

RESEARCH

Open Access



Predicting the risk and timing of major mood disorder in offspring of bipolar parents: exploring the utility of a neural network approach

Alysha Cooper¹, Julie Horrocks¹, Sarah Goodday^{2,3}, Charles Keown-Stoneman⁴ and Anne Duffy^{2,5*} 

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 major mood disorders in 1, 3 and 5-year intervals.

Results: Overall, for predictive performance, PLANN outperformed the more traditional discrete survival model for 3-year and 5-year predictions. PLANN was better able to discriminate or rank individuals based on their risk of developing a major mood disorder, better able to predict the probability of developing a major mood disorder and better able to identify individuals who would be diagnosed in future time intervals. The average AUC achieved by PLANN for 5-year prediction was 0.74, which indicates good discrimination.

Conclusions: This evaluation of PLANN is a useful step in the investigation of using neural networks as tools in the prediction of mood disorders in at-risk individuals and the potential that neural networks have in this field. Future research is needed to replicate these findings in a separate high-risk offspring sample.

Keywords: Bipolar spectrum, Major mood disorder, Prediction, Neural network, Longitudinal, High-risk offspring

Background

Bipolar disorder affects an estimated 2.5% of the population, with higher prevalence for spectrum conditions (Merikangas et al. 2007). The onset peaks in late adolescence and early adulthood (Manchia et al. 2008); however, delayed recognition and misdiagnosis remains a challenge. Untreated illness is associated with substantial morbidity and mortality early in the course (Kessing et al.

2015), and therefore timely and accurate diagnosis 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 (Duffy et al. 2017). Family studies have shown that bipolar-related mood disorders segregating in families includes major depressive disorder, bipolar I, II and schizoaffective bipolar disorder (Smoller and Finn 2003; McMahon et al. 2010; Craddock and Sklar 2013). The penetrance and spectrum of phenotypes vary between families and according to the subtype of bipolar illness. Longitudinal prospective studies of high-risk offspring have provided

*Correspondence: anne.duffy@queensu.ca

² Department of Psychiatry, University of Oxford, Oxford, UK

Full list of author information is available at the end of the article

strong evidence that the illness often debuts with depressive episodes years before any hypomanic or manic episodes (Duffy et al. 2019).

While key risk factors for the development of bipolar disorder have been identified such as parental age of onset and clinical course, early adversity, and antecedent clinically significant symptoms (Preisig et al. 2016; Duffy et al. 2016), translatable risk prediction tools for clinicians do not exist or are in the early stages of development. For example, to our knowledge there has only been one published individualized risk calculator based on data from the BIOS longitudinal study of children of bipolar parents (Hafeman et al. 2017), which is yet to be replicated on an independent sample of familial at-risk offspring.

For studying events that occur over time, specialized modelling techniques that accommodate censoring are required (Allison 2010; Collett 2015). At the time the data analysis is conducted, not all individuals will necessarily have experienced the event of interest (bipolar-related mood disorders), and methods of survival analysis have been developed to accommodate these “censored” individuals, without introducing bias. In a longitudinal study, information on exposures and symptoms that might affect the risk of experiencing the outcome is collected repeatedly over time. Specialized techniques are necessary to include these repeatedly measured time-varying predictors in the analysis, while ensuring that only those exposures which occur before the outcome event are counted as possibly affecting the outcome. These techniques allow us to include the most up-to-date information collected on exposures, study the effect of symptoms that may not be present at baseline but develop during the course of follow-up, and ensure that the timing of the exposures in relation to the event of interest is taken into account. The latter is important as risk may be increased immediately after the exposure onset compared to later in time. Also, these techniques allow us to include numerical covariates that change over time, such as cumulative number of exposures at each timepoint. The Cox model (Cox 1972) and the discrete survival model (Efron 1988) are two such techniques from statistical survival analysis.

Recently, the use of neural networks has become increasingly popular in research for risk prediction (Charati et al. 2018; Wunnavu et al. 2019; Li et al. 2017; LaFaro et al. 2015). An advantage of neural networks is that they do not rely on assumptions such as the distribution of the response variable or proportional hazards. Furthermore, they automatically accommodate non-linear relationships between the response and exposure variables (Biganzoli et al. 1998; Bourquin et al. 1997). Thus, rather than the researcher having to postulate a complicated, possibly non-linear model, the neural network can

just “learn” the relationship with minimal direction from the researcher.

