Published in final edited form as:

Acta Psychiatr Scand. 2008 April ; 117(4): 244–252. doi:10.1111/j.1600-0447.2008.01162.x.

Evaluation of clinical prognostic models for suicide attempts after a major depressive episode

H. C. Galfalvy, M. A. Oquendo, and J. J. Mann

Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA

Abstract

Objective—In this study, we compare the performance of prognostic models of increasing complexity for prediction of future suicide attempt.

Method—Using data from a 2-year prospective study of 304 depressed subjects, a series of Cox proportional hazard regression models were developed to predict future suicide attempt. The models were evaluated in terms of discrimination (the ability to rank subjects in order of risk), calibration (accuracy of predicted probabilities of attempt), and sensitivity and specificity of risk group stratification based on cross-validated predicted probabilities.

Results—Although an additive model with past attempt, smoking status, and suicidal ideation achieved 75% (cross-validated) sensitivity and specificity, models that performed best in terms of discrimination included interactions between predictor variables.

Conclusion—As several models had similar predictive power, clinical considerations and ease of interpretation may have a significant role in the final stage of model selection for assessing future suicide attempt risk.

Keywords

attempted suicide; prediction; Cox proportional hazards models

Introduction

Suicide attempts are among the most robust predictors of suicide, which is a leading cause of death in young people in several countries (1). In the USA, over 90% of suicides have a diagnosable psychiatric disorder at the time of death, with mood disorders accounting for approximately 60% of all suicides (2, 3). Lifetime suicide attempt rates of 29.2% in bipolar disorders (BD) and 15.9% in major depressive disorder (MDD) are reported in population studies (4). Accurate prognostic models for evaluating the risk of future suicide attempt in psychiatric patients would allow clinicians to identify patients in need of more intensive treatment without prescribing unnecessary treatment for patients less likely to attempt suicide.

Prospective studies have identified multiple significant risk factors for future suicide attempt in mood disorders, including: a past suicide attempt (PSA); substance use disorders; comorbid borderline personality disorder or antisocial personality disorder; a family history of suicidal behavior; greater subjective depression and hopelessness; melancholic depression

subtype; greater trait aggression and impulsivity; a history of childhood abuse; head injury or neurological disorder; and cigarette smoking (5). However, to be clinically useful, prognostic models need to be evaluated based on predictive accuracy and ability to discriminate between risk groups, rather than significance alone (6). For example, when stratifying patients into high- and low-risk groups, model evaluation generally includes some measures of predictive performance familiar from diagnostic models: sensitivity, specificity, positive and negative predictive values and area under the Receiver–Operator Characteristic (ROC) curve (7, 8).

Previous efforts to develop sensitive and specific predictive models for suicide attempts based on clinical predictors used retrospective data and most met with little success, reporting sensitivity between 0% and 38% (9–13). The one exception reported (14) 97% sensitivity and 69% specificity based on an artificial neural network model with 59 predictor variables, suggesting that more complicated models may improve prediction.

Aims of the study

This study aimed to evaluate the predictive power of a number of prognostic models for suicide attempt over the course of a 2-year follow-up, and to identify the most parsimonious model with adequate clinical validity, defined as at least 70% specificity and 70% sensitivity.

Material and methods

Patients

A total of 304 patients presenting for evaluation and treatment of a major depressive episode (MDE) who were diagnosed with MDD or BD based on the Structured Clinical Interview for DSM IV (SCID-I) and who had a minimum baseline Beck Depression Inventory (BDI) score of 10 were enrolled in the study after giving written informed consent as required by the Institutional Review Board. Exclusion criteria were current substance or alcohol abuse, neurological illness or active medical conditions that could confound diagnosis and clinical characterization of psychopathology. At baseline patients were aged 18–75 years (average age 38), 41% were male, 54% reported at least one prior suicide attempt, and 20% were diagnosed with BD. Following discharge, patients received treatment-as-usual in the community and were re-evaluated after 3 months, 1 year, and 2 years. Follow-up interviews included an in-depth assessment of suicidal behavior during the intervening time period. All 304 patients had data available from at least one follow-up visit. Data on these subjects has been previously reported (15, 16).

Instruments and candidate predictor variables

Patients were assessed on a variety of diagnostic and psychiatric instruments as previously described (16). Severity of objective depression was assessed using the clinician-rated 17-item Hamilton Depression Rating Scale (HDRS) and subjective depression using the self-report BDI. Hopelessness was rated with the Beck Hopelessness Scale (BHS), and suicidal ideation at intake and in prior 2 weeks was rated using the Scale for Suicidal Ideation (SSI). The Reasons for Living Inventory (RFLI) assessed possible protective factors. Lifetime aggression was rated using the Brown Goodwin Lifetime Aggression History Scale (BGLAH), hostility using the Buss-Durkee Hostility Inventory (BDHI), and impulsivity using the Barratt Impulsivity Scale (BIS). Life stressors were recorded using the St Paul Ramsey Questionnaire (SPRQ).

