EClinicalMedicine 39 (2021) 101083

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Research Paper

Development and validation of a risk calculator for major mood disorders among the offspring of bipolar parents using information collected in routine clinical practice

Charles D.G. Keown-Stoneman[a,b], Sarah M. Goodday[c,d], Martin Preisig[e], Caroline Vandeleur[e], Enrique Castelao[e], Paul Grof[f,g], Julie Horrocks[h], Nathan King[i], Anne Duffy[c,f,j,*]

[a] Applied Health Research Centre (AHRC), Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, ON, Canada
[b] Dalla Lana School of Public Health, University of Toronto, Ontario, Canada
[c] Department of Psychiatry, University of Oxford, Oxford, UK
[d] 4YouandMe, Seattle, USA
[e] Psychopathology Research Center, Department of Psychiatry, Lausanne University Hospital and University of Lausanne. Lausanne, Switzerland
[f] Mood Disorders Centre of Ottawa, Ottawa, Ontario, Canada
[g] Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada
[h] Department of Mathematics and Statistics, Guelph University, Ontario, Canada
[i] Department of Public Health Sciences, Queen's University, Ontario, Canada
[j] Department of Psychiatry, Queen's University, Ontario, Canada

ARTICLE INFO

Article History:
Received 28 May 2021
Revised 13 July 2021
Accepted 23 July 2021
Available online xxx

ABSTRACT

Background: Family history is a significant risk factor for bipolar disorders (BD), but the magnitude of risk varies considerably between individuals within and across families. Accurate risk estimation may increase motivation to reduce modifiable risk exposures and identify individuals appropriate for monitoring over the peak risk period. Our objective was to develop and independently replicate an individual risk calculator for bipolar spectrum disorders among the offspring of BD parents using data collected in routine clinical practice.
Methods: Data from the longitudinal Canadian High-Risk Offspring cohort study collected from 1996 to 2020 informed the development of a 5 and 10-year risk calculator using parametric time-to-event models with a cure fraction and a generalized gamma distribution. The calculator was then externally validated using data from the Lausanne–Geneva High-Risk Offspring cohort study collected from 1996 to 2020. A time-varying C-index by age in years was used to estimate the probability that the model correctly classified risk. Bias corrected estimates and 95% confidence limits were derived using a jackknife resampling approach.
Findings: The primary outcome was age of onset of a major mood disorder. The risk calculator was most accurate at classifying risk in mid to late adolescence in the Canadian cohort (n = 285), and a similar pattern was replicated in the Swiss cohort (n = 128). Specifically, the time-varying C-index indicated that there was approximately a 70% chance that the model would correctly predict which of two 15-year-olds would be more likely to develop the outcome in the future. External validation within a smaller Swiss cohort showed mixed results.
Interpretation: Findings suggest that this model may be a useful clinical tool in routine practice for improved individualized risk estimation of bipolar spectrum disorders among the adolescent offspring of a BD parent; however, risk estimation in younger high-risk offspring is less accurate, perhaps reflecting the evolving nature of psychopathology in early childhood. Based on external validation with a Swiss cohort, the risk calculator may not be as predictive in more heterogenous high-risk populations.
Funding: The Canadian High-Risk Study has been funded by consecutive operating grants from the Canadian Institutes for Health Research, currently CIHR PJT Grant 152796 he Lausanne-Geneva high-risk study was and is supported by five grants from the Swiss National Foundation (#3200–040,677, #32003B-105,969, #32003B-118,326, #3200–049,746 and #3200–061,974), three grants from the Swiss National Foundation for the National Centres of Competence in Research project "The Synaptic Bases of Mental Diseases" (#125,759, #158,776, and #51NF40 – 185,897), and a grant from GlaxoSmithKline Clinical Genetics.

* Corresponding author.
E-mail addresses: anne.duffy@queensu.ca, anne.duffy@psych.ox.ac.uk (A. Duffy).

https://doi.org/10.1016/j.eclinm.2021.101083

Research in context

Evidence before this study

Family history is arguably the most robust risk factor for bipolar disorder (BD), yet there are limited clinical tools available to accurately predict risk for a given individual. Accordingly, BD remains one of the most challenging disorders to predict and accurately identify. To our knowledge only one group (BIOS study) has developed a risk calculator using research data collected from offspring of BD parents based on measures not available to most clinicians and as yet not independently replicated on another high-risk offspring cohort.

Added value of this study

This study developed, and externally validated a novel individual-level risk calculator for major mood disorder in high-risk children of parents with confirmed BD, using variables collected in routine clinical practice. The approach taken has methodological advantages to the BIOS risk calculator in that it employs a parametric time-to-event model with a cure fraction and a flexible generalized gamma distribution. The addition of a cure fraction negates the erroneous assumption from traditional parametric time to event models that everyone will eventually develop the outcome.

Implications of all the available evidence

Having a clinical tool which incorporates readily available information that is collected in routine clinical practice to generate more precise 5 and 10-year risk predictions for individual adolescent offspring of BD parents is of critical importance. Precise individual risk prediction can be used to identify young people at familial risk most appropriate for prevention and early intervention, as well as self and clinical monitoring.

1. Introduction

There are several known risk factors associated with onset of bipolar disorder (BD) that should be reliably collected as part of a comprehensive clinical assessment including: family history [1], history of early maltreatment [2], antecedent disorders and sub-threshold symptoms [3], circadian disruption [4] and substance use [5]. However, there is limited information as to how to combine these risk factors to make an accurate risk prediction for an individual at confirmed familial risk. Given the substantial genetic and phenotypic complexity, including the variability in penetrance and emergent course of illness, tailored more precise risk prediction would be a major clinical advance, improving time to diagnosis and reducing the associated morbidity and excess mortality evident already early in the illness course [6,7]. As in cardiovascular medicine and oncology, individualized risk prediction in psychiatry could be used to identify those suitable for more intensive self and clinical monitoring, prevention, and low-intensity early interventions; for example, those targeting improving sleep quality, healthy coping strategies and positive lifestyle changes [8,9].