The purpose of this article is to explore the use of a neural network known as Partial Logistic Artificial Neural Network (PLANN) (Biganzoli et al. 1998), to predict the time to diagnosis of bipolar-related major mood disorders in the offspring of parents with confirmed bipolar disorder. For context, we compare PLANN to the more traditional discrete survival model. Both approaches accommodate censoring and time-varying predictors. The models are compared using several measures that assess the accuracy of the predictions. The prediction of which offspring are at greater risk of major mood disorder over time is important for clinical researchers, as it may allow for more proactive monitoring and prevention.

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. 2014, 2019). 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-related major mood 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 then 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 and variables

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-IV diagnosis of bipolar-related major mood disorder including: 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. Participants who did not have a diagnosis of major mood disorder before their last follow-up visit were considered censored at their last visit.

We limited variables in the model to those that would be relevant and routinely collected by clinicians in an

office setting. Time-fixed predictors included sex at birth, parental response to lithium prophylaxis, parental age of onset of bipolar diagnosis, and childhood physical/sexual abuse. Time-varying predictors included the absence or presence at a given age of antecedent clinically significant symptoms and non-mood disorders (occurring prior to the outcome). In addition, the cumulative number of antecedent major and minor mood episodes at each age (occurring prior to the outcome) were included as time-varying predictors. Clinically significant activation (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 (Duffy et al. 2019) (see Additional file 1). Childhood physical and sexual abuse was determined in offspring 13 years of age and older using the Childhood experiences of care and Abuse Scale (Bifulco et al. 2005). Note that only information up until time of diagnosis of major mood disorder or last follow-up visit was used in constructing the time-varying predictors. The time scale for both models was age of participant. In all analyses, covariates were lagged by one time interval in order to reduce the risk of reverse-causality and to enable predictions in the future time interval.

Statistical analysis

PLANN and the discrete survival model are similar in many respects. Both approaches can accommodate time-fixed and time-varying predictors and use the same data setup (see Additional file 1). Both approaches predict the probability that an individual will experience the outcome within a given time frame, conditional on the individual not yet having experienced the outcome. However, the internal calculations performed to make the predictions are different (see Additional file 1), resulting in different values for the predicted probabilities. The discrete survival model assumes that odds are proportional at each time point and that the relationship between the hazard (risk) of the event and covariates is linear on the logistic scale, whereas PLANN makes no such assumptions (Allison 2010; Biganzoli et al. 1998). Details of the two models are included in the Additional file 1.

Models were evaluated using several assessment measures. The time-dependent c-index (Antolini et al. 2005) was used to quantify how well the model can rank individuals on their time to developing the outcome. The area under the receiver operating curve (AUC) is often used as a measure of the model's ability to discriminate low and high-risk individuals (Zhou et al. 2011). Both the c-index and AUC range between 0 and 1, with 1 being best and a value greater than 0.5 being better than chance. The Brier score (Graf et al. 1999) 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 lower scores being better, 0 indicating perfect calibration and 0.25 indicating a non-informative model that is no better than chance. Common measures of prediction performance were used including accuracy, sensitivity, specificity, and positive predictive value. We used tenfold stratified cross-validation (see Additional file 1) to evaluate the predictive performance of the two models over 1-year, 3-year, and 5-year time intervals. For percentages, 95% confidence intervals were calculated as $100 \times (p \pm 1.96 \times \sqrt{p \times (1-p)/n})$ where p is the observed proportion and n is the sample size. For means m , 95% confidence intervals were calculated as $m \pm 1.96 \times (sd/\sqrt{n})$ where sd is the standard deviation.

Results

Table 1 presents the observed percent or means (and 95% confidence intervals) of the predictor variables included in the analyses. About 41% of participants were male. The mean age (at the time of the outcome event or censoring) was 21 years, and the mean age of parental onset was about 25 years. Childhood abuse was reported by 10% of offspring, while 38% were missing information on this variable as result of not being age appropriate for the measure or not yet completing the measure on next research visit. For the time-varying predictors, reported percentages/means are for disorders/episodes experienced before the outcome event or censoring. Over 41% of individuals had at least one clinically significant sub-threshold symptom presentation, with the most prevalent being subthreshold anxiety at 15.4%. The most prevalent disorder was anxiety disorder at close to 30%.

