

## Original Article

**Cite this article:** Machado CdosS, Ballester PL, Cao B, Mwangi B, Caldieraro MA, Kapczinski F, Passos IC (2022). Prediction of suicide attempts in a prospective cohort study with a nationally representative sample of the US population. *Psychological Medicine* 52, 2985–2996. <https://doi.org/10.1017/S0033291720004997>

Received: 13 September 2020

Revised: 27 November 2020

Accepted: 1 December 2020

First published online: 14 January 2021


### Key words:

Suicide; machine learning; prediction; NESARC; depression

### Author for correspondence:

Ives Cavalcante Passos,  
E-mail: [ivescp1@gmail.com](mailto:ivescp1@gmail.com)

# Prediction of suicide attempts in a prospective cohort study with a nationally representative sample of the US population

Cristiane dos Santos Machado<sup>1,2</sup> , Pedro L. Ballester<sup>3</sup>, Bo Cao<sup>4</sup>, Benson Mwangi<sup>5</sup>, Marco Antonio Caldieraro<sup>1,2</sup>, Flávio Kapczinski<sup>1,2,6</sup> and Ives Cavalcante Passos<sup>1,2</sup>

<sup>1</sup>Laboratory of Molecular Psychiatry, Centro de Pesquisa Experimental (CPE) e Centro de Pesquisa Clínica (CPC), Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, RS, Brazil; <sup>2</sup>Department of Psychiatry, Faculty of Medicine, Graduate Program in Psychiatry and Behavioral Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; <sup>3</sup>Neuroscience Graduate Program, McMaster University, Hamilton, ON, Canada; <sup>4</sup>Department of Psychiatry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada; <sup>5</sup>Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center at Houston, Houston, Texas, USA and <sup>6</sup>Department of Psychiatry and Behavioural Neurosciences, McMaster University and St. Joseph's Healthcare Hamilton, Hamilton, ON, Canada

## Abstract

**Background.** There is still little knowledge of objective suicide risk stratification.

**Methods.** This study aims to develop models using machine-learning approaches to predict suicide attempt (1) among survey participants in a nationally representative sample and (2) among participants with lifetime major depressive episodes. We used a cohort called the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) that was conducted in two waves and included a nationally representative sample of the adult population in the United States. Wave 1 involved 43 093 respondents and wave 2 involved 34 653 completed face-to-face reinterviews with wave 1 participants. Predictor variables included clinical, stressful life events, and sociodemographic variables from wave 1; outcome included suicide attempt between wave 1 and wave 2.

**Results.** The model built with elastic net regularization distinguished individuals who had attempted suicide from those who had not with an area under the ROC curve (AUC) of 0.89, balanced accuracy 81.86%, specificity 89.22%, and sensitivity 74.51% for the general population. For participants with lifetime major depressive episodes, AUC was 0.89, balanced accuracy 81.64%, specificity 85.86%, and sensitivity 77.42%. The most important predictor variables were a diagnosis of borderline personality disorder, post-traumatic stress disorder, and being of Asian descent for the model in all participants; and previous suicide attempt, borderline personality disorder, and overnight stay in hospital because of depressive symptoms for the model in participants with lifetime major depressive episodes. Random forest and artificial neural networks had similar performance.

**Conclusions.** Risk for suicide attempt can be estimated with high accuracy.

## Introduction

About 800 000 people die by suicide every year making suicide the 15 leading cause of death worldwide according to the World Health Organization (2014), and the second among 15–29 year-olds (WHO, 2018). In the United States, suicide rates increased from 1999 through 2017, and the age-adjusted suicide rate was 33% higher in 2017 than in 1999 (Hedegaard, Curtin, & Warner, 2018). Despite these findings, there is still little awareness in medical practice of objective suicide risk stratification, which has led to suicide being referred to as ‘the quiet epidemic’ (Turecki, 2014).

A growing body of knowledge has put forward several sociodemographic and clinical risk factors associated with individuals who attempt suicide (Borges et al., 2007, 2010; Nock et al., 2008). For instance, gender, age, race, marital status, education, income, prior suicide attempt, stressful life events, and body mass index (BMI) are all variables associated with suicide attempts (Borges et al., 2007, 2010; Heikkinen, Aro, & Lönnqvist, 1992; Johnston, Pirkis, & Burgess, 2009; Nock et al., 2008; Oquendo et al., 2014; Perera et al., 2016; Zhang, Yan, Li, & McKeown, 2013). Additionally, retrospective studies with psychological autopsies have shown that 90% of the subjects who died by suicide had a psychiatric disorder, including major depressive disorder, substance-related disorders, and/or personality disorders (Arsenault-Lapierre, Kim, & Turecki, 2004). These efforts have largely reported average group-level differences between suicide attempters and non-attempters. However, what was not

known until recently is how to integrate these variables to build models to estimate the probability of an individual attempting suicide. Importantly, this problem should be approached with caution, focusing on generating models that can generalize well for future instances and can create proper sparse representations to reduce data collection efforts. This is an important question because suicide is a highly preventable event (Zalsman et al., 2016). It is known that interventions such as cognitive behavior therapy (Morey, Lowmaster, & Hopwood, 2010), and lithium (Cipriani, Hawton, Stockton, & Geddes, 2013) can significantly reduce suicide attempts.

Over the past 5 years, our group and others started to build machine-learning models to predict suicide attempts (Belsher et al., 2019; Kessler et al., 2015; Passos et al., 2016; Walsh, Ribeiro, & Franklin, 2017). However, these studies had three limitations. First, most studies had only a few months of follow-up or relied on a retrospective (Choi, Lee, Yoon, Won, & Kim, 2018) or cross-sectional design (Borges et al., 2010). Second, some of the studies aimed to build suicide prediction models within the general population (Borges et al., 2007, 2010), but they did not comprise nationally representative samples, which may have biased their findings. Third, some of the studies had a small sample size (Galfalvy, Oquendo, & Mann, 2008; Passos et al., 2016). It has also been stated recently that future studies should address specific populations with higher rates for suicide attempts, such as individuals with depressive episodes (Passos & Ballester, 2019).