The children of BD parents are an informative and identifiable high-risk population. A Danish national registry-based study reported that the cumulative incidence of BD by age 52 years was negligible for individuals in which neither parent was admitted to psychiatric care (inpatient or outpatient) with a diagnosis of BD; however, this risk rose to 4.4% when one parent was admitted and 24.9% when both parents were admitted with BD [10]. The risk increased to 36% with the inclusion of major depressive disorder (MDD) as an offspring outcome [10]. Evidence from family and genetic studies support that major depression is part of the bipolar spectrum in those at confirmed familial risk [11,12]. Several published longitudinal studies describing the onset and early course of BD in the offspring of BD parents (for a review see [13]) have reported an elevated risk of both BD and MDD in high-risk children, consistent with earlier family studies [14,15]. Furthermore, the majority of cases of BD in prospectively studied high-risk offspring debuted with MDD years prior to the first hypomanic or manic episode [16,17].

Recently, Hafeman et al. [18] published a risk calculator (http://www.pediatricbipolar.pitt.edu) to predict the 5-year risk of BD spectrum disorder in 412 high-risk children between 6 and 17 years. A "baseline-resetting" Cox proportional hazards model treated each research visit separately and frailty random effects was used to accommodate non-independence of assessments for the same participant. Predictors included summary symptom scores from parent and child KSADS-PL interviews, and parent and/or offspring rating scales measuring mood lability, anxiety, and functioning. The model yielded clinically relevant discrimination (AUC 0.76) between high-risk offspring from this cohort who did versus did not develop a BD spectrum disorder within 5-years. While this model has not been tested on an independent high-risk offspring cohort, it was examined in predicting progression to BD I or II in a referred sample of youth with sub-threshold bipolar symptoms [19].

In the present study, we developed and externally validated an individual risk calculator that would be relevant for use in routine clinical practice settings. We used data from the Canadian Flourish Longitudinal High-Risk Offspring Study [16] and then independently tested the model using data from the Lausanne–Geneva Longitudinal High-Risk Offspring study [20]. Both longitudinal observational studies identified high-risk offspring through a parent with confirmed BD and repeatedly assessed the offspring over a period up to 20 (Flourish) and 13 (Lausanne-Geneva) years, respectively. We selected candidate risk factors that would be collected in routine practice by non-research clinicians. Our aim was to determine whether we could accurately estimate the likelihood of developing a BD spectrum disorder at 5 and 10-years into the future in children and adolescents at confirmed familial risk.

2. Methods

2.1. Design

The Canadian Flourish Longitudinal High-risk Offspring study is an open, dynamic prospective cohort study that started over two decades ago and described in detail elsewhere [3,21]. Briefly, high-risk offspring were identified from a parent recruited from mood disorders outpatient specialty clinics in Ottawa and Halifax and were invited to consent to participate, or assent with parental authorization if a minor. Parental diagnosis of BDI was based on SADS-L interview conducted by a research psychiatrist and confirmed on the basis of blind consensus review of all available research and clinical information. Subsequently, pedigrees were expanded using the same methods to include consenting first degree relatives of the original probands with a confirmed BD spectrum disorder (BD I, II, MDD, schizoaffective BD). High-risk offspring who were between the ages of 5–25 years at entry completed research assessments by a research psychiatrist on average 1.6 years using KSADS-PL/SADS-PL format interviews and validated self and clinician reported measures of symptoms and psychosocial risk factors [16]. Offspring DSM-IV diagnoses were based on blind consensus review and best estimate procedure using all available research and clinical information [16].

The Lausanne High-Risk Study is a prospective cohort study that used a 3-year interval panel design for both parents and offspring as described in detail elsewhere [20]. Offspring were invited to consent to participate, or assent with parental authorization if a minor.

Briefly, high-risk offspring were recruited among inpatients and outpatients treated for BD I or II, schizoaffective BD or MDD in the psychiatric departments of Lausanne and Geneva. Only offspring of patients with BD or schizoaffective BD were included in the present paper. The diagnostic assignment of probands relied on a best-estimate procedure [22] including the semi-structured interviews using the Diagnostic Interview for Genetic Studies (DIGS [23]), family history reports from first-degree relatives and medical records. High-risk offspring between the ages of 6 to 17.9 years at study entry completed semi-structured assessments using the KSADS-E interview and during 3-year follow-up intervals between 7 and 17.9 years of age. In offspring >18 years, the DIGS was used to establish diagnoses. As in the Canadian Flourish High-Risk Cohort, self and clinician-reported validated measures of symptoms and psychosocial factors were collected at baseline and during follow-up assessments [20].

The two studies have been approved by respective research ethics boards at each site: the Ottawa Independent Research Ethics Board (Pro00011514), the Queens University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (PSIY-561−17) and the Ethics Committee for Clinical Research of the Canton de Vaud. This manuscript adheres to the reporting guidelines of the TRIPOD Checklist for Prediction Model Development and Validation.

2.2. Study sample selection

For this analysis, offspring from the high-risk cohorts were included if they met the following criteria: i) complete data on all but one of the predictors of interest (specifically, those missing childhood physical/sexual abuse were not excluded) and ii) predictor assessments occurred prior to the outcome. The resulting analysis sample included 285 high-risk offspring of a total of 292 from the Canadian cohort and 128 high-risk offspring of a total of 155 from the Swiss cohort were used to assess predictions from the model. From the 155 subjects in the Swiss cohort, 7 were dropped for only having 1 assessment available, making it impossible to assess their predictions, and 20 were omitted because they were diagnosed with a major mood disorder before or at their first visit. Although relatively high in missingness in the Canadian cohort (36% missing), it was decided that childhood physical/sexual abuse was likely an important predictor of future mental health, and as such was included in the risk predictions. In order to include the variable in the models a “missing” level of the childhood physical/sexual abuse variable was introduced. While it is expected that predictions based on the “missing” category will include more noise and uncertainty, this should be properly captured within the model, even if the missingness is not at random.

2.3. Measures

Measures are described in detail elsewhere [16] and in the Supplemental Methods. Briefly, predictors were limited to candidate risk factors that would be routinely collected by clinicians in an office setting. These included the following: high-risk offspring age in years, sex assigned at birth, socio-economic status (SES) [24] of the offspring’s family at recruitment and offspring global assessment of functioning; parent response or non-response to lithium maintenance treatment based on parent self-report in the Swiss cohort and clinical research interview and use of the Alda scale [25] in the Canadian cohort; BD parent age of onset defined as the age (in years) of first major mood episode (MDD, hypomania, mania); reported childhood physical and sexual abuse occurring <10 years of age; antecedent clinically significant symptoms (following previously published clinically determined criteria [16] and described in Supplementary methods); and full DSM-IV threshold antecedent non-mood diagnoses and number of minor mood (depression NOS, mood NOS) episodes. With the exception of, sex, proband lithium response and abuse, all predictors were collected repeatedly at each research assessment. Where possible, time-varying predictors were included in the model that could change at each new assessment.

