models was not feasible, such as the type of fine bipolar disorder, disease severity, and location of residence, which can be attributed to a lack of data. Additionally, the term "mixed mania" is frequently used to describe irritated patients rather than euphoric patients. In contrast, in the DSM-IV and subsequent papers describing adults with bipolar disorder, the word "mixed" is used to refer to people who are irritable but also gloomy or melancholic. Thus, researchers and doctors should be particularly cautious when making diagnoses. Next, a previous large network meta-analysis [32] showed that most patients with manic/hypomania/mixed episodes receiving a combination of lithium, anticonvulsants, and second-generation antipsychotics had the lowest relapse/recurrence rates and the lowest incidence of all-cause discontinuation. As a result, the hope of the

present authors is that the prediction of combined medication treatment can be further discussed in future studies. Finally, a lack of large samples, the number of "events per variable" used for prediction [33, 34], and external validation for populations with varying demographic features may have led to biases in the present study. Both physicians and researchers face significant hurdles in diagnosing mania and selecting therapy choices based on diverse action mechanisms. The diversity of the disease and the effect of developmental variables on its clinical presentation in the child-adolescent population further restrict the existing limits of diagnostic taxonomy, prompting researchers to propose the formulation of a more precise and comprehensive phenotypic system. Additionally, the present results may not be representative of the general population, as the populations in the dataset used were highly concentrated geographically. The lack of diversity or inaccuracy in the dataset reflected in the predictions when using machine learning to predict medication persistence may have caused ethical issues related to bias, trust, and impact on care to emerge. The present authors will explore the ethical concerns involved with the future deployment of such technology in young patients, considering the advantages and risks related to the technology and care delivery, as well as organizational and legal issues [35, 36]. Nevertheless, a clinical prediction model for medication therapy persistence based on machine learning was provided in the present study. The present authors also hope to contribute to the advancement of understanding and treating the long-term condition of bipolar disorder and to provide fresh ideas regarding the pharmaceutical treatment of mania from both a clinical and administrative standpoint, thereby improving our modeling capabilities and adaptation to a larger population in future studies.

Conclusions

In summary, compared to other algorithms used in predicting risperidone treatment continuation over one year in young patients with mania, the performance and analysis of the SL and traditional GLM algorithms demonstrated significant discriminatory power in the training dataset and adequate calibration power in the validation set. The number of emergency room visits for non-mental health disorders emerged as the most crucial predictor, warranting careful consideration in the continued treatment of youth mania. However, due to the relatively small population in the utilized dataset, imbalanced and biased results may have occurred, which is a factor that will be improved in future studies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-024-02622-z.

Supplementary Material 1

Author contributions
X.Y. and W.H. were in charge of the study design, paper preparation, and manuscript writing. W.H. participated in the data analysis and statistical modeling, had full access to the raw data, and edited the manuscript. L.L., L.L., S.Q., J.Z., and K.Y. created the concept, research design, and manuscript revision. K.Y. was in charge of the final revision after revising each stage of the job. All of the writers reviewed and approved the final version of the paper.

Data availability
The datasets used to train and evaluate the ML algorithms include personal information and are not publicly available. Researchers interested in obtaining the data for research reasons can contact Mr. Kai Yang (a15828075272@163.com).

Code availability
The statistical coding and machine learning methods used in this study will be made available to the author upon reasonable request.

Declarations

Ethics approval and consent to participate
This study was performed in accordance with the good research practices and Ethics and the protocol was approved by the research ethics committee of Chengdu Secondary People's Hospital (2022CYFYIRB-BA). All participants among health professionals, case managers, community health workers and patients were asked consent to participate, having signed a written declaration of informed consent. Informed consent was also obtained from all subjects and/or their legal guardian(s).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Grande I, Berk M, Birmaher B, Vieta E. Bipolar disorder. Lancet. Apr 9. 2016;387(10027):1561–1572. https://doi.org/10.1016/s0140-6736(15)00241-x
2. National Collaborating Centre for Mental H. National Institute for Health and Clinical Excellence: Guidance. Bipolar disorder: the management of bipolar disorder in adults, children and adolescents, in primary and Secondary Care. British psychological society copyright © 2006. The British Psychological Society & The Royal College of Psychiatrists.; 2006.
3. McIntyre RS, Berk M, Brietzke E, et al. Bipolar disorders. Lancet Dec. 2020;5(10265):1841–56. https://doi.org/10.1016/s0140-6736(20)31544-0.
4. Judd LL, Akiskal HS, Schettler PJ, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry Jun. 2002;59(6):530–7. https://doi.org/10.1001/archpsyc.59.6.530.
5. Runge C, Grunze H. [Annual costs of bipolar disorders in Germany]. Nervenarzt Sep. 2004;75(9):896–903. https://doi.org/10.1007/s00115-004-1691-x. Jährliche Krankheitskosten bipolarer Störungen in Deutschland.
6. Janssen Pharmaceutica L. The economic burden of bipolar disease. J Clin Psychiatry. 2000;61(13):38–41.

7. Angst J. Bipolar disorder–a seriously underestimated health burden. Eur Arch Psychiatry Clin Neurosci Apr. 2004;254(2):59–60. https://doi.org/10.1007/s00406-004-0502-5.

8. Baldessarini RJ, Tondo L, Vázquez GH. Pharmacological treatment of adult bipolar disorder. Mol Psychiatry Feb. 2019;24(2):198–217. https://doi.org/10.1038/s41380-018-0044-2.