The current study, therefore, aims to develop models to predict suicide attempts in the general population (aim 1) and in participants with lifetime major depressive episodes (aim 2) by using machine-learning techniques coupled with sociodemographic and clinical data. To address the limitations of previous studies, we used a nationally representative cohort publicly available by request with 43 093 participants and a follow-up period of 3 years (Hasin & Grant, 2015). Of note, we used easily accessible clinical variables to achieve our aims.

## Methods

### Data collection, study design, and participants

We used sociodemographic, clinical, and stressful life events data from a large 3-year follow-up study called the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) (National Institute on Alcohol Abuse and Alcoholism, 2006). NESARC was collected in two waves. Wave 1 was conducted in 2001–2002 and surveyed a representative sample of the adult population of the United States, oversampling black people, Hispanic individuals, and young adults aged 18–24 years. The target population was the civilian non-institutionalized population, 18 years and older, residing in households and group quarters. Face-to-face interviews were conducted with 43 093 respondents, yielding an overall response rate of 81%. Weighted data were adjusted to be representative of the civilian population of the United States on socioeconomic variables based on the 2000 Decennial Census. The mean interval between wave 1 and wave 2 interviews was 36.6 (s.e. = 2.62) months. Wave 2 of the NESARC was conducted in 2004–2005 and involved face-to-face reinterviews with all participants in the wave 1 interview. Excluding respondents ineligible for the wave 2 interview because they were deceased ( $n = 1403$ ), deported, mentally or physically impaired ( $n = 781$ ), or on active duty in the armed forces throughout the follow-up period ( $n = 950$ ), the wave

2 response rate was 86.7%, reflecting 34 653 completed interviews. The cumulative response rate at wave 2 was the product of the wave 2 and wave 1 response rates, or 70.2%. The mean interval between wave 1 and wave 2 interviews was 36.6 (s.e. = 2.62) months. Wave 2 NESARC data were weighted to reflect design characteristics of the NESARC and account for oversampling. More information about NESARC can be found elsewhere (Hasin & Grant, 2015; National Institute on Alcohol Abuse and Alcoholism, 2006).

All potential NESARC respondents were informed in writing about the nature of the survey, the statistical uses of the data to be collected, the voluntary nature of their participation, and the federal laws that rigorously provide for the confidentiality of identifiable survey information. Only respondents consenting to participate after securing this information were interviewed. The research protocol for the initial NESARC survey and the follow-up survey (wave 2), including informed consent procedures, received full-ethical review and approval from the US Census Bureau and the Office of Management and Budget.

### Assessments

The Alcohol Use Disorder and Associated Disabilities Schedule – *Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition* (AUDADIS-IV) was used (Hasin & Grant, 2015). AUDADIS-IV is a fully structured diagnostic interview designed to assess alcohol, drug, and mental disorders according to DSM-IV diagnostic criteria in both clinical and general populations, with good to excellent reliability for most variables shown in test–retest studies (Hasin & Grant, 2015).

### Specific aims

Aim 1 was to build a tool for predicting future suicide attempt in the general population that would be able to objectively stratify the risk at an individual level. To achieve this, we built machine-learning models by using easily accessible predictor variables from wave 1. The outcome was attempted suicide in the follow-up period and this was assessed in wave 2, approximately 3 years later.

Aim 2 was to investigate whether a specific predictive clinical signature derived from a sample of this population, with lifetime major depressive episodes, could be created using a similar approach.

### Selection of predictor variables

Selection of predictor variables to be utilized in ‘training’ an algorithm is a challenge in machine learning. However, a recommended method of selecting relevant predictor variables is to use expert domain knowledge – largely from previously published literature (Passos et al., 2019). We selected predictor variables using *a priori* knowledge, through hypothesis-driven approaches. It is worth mentioning that these variables were decided *a priori* and approved by the US Census Bureau before the analysis.

Predictor variables comprised of psychiatric diagnoses [alcohol and drug use disorders, panic disorder, generalized anxiety disorder, specific phobia, social phobia, post-traumatic stress disorder (PTSD), major depressive disorder, dysthymic disorder, bipolar disorder, schizophrenia, and personality disorders]; stressful life events in the past 12 months (e.g. death of a family member or a close friend, being fired or laid off from a job, getting separated or divorced, being a victim of any type of crime);

sociodemographic variables (age, gender, race, marital status, education, income, being raised by biological parents or not); and BMI. Additional details on variables used are provided in the online Supplementary methods. Notably, the majority of variables selected were related to psychiatric comorbidities, given that most individuals who attempt suicide are affected by a psychiatric disorder (Hoertel et al., 2015; Nock, Hwang, Sampson, & Kessler, 2010). Recent findings have demonstrated that the effects of mental disorders on suicide risk can be exerted almost exclusively through a general psychopathology factor representing the shared effect across all mental disorders (Hoertel et al., 2015). In addition, all selected sociodemographic variables were associated with suicide attempts in previous studies (Borges et al., 2007, 2010; Heikkinen et al., 1992; Johnston et al., 2009; Nock et al., 2008; Oquendo et al., 2014; Perera et al., 2016; Zhang et al., 2013), as well as being raised by biological parents (Borczykowski, Hjern, Lindblad, & Vinnerljung, 2006; Keyes, Malone, Sharma, Iacono, & McGue, 2013; Slap, Goodman, & Huang, 2001), and BMI variables (Perera et al., 2016; Zhang et al., 2013). Suicidal crises are typically triggered by recent life events (Turecki & Brent, 2015), but how stressful events interact with individual susceptibility to suicidal behavior or trait-like diathesis is as yet unclear (Van Heeringen & Mann, 2014). Moreover, the specific nature of stressful life events can impact an individual in different ways (Oquendo et al., 2014) and a greater understanding of this phenomenon is required.

For aim 2, besides the predicting variables used in the first aim, we included another four predictor variables assessed only in participants with lifetime major depressive episodes: prior hospitalization because of depressive symptoms, past-suicide attempts, age at onset of first episode of major depression, and suicidal ideation (Holma et al., 2010; Isometsä, 2014; Oquendo et al., 2004; Schaffer et al., 2014; Tondo, Lepri, & Baldessarini, 2007).

### Statistical analysis

Descriptive analyses were reported as means (with standard deviations) or absolute and relative frequencies. We divided participants into two groups based on the outcome (participants who attempted suicide *v.* participants who did not between wave 1 and wave 2) for each aim, and we used chi-squared ( $\chi^2$ ) or Student's *t* tests to analyze sociodemographic and clinical variables among these groups.