2.4. Outcome

The outcome was defined as the age of onset when offspring first met criteria for full DSM-IV diagnosis of a major mood disorder including MDD or BD: (BD1, BDII, BDNOS, Schizoaffective BD). As in prior analyses [3,16] we included MDD as part of the BD spectrum disorder outcomes given prior family studies showing that in the first degree relatives of BD patients, MDD is related to the BD diathesis in the majority of cases and that in over 85% of the cases of BD in the Canadian and Dutch high-risk studies, BD debuted as depression [13]. In an exploratory analysis, we also examined a narrower outcome definition of BD outcome (BDI, BDII, BDNOS), although as expected given the age of the cohort at last observation the number of events was low contributing to less precise estimates (please see Supplementary materials).

2.5. Analysis

All analyses were performed in R version 4.0.2 for Windows 10 64bit [26]. The models used to make predictions are parametric time-to-event models with a cure fraction and a generalized gamma distribution. All models were fit in R using the flexsurvcure package [27] which makes use of the flexsurv package [28] and the survival package [29,30] to fit parametric time-to-event models with a cure fraction. Birth was used as the origin for the time-scale in all of the time-to-event models. All predictors were either fixed at birth or were only included if they occurred prior to the age of onset of outcome. At the time of the analysis, there were no known competing risk events such as death in either cohort.

In traditional parametric time-to-event models, it is assumed that all subjects will eventually experience the event of interest, even if the event was not observed during the course of the study. While this assumption is reasonable for an inevitable event such as death, this assumption is unlikely to hold for events such as a diagnosis of BD. For that reason, a cure fraction can be introduced into the modeling process, which represents an unknown proportion of the population that will never experience the event of interest. For this risk calculator, we assessed the necessity of a cure-fraction in the model using AIC (supplementary Table 1), and visually (supplementary Figure 1), and based on those assessments, subsequently included a cure-fraction approach for modeling.

The generalized gamma distribution was chosen for the time-to-event outcomes for its flexibility, and because many commonly used distributions are special cases of the generalized gamma distribution, such as the Weibull and exponential distributions [31]. For the models used in this calculator, it was assumed that the population shares a generalized gamma distribution for the at-risk population (i.e., those not in the cure fraction) and the covariates are associated with the probability of being in the cure fraction through a logistic link function. When making predictions, as age increases, this model approaches a logistic regression model for the odds of getting the outcome of interest. To assess if the generalized gamma distribution is unnecessarily complex for this application, the generalized gamma models were compared to similar models assuming a Weibull distribution.

Two versions of the risk calculator were estimated, one with known lithium response of the affected parent, and another with unknown lithium response of the affected parent. The second model was estimated using the same methods and data as the first but omitted the information on lithium response of the parent. This was done since all participants in the Canadian high-risk study have information on the lithium response of the affected parent; however, this information would not necessarily be available in other clinical

settings (supplementary Figure 3). The exclusion of the added information on the lithium response of the affected parent should not bias the risk predictions, given the proportion of lithium responders in the sample is reflective of the target population of interest, but will likely introduce additional uncertainty, which will be reflected in larger confidence intervals around predictions.

We present a time-varying C-index by age in years as a measure of probability that the model correctly classifies two randomly chosen high-risk individuals. A C-index of 0.50 indicates random probability; 0.70 is good; and 0.80 is excellent. This C-index is based on the C-index developed for survival analysis by Frank Harrell [32], and further developed to include time-varying covariates by Antolini et al. [33]. It shows the probability of correctly classifying which of two random individuals has the higher risk of the outcome given they were under observation at a given age. Bias corrected estimates and 95% confidence limits were derived using a jackknife resampling approach.

To assess the external validity of the final risk prediction model, the risk predictions were replicated using similar data collected from the Swiss high-risk offspring cohort. Since the model was fit using data from a Canadian cohort, applying it to a Swiss cohort may help understand the generalizability of the final model. The time-varying C-index was then calculated using the data from the Swiss High-Risk cohort and the model parameters estimated using the Canadian High-Risk cohort. Because sleep disorders were not assessed in the Swiss cohort, population level estimates/survival plots assumed that sleep disorder incidence was the same as in the Canadian cohort. For C-index calculations in the Swiss cohort, since sleep disorders information was unavailable, the presence or absence of sleep disorders could not be used to compare estimated risk between subjects and therefore this information was left out of the predictions. In practice, this was achieved by setting the binary sleep disorders variable to 0 for everyone, which represents no sleep disorders.

2.6. Role of the funding source

The funding sources had no involvement in the study design, collection, analysis, or interpretation of the data, writing of the report, or the decision to submit the paper for publication.

3. Results

3.1. Description of the two cohorts

This analysis included data from the Canadian Flourish high-risk cohort and an independent replication in the Swiss high-risk cohort. The Canadian cohort included 285 high-risk offspring up to 47.8 years of age (min = 4.3, max = 47.8, median = 25.4). 107(37.5%) developed a DSM-IV diagnosis of a major mood disorder by last observation (MDD, BDI, BDII, BDNOS, Schizoaffective), while 38(13.3%) developed a DSM-IV diagnosis of BD (narrow definition: BDI BDII, BDNOS) by last observation. The Swiss high-risk cohort included 128 high-risk offspring up to an age of 38.4 years of age (min = 7.1, max = 38.4, median = 25.8) in which 65 (50.7%) developed a DSM-IV diagnosis of major mood disorder by last observation, while 19 (14.8%) developed a DSM-IV diagnosis of BD by last observation.

Risk calculator link: https://www.queensu.ca/u-flourish/mood-disorder-calculator

3.2. Model fit

Comparisons between time-to-event models fit with a generalized gamma, and Weibull distribution, as well as with cure fraction and those without (traditional) showed an improvement in the model fit when including a generalized gamma cure fraction (as measured by AIC, and visual inspection of predicted survival curves compared to non-parametric Kaplan-Meier plots) (Supplementary Figures 1, 2 and Table 1). Specifically, there was a clear improvement in model fit between the cure fraction models and the traditional parametric models. The risk calculator model parameter estimates can be found in Supplementary Tables 2 and 3. Variance inflation factors (VIF) were used to assess any potential multi-collinearity between predictor variables in all of the risk prediction models tested, with no evidence of multi-collinearity found (VIF$<$2 for all covariates).

