Predicting The Risk and Timing of Bipolar Disorder In Offspring of Bipolar Parents: Exploring The Utility of a Neural Network Approach

Alysha Cooper
University of Guelph

Julie Horrocks
University of Guelph

Sarah Margaret Goodday
University of Oxford https://orcid.org/0000-0003-2159-1754

Charles Keown-Stoneman
University of Toronto

Anne Duffy (✉anne.duffy@queensu.ca)
University of Oxford https://orcid.org/0000-0002-5895-075X

Research

**Keywords:** Bipolar disorder, prediction, neural network, longitudinal

**Posted Date:** December 29th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-134834/v1

**License:** ☝️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

Bipolar disorder onset peaks over early adulthood and confirmed family history is a robust risk factor. However, penetrance within families varies and most children of bipolar parents will not develop the illness. Individualized risk prediction would be helpful for identifying those young people most at risk and to inform targeted intervention. Using prospectively collected data from the Canadian Flourish High-Risk Offspring cohort study available in routine practice, we explored the use of a neural network, known as the Partial Logistic Artificial Neural Network (PLANN) to predict the time to diagnosis of bipolar spectrum disorders.

Results

Overall, for predictive performance, PLANN outperformed the more traditional logistic model for one year, three year and five-year predictions. PLANN was better able to discriminate or rank individuals based on their risk of developing bipolar disorder, better able to predict the probability of developing bipolar disorder and had higher accuracy than the logistic model.

Conclusions

This evaluation of PLANN is a useful step in the investigation of using neural networks as tools in the prediction of diagnosis of mental health for at-risk individuals and demonstrated the potential that neural networks have in this field. Future research is needed to replicate these findings in a separate high-risk sample.

Background

Bipolar disorder affects an estimated 2.5% of the population, with higher prevalence for spectrum conditions\(^1\). The onset peaks in late adolescence and early adulthood\(^2\), however, delayed recognition and misdiagnosis remains a challenge. Untreated illness is associated with substantial morbidity and mortality early in the course\(^3\), and therefore timely and accurate diagnosis of bipolar disorder is critical to facilitate prompt treatment.

Bipolar disorder runs in families, and therefore the children of bipolar parents are an identifiable high-risk group ideally suited for risk prediction studies\(^4\). Family studies have shown that the bipolar trait segregating in families includes major depressive disorder, bipolar I, II and schizoaffective bipolar disorder\(^5\). The penetrance and spectrum varies between families and according to the subtype of illness. Furthermore, longitudinal prospective studies of high-risk offspring have provided strong evidence that the illness often debuts with depressive episodes\(^6\).
While key risk factors for the development of bipolar disorder have been identified such as characteristics of the parental age of onset and clinical course, early adversity, and antecedent clinically significant symptoms\(^7,8\), translatable risk prediction tools for clinicians do not exist or are in the early stages of development.

Given the heterogeneity in age of onset, it is imperative to use survival models rather than logistic regression or classification methods. Time-varying covariates are exposure variables that can vary with time across individuals, such as level of anxiety, or antecedent symptoms. Given the importance of antecedent risk factors contributing to the risk of bipolar disorder together with the variable age of onset, it is important to use methods which accommodate time-varying covariates, such as the Cox model or discrete-time survival models with time-dependent covariates\(^9\). In addition, including time-varying covariates allows the model to use the most recently available information for each individual.

Risk prediction tools attempt to incorporate multiple risk factors into a single model to estimate the probability that an individual will develop an outcome in the future. More recently, the use of neural networks has become increasingly popular in research for risk prediction. The goal of neural networks is to learn the relationship between a set of predictors and response(s) (i.e. target outcome(s)). The building blocks of neural networks are known as nodes, which are organized into layers and connected to one another through weights. Feed-forward neural networks, a common type of neural network, often have an input layer, one or more hidden layer(s) and an output layer. The information is distributed through the neural network in one direction, beginning at the input layer and finishing at the output layer\(^10,11\). (See supplemental material: Additional Methods – Neural networks and the discrete survival model).