The statistical summaries reported in this document have been cleared by the US Census Bureau's Disclosure Review Board release authorization number CBDRB-FY20-094.

### Machine-learning analysis

We used R software (Version R 3.3.1), RStudio (Version 0.99.902), and the following packages: *caret*, *glmnet*, *randomForest*, and *nnet* for this step (Kuhn, 2008). Machine-learning approaches are usually superior to traditional multiple regression analyses, especially in contexts where coefficients would be unstable due to high correlations of predictors (Zou & Hastie, 2005). The elastic net is a machine-learning method that uses regularization with an embedded feature selection procedure. Through a cost function composed of both L1 (least absolute shrinkage and selection operator, i.e. Lasso regression) and L2 (ridge regression) weight magnitude penalties, the method can remove predictors with low impact on the outcome while regularizing for improved generalization. The coefficients of features less predictive to the outcomes shrunk toward zero simplifying the model, and reducing

overfitting. As our dataset is composed of several attributes, identifying the most important of these enables wider applicability and more practical use of our predictive models.

As supplementary analysis, we also built models with two other machine-learning models called random forest and artificial neural networks (ANNs), because they can analyze complex relationships between variables, including nonlinear patterns (Passos et al., 2019). Random forest (or decision tree forests) is an ensemble-based method that builds multiple decision trees (Breiman, 2001). The method combines the base principles of 'bagging' with random feature selection to add additional diversity to the decision tree models. ANNs model the relationship between a set of input and output signals using a model derived from our understanding of how a biological brain responds to stimuli from sensory inputs (Cross, Harrison, & Kennedy, 1995). We only used ANNs with a single hidden layer.

To build the model, we randomly split the dataset into two parts: (1) a training dataset with 75% of the whole sample and (2) test datasets with 25% of the sample. We removed all instances with missing data. After this, we used a standard machine-learning protocol with 10-fold cross-validation, hyperparameter tuning, and class imbalance correction in the training dataset (Fig. 1).

### Class imbalance

Class imbalance introduces a bias toward classifying all the data as the majority class (i.e. did not attempt suicide in the current study), which usually leads to poor detection of the infrequent class. For the elastic net model, we implemented a class weighting technique instead of under-sampling. Each instance of the dataset was reweighted according to the inverse of the frequency of their class, as follows:

$$w_i = c_i \times \frac{p(n)}{p(y) + p(n)} + (1 - c_i) \times \frac{p(y)}{p(y) + p(n)},$$

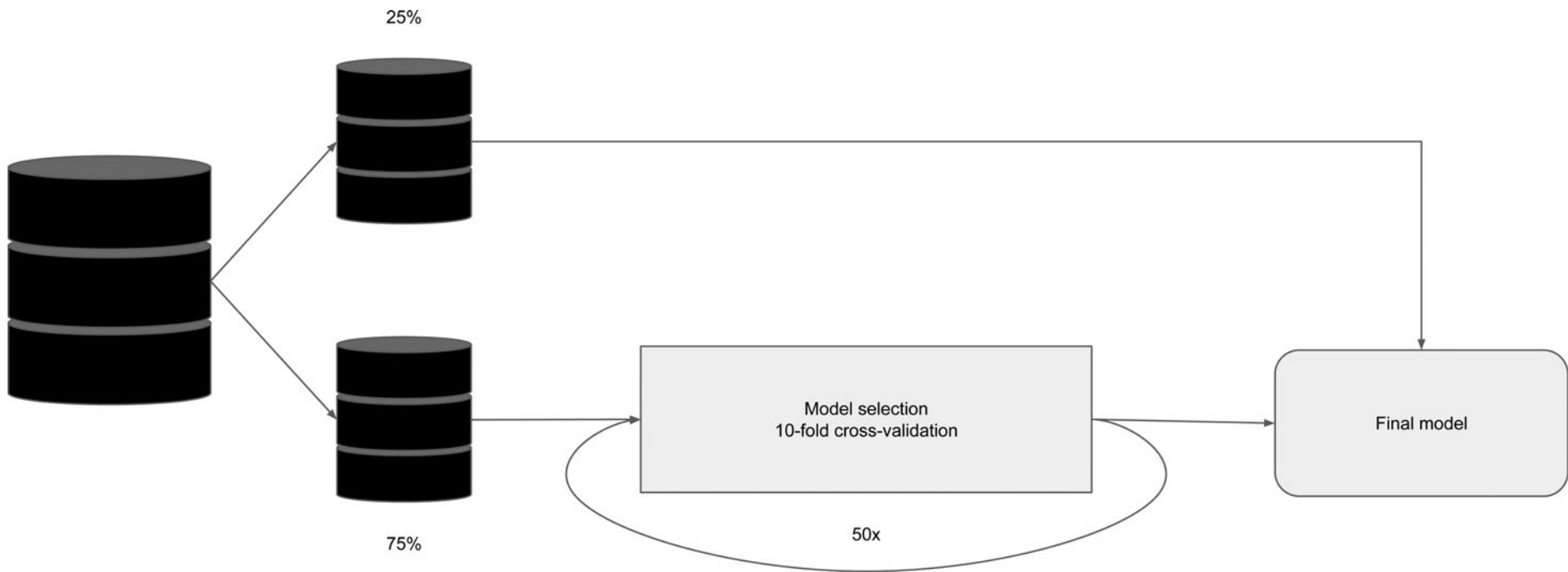
where  $w_i$  is the weight for the instance  $i$ ,  $c_i \in \{0,1\}$  is the class of the instance  $i$ , and  $p(y)$  and  $p(n)$  are the marginal probabilities for the positive and negative class, respectively. Class imbalance for random forest and ANN was addressed through a resampling step, which entailed randomly under-sampling the majority class so that both classes match the prevalence on the sample without further stratification of other confounding factors in each analysis followed by model training. The whole process was repeated in 50 iterations. The algorithm-predicted probabilities were averaged over the resampling iterations.