Based on the literature, and considering the need for both accuracy and parsimony in a prediction model, 15 candidate predictor variables were selected for model building: age,

gender, diagnosis (BD vs. MDD), co-morbid borderline personality disorders (BPD), history of PSA, smoking, and baseline scores on nine rating scales. Missing baseline data were imputed using predictive mean matching from a flexible additive imputation model (17) implemented in the acepack library of the statistical software R (<http://www.r-project.org/>).

Statistical models

Due to sample attrition over the 2-year study period, Cox proportional hazards regression was used to analyze the time to the first suicide attempt taking censoring into account (18). The proportional hazard assumption was checked for each predictor variable by plotting the scaled Schoenfeld residuals against time (19). No significant departures from the assumption were detected for any of the models discussed herein. First, each predictor variable was entered individually into a Cox model. Then, several Cox models with multiple predictors were built as follows.

In model 1, all 15 variables were entered simultaneously. This model assumes that hazard ratios for individual predictors, when adjusted for the effect of all others, are multiplicative. As predictor variables are not independent from each other, the adjusted hazard ratios will be different from those computed from single-predictor models. Using multiple predictors may improve the predictive accuracy by explaining additional variability. However, large models may overfit the data and lead to poor out-of-sample prediction, as well as violating the criteria of parsimony. Statistical studies assessing the effect of low 'Event-Per-Variable' numbers on the accuracy of parameter estimates, such as odds ratios or hazard ratios, conclude that it can seriously affect the quality of parameter estimates and thus predictive power (20). Variable selection based on likelihood ratio tests balances goodness-of-fit against parsimony and is a widely used method. While there remain questions as to its appropriateness in hypothesis testing contexts, we investigated whether it improves predictive accuracy. In model 2, a stepwise model selection procedure based on a modified version of the Bayes Information Criterion (BIC) (21) was used to reduce the number of predictor variables. This method assures that, regardless of the number of subjects in the study, when the number of suicide attempts is small, fewer variables are selected than would be for a more common event. In model 3, flexibility in the functional form of the predictor variables was introduced into the Cox model by using univariate smoothing splines as substitute predictors for some scale variables. Smoothing splines are piecewise polynomial curves used to model data where the relationship between predictor and response is unknown (not linear). This model allows a better fit and more accurate predictions when quantitative variables have non-linear effects on suicide risk, for example, threshold or saturation effects or a quadratic form where extreme values confer higher risk. To achieve flexibility, smoothing splines can take up a large number of degrees of freedom and can lead to overfitting. For this reason, only two spline predictors were entered into the model, corresponding to BDI and RFLI. These were selected because they demonstrated some nonlinearity on visual inspection of the plots of partial risk function in single-predictor models. In particular, the hazard function appeared to be sensitive to differences in the BDI only for scores between 20 and 40, and did not change for differences in the RFLI scores between 100 and 190. In the interest of parsimony, the remaining terms were entered in a stepwise fashion as in model 2, retaining the two spline terms in the model (model 3).

Models 4 and 5 included interactions between the predictor variables. Interactions in prognostic models denote departures from the multiplicative hazard model where an individual's overall risk is based on the product of the hazard ratio due to each risk factor. For 15 predictors, there are 105 possible second-order interactions and thus a method of model selection is required. We used two forward stepwise selections based on the Akaike Information Criterion (AIC) and the BIC beginning with the full 15-variable model. Model 4

included 40 terms, and model 5 had nine terms. Using two models allowed better assessment of the cumulative effect of complexity in the model on predictive performance measures.

Models based on six initial predictors

Because a small initial number of variables is desirable for clinical purposes, we assessed the loss of predictive power incurred when running all models with only six initial variables. Beginning with all possible six-variable combinations of our predictors (5005 combinations), models were built by stepwise selection of variables and their interactions using the BIC criterion. Each model was evaluated based on discrimination as measured by Harrell's C index, and one of the top five models (labeled model 6, selected based on clinical interest) was further examined to compare model terms to those of the similar-sized model 5 and the theoretically derived model described below.