3.3. Model performance

Fig. 1 shows the complete time-varying C-index estimates from birth to 30 years of age. As an example interpretation, the plot suggests that in the Canadian high-risk cohort given 2 randomly selected 15-year-old offspring, there is approximately a 71% (95% CI: 0.65, 0.77) chance that the model would correctly predict which of them is more likely to develop a major mood disorder in the future (if at all). Based on the C-index estimates, the model performed poorest in the Canadian cohort at the age of 9 (C-index$_{9yo}$ = 0.65; 95% CI: 0.58, 0.70), and best at age 29 (C-index$_{29yo}$ = 0.83; 95% CI: (0.66, 1.00)). In the Canadian cohort, the C-index suggests that the model starts to improve and does a better job at classifying the risk of the outcome in early adolescence ($>$11 years of age) and improves further in late adolescence. Thereafter ($>$20 years of age), the 95% confidence bands become wider, suggesting less precision in classifying risk.

In the Swiss cohort, the plot suggests that given 2 randomly selected 15-year-old offspring, there is approximately a 57% chance (95% CI: 0.45, 0.69) that the model would correctly predict which of them is more likely to develop a major mood disorder in the future. In the Swiss cohort, the 95%CI are wider, likely due to the smaller size of the Swiss as compared to the Canadian cohort. Of note, the external nature (e.g., used to replicate) of the Swiss cohort may impact accuracy (C-index level), but not necessarily precision (confidence intervals on C-indices). In both cohorts, the time-varying C-index starts to improve after 11 years of age up until about age 20, showing evidence of higher probability of correctly classifying risk of BD-related mood disorders with tighter 95%CIs compared to other ages. It should also be noted that the Swiss cohort has substantial left-truncation in relation to the Canadian cohort. That is, Swiss subjects did not all enter the study at age 0, as evident by the increasing sample under observation in the C-index calculations from age 0 to age 15 (Fig. 1). The low number of subjects under observation in the Swiss cohort at early ages leads to extremely wide 95% confidence intervals, and is also associated with poor model performance (C-index near or below 0.5).

Fig. 2 presents the estimated survival curve for major mood disorder from the model demonstrating that the parametric generalized gamma cure-fraction model fits the observed data from the Canadian cohort well. Note that the Kaplan-Meier estimate here gives the probability of not meeting diagnostic criteria for a major mood disorder by the age on the x axis. There is marginal evidence from the plot that the model may slightly overestimate the probability of major mood disorder around ages 12–15 years old; however, for the majority of the observed ages, the model appears to fit the raw data very well. Based on time-varying C-indices, the model taking into account lithium response of the affected parent and the model assuming missing lithium response of the affected parent performed very similarly (Supplementary Figure 3). The estimated survival curve for major mood disorder in the Swiss cohort can be found in Supplementary Figure 2 and while the overall shape of the survival curve appeared to match the Swiss cohort well, the model demonstrated evidence of underestimating the cumulative probability of major mood disorder.

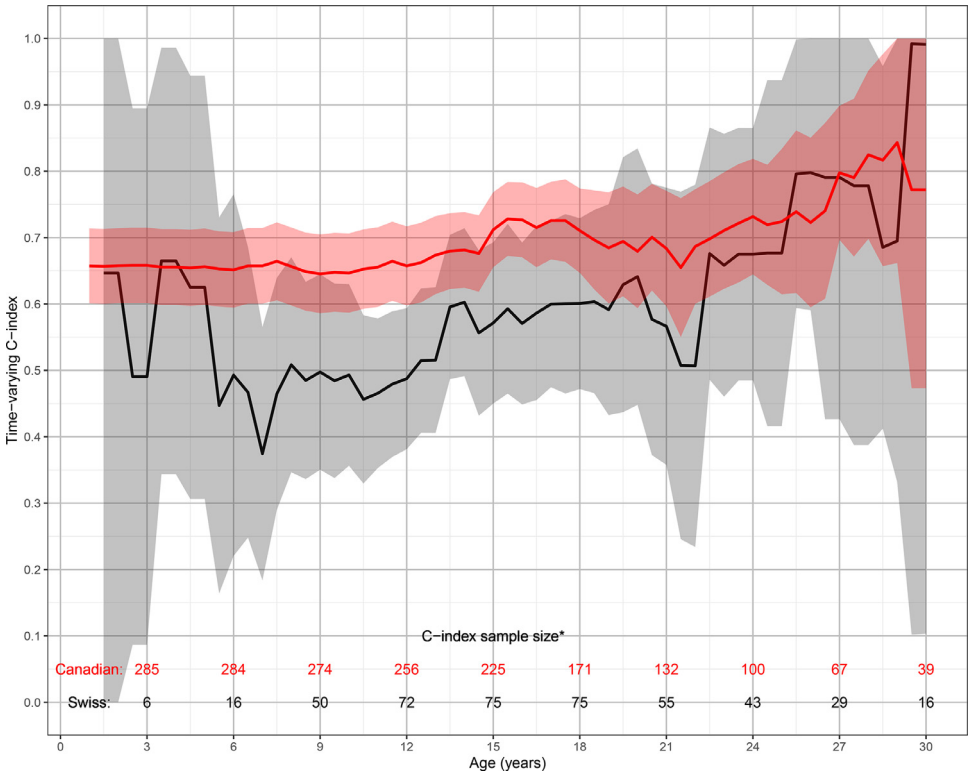


Fig. 1. Bias corrected time varying C-indices** for the risk calculator using the Canadian High-Risk Cohort (red) and Swiss High-Risk Cohort (black) by age in years. Shaded areas indicate 95% confidence bands. Bias correction, standard errors, and point-wise 95% confidence intervals were estimated using a Jackknife approach.

Notes: (1) *C-index sample size of subjects under observation for that age (shown every 3 years); (2) **The C-index shows the probability of correctly classifying which of two random individuals has the higher risk of the outcome given they were under observation at a given age.