### Model performance measures

The validity of the models to predict 'unseen' subjects in test dataset was evaluated using sensitivity, specificity, balanced accuracy, positive predictive value (PPV), negative predictive value (NPV), and area under the ROC curve (AUC). We used a cutoff of 0.5 as the boundary for the class decision, that is, the algorithm classified probabilities above 50% as belonging to the positive outcome level (i.e. subject attempted suicide) and those below 50% to the negative outcome level (i.e. subject did not attempt suicide).

### Variable importance

Variable importance was estimated using the standard procedures from the *caret* package. For elastic net, the values of the coefficients are used. For random forest, the model sensitivity to



**Fig. 1.** Machine-learning protocol. First, we split the dataset into two parts: (1) training dataset with 75% of the whole sample and (2) test datasets with 25% of the sample. After this, we used a standard machine-learning protocol with 10-fold cross-validation, hyperparameter tuning, and class imbalance correction in the training dataset and we repeated the whole process in 50 iterations.

removing a predictor from its trees is used as a proxy for variable importance. For neural networks, the method described in Gevrey, Dimopoulos, and Lek (2003) is used.

#### Hyperparameter tuning

The standard grid search for the *caret* package was used. We changed the default search strategies of each algorithm such as: *Elastic net* searched for *alpha* from 0.1 to 1.0 with 0.1 intervals and *lambda* from 0.001 to 0.51 with 0.05 intervals; *random forest* searched for *mtry* from 1 to the total number of variables; *neural networks* searched for *size* from 1 to 100 with intervals of 5 and

*decay* from 0.1 to 0.5 with intervals of 0.1. The selection of the best model was performed independently for each approach following the AUC.

#### Results

A total of 32 700 subjects were included in aim 1 of this study and 6350 in aim 2. Tables 1 and 2 summarize the clinical and socio-demographic characteristics among participants who attempted suicide *v.* participants who did not between wave 1 and wave 2 for the general population and for participants with lifetime

**Table 1.** Sociodemographic and clinical characteristics in all participants

	Suicide attempt ( <i>n</i> = 200)	No suicide attempt ( <i>n</i> = 32,500)	Analysis	<i>p</i> value
Age (years)	36.8 (s.d. = 12.4)	46.1 (s.d. = 17.4)	<i>t</i> = 8.13	<0.001
Gender			$\chi^2(1) = 13.2$	<0.001
Male	70 (30.4%)	14 000 (42.6%)		
Female	150 (69.6%)	18 500 (57.4%)		
Race			$\chi^2(4) = 20.1$	<0.001
White	100 (51.8%)	19 000 (59%)		
Black	30 (14.3%)	6000 (18.5%)		
American Indigenous/Alaskan native	D	650 (2%)		
Asian	D	800 (2.5%)		
Hispanic	60 (25.4)	5700 (17.5%)		
Pacific Islander	D	100 (0.3%)		
Marital status			$\chi^2(1) = 22.5$	<0.001
Married <sup>a</sup>	80 (37.5)	17 500 (53.6%)		
Separated <sup>b</sup>	150 (62.5%)	15 000 (46.4%)		
Education			$\chi^2(5) = 22.1$	0.001
No school	D	150 (0.4%)		
Elementary school	D	1000 (3%)		
Middle school	D	1100 (3.2%)		
High school	100 (46.5%)	12 500 (38.3%)		
College	90 (39.9%)	14 000 (43.1%)		
Graduate school	D	4000 (12%)		
Income			$\chi^2(3) = 52.7$	<0.001
\$0–19 999	150 (70.5%)	14 000 (46.5%)		
\$ 20 000–34 999	40 (18.4%)	71 000 (23.5%)		
\$ 35 000–69 999	D	6800 (22.5%)		
>\$70 000	D	2300 (7.5%)		
BMI	27.8 (s.d. = 7.1)	27.2 (s.d. = 5.7)	<i>t</i> = 1.31	0.193
Substance use disorder				
Alcohol use disorder	100 (43.8%)	9400 (29%)	$\chi^2(1) = 22.6$	<0.001
Tobacco dependence	100 (44.6%)	5400 (16.5%)	$\chi^2(1) = 124.0$	<0.001
Drug use disorder	70 (29%)	3299 (9.9%)	$\chi^2(1) = 87.3$	<0.001
Mood disorder				
Dysthymia	80 (36.2%)	1600 (5%)	$\chi^2(1) = 427.9$	<0.001

(Continued)



**Table 1.** (Continued.)

	Suicide attempt ( <i>n</i> = 200)	No suicide attempt ( <i>n</i> = 32,500)	Analysis	<i>p</i> value
Bipolar disorder	70 (30.8%)	2000 (6.2%)	$\chi^2(1) = 222.4$	<0.001
Depressive disorder	150 (60.3%)	6200 (19.1%)	$\chi^2(1) = 238.5$	<0.001
Anxiety disorder				
Panic disorder	60 (26.3%)	1800 (5.6%)	$\chi^2(1) = 175.0$	<0.001
Social phobia	40 (16.5%)	1700 (5.1%)	$\chi^2(1) = 56.2$	<0.001
Specific phobia	50 (20.1%)	3200 (10%)	$\chi^2(1) = 23.7$	<0.001
General anxiety disorder	50 (23.2%)	1500 (4.7%)	$\chi^2(1) = 161.7$	<0.001
PTSD	90 (37.9%)	2700 (8.3%)	$\chi^2(1) = 245.9$	<0.001
ADHD	40 (17.9%)	750 (2.3%)	$\chi^2(1) = 227.8$	<0.001
Psychotic disorders				
Schizophrenia or psychotic illness or episode	20 (10.7%)	250 (0.8%)	$\chi^2(1) = 241.4$	<0.001
Personality disorders				
Borderline	150 (61.6%)	1900 (6%)	$\chi^2(1) = 1150$	<0.001
Schizotypal	80 (33.9%)	1300 (4.1%)	$\chi^2(1) = 468.3$	<0.001
Narcissistic	50 (23.7%)	2200 (6.9%)	$\chi^2(1) = 93.2$	<0.001
Avoidant	50 (21%)	750 (2.3%)	$\chi^2(1) = 321.8$	<0.001
Antisocial	40 (15.6%)	1100 (3.3%)	$\chi^2(1) = 97.8$	<0.001
Dependent	D	100 (0.4%)	D	
Obsessive-compulsive	50 (20.5%)	2600 (8.1%)	$\chi^2(1) = 44.6$	<0.001
Paranoid	60 (28.6%)	1500 (4.8%)	$\chi^2(1) = 263.0$	<0.001
Schizoid	40 (16.1%)	1100 (3.3%)	$\chi^2(1) = 108.1$	<0.001
Histrionic	20 (8.5%)	600 (1.9%)	$\chi^2(1) = 47.6$	<0.001

ADHD, attention deficit hyperactivity disorder; PTSD, post-traumatic stress disorder.