Some advantages of neural networks are that they do not rely on strict assumptions and that they can accommodate non-linear relationships in the data\(^10,12\). Recently, recurrent neural networks have been used for survival analysis applications with time-dependent covariates\(^13\). Alternatively, discrete survival analysis has been extended to the field of neural networks in order to accommodate time-dependent covariates in prediction of survival\(^10,14-16\). The neural networks which extend upon discrete survival analysis lead to simpler implementation and interpretation compared to more complex methods such as the use of recurrent neural networks. That being said, the Ohno-Machado (1996)\(^15\) and Ohno-Machado (1995)\(^16\) neural networks involve the use of multiple neural networks, which adds complexity and can become computationally intensive.

The purpose of this article is to explore the use of a neural network, known as the Partial Logistic Artificial Neural Network (PLANN)\(^10\) to predict the time to diagnosis of bipolar spectrum disorders in the offspring of parents with confirmed bipolar disorder. PLANN is based on the logistic model for discrete survival analysis\(^9\). In this paper, we compare the two approaches. Both PLANN and the logistic model for discrete survival analysis predict the probability of an individual experiencing an event within a given time frame conditional on the individual not yet having experienced the event, which can be useful information for
clinicians. The prediction of which offspring are at greater risk of bipolar disorder over time may allow for more proactive monitoring and prevention (reducing stress, improving sleep, healthy lifestyle choices).

**Methods**

**Study design**

For this study, we used the data collected as part of the ongoing Canadian longitudinal high-risk offspring study described in detail elsewhere (Duffy et al., BJP, AJP)\(^6\,\text{17}\). The study design is a dynamic, prospective cohort study. Briefly, original study families were identified through parents with bipolar I disorder confirmed by SADS-L interview and blind consensus review of all available clinical information. Subsequently, pedigrees were expanded and included first degree relatives of the original probands, who themselves were affected with bipolar spectrum disorders (bipolar I, II, recurrent major depression). Agreeable offspring ages 5–25 years were enrolled and completed face to face research interviews following KSADS-PL format and study measures at baseline and the followed-up prospectively on average annually. This study has been reviewed for ethical compliance by the Ottawa Independent Research Ethics Board and the Queen’s University Health Science Research Ethics Board.

**Characteristics of participants**

In this analysis, we included 304 high-risk offspring from the Canadian high-risk cohort. The final data analysis was based on 292 high-risk individuals with no missing data for the predictors of interest. The outcome was defined as a DSM-V diagnosis of bipolar disorder (Bipolar I, II, NOS), major depressive disorder and/or schizoaffective disorder based on semi-structured KSADS-PL format interviews and blind consensus review based on all available clinical and research material.

We limited variables in the model to those that would be relevant and routinely collected by clinicians in an office setting (i.e., sex at birth, age at last observation, history of childhood abuse, antecedent clinically significant symptoms and non-mood syndromes. Clinically significant hypomanic, depressive, anxiety symptoms falling short of diagnostic criteria, as well as substance misuse and sleep problems were quantified based on clinical research interview and previously published consensus criteria\(^6\). Childhood physical and sexual abuse was determined in offspring 13 + years of age using the Childhood experiences of care and Abuse Scale\(^18\), while the (children’s) Global Assessment of Functioning\(^19\) was used at each assessment.

**Statistical analysis**

We compared the ability of Partial Logistic Artificial Neural Network (PLANN)\(^10\) to predict the time to diagnosis of bipolar spectrum disorders. We compared this approach to the more traditional logistic model for discrete survival analysis\(^9\). Models were evaluated using the time-dependent c-index\(^20\), the Brier score, and additional common measures of prediction accuracy. The time-dependent c-index is a measure of discrimination performance (See supplemental material – Additional methods – Model
Evaluation), that is, how well the model can rank individuals on their time to developing the outcome. The Brier score\textsuperscript{21} is a measure of calibration performance, which measures how well the model predicts the observed response. The Brier score measures the difference between the predicted probability of the event not occurring by a given follow-up time and the observed status of the individual at that time. The Brier score ranges between 0 and 1, with 0 indicating perfect calibration and 0.25 indicating a non-informative model that is no better than chance. Additional common measures of accuracy of prediction were used including the true positive rate (sensitivity), false positive rate (1 – specificity) and positive predictive value (probability that the those predicted to have the outcome actually experienced the outcome). We evaluated the predictive performance of the two approaches when the follow up time was divided into one year, three year, and five-year time intervals.