Out of the 292 individuals included in the analysis, 112 (38.4%) developed a major mood disorder by their last follow-up visit, while 180 (61.6%) did not and were censored. As shown in the right-hand column of Table 1, proportionately fewer males experienced the outcome than were present in the full sample (31% versus 41%), and those who experienced the outcome were slightly younger on average (19.6 versus 21.0 years). More of those diagnosed with a major mood disorder experienced substance use, anxiety, and subthreshold activation and depression than in the full at-risk offspring sample.

Model comparisons

Table 2 compares PLANN and the discrete survival model in terms of the assessment measures, averaged across time intervals and using tenfold cross-validation, for 1-, 3- and 5-year predictions. Recall that for the Brier score, lower is better, while for the c-index and

Table 1 Percent observed or mean of all variables included as covariates in the models and 95% confidence intervals for the percent or mean, for full sample (n = 292) and individuals with the outcome (n = 112)

Variable	Full sample (n = 292)		Individuals with outcome (n = 112)	
	n (%) or mean	[95% CI]	n (%) or mean	[95% CI]
Time-fixed				
Sex (male)	121 (41.4)	[35.8, 47.1]	35 (31.3)	[22.6, 39.9]
Age (at outcome or censoring)	21.0	[20.2, 21.9]	19.6	[18.6, 20.6]
Parental lithium response	129 (44.2)	[38.5, 49.9]	47 (42.0)	[32.8, 51.1]
Parental onset age	25.4	[24.4, 26.5]	25.5	[23.7, 27.4]
Physical/sexual abuse				
Yes	30 (10.3)	[6.1, 13.8]	14 (12.5)	[6.3, 18.7]
Missing ^a	111 (38.0)	[32.4, 43.6]	32 (28.6)	[20.2, 37.0]
Time-Varying				
Disorders				
Subthreshold ^b activation	29 (9.9)	[6.5, 13.4]	19 (17.0)	[10.0, 23.9]
Subthreshold depression	31 (10.6)	[7.1, 14.2]	17 (15.2)	[8.5, 21.6]
Subthreshold sleep	9 (3.1)	[1.1, 5.1]	2 (1.8)	[0.0, 4.2]
Subthreshold substance use	36 (12.3)	[8.6, 16.1]	13 (11.6)	[5.6, 17.6]
Subthreshold anxiety	45 (15.4)	[11.3, 19.6]	17 (15.2)	[8.5, 21.9]
Substance use	51 (17.5)	[13.1, 21.8]	29 (25.9)	[17.7, 34.0]
Sleep	54 (18.5)	[14.0, 23.0]	18 (16.1)	[9.2, 22.9]
Anxiety	87 (29.8)	[24.5, 35.0]	39 (34.8)	[26.0, 43.7]
Neurodevelopmental	33 (11.3)	[7.7, 14.9]	12 (10.7)	[5.0, 16.5]
Number of episodes				
Major mood	0.2	[0.1, 0.3]	0.2	[0.1, 0.3]
Minor mood	0.2	[0.1, 0.3]	0.1	[0.0, 0.2]

^a Missing refers to having not completed the measure either due to not being age appropriate to complete or not yet completing the measure; ^bsee Additional file 1 for definitions of subthreshold conditions

Table 2 Comparison of the evaluation metrics for PLANN and the discrete survival model with 1-year, 3-year and 5-year predictions

	One-year		Three-year		Five-year	
	PLANN	survival	PLANN	survival	PLANN	survival
Mean Brier Score ^a	0.177	0.180	0.182	0.188	0.179	0.190
C-index	0.633	0.556	0.549	0.513	0.590	0.520
Mean AUC(SD) ^b	0.633 (0.170)	0.740 (0.144)	0.708 (0.123)	0.655 (0.124)	0.745 (0.076)	0.669 (0.063)

^a The mean Brier score is the average Brier score across 15, 20, and 25 years

^b The mean AUC is the average AUC across the time intervals measured

AUC, higher is better. PLANN outperformed the discrete survival model on the Brier score which measures the difference between observed and predicted values. The c-index indicates how well the model ranks individuals in terms of their event times. PLANN outperformed the discrete survival model at all three interval lengths. On mean AUC, PLANN outperformed the discrete survival model in 3-year and 5-year predictions, but the opposite is true in 1-year predictions. PLANN does better at 5-year predictions than at 1- or 3-year

predictions according to mean AUC. The 5-year mean AUC for PLANN was 0.74.