Statistical model based on the stress-diathesis model of suicide

The stress-diathesis model of suicidal behavior proposes that individuals engage in suicidal behavior when stressors, such as life events or a depressive episode, exceed a threshold set by the individual's diathesis (i.e. lifetime traits such as impulsivity/aggression and a tendency for pessimism) for suicidal acts (22). According to this model, risk is driven not only by individual risk factors, but also by their interactions. To compare the results of data-driven model selection methods (models 2–6) to a theory-based method, we fit a Cox regression model (model 7) where the risk of suicide attempt was modeled as a function of aggression (BGLAH), impulsiveness (BIS) and their interaction with life stress (SPRQ), hopelessness (BHS) and depression severity (HDRS) and their interaction. History of past attempt (PA) was added to the model as an additive factor (six variables, nine terms total). This model is one of many that can be derived from the stress-diathesis principle, and the number of terms it contains is consistent with those from the data-driven models model 5 and 6.

Probability of future suicide attempt

Clinically, the important information is whether a patient is at low or high risk of making an attempt 'in the near future', corresponding to binary prediction. However, censoring must be taken into account. Cox Proportional Hazards regression does not estimate the baseline hazard and thus does not provide a direct estimate of individual conditional survival functions. These were estimated using the non-parametric Breslow estimate (19). Computing the probabilities on the same data used to build the model leads to overly optimistic performance claims. Therefore, we used leave-one-out cross-validation to arrive at a more realistic probability that the subject will make a suicide attempt during the 2 years after the index MDE (day 737). For each subject, prediction was based on a Cox model where coefficients were estimated using data from the remaining subjects with the current subject excluded.

Classification of subjects into risk groups

As the clinical validity of the models is assessed partly in terms of sensitivity and specificity, patients were classified into high- or low-risk groups based on the cross-validated probability of suicide attempt by the end of the follow-up period. The cut-off value for classification reflects the costs of false-negative and false-positive predictions. A cut-off value of 0.5 reflects equal cost, while a cut-off of 0.25 is equivalent to assigning three times as much cost to false negatives as to false positives – a more clinically realistic, albeit arbitrary, assumption in the case of suicide attempt.

Measures of statistical validity

Discrimination, that is, the ability to rank patients in order of risk, was measured by Harrell's C index (23, 24). This measure is related to the area under the ROC curve in the case of binary prediction (logistic regression). The concept has been extended to the survival analysis case by defining the overall C as the concordance probability: the ratio of the number of pairs of subjects where the observations and predictions are concordant, out of all usable pairs. Here, usable pairs of subjects are those with at least one event, and concordance means longer predicted survival time for the subject with longer observed survival time. Models with a concordance probability of 1 have perfect discrimination, 0.5 indicates a model is no better than a coin toss. This measure offers better validation of survival regression models than a simple sensitivity/specificity value as it can accommodate censored observations. Bootstrap estimates of the C were computed using $k = 200$ bootstrap samples.

Calibration, or agreement of predicted probabilities and observed frequencies of events, was measured by the Brier score (25), which gives the mean squared difference between the observed patient outcomes and the corresponding predicted probabilities. Predicted probabilities used in this measure were based on leave-one-out cross-validation and were calculated at the end of the 2-year follow-up period. Thus, it measures validity at one time-point only.

Clinical validity

The criteria for clinical validity used in this study are based on clinician judgment, supported by the findings of past studies described in the introduction.

Results

Suicide attempts during follow-up period

Fifty-two individuals (17%) made suicide attempts in the follow-up period, four of whom died by suicide. Of the 52, 34 made one attempt, the other 18 subjects made up to seven attempts. Only the time to the first attempt was analyzed. Figure 1 shows the non-parametric (Kaplan–Meier) estimate of the survivor curve. Thirty individuals (58% of attempters) reported suicide attempts in the first 6 months following the baseline assessment. There was significant censoring around the 1-year assessment time-point. One hundred and twenty subjects (40%) had left the study before the 2-year follow-up assessment. The length of time subjects stayed in the study was not associated with whether they had an attempt during follow-up, nor was it related to any of the 15 baseline variables, with the exception of a weak positive correlation with the RFLI (Spearman correlation $r = 0.11$, P -value = 0.042).

Risk estimates and predictive power of single predictor models

In the Cox proportional hazard models fitted individually for the 15 candidate variables, ten of the 15 variables were significant risk or protective factors (Table 1). The measures of discrimination (Harrell's C) are poor to mediocre (between 0.51 and 0.67) and are strongly associated with the statistical significance of the respective hazard ratio. Calibration measures (Brier scores) are weakly associated with the respective p values. Sensitivity values are very low even with a low cut-off of 0.25 for the classification into risk groups, with the exception of the PA model (sensitivity = 85%, specificity = 52%). None of the models met our criteria for clinical validity.