Cross validation is a technique used to evaluate how the results might generalize to an independent data set. For all these measures we used 10-fold cross-validation, which divides the data set into 10 subsets. The model is trained on (fit to) the data 10 times, each time leaving out a different subset which is used as the test set, on which the evaluation measures are calculated. Finally, the evaluation measures are averaged across the 10 folds\textsuperscript{22}.

**Predictor variables**

The time-fixed predictors included: sex (male, female), parental response to lithium prophylaxis (yes, no), parental age of onset (first meeting bipolar diagnosis) and physical/sexual abuse (yes, no, unsure). In addition, over the follow-up period, time-dependent measures were recorded for the individuals. Binary time-dependent variables indicating diagnosis of subthreshold symptoms or full-threshold clinical diagnoses of various disorders (i.e. subthreshold activation, subthreshold depression, subthreshold sleep, subthreshold substance use, subthreshold anxiety, substance use, sleep, anxiety, and neurodevelopmental) were each equal to 0 prior to diagnosis and 1 after diagnosis. In addition, the cumulative number of major mood and minor mood episodes were measured for individuals over the follow-up period.

**Partial Logistic Artificial Neural Network (PLANN)**

A neural network approach to discrete survival analysis was developed by Biganzoli et al. (1998)\textsuperscript{10}, known as the Partial Logistic Artificial Neural Network (PLANN). This approach divides the follow-up time into discrete intervals. The inputs to the PLANN model are the values of the covariates (e.g., predictors such as sex, occurrence of prior minor mood episode) in each time interval. The time interval itself is an additional input to the model. There is a single hidden layer. The predicted discrete hazard of the event (developing diagnosable bipolar disorder) is estimated for each time interval (1, 3 and 5 years) (See supplemental material – Additional Methods: PLANN).

For training the data, the observed or target response must be known. The target response of the neural network is the event indicator (outcome of interest i.e. bipolar spectrum disorder), $\delta_{ik}$, which is equal to 1 if the event occurs for the $i^{th}$ subject in the $k^{th}$ time interval and 0 if the event does not occur. For censored
individuals, the target response is equal to 0 for each time interval in which the subject is observed. For uncensored subjects, the target response is equal to 1 for the time interval in which the event occurs and is equal to 0 for the previous time intervals(10).

Hyper-parameters are parameters of the neural network which are selected by the researcher prior to training the neural network. We selected the hyper-parameters which enhanced discrimination performance via the time-dependent c-index. The learning rate, momentum, and the ridge regularization parameter were the hyper-parameters that were optimized. The learning rate and momentum are both used in the process of minimizing the loss function\(^{23}\) and the ridge regularization parameter determines to what degree the weights will be shrunk towards zero in order to avoid overfitting the model to the training set\(^{24}\).

**Results**

Table 1 presents the percent observed or means of the predictors included in the analyses. 112 (38.36\%) individuals developed bipolar spectrum disorder (outcome of interest) by last observation, while 180 (61.64\%) did not and were censored. Figure 1 presents the Kaplan-Meier curve showing the probability that the outcome has not occurred by the age on the x-axis.
Table 1  
Percent observed or mean of all variables included as covariates in models

| Variable                                      | Percentage Observed or Mean |
|-----------------------------------------------|----------------------------|
| Number of individuals                         | 292                        |
| Event observed                                | 38.36%                     |
| Censored                                      | 61.64%                     |
| Sex                                           | 41.10% male                |
| Age at last visit                             | 20.03 years                |
| Parental Lithium Response                     | 44.52%                     |
| Parental Onset Age                            | 25.47 years                |
| Physical/Sexual abuse                         |                            |
| Yes                                           | 10.24%                     |
| Missing                                       | 25.45%                     |
| Disorders                                     |                            |
| Subthreshold activation                       | 9.93%                      |
| Subthreshold depression                       | 10.62%                     |
| Subthreshold sleep                            | 3.08%                      |
| Subthreshold substance use                    | 12.33%                     |
| Subthreshold anxiety                          | 15.41%                     |
| Substance                                     | 17.47%                     |
| Sleep                                         | 18.49%                     |
| Anxiety                                       | 29.79%                     |
| Neurodevelopmental disorder                  | 11.30%                     |
| Number of major mood episodes                 | 0.17                       |
| Number of minor mood episodes                 | 0.21                       |