Table 3 shows mean accuracy, specificity, sensitivity, and positive predictive value (PPV), for 1-year, 3-year, and 5-year predictions for the two models, where the mean is taken across time intervals and tenfold cross-validation was used (see Additional file 1 for more details). For 1-year prediction, PLANN outperformed the discrete survival model with higher sensitivity, specificity, and PPV when the optimal threshold was used. However, for

Table 3 Mean accuracy, specificity, sensitivity, and positive predictive value (PPV) across time intervals and 10 CV-folds for 1-year, 3-year, and 5-year predictions for PLANN and the discrete survival model

Model	Threshold	Accuracy	Specificity	Sensitivity	PPV
One year					
PLANN ^a					
	0.05	0.751	0.762	0.272	0.064
	0.10	0.948	0.986	0.019	0.051
	0.15	–	–	–	–
	0.20	–	–	–	–
	Optimal ^b	0.681	0.683	0.488	0.081
Survival					
	0.05	0.772	0.783	0.360	0.073
	0.10	0.895	0.924	0.186	0.100
	0.15	0.927	0.960	0.128	0.133
	0.20	0.940	0.976	0.083	0.160
	Optimal ^b	0.556	0.560	0.383	0.043
Three year					
PLANN					
	0.05	0.269	0.186	0.822	0.107
	0.10	0.588	0.583	0.443	0.111
	0.15	0.762	0.808	0.257	0.167
	0.20	0.818	0.899	0.096	0.143
	Optimal ^b	0.591	0.591	0.506	0.134
Survival					
	0.05	0.270	0.181	0.791	0.101
	0.10	0.533	0.513	0.483	0.111
	0.15	0.760	0.813	0.218	0.125
	0.20	0.841	0.919	0.128	0.134
	Optimal ^b	0.572	0.582	0.474	0.130
Five year					
PLANN					
	0.05	0.222	0.070	0.934	0.162
	0.10	0.425	0.335	0.709	0.169
	0.15	0.582	0.546	0.565	0.210
	0.20	0.687	0.722	0.360	0.213
	0.607	0.617	0.577	0.230	0.607
Survival					
	0.05	0.264	0.116	0.893	0.161
	0.10	0.410	0.300	0.662	0.143
	0.15	0.548	0.511	0.508	0.160
	0.20	0.709	0.738	0.420	0.238
	Optimal ^b	0.601	0.615	0.515	0.224

^a One-year predictions made by PLANN were all below 0.15 and therefore, thresholds of 0.15 or greater could not be evaluated for accuracy metrics

^b The Optimal rows present the average accuracy metrics across time intervals when the optimal threshold of each time interval is used (see Additional file 1)

3- and 5-year predictions, PLANN and the discrete survival model gave similar results for the optimal thresholds. A trade-off between specificity and sensitivity was