D: Statistic is based upon fewer than 15 observations.

The sum of some variables may vary because estimates on released outputs were rounded to minimize disclosure risk within and between projects.

$\chi^2$  tests with more than 1 degree of freedom (df) used Fisher's exact corrections, and the  $\chi^2$  tests with 1 df used the Yates exact correction to *p* values.

Authorization number: CBDRB-FY20-094.

<sup>a</sup>Married or living with another as if married.

<sup>b</sup>Widowed, separated, divorced, or never married.

*p* values in the table are not adjusted for multiple comparisons.

major depressive episodes, respectively. All variables showed differences between groups, except for BMI in the general population and gender, BMI, and specific phobia in the sample with lifetime major depressive episodes.

Figure 2 shows the ROC of all machine-learning algorithms used in the analyses performed on both samples.

### Elastic net regularization

The model built with elastic net regularization distinguished individuals who attempted suicide from those who did not with an AUC of 0.89 for aim 1 and 0.89 for aim 2. Balanced accuracy was 81.86% for aim 1 and 81.64% for aim 2. Other performance measures can be found in Table 3. The most important variables were borderline personality disorder, PTSD, and being of Asian descent for the model in all participants and previous suicide attempt, borderline personality disorder, and overnight stay in hospital because of depressive symptoms for the model in participants with lifetime major depressive episodes (online Supplementary Fig. S1).

Performance measures for random forest and ANN can be found in Table 3, while variable importance for these models is provided in online Supplementary Fig. S2.

### Discussion

This is the first study to evaluate the prediction of suicide attempt in a nationally representative sample of the US population. Our models achieved good performance and all algorithms achieved greater than chance (>50%) accuracy in distinguishing attempters from non-attempters, with balanced accuracy for suicide attempt exceeding 0.80 in all models. As our primary analysis, elastic net found the most relevant predictive variables that distinguished those who attempted suicide from those who did not in the general population, to be, in descending order, borderline personality disorder, PTSD, and being of Asian descent. Similarly, in the sample with lifetime major depressive episode, the most relevant predictor variables were, in descending order, previous suicide attempt, borderline personality disorder, and overnight stay in hospital because of depressive symptoms.

**Table 2.** Sociodemographic and clinical characteristics in participants with lifetime major depressive episodes

	Suicide attempt ( <i>n</i> = 150)	No suicide attempt ( <i>n</i> = 6200)	Analysis	<i>p</i> value
Age (years)	37.0 (s.d. = 12.2)	43.7 (s.d. = 15.3)	<i>t</i> = 5.05	<0.001
Gender			$\chi^2(1) = 2.03$	0.155
Male	30 (24.2%)	1900 (30.4%)		
Female	100 (75.8%)	4300 (69.6%)		
Race			$\chi^2(4) = 10.4$	0.025
White	80 (56.8%)	4200 (67.2%)		
Black	20 (14.4%)	850 (13.4%)		
American Indigenous/Alaskan native	D	200 (3%)		
Asian	D	80 (1.4%)		
Hispanic	30 (21.2%)	900 (14.8%)		
Pacific Islander	D	20 (0.26%)		
Marital status			$\chi^2(1) = 6.19$	0.015
Married <sup>a</sup>	50 (34.8%)	2900 (46.1%)		
Separated <sup>b</sup>	90 (65.2%)	3300 (53.9%)		
Education			$\chi^2(5) = 23.5$	<0.001
No school	D	D		
Elementary school	D	150 (2.1%)		
Middle school	D	150 (2.5%)		
High school	60 (42.6%)	2300 (37%)		
College	60 (43.4%)	2800 (45.2%)		
Graduate school	D	800 (13%)		
Income			$\chi^2(3) = 24.3$	<0.001
\$0–19 999	90 (72.4%)	3000 (50.7%)		
\$20 000–34 999	20 (15%)	1400 (23.1%)		
\$35 000–69 999	D	1200 (20.4%)		
>\$70 000	D	350 (5.8%)		
BMI	28.0 (s.d. = 7.4)	27.9 (s.d. = 6.7)	<i>t</i> = 0.22	0.823
Substance use disorder				
Alcohol use disorder	70 (53.8%)	2600 (42.3%)	$\chi^2(1) = 6.47$	0.011
Tobacco dependence	80 (56.8%)	1900 (30.4%)	$\chi^2(1) = 41.0$	<0.001
Drug use disorder	50 (35.6%)	1200 (19.8%)	$\chi^2(1) = 19.0$	<0.001
Mood disorder				
Dysthymia	70 (55.3%)	1400 (21.9%)	$\chi^2(1) = 80.5$	<0.001
Bipolar disorder	60 (46.2%)	1300 (20.4%)	$\chi^2(1) = 50.3$	<0.001
Depressive disorder	100.0% (sample requirement)	100.0% (sample requirement)		
Anxiety disorder				
Panic disorder	50 (40.9%)	1100 (17.1%)	$\chi^2(1) = 48.8$	<0.001
Social phobia	30 (25%)	900 (14.3%)	$\chi^2(1) = 11.1$	<0.001
Specific phobia	40 (28%)	1400 (21.7%)	$\chi^2(1) = 2.68$	0.102
General anxiety disorder	50 (34.8%)	1100 (17.9%)	$\chi^2(1) = 23.9$	<0.001
PTSD	60 (43.2%)	1300 (20.8%)	$\chi^2(1) = 37.5$	<0.001
ADHD	30 (22.7%)	300 (5%)	$\chi^2(1) = 75.6$	<0.001

(Continued)

Table 2. (Continued.)