Prediction models with multiple predictor variables

Table 2 details predictive performance measures for the five models with multiple predictor variables. Predictive performance measures for model 1 (15-variable Cox Proportional Hazards Regression) showed better discrimination compared with the univariate models (Harrell's $C = 0.76$, sensitivity: 56% at 0.25 cut-off), but no improvement in calibration (Brier score = 0.136). For model 2 (stepwise selection of variables) the reduced Cox model included three variables (history of PSA, smoking status, and baseline suicidal ideation score) and had good predictive properties ($C = 0.76$, Brier score = 0.125). Sensitivity and specificity satisfied our clinical validity criteria (75% each, with 0.25 cut-off). This model correctly classified 39 of 52 follow-up attempters (including three of four suicide completers) and 184 of 252 non-attempters. Correctly classified attempters made follow-up attempts of slightly higher medical severity than incorrectly classified attempters (median lethality = 4 vs. 3, on a 0–8 scale where 0 means no damage and 8 means death, Wilcoxon $W = 174$, $P = 0.092$).

Model 3 included univariate penalized (smoothing) splines. Figure 2 shows spline estimates of the log hazard for some of candidate variables. Spline predictors for the BDI and RFLI were retained and stepwise selection added the same three variables included in model 2 (PSA, smoking status and baseline suicidal ideation). This model did not improve on the discrimination of the more parsimonious model 2 (Harrell's $C = 0.77$, sensitivity = 69%), and seriously under-performed in terms of calibration (Brier score = 0.138 vs. 0.125) (Table 2).

Model 4 with the interactions selected by the AIC criterion had 40 terms and should have been a clear case of over-fitting. However, its predictive validity measures were excellent ($C = 0.90$, Brier score = 0.115) and its cross-validated sensitivity was 71%.

Model 5 was a result of stepwise selection based on the BIC criterion with nine terms based on seven predictor variables. This smaller model did not achieve the performance of the larger model 4 but outperformed model 2 in terms of overall discrimination (Table 2) and met criteria of clinical validity (sensitivity = 71%, specificity 77%). ROC curves based on cross-validated probabilities of suicide attempt for models 1, 4, and 5 are charted on Fig. 3, with the cut-off values (0.25 and 0.5) printed on each curve.

Models based on six initial predictors

The more parsimonious models with six initial variables had mediocre to very good discriminatory power ($C = 0.63$ – 0.81 , data available upon request). Performance generally improved with the number of terms (interactions) included in the final model. There were no obvious outliers in performance, indicating a continuous gradual improvement. However, certain variable combinations occurred more frequently in the top performing model: smoking, history of PSA, age at presentation, and the past attempt*age interaction. Table 3 compares the terms included in one of the top 5 performing models (model 6, $C = 0.81$) and model 5. Older age was protective against future suicidal attempt for those with no history of suicide attempt, but was not protective for those with a history of PAs. Some of the interactions appear to impose a 'ceiling effect' on the predicted risk of suicide, indicating that predicted risk of suicide attempt cannot be increased beyond a certain limit. In model 6 (Table 3) higher subjective depression (BDI) is a risk factor for non-smokers, but smokers are at high risk regardless of BDI score. The interaction between age and life events (SPRQ) indicates that older subjects cope better with life stress than an additive model would predict.

Theoretical model

Model 7 underperformed compared to the similarly sized models 5 and 6 in terms of predictive accuracy as measured by Harrell's C ($C = 0.70$) and Brier score (0.137). Only one of the interactions was statistically significant (BIS and SPRQ), indicating a departure from the additive model with lower than expected risk being conferred by high impulsivity and high life stress. Other than impulsivity and its interaction with life stress, only one term in the model achieved statistical significance, history of PA.

Discussion

Statistical models can be built for three different reasons: estimation, testing and prediction. The purpose of this study was to build models that have satisfactory predictive accuracy. There is a clinical value for such models irrespective of their ability to provide or sustain scientific insight although a review of the best models provides novel ideas, for example the role of ceiling effect, for future study.

There is a consensus that (scientific) theory-based statistical models are superior to data-driven models (built for example by stepwise methods) for purposes of estimation and hypothesis testing. However, data-driven methods can be useful for discovering possible predictors, or for building predictive models, even so-called black-box models. This is especially true in situations where there are gaps in scientific knowledge and/or outcomes and measurement methods are too noisy; as long as due care is taken to avoid overfitting the data. This study demonstrates that even relatively complex theoretical models can severely underperform similar-sized models generated by data-driven selection techniques, indicating our as yet imperfect understanding of suicidal behavior and the difficulty of translating theory into a sensitive and specific statistical model.