Predictive model using PLANN

The hyper-parameters selected for one-year, three-year and five-year predictions can be found in Table 2. The selected combination of hyper-parameters attained a mean time-dependent c-index of 0.6294,
0.5700, and 0.5841 during hyper-parameter optimization for one year, three year and five-year predictions, respectively.

| Table 2  | Selected hyper-parameters for PLANN for each time interval |
|----------|----------------------------------------------------------|
| Hyper-Parameter | One year | Three year | Five year |
| λ          | 0.0001    | 0.0001    | 0.0005    |
| Learning Rate | 0.1       | 0.1       | 0.05      |
| Momentum   | .90       | .90       | .99       |

For one-year predictions, PLANN performed well in terms of discrimination performance with a mean time-dependent c-index of 0.6325 across the 10 folds, indicating that the model can rank individuals better than chance. For three and five years, the mean time-dependent c-index across 10 folds was 0.5468 and 0.5902, respectively indicating relatively weaker performance at three and five years.

The results for the more traditional prediction measures such as accuracy, false positive rate (FPR), true positive rate (TPR), and positive predictive value (PPV) can be seen in Supplemental Tables 1–3 for one, three and five-year time intervals, respectively. PLANN had good accuracy overall for one-year predictions with an average accuracy across time of 0.6804. However, the mean accuracy across time decreased when three year and five-year predictions were made. Although PLANN had a high true positive rate in the majority of the time intervals for one, three and five-year predictions, it additionally had fairly high false positive rates and fairly low positive predictive values (probability that offspring predicted to have the outcome actually experienced the outcome). It should be noted that the time intervals in which PLANN had particularly low positive predictive values are those which had only a small number of events observed.

It was of interest to assess whether PLANN could distinguish between three individuals in the test set. The three individuals were selected based on their observed diagnosis and censoring time. The individual in the test set with the earliest diagnosis time was considered the ‘earlier-onset’ individual, the individual with the median diagnosis time was considered the ‘mid-onset’ individual and the individual with the highest censoring time (longest survival time) was the ‘no onset’ individual. The ‘earlier-onset’ individual experienced a diagnosis at 11.64 years, the ‘mid-onset’ individual was diagnosed at 19.85 years and the ‘no onset’ individual was censored at 39.79 years. For these three individuals, the predicted survival curves were plotted for one year, three year and five-year predictions (Fig. 2). As seen in Fig. 2, PLANN predicted that the ‘earlier-onset’ individual had the lowest survival probability over time. When making one-year predictions, PLANN could predict that the ‘mid-onset’ individual had a lower survival probability than the ‘no onset’ individual. However, PLANN had more difficulty distinguishing between the ‘mid-onset’ and ‘no onset’ individuals when three and five-year predictions were made.
For comparison with PLANN, the logistic model for discrete survival analysis was assessed in terms of its predictive performance for one, three and five-year predictions. The mean time-dependent c-index and Brier scores across the 10 folds can be found in Table 3. In terms of discrimination performance (i.e. ranking individuals based on their risk), the logistic model performed only slightly better than chance when making one year, three year and five-year predictions, as indicated by the fact that the time-dependent c-indices were each close to 0.5. The logistic model had the best discrimination performance (i.e. highest c-index) when making one-year predictions compared to three and five-year predictions.

The more traditional measures of prediction can be seen in Supplemental Tables 4–6 for one year, three year, and five-year intervals, respectively. As can be seen, the mean accuracy across all time intervals increased as the time interval width increased. For one-year predictions, the time intervals which had higher true positive rates additionally had higher false positive rates and the positive predictive values were all fairly low for the logistic model with one-year predictions. Overall, the false positive rates dropped when the time interval width increased.

Finally, it was of interest to assess whether the logistic model could distinguish between the ‘earlier-onset’, ‘mid-onset’ and ‘no onset’ individuals on a single test set, defined previously. For these individuals, the predicted survival curves were plotted for the logistic model with one-year predictions, three-year predictions and five-year predictions. See Fig. 3 for the predicted survival curves. As seen in Fig. 3, the logistic model predicted that the ‘earlier-onset’ individual had a higher probability of diagnosis (i.e. lower survival probability) over time compared to the ‘mid-onset’ and ‘no onset’ individuals. However, for one year, three year and five-year predictions, the logistic model predicted that the ‘mid-onset’ individual had a higher probability of not being diagnosed than the ‘no onset’ individual.