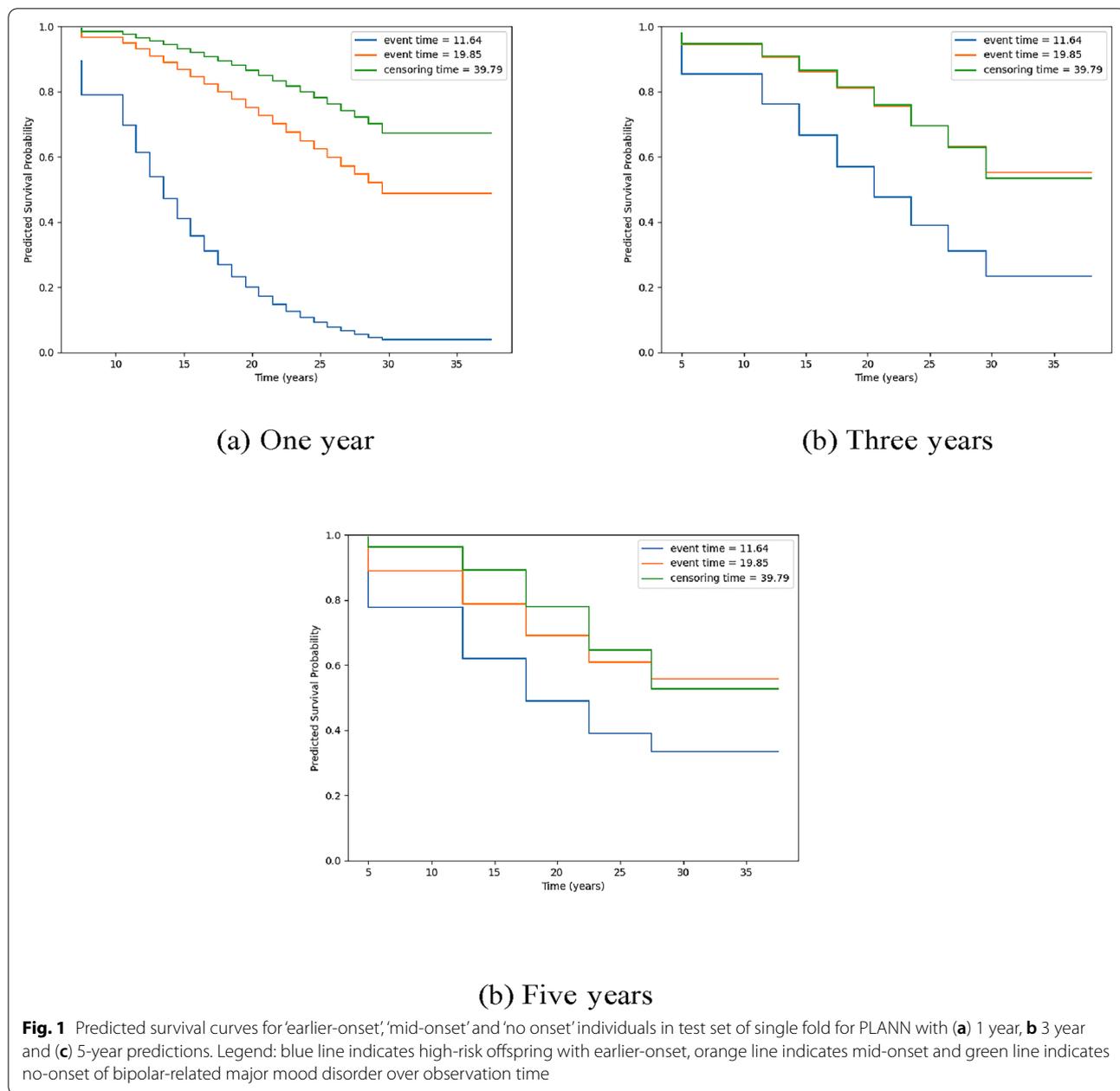
demonstrated for both models across all prediction intervals when examining performance across different thresholds. With lower thresholds, sensitivity is high and specificity is low, whereas the opposite was true for higher thresholds. Optimal thresholds achieve a balance between sensitivity and specificity.

Finally, it was of interest to assess whether the models could distinguish between three individuals in the test set 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 major mood disorder 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 1 year, 3 year and 5-year predictions made by PLANN (Fig. 1). PLANN predicted that the ‘earlier-onset’ individual had the lowest survival probability over time.

When making 1-year predictions of major mood disorder, 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 these individuals, the predicted survival curves were additionally plotted for the discrete survival model with one-year predictions, three-year predictions and five-year predictions. As seen in Fig. 2, the discrete survival 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 1-year, 3-year and 5-year predictions, the discrete survival model predicted that the ‘mid-onset’ individual had a higher probability of not being diagnosed than the ‘no onset’ individual. These findings further demonstrate that PLANN outperforms the discrete survival model in discrimination of higher-risk versus lower-risk offspring.