	Suicide attempt (n = 150)	No suicide attempt (n = 6200)	Analysis	p value
Psychotic disorders				
Schizophrenia or psychotic illness or episode	20 (15.9%)	200 (2.9%)	$\chi^2(1) = 66.4$	<0.001
Personality disorders				
Borderline	90 (65.9%)	950 (15.1%)	$\chi^2(1) = 241.9$	<0.001
Schizotypal	50 (34.1%)	550 (9.1%)	$\chi^2(1) = 90.6$	<0.001
Narcissistic	30 (24.2%)	650 (10.8%)	$\chi^2(1) = 22.3$	<0.001
Avoidant	40 (32.6%)	500 (8.4%)	$\chi^2(1) = 90.2$	<0.001
Antisocial	30 (22.1%)	500 (7.8%)	$\chi^2(1) = 29.2$	<0.001
Dependent	D	90 (1.4%)	D	
Obsessive-compulsive	40 (28.8%)	1200 (18.9%)	$\chi^2(1) = 7.47$	0.006
Paranoid	60 (43.9%)	900 (13.9%)	$\chi^2(1) = 91.0$	<0.001
Schizoid	30 (25%)	600 (9.1%)	$\chi^2(1) = 36.4$	<0.001
Histrionic	20 (13.6%)	350 (5.6%)	$\chi^2(1) = 14.1$	<0.001
Specific variables to aim 2				
Wave 1 suicide attempt	100 (72%)	700 (10.9%)	$\chi^2(1) = 444.2$	<0.001
Hospitalized for depression	80 (58.3%)	750 (11.9%)	$\chi^2(1) = 243.6$	<0.001
Age at depression onset	23.3 (s.d. = 10.9)	30.3 (s.d. = 14.5)	$t = 7.48$	<0.001
Suicidal ideas	100 (80.3%)	2400 (38.7%)	$\chi^2(1) = 91.5$	<0.001

ADHD, attention deficit hyperactivity disorder; PTSD, post-traumatic stress disorder.

D: Statistic is based upon fewer than 15 observations.

The sum of some variables may vary because estimates on released outputs were rounded to minimize disclosure risk within and between projects.

$\chi^2$  tests with more than 1 df used Fisher's exact corrections, and the  $\chi^2$  tests with 1 df used the Yates exact correction to p values.

Authorization number: CBDRB-FY20-094.

<sup>a</sup>Married or living with another as if married.

<sup>b</sup>Widowed, separated, divorced, or never married.

p values in the table are not adjusted for multiple comparisons.

Psychopathology is strongly associated with suicidal behavior (Arsenault-Lapierre et al., 2004; Borges et al., 2010), and personality disorders, including borderline personality disorder, are also associated with premature mortality (Temes, Frankenburg, Fitzmaurice, & Zanarini, 2019; Tyrer, Reed, & Crawford, 2015). For borderline personality disorder, the presence of suicide attempt or self-injurious behavior is one of the diagnostic criteria (APA, 2013) and a defining feature of the disorder, with over 60% reporting multiple suicide attempts (Zanarini et al., 2008). An 8-year longitudinal follow-up study of 123 subjects with borderline personality disorder showed an increased risk of suicide attempt associated with illness severity and socioeconomic status, including minority race and frequent changes in employment (Soloff & Chiappetta, 2017). PTSD is considered an independent predictor of attempted suicide (Sareen et al., 2007; Wilcox, Storr, & Breslau, 2009). A cohort study of 1698 young adults showed an adjusted relative risk between PTSD and suicide attempt of 2.7, even after adjustment for a prior major depressive episode, alcohol and drug abuse or dependence, whereas exposure to traumatic events without PTSD was not associated with an increased risk of attempted suicide (Wilcox et al., 2009). A traumatic experience is required for a diagnosis of PTSD and it is highly prevalent in the childhood of those who develop a borderline personality disorder (Leichsenring, Leibing, Kruse, New, & Leweke, 2011). Our results, combined with those of previous studies, may indicate that trauma is a significant predictor of a suicide attempt, but only for those who develop a trauma related disorder. A meta-analysis

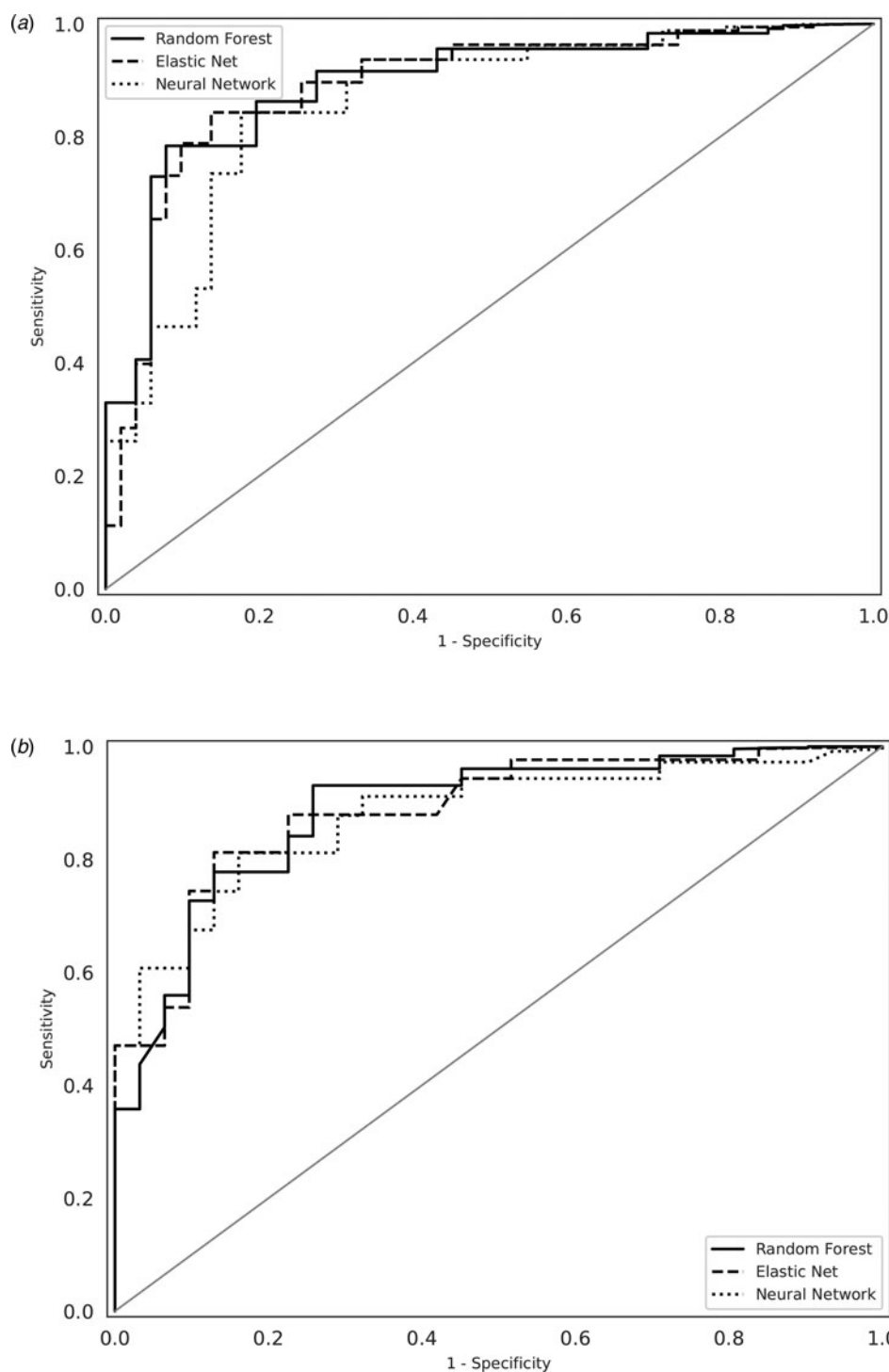
reinforced the evidence that a PTSD diagnosis is associated with increased suicidality and supported an important role of comorbid major depression in the etiology of suicidality in PTSD (Panagioti, Gooding, & Tarrier, 2012).