Table 3: Comparison of evaluation metrics for PLANN and the logistic model with one year, three year and five-year predictions.
## Discussion

In this study we explored the potential utility of using partial logistic artificial neural network (PLANN), an extension of discrete survival analysis, to predict time to diagnosis of bipolar disorder at 1, 3 and 5 years into the future in a well-characterized prospectively followed cohort of high-risk individuals identified based on a parent with bipolar spectrum disorder. We limited fixed and time varying covariates in the model to data that would be routinely collected and available in clinical practice (i.e., sex, age, childhood abuse, subthreshold antecedent clinically significant symptoms and lifetime antecedent non-mood diagnoses).

PLANN was compared to a traditional logistic model for discrete survival analysis to assess whether the use of a neural network provides any benefit over a traditional statistical modeling approach. While PLANN and the logistic model have common advantages, such as enabling the incorporation of time-varying covariates due to the use of discrete time intervals, both models also have distinct advantages.

|                | One Year | Three Years | Five Years |
|----------------|----------|-------------|------------|
|                | PLANN    | Logistic    | PLANN      | Logistic    | PLANN      | Logistic    |
| Mean Brier Score\(^a\) | 0.1677   | 0.1796      | 0.1823     | 0.1876      | 0.1785     | 0.1900      |
| C-index        | 0.6325   | 0.5562      | 0.5486     | 0.5132      | 0.5902     | 0.5204      |
| Mean Accuracy  | 0.6804   | 0.5557      | 0.5905     | 0.5719      | 0.6066     | 0.6006      |
| (SD, Accuracy) | (0.1600) | (0.1902)    | (0.0973)   | (0.0943)    | (0.0580)   | (0.0646)    |
| Mean FPR\(^b\) | 0.3173   | 0.4401      | 0.4094     | 0.4178      | 0.3826     | 0.3852      |
|                | (0.1645) | (0.1932)    | (0.1004)   | (0.1113)    | (0.0774)   | (0.0764)    |
| Mean TPR\(^b\) | 0.4883   | 0.3826      | 0.5058     | 0.4743      | 0.5773     | 0.5154      |
|                | (0.2645) | (0.2402)    | (0.1875)   | (0.1551)    | (0.0849)   | (0.0857)    |
| Mean PPV\(^b\) | 0.0811   | 0.0432      | 0.1337     | 0.1301      | 0.2303     | 0.2239      |
|                | (0.0799) | (0.0428)    | (0.0930)   | (0.0822)    | (0.1215)   | (0.1340)    |

\(^a\)The mean Brier score is the average Brier score across 15, 20, and 25 years and mean accuracy is the average accuracy (standard deviation) over the time intervals measured.

\(^b\)FPR, TPR and PPV are the average values (standard deviation) over the time intervals measured.
over one another. The logistic model allows for the interpretation of the effect of covariates on the discrete hazard and the evaluation of whether or not the covariates have a significant effect on the discrete hazard. That is, it is possible to estimate the magnitude of effect of each unique predictor on the outcome, which is not possible using PLANN. On the other hand, PLANN has the ability to automatically detect non-linear relationships in the data. The importance of being enabled to automatically detect non-linear relationships in data, not possible by logistic models, is apparent for certain outcomes in bipolar disorder. For example, mood instability has been found to follow non-linear patterns\textsuperscript{25}.

Overall, for predictive performance, PLANN outperformed the logistic model for one year, three year and five-year predictions. PLANN was better able to discriminate or rank individuals based on their risk of developing bipolar disorder (i.e., higher time-dependent c-indices), better able to predict the probability of developing bipolar disorder (i.e., lower Brier scores) and had higher accuracy than the logistic model.