The models ranged from basic models of risk based on a small number of predictors to more advanced models using many terms and complex functional forms, making them difficult, if not impossible to interpret from an explanatory point of view or in terms of pathogenesis or causality. As a comparison, a statistical model based on the stress-diathesis model of suicidal behavior, which is an explanatory rather than a predictive model, was also included in the analysis. In a departure from other prediction studies, both the modeling and evaluation stage account for censored data due to early drop-out.

There are several obstacles to building more complex models for the prediction of suicidal behavior, the most serious of which is the disparity between sample size (or, more importantly, the number of suicide attempts in the sample) and the number of potential predictor variables. Sample sizes for homogeneous patient groups, where many measures have been obtained, are relatively low and, consequently, suicide attempts are few while there are many candidate predictor variables. Statistical studies assessing the effect of low event-per-variable numbers on the accuracy of parameter estimates such as odds ratios or hazard ratios conclude that it can seriously affect the quality of parameter estimates and thus predictive power (20). In this study, Cox proportional models with a single predictor had generally poor predictive performance. However, discrimination and accuracy improved with the addition of more predictor variables and/or interactions. This increased predictive power is likely due to the inclusion of complementary information from multiple variables and to the more complex functional form of these models. This finding is in agreement with previously published studies that show that risk factors can modify each other's effects (15, 13, 26–31), and supports the need for complex theoretical models of suicidal behavior (22) that include multiple factors that mediate and moderate each other's effect.

In the interest of clinical utility more parsimonious prediction models are desirable. Reducing the number of variables used for prediction with a stepwise selection method can lead to biased estimates in the final model and a predicted risk that is either too high or low (32). In this study, in terms of predictive performance the model with 3 predictor variables (model 2) constructed using stepwise selection outperformed the full linear model. Moreover, although for reasons of model size (the large number of possible interactions) we do not have a basis of comparison for the stepwise selection of interactions, this method resulted in the two best performing models of the study. This indicates that, with a relatively modest sample size and event-per-variable ratio, using the standard stepwise selection methods to reduce model size or introduce interactions may improve prediction.

The comparability of the top performing models is likely due to the high degree of correlation between predictor variables and the modest sample size. Considering the low event number to variable ratio, it is expected that different variable combinations will provide comparable information. However, there are important predictor variable combinations that have a higher likelihood of inclusion in a good model, in this case: history of suicide attempt, smoking, and age. History of suicide attempt is a well-documented risk factor for future attempts, and smoking is associated with suicidal ideation, suicide attempt, and suicide (33). There is evidence that suicide attempt rates are higher in young people (34), although rates of suicide increase with age in males (35). One possible interpretation is that older individuals who have not attempted suicide in the past have a higher stress threshold for the diathesis for suicidal behavior to be triggered, and thus are less likely to make a suicide attempt in the face of stress. Other variable combinations that complete these models appear to be interchangeable.

Some of the interactions between predictive variables appeared to impose a 'ceiling effect' on the predicted risk of suicide. It is an interesting question whether this is due to correlated predictors, for example subjects at high risk of suicide attempt have several risk-associated characteristics and their effect is essentially non-additive, or a limitation imposed by the amount of suicide-risk-related information in the dataset. The second explanation poses a further question: would inclusion of more relevant or more up-to-date (longitudinally measured) data in the model improve predictive performance or is there an amount of risk of suicidal behavior that is essentially random and impossible to predict. We plan to explore this question in a future study by including longitudinal information on state variables and possible stressors in prognostic models.

Ultimately, as our exploration of the six-variable-based models demonstrated, for this cohort there is no 'best' prognostic model in terms of statistical validity for determining the risk of future suicide attempts. Rather, several models provide comparable estimates of discrimination and calibration. This does not mean that all these models will be equally useful clinically. Indeed, the best model may be the one that requires the least and most readily obtainable data, minimizing patient and clinician burden. A model with 40 terms that is cumbersome to build and requires 15 variables from the outset is clearly not an optimal choice for clinical purposes. A more appropriate model might be the nine-term model where interactions were selected by the BIC approach based on seven initial variables (model 5), or even the parsimonious model (model 2) with no interaction terms and only three variables. Both of these models met our clinical validity criteria. If the primary clinical goal, is to classify patients into high- or low-risk groups, model 2 with history of PSA, smoking status, and baseline suicidal ideation score, and multiplicative hazard ratios provides adequate classification of follow-up attempters and non-attempters, even if this model is slightly outperformed by more complicated models on overall discrimination.