3.4. The risk calculator: An example of a lower risk offspring

Figs. 3 and 4 present a risk calculator using parametric cure fraction survival models for an offspring of a BD parent at a lower-risk and a higher-risk of developing a major mood disorder. Fig. 3 describes the probability of developing the outcome for a male child (observed at age 10 years) with a parent with BD responsive to (stabilized on) lithium whose age of onset was 29 years and coming from a low-middle SES background. This individual's clinician confirmed excellent functioning and did not experience any sub-threshold clinically significant symptoms or concerns and did not meet lifetime criteria for any psychiatric diagnoses or minor mood disorder. Information on whether childhood physical or sexual abuse occurred was missing for this child, and therefore unknown. For this particular lower-risk offspring, the probability of developing major mood disorder 5 years into the future is 1.0% (95%CI: <1.0, 4.0%) and 10 years into the future is 3.0% (95%CI: 1.0, 11.0%). The observed data show that this individual was last observed at age 20.03 years (censored from the analysis) and remained well (did not develop the outcome) by this age.

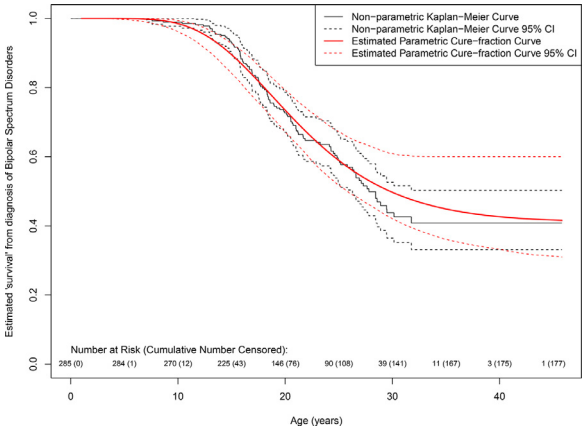


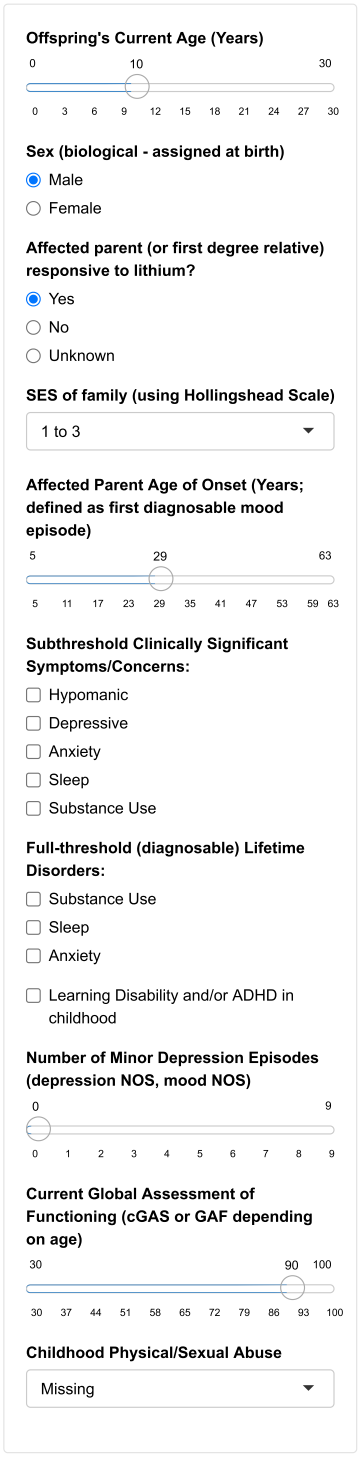
Fig. 2. Non-parametric Kaplan-Meier (black line), and Parametric Cure-fraction (red line) survival curves reflecting the probability of major mood disorders at a given age in the Canadian High-Risk Cohort.

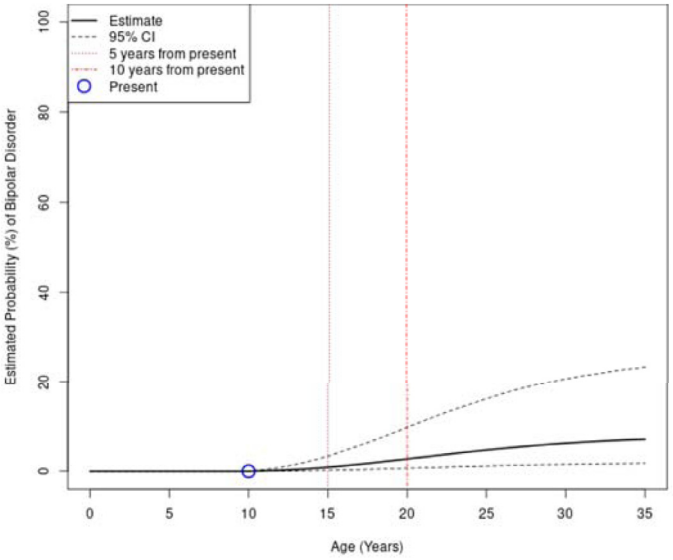
3.5. The risk calculator: An example of a higher risk offspring

Fig. 4 describes the probability of developing a major mood disorder for a female child (observed at age 10 years) with a parent with BD that did not stabilize on lithium whose age of onset was 15 years and coming from a high SES background. This individual's clinician confirmed good functioning and did experience sub-threshold clinically significant depressive symptoms and a lifetime diagnosis of anxiety disorder, neurodevelopmental disorder (learning disability and/or ADHD) and a minor mood disorder. It was confirmed that this individual did not experience any form of childhood physical or sexual abuse. For this particular offspring, the probability of developing a major mood disorder 5 years into the future is 30.0% (95%CI: 14.0, 47.0%) and 10 years into the future is 66.0% (95%CI: 35.0, 85.0%). The observed data show that this individual was diagnosed with a major mood disorder at age 16.03 years.