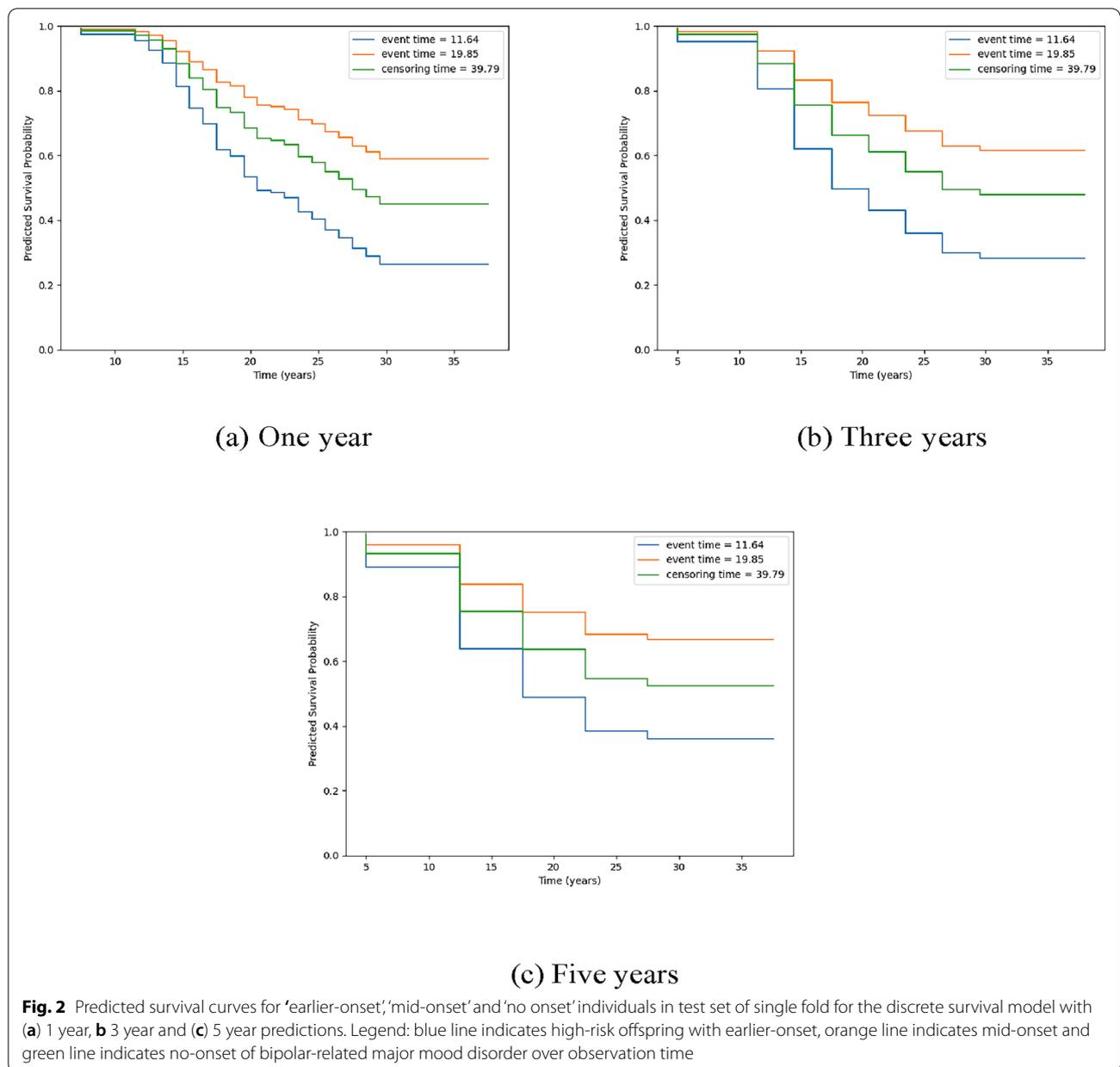
Discussion

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 major mood 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 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). We included major depressive disorder as part of the bipolar-related major mood disorders given (i) major depression is considered part of the bipolar spectrum in genetic studies (McMahon et al. 2010; McGuffin et al. 2003; Coleman et al. 2020) and (ii) bipolar disorder typically debuts as major depression in high-risk offspring of bipolar parents (Duffy et al. 2017; Mesman et al. 2013).