Acknowledgments

This study was supported by NIMH grants MH62185, MH074068 and the Stanley Medical Research Institute. The authors are grateful to Dr Dianne Currier for her advice and suggestions during the preparation of this manuscript.

References

1. WHO. Suicide. World Health Report 2001. 2001. Available at: <http://www.who.int/whr/2001/chapter2/en/index6.html>
2. Cavanagh JT, Carson AJ, Sharpe M, et al. Psychological autopsy studies of suicide: a systematic review. *Psychol Med*. 2003; 33:395–405. [PubMed: 12701661]
3. Bertolote JM, Fleischmann A, De LD, et al. Psychiatric diagnoses and suicide: revisiting the evidence. *Crisis*. 2004; 25:147–155. [PubMed: 15580849]
4. Chen YW, Dilsaver SC. Lifetime rates of suicide attempts among subjects with bipolar and unipolar disorders relative to subjects with other Axis I disorders. *Biol Psychiatry*. 1996; 39:896–899. [PubMed: 8860192]
5. Oquendo MA, Currier D, Mann JJ. Prospective studies of suicidal behavior in major depressive and bipolar disorders: what is the evidence for predictive risk factors? *Acta Psychiatr Scand*. 2006; 114:151–158. [PubMed: 16889585]
6. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000; 19:453–473. [PubMed: 10694730]
7. Loong TW. Understanding sensitivity and specificity with the right side of the brain. *BMJ*. 2003; 327:716–719. [PubMed: 14512479]
8. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978; 8:283–298. [PubMed: 112681]
9. Cohen J. Statistical approaches to suicidal risk factor analysis. *Ann N Y Acad Sci*. 1986; 487:34–41. [PubMed: 3471164]
10. Pokorny AD. Prediction of suicide in psychiatric patients. Report of a prospective study. *Arch Gen Psychiatry*. 1983; 40:249–257. [PubMed: 6830404]
11. Pokorny AD. Suicide prediction revisited. *Suicide Life Threat Behav*. 1993; 23:1–10. [PubMed: 8475527]
12. Goldstein RB, Black DW, Nasrallah A, et al. The prediction of suicide: sensitivity, specificity, and predictive value of a multivariate model applied to suicide among 1906 patients with affective disorders. *Arch Gen Psychiatry*. 1991; 48:418–422. [PubMed: 2021294]
13. Mann JJ, Currier D, Stanley B, et al. Can biological tests assist prediction of suicide in mood disorders? *Int J Neuropsychopharmacol*. 2006; 9:465–474. [PubMed: 15967058]
14. Modai I, Kurs R, Ritsner M, et al. Neural network identification of high-risk suicide patients. *Med Inform Internet Med*. 2002; 27:39–47. [PubMed: 12509122]
15. Oquendo M, Bongiovi-Garcia ME, Galfalvy H, et al. Sex differences in clinical predictors of suicidal acts after major depression: a prospective study. *Am J Psychiatry*. 2007; 164:134–141. [PubMed: 17202555]
16. Oquendo MA, Galfalvy H, Russo S, et al. Prospective study of clinical predictors of suicidal acts after a major depressive episode in patients with major depressive disorder or bipolar disorder. *Am J Psychiatry*. 2004; 161:1433–1441. [PubMed: 15285970]
17. Little R, An HG. Robust likelihood-based analysis of multivariate data with missing values. *Statistica Sinica*. 2004; 14:949–968.
18. Leon AC, Friedman RA, Sweeney JA, et al. Statistical issues in the identification of risk factors for suicidal behavior: the application of survival analysis. *Psychiatry Res*. 1990; 31:99–108. [PubMed: 2315425]
19. Therneau, TM.; Grambsch, PM. Modeling survival data: extending the Cox model. New York Berlin Heidelberg: Springer Verlag; 2000.
20. Peduzzi P, Concato J, Feinstein AR, et al. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995; 48:1503–1510. [PubMed: 8543964]