Both PLANN and the logistic model performed better in terms of discrimination (i.e. time-dependent c-index) and calibration performance (i.e. Brier scores) for more proximal predictions (i.e., one-year), compared to more distal predictions (three and five-year). Moreover, the calibration performance deteriorated over time, with poor performance for more distal predictions of survival probability (i.e. 25 years) compared to more proximal predictions (i.e. 15 years) for all time interval widths. This finding was corroborated when examining how well the models could distinguish between an individual who had an earlier-onset diagnosis (12 years) from individuals who did not experience a diagnosis (40 years) or experienced a mid-onset diagnosis (20 years) in terms of their survival probability. Both models predicted that the earlier-onset (i.e. higher risk) individual had the lowest survival probability, however had difficulty in distinguishing between the mid-onset (i.e. medium risk) and no onset (i.e. lower risk) individuals who had longer survival times.

Interestingly, the three and five-year models had higher true positive rates and higher positive predictive values compared to the one-year models. This difference in results can potentially be explained by the fact that the time-dependent c-index and the Brier score use the predicted probability of not being diagnosed by given follow-up times for their measures whereas the false positive rate, true positive rate and positive predictive values use the conditional probability of diagnosis within each time interval for their measures. Therefore, when making predictions of the probability of not being diagnosed at given follow-up times, the one-year predictions are preferable but if one is interested in the conditional risk within a given time interval, then wider time intervals such as three or five years are preferable.

Our risk prediction approach of using PLANN to predict onset of bipolar disorder differs from other published risk calculators (e.g., Hafeman et al (2017)\textsuperscript{26}), that have used a “baseline re-setting” Cox proportional hazards model. While both methods allow the inclusion of covariates measured at baseline and at follow-up visits and neither method requires an assumption about the distribution, the PLANN method we took does not require a proportional hazards assumption. In addition, we only included model variables that would be available in routine practice.
Strengths and Limitations

Strengths include the carefully assessed parental diagnoses based on longitudinal clinical observations confirming the risk status in the offspring, the measurement of diagnosis in high-risk offspring through semi-structured research clinical assessments and blind consensus reviews. However, the following limitations relevant to this analysis are worth noting. This sample size is small; additional breadth of data (e.g., genetic data, behavioural data) could have improved predictions; and external replication is needed.

Conclusion

This evaluation of PLANN is a useful step in the investigation of using neural networks as tools in the prediction of diagnosis of mental health for at-risk individuals and demonstrated the potential that neural networks have in this field. PLANN performed better than the traditional discrete time survival model in predicting the development of bipolar disorder in high-risk individuals. However, both approaches struggled in making more distal predictions into the future. Future research replicating these approaches in different samples with the inclusion of additional data will help inform the further utility of risk prediction models to aid in clinical decision making in patients with bipolar disorder.

Abbreviations

BD
Bipolar Disorder
CECA
Childhood experience of care and abuse
DSM
Diagnostic and Statistical Manual for Mental Disorders
K-SADS – PL
Kiddie Schedule for Affective Disorders – Present and Lifetime

Declarations

Ethics approval: Data used in this analysis came from the Canadian High-Risk Offspring of Bipolar Parents study that has been reviewed for ethical compliance and approved by the Independent Research Ethics Board (CIRBI Ottawa Pro00011514) and by the Queen’s University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (HSREB PSIY-561-17).

Consent for publication: all named authors have provided consent for publication.

Availability of data: This is an ongoing study and requests for access to analyze de-identified study data may be made to the principal investigator Dr Anne Duffy as per ethically reviewed study protocol.
Competing Interests: all authors confirm that they have no competing or conflicts of interest to declare in relation to this manuscript.

Funding: The high-risk offspring of bipolar parents study has and continues to be funded through operating grants from the Canadian Institutes of Health Research. PJT 152976.

Authors Contributions: Horrocks, Cooper designed the analysis with input from Keown-Stoneman, Goodday and Duffy. Duffy is the PI of the high-risk study and contributed with Goodday to data collection. Cooper supervised by Horrocks carried out the data analysis and discussed results and interpretation with the other authors listed. Cooper and Horrocks drafted the manuscript and Duffy, Keown-Stoneman and Goodday edited and revised iterative drafts.

Acknowledgements: The authors acknowledge the unprecedented commitment to research excellence and clinical care made by Dr Paul Grof who assessed and clinically treated many of the affected parents in this research. Without his dedication and attention to detail this work would not be possible. We also acknowledge the excellent technical assistance of Elizabeth Tetzlaff who prepared and vetted the database for this analysis.