3.6. The risk calculator: An example of an offspring in the Swiss cohort

Two examples of the risk calculator are described above using Canadian subjects (Fig. 3 and 4); however, the same approach could be used for the Swiss cohort. For example, A female subject was observed in the Swiss cohort at age 16 years old, with a parent with BD responsive to (stabilized on) lithium whose age of onset was 20 years and coming from a high SES background. This individual did





5 Year Estimated Risk of Bipolar-related Mood Disorder:

Estimate	Low 95% CI	High 95% CI
1%	<1%	4%

10 Year Estimated Risk of Bipolar-related Mood Disorder:

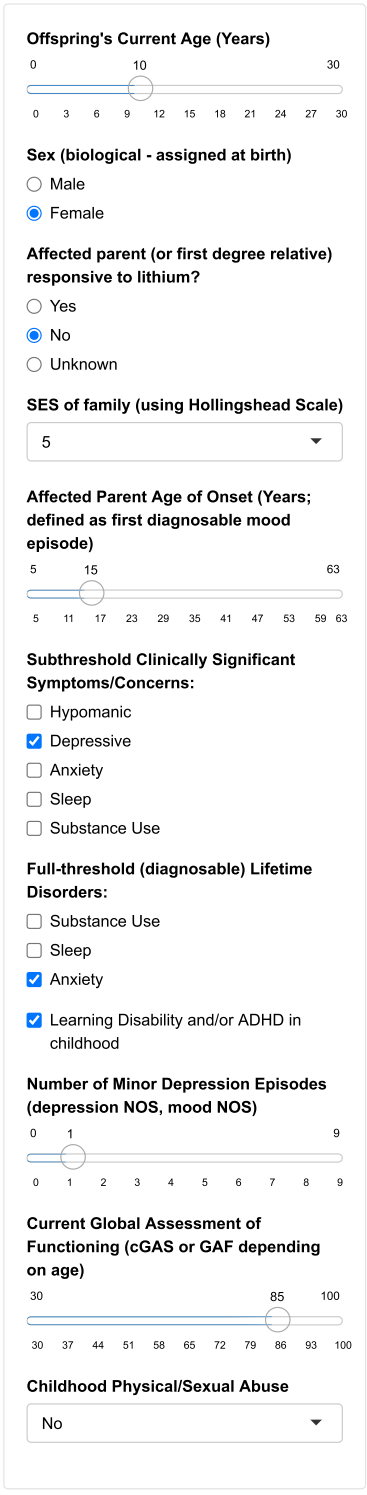
Estimate	Low 95% CI	High 95% CI
3%	1%	11%

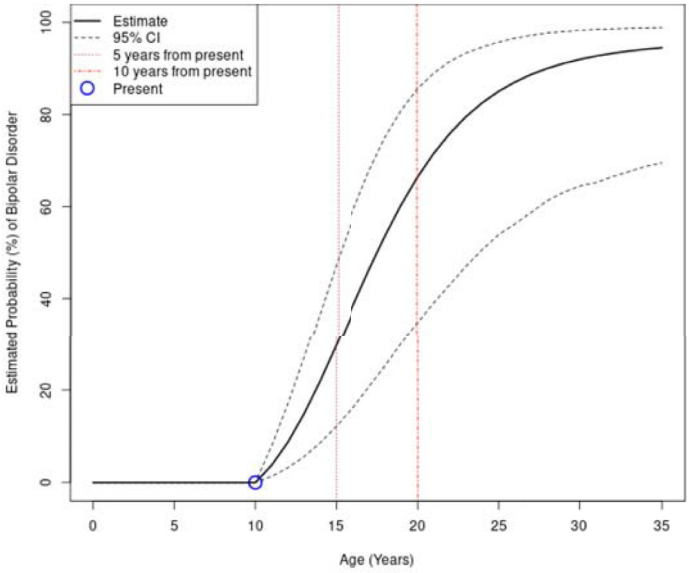
Fig. 3. Risk prediction example of a low-risk individual under observation at age 10 years using the Bipolar Spectrum Disorder Risk Calculator (for young people at familial risk of Bipolar Disorder).

not experience any sub-threshold clinically significant symptoms or concerns and did not meet lifetime criteria for any psychiatric diagnoses or minor mood disorder. There was no evidence childhood physical or sexual abuse occurred for this child. For this particular offspring, the probability of developing major mood disorder 5 years into the future is estimated to be 18% (95%CI: 10, 28%) and 10 years into the future is 33% (95%CI: 19, 48%). The observed data show that this individual was last observed at age 37.3 years and remained well (did not develop the outcome) by this age.

3.7. *Risk prediction at 5 and 10-years*

As would be expected, this risk calculator was most precise in predicting 5-year risk into the future. The 95% CI bands become wider with age reflecting less precise risk estimates 10 years into the future. For example, while the 10-year risk of developing the outcome for the presented higher-risk offspring was 66%, we are 95% confident that this risk lies anywhere between 35 and 85% (supplemental Figure 3).





5 Year Estimated Risk of Bipolar-related Mood Disorder:

Estimate	Low 95% CI	High 95% CI
30%	14%	47%

10 Year Estimated Risk of Bipolar-related Mood Disorder:

Estimate	Low 95% CI	High 95% CI
66%	35%	85%

Fig. 4. Risk prediction example of a high-risk individual under observation at age 10 years using the Bipolar Spectrum Disorder Risk Calculator (for young people at familial risk of Bipolar Disorder).

We produced risk calculator models to estimate the risk of developing a narrower definition of bipolar spectrum disorder (BDI, II, NOS) in both cohorts. Given that the number of events for this outcome was low, estimates are less precise (supplemental Figures 4 and 5).

4. Discussion

In this manuscript we present evidence that the individual risk of developing a major mood disorder 5 years into the future in young people at confirmed familial risk of BD can be accurately classified using a parametric time-to-event model restricted to data available in routine clinical practice. This risk calculator was most accurate at classifying risk in late adolescence to early adulthood in the Canadian high-risk offspring cohort. External validation results from a Swiss cohort were less conclusive. While there appeared to be a similar trend of better prediction at later ages, the model preformed no better than chance in the Swiss cohort in younger ages (<12 years), and suffered from small

sample sizes, leading to large confidence intervals around the time-varying C-index estimates.

`In this analysis, we tested an independent replication of the risk calculator by using the model estimates developed on the Canadian offspring cohort to estimate risk in an independent sample from a different country and continent. Not surprisingly, the model performance was strongest when tested on the Canadian cohort upon which it was derived compared to the Swiss offspring cohort. However, the risk calculator was able to predict major mood disorders in the Swiss cohort with good accuracy by late adolescence, while prediction during other ages were less accurate, although sample sizes were very small across some age categories. The Canadian offspring cohort was based on highly selected families identified through parents with a clinically confirmed BD1 diagnosis and who were prospectively evaluated in genetic and neurobiological studies repeatedly over the study period by research clinicians yielding a more homogeneous parental sample with stable diagnoses over many years of observation [16] . Further, response or non-response to long-term lithium treatment was based on prospective systematic treatment by the research team in a specialty clinical setting for the original Canadian parent probands [34]. Parental recruitment was subsequently extended within these families to include other first-degree relatives who met clinical research criteria for BD spectrum disorder and their lithium response profile was assessed. By contrast, in the Swiss families, offspring were recruited from parents with broader BD spectrum diagnoses that included schizoaffective BD and parents were assessed on the basis of cross-sectional structured interviews to confirm their diagnosis and lithium response. This, coupled with the smaller sample size and the fact that the model was derived from the Canadian cohort, may have contributed to the wider 95% CIs and lower C-index in the Swiss cohort. Based on the mixed results with wide confidence intervals when the model was applied to the Swiss cohort, future work should include applying the risk calculator to other populations to assess its predictive abilities outside a Canadian context.