PLANN was compared to the more traditional discrete survival model 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 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, although this was not pursued here. On the other hand, PLANN makes fewer assumptions about the relationship between the outcome and covariates and has the ability to automatically detect non-linear relationships in the data. The latter point is important, as for example, mood instability has been found to follow non-linear patterns (Bonsall et al. 2012). A drawback of PLANN, and machine learning models in general, is that they are computationally intensive, requiring high-performance computer clusters and many days of computation, whereas discrete survival models take only a few seconds to fit.

Overall, for predictive performance, PLANN outperformed the logistic model for 3-year and 5-year predictions. PLANN was better able to discriminate or rank individuals based on their risk of developing major mood disorder (i.e., higher time-dependent c-indices) and better able to predict the probability of developing major mood disorder (i.e., lower Brier scores). The results were mixed for 1-year predictions, with the discrete survival model outperforming PLANN for AUC. For five-year predictions, the average AUC from PLANN was 0.74, which indicates that the model shows good discrimination between high

and low risk individuals. This value is comparable to (Hafeman et al. 2017) who achieved an AUC of 0.76 for 5-year predictions.

Prediction was superior in the 3 and 5-year models compared to the 1-year models, as evidenced by higher AUC, sensitivity and positive predictive values. This is not surprising, as relatively few events occurred in specific 1-year intervals, compared to three or 5 year intervals (see Additional file 1: Tables S2 and S3). Also, it is intuitively more difficult to predict that an event will occur in a 1-year interval, than to say it will happen sometime over the next 5 years.

Individualized risk prediction has important implications for research and clinical practice. By identifying from among individuals at familial risk those most likely to develop illness, an ultra-high risk group can be identified that can inform research into the determinants of illness onset and prevention. Moreover, this ultra-high risk group would be suitable and would likely benefit from prospective surveillance and low-risk intervention and psychoeducation targeting sleep hygiene, healthy coping and stress reduction, healthy lifestyle and diet and avoidance of alcohol misuse and drug use. Our risk prediction approach of using PLANN to predict onset of bipolar-related major mood disorder differs from other published risk calculators such as (Hafeman et al. 2017), which used a “baseline re-setting” Cox proportional hazards model. Both the Cox model and PLANN allow the inclusion of covariates measured at baseline and at follow-up visits and neither method requires an assumption about the distribution of the outcome variable. However, unlike the Cox model, PLANN does not require a proportional hazards assumption. In addition, we only included model variables in PLANN 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. The sample size is small, particularly for neural networks, which typically require sample sizes in the thousands; it is notoriously difficult to make predictions in medicine, due to lack of relevant variables (Lawless 2010) and thus additional breadth of data (e.g., genetic data, behavioural data) may improve predictions.

Conclusion

This evaluation of PLANN is a useful step in the investigation of using neural networks as tools in the prediction of diagnosis of mood disorders 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 major mood disorders in high-risk individuals. 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 research and clinical decision making in individuals at familial risk of developing bipolar-related mood disorders.

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; PLANN: Partial Logistic Artificial Neural Network; SADS-L: Schedule for Affective Disorders–Lifetime.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40345-021-00228-2>.

Additional file 1: Table S1. Example of discrete survival data for hypothetical scenario with sex and blood pressure as covariates. **Figure S1.** Feed-forward neural network with one hidden layer. **Table S2.** PLANN one-, three-, five-year prediction assessments showing optimal threshold, accuracy, specificity (spec), sensitivity (sens), positive predictive value (PPV), total number of individuals in the risk set (n) and total number of events observed within each time interval, determined using tenfold cross-validation. **Table S3.** Discrete survival model one-, three-, five-year prediction assessments showing optimal threshold, accuracy, specificity (spec), sensitivity (sens), positive predictive value (PPV), total number of individuals in the risk set(n) and total number of events observed within each time interval, determined using tenfold cross-validation. **Table S4.** Number of subthreshold diagnoses for individuals in the full sample, individuals who experienced the major mood disorder outcome (diagnosis of BP, MDD and/or schizoaffective disorder), and individuals who did not experience the outcome.

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 and adult offspring 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.

Authors' contributions

JH, AC designed the analysis with input from CK-S, SG and AD. AD is the PI of the high-risk study and contributed with SG to data collection. AC supervised by JH carried out the data analysis and discussed results and interpretation with the other authors listed. AC and JH drafted the manuscript and AD, CK-S and SG edited and revised iterative drafts. All authors read and approved the final 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.