References

1. Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RMA, Petukhova M, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch Gen Psychiatry. 2007;64(5):543–52.
2. Manchia M, Lampus S, Chillotti C, Sardu C, Ardau R, Severino G, et al. Age at onset in Sardinian bipolar I patients: evidence for three subgroups. Bipolar disorders. 2008;10(3):443–6.
3. 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.
4. 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. International journal of bipolar disorders. 2017;5(1):37.
5. Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. In: American Journal of Medical Genetics Part C: Seminars in Medical Genetics. Wiley Online Library; 2003. p. 48–58.
6. 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. 2019;176(9):720–9.
7. Preisig M, Strippoli M-PF, Castelao E, Merikangas KR, Gholam-Rezaee M, Marquet P, et al. The specificity of the familial aggregation of early-onset bipolar disorder: a controlled 10-year follow-up study of offspring of parents with mood disorders. J Affect Disord. 2016;190:26–33.
8. Duffy A, Jones S, Goodday S, Bentall R. Candidate risks indicators for bipolar disorder: early intervention opportunities in high-risk youth. International Journal of Neuropsychopharmacology.
9. Cox DR. Regression models and life-tables. J Roy Stat Soc: Ser B (Methodol). 1972;34(2):187–202.
10. Biganzoli E, Boracchi P, Mariani L, Marubini E. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. Statistics in medicine. 1998;17(10):1169–86.
11. Warner B, Misra M. Understanding neural networks as statistical tools. The american statistician. 1996;50(4):284–93.
12. Bourquin J, Schmidli H, van Hoogevest P, Leuenberger H. Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development. Pharmaceutical development technology. 1997;2(2):95–109.
13. Zheng P, Yuan S, Wu X. Safe: A neural survival analysis model for fraud early detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019. p. 1278–85.
14. Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. Breast cancer research treatment. 1992;22(3):285–93.
15. Ohno-Machado L. Sequential use of neural networks for survival prediction in AIDS. In: Proceedings of the AMIA Annual Fall Symposium. American Medical Informatics Association; 1996. p. 170.
16. Ohno-Machado L, Walker MG, Musen MA. Hierarchical neural networks for survival analysis. Medinfo. 1995;8:828–32.
17. Duffy A, Horrocks J, Doucette S, Keown-Stoneman C, McCloskey S, Grof P. The developmental trajectory of bipolar disorder. The British Journal of Psychiatry. 2014;204(2):122–8.
18. Bifulco A, Bernazzani O, Moran PM, Jacobs C. The childhood experience of care and abuse questionnaire (CECA. Q): validation in a community series. Br J Clin Psychol. 2005;44(4):563–81.
19. Hall RCW. Global assessment of functioning: a modified scale. Psychosomatics. 1995;36(3):267–75.
20. Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. Statistics in medicine. 2005;24(24):3927–44.
21. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Statistics in medicine. 1999;18(17-18):2529–45.
22. Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural networks. Chemometrics intelligent laboratory systems. 1997;39(1):43–62.
23. Ruder S. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:160904747. 2016.
24. Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc: Ser B (Methodol). 1996;58(1):267–88.
25. Bonsall MB, Wallace-Hadrill SMA, Geddes JR, Goodwin GM, Holmes EA. Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. Proceedings of the Royal Society B: Biological Sciences. 2012;279(1730):916–24.
26. Hafeman DM, Merranko J, Goldstein TR, Axelson D, Goldstein BI, Monk K, 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(8):841–7.

Figures

**Figure 1**

Kaplan-Meier curve for diagnosis of bipolar disorder, major depressive disorder and/or schizoaffective
disorder for at-risk individuals.
Figure 2

Predicted survival curves for ‘earlier-onset’, ‘mid-onset’ and ‘no onset’ individuals in test set of single fold for PLANN with a) one year, b) three year and c) five-year predictions.
Figure 3

Predicted survival curves for ‘earlier-onset’, ‘mid-onset’ and ‘no onset’ individuals in test set of single fold for the logistic model with a) one year, b) three year and c) five-year predictions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NNRiskanalysisSuppMaterialDec21Submitted.docx