21. Volinsky CT, Raftery AE. Bayesian information criterion for censored survival models. *Biometrics*. 2000; 56:256–262. [PubMed: 10783804]
22. Mann JJ, Waternaux C, Haas GL, et al. Toward a clinical model of suicidal behavior in psychiatric patients. *Am J Psychiatry*. 1999; 156:181–189. [PubMed: 9989552]
23. Harrell FE Jr, Lee KL, Califf RM, et al. Regression modelling strategies for improved prognostic prediction. *Stat Med*. 1984; 3:143–152. [PubMed: 6463451]
24. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. 2004; 23:2109–2123. [PubMed: 15211606]
25. Graf E, Schmoor C, Sauerbrei W, et al. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999; 18:2529–2545. [PubMed: 10474158]
26. Amsel, L.; Mann, JJ.; van Heeringen, K. Understanding suicidal behaviour: the suicidal process approach to research and treatment. Sussex: John Wiley and Sons, Ltd; 2001. Suicide risk assessment and the suicidal process approach; p. 163-181.
27. Young MA, Fogg LF, Scheftner WA, et al. Interactions of risk factors in predicting suicide. *Am J Psychiatry*. 1994; 151:434–435. [PubMed: 8109656]
28. Dieserud G, Roysamb E, Ekeberg O, et al. Toward an integrative model of suicide attempt: a cognitive psychological approach. *Suicide Life Threat Behav*. 2001; 31:153–168. [PubMed: 11459248]
29. Jeglic EL, Sharp IR, Chapman JE, et al. History of family suicide behaviors and negative problem solving in multiple suicide attempters. *Arch Suicide Res*. 2005; 9:135–146. [PubMed: 16020157]
30. Hovey JD. Moderating influence of social support on suicidal ideation in a sample of Mexican immigrants. *Psychol Rep*. 1999; 85:78–79. [PubMed: 10575976]
31. Gould MS, Fisher P, Parides M, et al. Psychosocial risk factors of child and adolescent completed suicide. *Arch Gen Psychiatry*. 1996; 53:1155–1162. [PubMed: 8956682]
32. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996; 15:361–387. [PubMed: 8668867]
33. Oquendo MA, Currier D, Mann JJ. Prospective studies of suicidal behavior in major depressive and bipolar disorders: what is the evidence for predictive risk factors? *Acta Psychiatr Scand*. 2006; 114:151–158. [PubMed: 16889585]
34. De LD, Cerin E, Spathonis K, et al. Lifetime risk of suicide ideation and attempts in an Australian community: prevalence, suicidal process, and help-seeking behaviour. *J Affect Disord*. 2005; 86:215–224. [PubMed: 15935241]
35. Center for Disease Control. CDC recommendations for a community plan for the prevention and containment of suicide clusters. *MMWR*. 1988; 37:1–12.

Significant outcomes

- For prognostic models of suicide attempts, model selection needs to combine the usual goodness-of-fit criteria with measures of predictive validity.
- Suicide attempts in the first two years after a major depressive episode can be predicted with acceptable (above 70%) sensitivity and specificity.
- Including interactions between predictors using a stepwise selection process can improve the discriminative performance of the prognostic model.

Limitations

- The 15 initial predictor variables were selected based on the literature, models may be outperformed if other predictor variables were included.
- The sample did not include subjects with current substance abuse. In addition, our sample is representative of the high-risk patients of a psychiatric research center and our results may not generalize directly to patients presenting with a milder major depressive episode to their primary care provider.
- Prognostic models are most reliable when externally validated on a new dataset. We used cross-validation and the bootstrap which are forms of internal validation.

Fig. 1. Non-parametric estimate of the survivor curve for time to first suicide attempt in the 2-year follow-up period. Dashed lines denote the 95% confidence interval.

Fig. 2. Univariate penalized spline estimates of the log hazard function for four predictor variables. Dashed lines denote point-wise standard errors, 1-dimensional histograms of the predictor values are added to the bottom of each plot.

Fig. 3.

ROC curves based on cross-validated probabilities of suicide attempt at the end of the follow-up period for the full linear model (Model 1, 15 terms) and the models with interactions selected by the AIC (Model 4, 40 terms) and BIC (Model 5, 12 terms) methods. The cut-off values of 0.25 and 0.5 for the classification have been marked on each curve.

Table 1

Summaries of candidate predictor variables and their effect on risk of future suicide attempt. Cox proportional hazard regression was used to model the risk of suicide attempt separately for each candidate predictor. Models were evaluated based on calibration and discrimination. $N = 304$ with $n = 52$ suicide attempters