Availability of data and materials

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.

Declarations**Ethics approval and consent to participate**

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 PS1Y-561-17).

Consent for publication

All named authors have provided consent for publication.

Competing interests

All authors confirm that they have no competing or conflicts of interest to declare in relation to this manuscript.

Author details

¹Department of Mathematics and Statistics, University of Guelph, Guelph, ON, Canada. ²Department of Psychiatry, University of Oxford, Oxford, UK. ³4YouandMe, Seattle, USA. ⁴Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. ⁵Department of Psychiatry, Queen's University, Kingston, ON, Canada.

Received: 21 December 2020 Accepted: 27 May 2021

Published online: 01 July 2021

References

- Allison PD. Survival analysis using SAS: a practical guide. Cary: Sas Institute; 2010.
- Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. *Stat Med*. 2005;24(24):3927–44.
- 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.
- Biganzoli E, Boracchi P, Mariani L, Marubini E. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Stat Med*. 1998;17(10):1169–86.
- 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. *Proc R Soc b Biol Sci*. 2012;279(1730):916–24.
- Bourquin J, Schmidli H, van Hoogevest P, Leuenberger H. Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development. *Pharm Dev Technol*. 1997;2(2):95–109.
- Charati JY, Janbabaie G, Alipour N, Mohammadi S, Gholiabad SG, Fendereski A. Survival prediction of gastric cancer patients by Artificial Neural Network model. *Gastroenterol Hepatol Bed to Bench*. 2018;11(2):110.
- Coleman JRI, Gaspar HA, Bryois J, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 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(2):169–84.
- Collett D. Modelling survival data in medical research. Boca Raton: CRC Press; 2015.
- Cox DR. Regression models and life-tables. *J R Stat Soc Ser B (methodol)*. 1972;34(2):187–202.
- Craddock N, Sklar P. Genetics of bipolar disorder. *Lancet*. 2013;381(9878):1654–62.
- Duffy A, Horrocks J, Doucette S, Keown-Stoneman C, McCloskey S, Grof P. The developmental trajectory of bipolar disorder. *Br J Psychiatry*. 2014;204(2):122–8.
- Duffy A, Jones S, Goodday S, Bental R. Candidate risks indicators for bipolar disorder: early intervention opportunities in high-risk youth. *Int J Neuropsychopharmacol*. 2016;19(1):pyy071.
- 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(1):37.
- 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.
- Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *J Am Stat Assoc*. 1988;83(402):414–25.
- Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17–18):2529–45.
- 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 Psychiat*. 2017;74(8):841–7.
- 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.
- LaFaro RJ, Pothula S, Kubal KP, Inchiosa ME, Pothula VM, Yuan SC, et al. Neural network prediction of ICU length of stay following cardiac surgery based on pre-incision variables. *PLoS One*. 2015;10(12):e0145395.
- Lawless JF, Yuan Y. Estimation of prediction error for survival models. *Stat Med*. 2010;29(2):262–74.
- Li H, Luo M, Zheng J, Luo J, Zeng R, Feng N, et al. An artificial neural network prediction model of congenital heart disease based on risk factors: a hospital-based case-control study. *Medicine*. 2017;96(6):e6090.
- Manchia M, Lampus S, Chillotti C, Sardau C, Ardu R, Severino G, et al. Age at onset in Sardinian bipolar I patients: evidence for three subgroups. *Bipolar Disord*. 2008;10(3):443–6.
- McGuffin P, Rijdsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry*. 2003;60(5):497–502.
- McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F, Detera-Wadleigh SD, et al. Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21. 1. *Nat Genetics*. 2010;42(2):128.
- 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.
- Mesman E, Nolen WA, Reichart CG, Wals M, Hillegers MHJ. The Dutch bipolar offspring study: 12-year follow-up. *Am J Psychiatry*. 2013;170(5):542–9.
- 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.
- 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.
- Wunnava S, Qin BIX, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. *Drug Saf*. 2019;42(1):113–22.
- Zhou X-H, Obuchowski NA, McClish DK. Measures of Diagnostic Accuracy. In: *Statistical Methods in Diagnostic Medicine*. Wiley; 2011. p. 13–55. <https://doi.org/10.1002/9780470906514.ch2>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.