Hafeman et al. [18] previously published a risk calculator based on data collected from the Pittsburgh BIOS high-risk offspring cohort. Their analysis approach used baseline-resetting Cox proportional hazards regression to model the time to event from each index assessment using a frailty model parametrization to account for clustering of visits within individual. The Cox model requires the assumption that hazards are proportional and requires some method for estimation of the baseline hazard [35]. It is unclear how the baseline hazard was estimated in the Hafeman et al. [18] risk calculator. Predictions of interest beyond hazard ratios require that the baseline hazard must be estimated in some way in conjunction with the Cox model. This can be done in many different ways (i.e. step functions, spline functions etc.) which was not specified by Hafeman et al. That said, how the baseline hazard was estimated, and how it is used in the prediction model are important pieces of information for any future replication or external validation of the risk prediction model. Another BD risk calculator developed to predict conversion rate from major depressive disorder, looked at 1-year fixed follow-up in a Chinese outpatient population [36]. This calculator was constructed using multivariable logistic regression modeling and forward variable selection. In addition to the limitations related to a fixed time-horizon of 1-year, forward variable selection techniques have been shown to lead to unstable and irreproducible risk calculators [37,38], and is therefore not recommended. Subjects with loss to follow-up in the 1-year were also removed from all analyses, which may lead to bias in the parameter estimates and risk predictions.

Strengths of our modeling approach include that we estimated 95% CIs which lend important information about the uncertainty of our estimates. Importantly, the inclusion of these estimates shows how imprecise certain predictions can be. We have also explicitly included a cure fraction in our model that negates the assumption that all participants will eventually experience the outcome. Further, we have replicated the model by testing it in an independent high-risk offspring sample that is more heterogeneous and generalizable then the highly selected Canadian families.

A potential limitation of our approach is the assumption that the population of individuals that will ever experience a BD spectrum disorder follows a shared generalized gamma distribution; however, based on model diagnostics, it appears the model making this assumption fits the raw data well. In addition, the precision of the model is reduced in high-risk offspring under age 10, but this agrees with clinical observations that symptoms and brain development are very much in flux and plastic, respectively during this developmental stage [39]. Another limitation is that in the replication analysis in the Swiss cohort a good C-index for the prediction of major mood disorders was only found from late adolescence. Our prediction model may not be generalizable beyond offspring of parents with confirmed BD. Further, while we attempted to select variables that are captured in routine clinical practice, these factors may differ by clinic and geography and some variables may not be readily available. For example, lithium response of the affected parent. However, we did estimate our model without lithium response and found a similar C-index pattern. Censoring was assumed to be non-informative, that is the length of follow-up was not directly associated with the probability of the outcome after accounting for covariates in the model. However, it is theoretically possible that subjects that contributed more information to the model (longer follow-up) may have more accurate risk predictions than those that contributed less information (shorter follow-up). Finally, the risk calculator approach could be dramatically improved with the inclusion of other variables in more of a research-based risk prediction model, including for example biomarker data, penetrance and other risk exposures.

Concerns from affected parents about the likelihood of their children developing a similar disorder was a major impetus for starting the Canadian high-risk offspring study over two decades ago. There was a clear need to improve upon risk estimation for individual children from these highly heterogeneous families. Here, we present an individual risk prediction model that uses clinical information available in routine practice and advanced statistical techniques to improve the precision of 5 and 10-year risk estimates for individual children at confirmed familial risk with good accuracy from the start of adolescence ($>$ 11 years of age). Individual risk prediction identifies from amongst young people at confirmed familial risk who might warrant closer observation and self-monitoring and benefit from psycho education and prevention of modifiable risk exposures (i.e., avoiding recreational drugs, moderating alcohol, adopting a healthy lifestyle and developing socio-emotional coping resources), while improving timely early intervention. In addition, prospective study of this high-risk group might support future targeted research into the mechanisms of illness onset thereby contributing to important advances in both clinical and research spheres. This risk model would benefit from further refinement including factoring in penetrance and should be tested on other high-risk, clinical and epidemiological populations.

Declaration of Competing Interest

All authors declare no conflicts of interest relating to this study. All funding for these studies came from the operating grants referenced in the Funding statement. In addition, Dr. Duffy served as the Chair of the Janssen Independent Data Safety Monitoring Board, which did not pose a conflict of interest relating to this work.

Contributors

AD, PG and SG (Canada), and MP (Switzerland) contributed to data collection and curation. EC contributed to data curation (Switzerland). CK conducted the analysis and contributed to data curation (Canada). CK, AD, and SG wrote the original draft of the manuscript. All authors contributed to editing the manuscript and interpreting

the results. MP and CV provided data for the independent replication. AD (Canada) and MP (Switzerland) verify and are responsible for the data underpinning the study.

Data sharing statement

Data may be shared upon request. All data inquiries can be made to Dr. Anne Duffy (anne.duffy@queensu.ca)

Acknowledgements

We would like to thank all of the high-risk families who participated in this research for their dedication to this work over many years. We also would like to acknowledge Elizabeth Tetzlaff for her help with the data collection.

Funding

The Canadian High-Risk Study has been funded by consecutive operating grants from the Canadian Institutes for Health Research, currently CIHR PJT Grant 152976

The Lausanne-Geneva high-risk study was and is supported by five grants from the Swiss National Foundation (#3200−040677, #32003B-105969, #32003B-118326, #3200−049746 and #3200−061974), three grants from the Swiss National Foundation for the National Centres of Competence in Research project "The Synaptic Bases of Mental Diseases" (#125759, #158776, and #51NF40 − 185897), and a grant from GlaxoSmithKline Clinical Genetics.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eclinm.2021.101083.