	Mean $\pm$ SD (or % yes)	Hazard ratio* (P-value)	Brier score [†]	Discrimination		
				Harrell's C statistic [‡]	Sensitivity (%; $P = 0.25$) [§]	Specificity (%; $P = 0.25$) [§]
Age (years)	38 $\pm$ 12	0.97 (0.0130)	0.1398	0.59	33	79
Male	41 %	0.60 (0.0520)	0.1405	0.56	0	100
Smoker (yes/no)	39 %	2.80 (0.0003)	0.1365	0.63	62	65
Past attempt	54 %	5.14 (0.0001)	0.1312	0.66	85	52
Bipolar disorder	20 %	1.17 (0.6400)	0.1426	0.51	0	100
Borderline Personality Disorder (BPD)	32 %	2.72 (0.0003)	0.1373	0.62	52	73
Hamilton Depression Rating Scale (HDRS)	20 $\pm$ 6	1.03 (0.2300)	0.1421	0.55	4	94
Beck Depression Inventory (BDI)	29 $\pm$ 10	1.04 (0.0038)	0.1386	0.63	31	83
Beck Hopelessness Scale (BHS)	12 $\pm$ 6	1.01 (0.6300)	0.1424	0.51	0	100
Scale for Suicidal Ideation (SSI)	14 $\pm$ 11	1.06 (0.0001)	0.1360	0.67	54	75
Reasons for Living Inventory (RFLI)	154 $\pm$ 44	0.99 (0.0016)	0.1396	0.62	35	79
Aggression History (BGLAH)	18 $\pm$ 6	1.05 (0.0170)	0.1413	0.60	25	83
Barratt Impulsivity Scale (BIS)	52 $\pm$ 17	1.01 (0.1500)	0.1414	0.56	10	90
Buss-Durkee Hostility Index (BDHI)	36 $\pm$ 12	1.03 (0.0064)	0.1386	0.61	33	82
St. Paul-Ramsey (SPRQ)	4 $\pm$ 1	1.10 (0.3400)	0.1425	0.53	0	98

* For quantitative variables, HR shown is per 1 point increase.

[†] The smaller the score the closer the predictions are to observed values. Predictions were calculated using leave-one-out cross-validation and refer to the end of the 2-year period.

[‡] Corresponds to the area under the ROC curve, but takes censoring into account.

[§] Risk groups defined with probability cut-off of 0.25 using leave-one-out cross-validated predicted probabilities of attempt at the end of the 2-year period.

Measures of predictive accuracy for five candidate survival regression models. All predictions were computed by cross-validation, Harrell's C statistic by bootstrap

Table 2

	Brier score	Harrell's C statistic	Sensitivity [‡]				Specificity [‡]			
			0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25
Model 1 (15 terms [*])	0.136	0.76	25%	56%	90%	75%	90%	75%	75%	75%
Model 2 (3 terms)	0.125	0.76	27%	75%	92%	75%	92%	75%	75%	75%
Model 3 (df [†] = 9)	0.138	0.77	25%	69%	90%	76%	90%	76%	76%	76%
Model 4 (40 terms)	0.115	0.90	63%	71%	91%	80%	91%	80%	80%	80%
Model 5 (9 terms)	0.125	0.81	31%	71%	92%	77%	92%	77%	77%	77%
Model 6 (10 terms)	0.122	0.81	27%	69%	89%	74%	89%	74%	74%	74%
Model 7 (9 terms)	0.137	0.70	6%	58%	97%	67%	97%	67%	67%	67%

^{*}The intercept was not counted as a term although it was included in each model.

[†]Degrees of freedom, corresponds to the number of terms. In addition to two 3-df spline predictors based on BDI and RFLI, this model also contains the three terms from model 2.

[‡]Two sets of high-low risk groups defined with probability cutoff of 0.25 and 0.5 respectively.

Table 3

Estimates of the hazard ratios from three Cox regression models^{*}

Model 5 (C = 0.81)		Model 6 (C = 0.81)		Model 7 (C = 0.70)	
Terms	HR [†]	Terms	HR [†]	Terms	HR [†]
PA	5.74	PA	5.17	PA	4.8
Smoker	2.99	Smoker	3.32	BIS	1.05
Age	0.88	Age	0.90	SPRQ	0.96
PA × age	1.14	PA × age	1.12	BIS × SPRQ	0.98
Male	0.45	BDI	1.08	BGAH	0.89
SSI	1.03	SSI	1.04	BGAH × SPRQ	1.04
BDHI	1.00	SPRQ	1.14	BHS	1.01
Bipolar	0.56	Smoker × BDI	0.93	HDRS17	1.02
Bipolar × BDHI	1.09	BDI × SPRQ	0.98	BHS × HDRS17	1.01
		Age × SPRQ	0.98		

PA, past attempt; BIS, Barratt Impulsivity Scale; SPRQ, St Paul Ramsey Questionnaire; BDI, Beck Depression Inventory; BHS, Beck Hopelessness Scale; HDRS, Hamilton Depression Rating Scale; BDHI, Buss-Durkee Hostility Inventory; SSI, Scale for Suicidal Ideation.

^{*} Quantitative variables were centered, because interactions were present.

[†] For quantitative variables HR is estimated for 1 point increase.