9. Mukherjee S, Sackeim HA, Schnur DB. Electroconvulsive therapy of acute manic episodes: a review of 50 years' experience. Am J Psychiatry Feb. 1994;151(2):169–76. https://doi.org/10.1176/ajp.151.2.169.

10. 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 Mar. 2018;20(2):97–170. https://doi.org/10.1111/bdi.12609.

11. Fountoulakis KN, Vieta E, Young A, et al. The International College of Neuropsychopharmacology (CINP) treatment guidelines for bipolar disorder in adults (CINP-BD-2017), part 4: unmet needs in the treatment of bipolar disorder and recommendations for future research. Int J Neuropsychopharmacol. 2017;20(2):196–205.

12. Fountoulakis KN, Vieta E, Sanchez-Moreno J, Kaprinis SG, Goikolea JM, Kaprinis GS. Treatment guidelines for bipolar disorder: a critical review. J Affect Disord May. 2005;86(1):1–10. https://doi.org/10.1016/j.jad.2005.01.004.

13. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12–22.

14. Reeves H, Batra S, May RS, Zhang R, Dahl DC, Li X. Efficacy of risperidone augmentation to antidepressants in the management of suicidality in major depressive disorder: a randomized, double-blind, placebo-controlled pilot study. J Clin Psychiatry Aug. 2008;69(8):1228–36. https://doi.org/10.4088/jcp.v69n0805.

15. Crespo-Facorro B, Pérez-Iglesias R, Ramirez-Bonilla M, Martínez-García O, Llorca J, Luis Vázquez-Barquero J. A practical clinical trial comparing haloperidol, risperidone, and olanzapine for the acute treatment of first-episode nonaffective psychosis. J Clin Psychiatry Oct. 2006;67(10):1511–21. https://doi.org/10.4088/jcp.v67n1004.

16. Merlo MC, Hofer H, Gekle W, et al. Risperidone, 2 mg/day vs. 4 mg/day, in first-episode, acutely psychotic patients: treatment efficacy and effects on fine motor functioning. J Clin Psychiatry Oct. 2002;63(10):885–91. https://doi.org/10.4088/jcp.v63n1006.

17. Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet Sep. 2018;17(Suppl 1):65. https://doi.org/10.1186/s12863-018-0633-8.

18. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. Brief Bioinform Mar. 2019;22(2):492–503. https://doi.org/10.1093/bib/bbx124.

19. Fonti V, Belitser E. Feature selection using lasso. VU Amsterdam Res Paper Bus Analytics. 2017;30:1–25.

20. Van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol. 2007;6(1).

21. Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell. 2009;32(3):569–75.

22. Dixon PM. Bootstrap resampling. Encyclopedia of environmetrics. 2006.

23. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med Jan. 2015;6(1):W1–73. https://doi.org/10.7326/m14-0698.

24. Atagün M, Oral T. Acute and Long Term Treatment of Manic episodes in bipolar disorder. Noro Psikiyatr Ars. 2021;58(Suppl 1):S24–30. https://doi.org/10.29399/npa.27411.

25. Bonnín CDM, Reinares M, Martínez-Arán A, et al. Improving Functioning, Quality of Life, and well-being in patients with bipolar disorder. Int J Neuropsychopharmacol Aug. 2019;1(8):467–77. https://doi.org/10.1093/ijnp/pyz018.

26. Jeon SM, Cho J, Lee DY, Kwon JW. Comparison of prediction methods for treatment continuation of antipsychotics in children and adolescents with schizophrenia. Evid Based Ment Health Apr. 2022;13. https://doi.org/10.1136/ebmental-2021-300404.

27. Vieta E. Guide to assessment scales in bipolar disorder. Springer Science & Business Media; 2011.

28. Polley EC, Van Der Laan MJ. Super learner in prediction. 2010.

29. Jiang Y, Ni W. Estimating the impact of adherence to and persistence with atypical antipsychotic therapy on Health Care costs and risk of hospitalization. Pharmacotherapy Sep. 2015;35(9):813–22. https://doi.org/10.1002/phar.1634.

30. O'Reilly LM, Dalal AI, Maag S, et al. Computer adaptive testing to assess impairing behavioral health problems in emergency department patients with somatic complaints. J Am Coll Emerg Physicians Open Oct. 2022;3(5):e12804. https://doi.org/10.1002/emp2.12804.

31. Keks N, Schwartz D, Hope J. Stopping and switching antipsychotic drugs. Australian Prescriber. 2019;42(5):152.

32. Kishi T, Ikuta T, Matsuda Y, et al. Mood stabilizers and/or antipsychotics for bipolar disorder in the maintenance phase: a systematic review and network meta-analysis of randomized controlled trials. Mol Psychiatry Aug. 2021;26(8):4146–57. https://doi.org/10.1038/s41380-020-00946-6.

33. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373–9.

34. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48(12):1503–10.

35. Leibenluft E, Charney DS, Towbin KE, Bhangoo RK, Pine DS. Defining clinical phenotypes of juvenile mania. Am J Psychiatry. 2003;160(3):430–7.

36. González-Pinto A, Aldama A, Mosquera F, Gómez CG. Epidemiology, diagnosis and management of mixed mania. CNS Drugs. 2007;21(8):611–26.

Publisher's Note