References

[1] Craddock N, Sklar P. Genetics of bipolar disorder. Lancet 2013;381:1654–62.
[2] Agnew-Blais J, Danese A. Childhood maltreatment and unfavourable clinical outcomes in bipolar disorder: a systematic review and meta-analysis. Lancet Psychiatry 2016;3:342–9.
[3] Duffy A, Horrocks J, Doucette S, Keown-Stoneman C, McCloskey S, Grof P. The developmental trajectory of bipolar disorder. Br J Psychiatry: J Ment Sci 2014;204:122–8.
[4] Lyall LM, Wyse CA, Graham N, et al. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. Lancet Psychiatry 2018;5:507–14.
[5] Gibbs M, Winsper C, Marwaha S, Gilbert E, Broome M, Singh SP. Cannabis use and mania symptoms: a systematic review and meta-analysis. J Affect Disord 2015;171:39–47.
[6] Kessing LV, Vradi E, McIntyre RS, Andersen PK. Causes of decreased life expectancy over the life span in bipolar disorder. J Affect Disord 2015;180:142–7.
[7] Kessing LV, Vradi E, Andersen PK. Life expectancy in bipolar disorder. Bipolar Disord 2015;17:543–8.
[8] Vieta E, Salagre E, Grande I, et al. Early Intervention in Bipolar Disorder. Am J Psychiatry 2018;175:411–26.
[9] Silva Ribeiro J, Pereira D, Salagre E, et al. Risk Calculators in Bipolar Disorder: a Systematic Review. Brain Sci 2020:10.
[10] Gottesman II, Laursen TM, Bertelsen A, Mortensen PB. Severe mental disorders in offspring with 2 psychiatrically ill parents. Arch Gen Psychiatry 2010;67:252–7.
[11] O'Donovan C, Alda M. Depression Preceding Diagnosis of Bipolar Disorder. Front Psychiatry 2020;11:500.
[12] Coleman JRI, Gaspar HA, Bryois J. Bipolar Disorder Working Group of the Psychiatric Genomics C, Major Depressive Disorder Working Group of the Psychiatric Genomics C and Breen G. The Genetics of the Mood Disorder Spectrum: genome-wide Association Analyses of More Than 185,000 Cases and 439,000 Controls. Biol Psychiatry 2020;88:169–84.
[13] Duffy A, Vandeleur C, Heffer N, Preisig M. The clinical trajectory of emerging bipolar disorder among the high-risk offspring of bipolar parents: current understanding and future considerations. Int J Bipolar Disord 2017;5:37.
[14] Rice J, Reich T, Andreasen NC, et al. The familial transmission of bipolar illness. Arch Gen Psychiatry 1987;44:441–7.
[15] Blacker D, Faraone SV, Rosen AE, et al. Unipolar relatives in bipolar pedigrees: a search for elusive indicators of underlying bipolarity. Am. J. Med. Genet. 1996;67:445–54.
[16] Duffy A, Goodday S, Keown-Stoneman C, Grof P. The emergent course of bipolar disorder: observations over two decades from the Canadian high-risk offspring cohort. Am J Psychiatry 2018 appiajp201818040461.
[17] Mesman E, Nolen WA, Reichart CG, Wals M, Hillegers MH. The Dutch bipolar offspring study: 12-year follow-up. Am J Psychiatry 2013;170:542–9.
[18] Hafeman DM, Merranko J, Goldstein TR, et al. Assessment of a person-level risk calculator to predict new-onset bipolar spectrum disorder in youth at familial risk. JAMA Psychiatry 2017;74:841–7.
[19] Birmaher B, Merranko JA, Goldstein TR, et al. A risk calculator to predict the individual risk of conversion from Subthreshold Bipolar symptoms to bipolar disorder I or II in youth. J Am Acad Child Adolesc Psychiatry 2018;57:755–63 e4.
[20] Vandeleur CL, Strippoli MF, Castelao E, et al. The Lausanne-Geneva cohort study of offspring of parents with mood disorders: methodology, findings, current sample characteristics, and perspectives. Soc Psychiatry Psychiatr Epidemiol 2017.
[21] Duffy A, Alda M, Kutcher S, Fusee C, Grof P. Psychiatric symptoms and syndromes among adolescent children of parents with lithium-responsive or lithium-nonresponsive bipolar disorder. Am J Psychiatry 1998;155:431–3.
[22] Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM. Best estimate of lifetime psychiatric diagnosis: a methodological study. Arch Gen Psychiatry 1982;39:879–83.
[23] Nurnberger Jr. JI, Blehar MC, Kaufmann CA, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry 1994;51:849–59 discussion 63-4.
[24] Hollingshead A. Commentary on the indiscriminate state of social class measurement. Soc Forces 1970;49:563.
[25] Manchia M, Adli M, Akula N, et al. Assessment of Response to Lithium Maintenance Treatment in Bipolar Disorder: a Consortium on Lithium Genetics (ConLiGen) Report. PLoS ONE 2013;8:e65636.
[26] R. R. A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020 https://wwwR-projectorg/.
[27] Amdahl J. Flexsurvcure: Flexible Parametric Cure Models. *R package version 120*. 2020 https://CRANR-projectorg/package=flexsurvcure.
[28] flexsurv Jackson C. A platform for parametric survival modeling in R. J Stat Softw 2016;70:1–33.
[29] TM Therneau, Grambsch PM. Modeling survival data: extending the cox model. statistics for biology and health. New York: Springer; 2000.
[30] Therneau TM. A Package for Survival Analysis in R. R Package Version 3.1-12. <*URL:* https://CRANR-projectorg/package=survival >.
[31] Mihram. EWSGA. Parameter estimation for a generalized gamma distribution. Technometrics 1965;7:349–58.
[32] Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modeling strategies for improved prognostic prediction. Stat Med 1984;3:143–52.
[33] Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. Stat Med 2005;24:3927–44.
[34] Grof P, Duffy A, Alda M, Hajek T. Lithium response across generations. Acta Psychiatr Scand 2009;120:378–85.
[35] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013:13.
[36] Gan Z, Diao F, Wei Q, et al. A predictive model for diagnosing bipolar disorder based on the clinical characteristics of major depressive episodes in Chinese population. J Affect Disord 2011;134:119–25.
[37] Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer International Publishing AG; 2015.
[38] PC Austin, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. J Clin Epidemiol 2004;57:1138–46.
[39] Rutter M, Kim-Cohen J, Maughan B. Continuities and discontinuities in psychopathology between childhood and adult life. J Child Psychol Psychiatry 2006;47:276–95.