A literature overview about suicide risk among immigrants and ethnic minorities showed a positive correlation between suicidal behavior and specific countries of origin. Non-European immigrant women demonstrated the highest risk for suicide attempt, a group that included young women of South Asian and black African origin (Forte et al., 2018).

Suicide attempt and hospitalization are risk factors for subsequent suicide attempts and suicide in participants with mood disorders (Tondo et al., 2007). A meta-analysis showed that the risk of suicide in people who presented to health care services after an incident of self-harm was 1.6% after 1 year and 3.9% after 5 years, and the estimated rate of repetition of non-fatal self-harm was 16.3% at 1 year, 16.8% at 2 years, and 22.4% at 5 years (Carroll, Metcalfe, & Gunnell, 2014). In a 5-year prospective study, 249 patients with major depressive disorder were assessed and history of suicide attempts showed a hazard ratio of 4.39 to predict suicide during the follow-up (Holma et al., 2010).

There is conflicting evidence regarding the association between BMI and attempted suicide (Perera et al., 2016). A critical review demonstrated that among men, a high BMI was associated with a low risk of attempted or completed suicide, while there was a paradox among women, namely, a high BMI was associated with an elevated risk of attempted suicide but a low risk of





**Fig. 2.** ROC of the different algorithms. (a) ROC in all participants. (b) ROC in participants with lifetime major depressive episodes.

completed suicide (Zhang et al., 2013). BMI was among the most important predictive variables only in the random forest model (a nonlinear algorithm), which may highlight the complexity of the relationship between BMI and suicide attempt.

A recent systematic review has discussed the finding that prediction models of suicide death and suicide attempt achieved good accuracy but the PPV were low with high false-positive rates (Belsher et al., 2019). Unfortunately, prevalence imposes a ceiling on PPV, so low PPV is expected because these models work with rare outcomes. Due to the higher prevalence of suicide attempt in the depressed sample, PPV was also higher (10.48%)

compared to the general population (4.55%). These results are higher than most prior studies (Belsher et al., 2019). We recommend that the model for the general population (aim 1) should be used as a screening tool to identify people at higher risk to attempt suicide. Health authorities should contact these people (or their relatives) to suggest more specific mental health assessments in the upcoming years. For people that already have a major depressive episode (the model built in aim 2) and were identified as positives for suicide attempts in the future, preventive strategies, such as the use of lithium or CBT, for instance, should be implemented.

**Table 3.** Model performance measures

	Elastic net	Random forest	Neural network
<i>(a)</i>			
Balanced accuracy	0.8186	0.8339	0.8244
Specificity	0.8922	0.8443	0.8449
Sensitivity	0.7451	0.8235	0.8039
NPV	0.9980	0.9985	0.9984
PPV	0.0455	0.0352	0.0345
AUC	0.89	0.89	0.86
<i>(b)</i>			
Balanced accuracy	0.8164	0.8211	0.8045
Specificity	0.8586	0.8034	0.7703
Sensitivity	0.7742	0.8387	0.8387
NPV	0.9944	0.9957	0.9955
PPV	0.1048	0.0836	0.0724
AUC	0.89	0.89	0.88

AUC, area under the ROC curve.

*(a)* Model performance measures in all participants. *(b)* Model performance measures in participants with lifetime major depressive episodes.

Authorization number: CBDRB-FY20-094.

The current study has some potential limitations. First, although our study has a longer follow-up period compared to prior literature [in Belsher's systematic review (Belsher et al., 2019) only one of the included studies had a follow-up of more than 2 years], death by suicide or suicide attempt could still be ahead for people considered as false positives. Second, we are missing suicide attempts that resulted in deaths and all the individuals who died between wave 1 and wave 2. It is also noteworthy that a history of attempted suicide is associated with an increased rate of all-cause death and the life expectancy is reduced in these individuals (Al-Sayegh et al., 2015; Jokinen, Talbäck, Feychting, Ahlbom, & Ljung, 2018). Third, we are only reporting the self-reported suicide attempts, so we are missing the ones that could be found in administrative data. Fourth, we did not include exposure to early-life adversity, another well-characterized risk factor associated with suicidal behavior (Almeida et al., 2012; Turecki & Brent, 2015), because these data were not collected in wave 1. Past suicide attempts are also strongly associated with suicidal behavior (Carroll et al., 2014), but this was not included in the analysis with the general population because it was only assessed in wave 1 in individuals with lifetime major depressive episodes. Fifth, the models built in the current study may be useful for the US population; however, their accuracy should be assessed in other countries before implementation, as suicide attempts may vary according to culture and other population variables, such as religion (WHO, 2018). Sixth, data analyzed in the current study are more than 10 years old; however, the association between the variables assessed in the current study and the outcome do not change over time. Finally, regarding the machine-learning analysis, we failed to conduct calibration experiments to ensure that predicted probabilities are representative of actual suicide attempt probabilities. Future research on the same lines

needs to ensure calibration is in-place before predictive models can be employed large-scale at the population level.

In summary, we report a highly accurate algorithm that is able to identify suicide attempts in the general population and in individuals with lifetime major depressive episodes using clinical, sociodemographic, and stressful life events' data in a nationally representative sample. These results suggest that it is possible to utilize clinical measures to identify individuals at greater risk of attempting suicide. Future studies integrating data from different biological levels, such as genetics, metabolomics, and digital health data (Torous & Walker, 2019) could potentially help to build more accurate models. Additionally, future studies should have even longer follow-up periods to increase PPV.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291720004997>.

**Acknowledgements.** The statistical summaries reported in this document have been cleared by the US Census Bureau's Disclosure Review Board release authorization number CBDRB-FY20-094. Any opinions and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the US Census Bureau. In addition, all results have been reviewed to ensure that no confidential information is disclosed. We also acknowledge the statistician from the US Census Bureau, Dr Jahn K. Hakes.

#### Author contributions.

All authors contributed to the study design. The statistician from the US Census Bureau, Dr Jahn K. Hakes, did the analyses. CSM, PB, and ICP participated in the data analysis. CSM, PB, BC, BW, MAC, FK, and ICP were responsible for the interpretation of findings. CSM and PB were responsible for the figures. CSM, PB, BC, BW, MAC, FK, and ICP did the scientific literature search. CSM, PB, BC, BW, MAC, FK, and ICP participated in writing of the report, and all authors approved the final version of the manuscript.

**Financial support.** This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. IC Passos receives research support from CAPES, FIPE, and CNPq. CS Machado received master scholarship from CAPES, during the conduct of the study.

**Conflict of interest.** P Ballester, B Can, B Mwangi, M Caldieraro, F Kapczinski have nothing to disclose. IC Passos receives research support from CAPES (Finance Code 001), FIPE, and CNPq. CS Machado received master scholarship from CAPES (Finance Code 001), during the conduct of the study.

#### References

- Al-Sayegh, H., Lowry, J., Polur, R. N., Hines, R. B., Liu, F., & Zhang, J. (2015). Suicide history and mortality: A follow-up of a national cohort in the United States. *Archives of Suicide Research*, 19(1), 35–47. doi: 10.1080/13811118.2013.855154.
- Almeida, O. P., Draper, B., Snowdon, J., Lautenschlager, N. T., Pirkis, J., Byrne, G., ... Pfaff, J. J. (2012). Factors associated with suicidal thoughts in a large community study of older adults. *British Journal of Psychiatry*, 201(6), 466–472. doi: 10.1192/bjp.bp.112.110130.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (Fifth Edition). Arlington, VA: American Psychiatric Association.
- Arsenault-Lapierre, G., Kim, C., & Turecki, G. (2004). Psychiatric diagnoses in 3275 suicides: A meta-analysis. *BMC Psychiatry*, 4, 1–11. doi: 10.1186/1471-244X-4-37.
- Belsher, B. E., Smolenski, D. J., Pruitt, L. D., Bush, N. E., Beech, E. H., Workman, D. E., ... Skopp, N. A. (2019). Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*, 76(6), 642–651. doi: 10.1001/jamapsychiatry.2019.0174.
- Borczykowski, A., Hjern, A., Lindblad, F., & Vinnerljung, B. (2006). Suicidal behaviour in national and international adult adoptees: A Swedish cohort study. *Social Psychiatry and Psychiatric Epidemiology*, 41(2), 95–102. doi: 10.1007/s00127-005-0974-2.

- Borges, G., Angst, J., Nock, M. K., Ruscio, A. M., Walters, E. E., & Kessler, R. C. (2007). Risk factors for twelve-month suicide attempts in the National Comorbidity Survey Replication (NCS-R). *Psychological Medicine*, 36(12), 1747–1757. doi: 10.1017/S0033291706008786.
- Borges, G., Nock, M. K., Abad, J. M. H., Sampson, N. A., Alonso, J., Helena, L., ... Williams, D. R. (2010). Twelve month prevalence of and risk factors for suicide attempts in the WHO world mental health surveys. *The Journal of Clinical Psychiatry*, 71(12), 1617–1628. doi: 10.4088/JCP.08m04967blu.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi: 10.1023/A:1010933404324.
- Carroll, R., Metcalfe, C., & Gunnell, D. (2014). Hospital presenting self-harm and risk of fatal and non-fatal repetition: Systematic review and meta-analysis. *PLoS ONE*, 9(2), e89944. doi: 10.1371/journal.pone.0089944.
- Choi, S. B., Lee, W., Yoon, J. H., Won, J. U., & Kim, D. W. (2018). Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea. *Journal of Affective Disorders*, 231, 8–14. doi: 10.1016/j.jad.2018.01.019.
- Cipriani, A., Hawton, K., Stockton, S., & Geddes, J. R. (2013). Lithium in the prevention of suicide in mood disorders: Updated systematic review and meta-analysis. *BMJ*, 346, f3646. doi: 10.1136/bmj.f3646.
- Cross, S. S., Harrison, R. F., & Kennedy, R. L. (1995). Introduction to neural networks. *The Lancet*, 346(8982), 1075–1079. doi: 10.1016/S0140-6736(95)91746-2.
- Forste, A., Trobia, F., Gualtieri, F., Lamis, D. A., Cardamone, G., Giallonardo, V., ... Pompili, M. (2018). Suicide risk among immigrants and ethnic minorities: A literature overview. *International Journal of Environmental Research and Public Health*, 15(7), 1438. doi: 10.3390/ijerph15071438.
- Galfalvy, H. C., Oquendo, M. A., & Mann, J. J. (2008). Evaluation of clinical prognostic models for suicide attempts after a major depressive episode. *Acta Psychiatrica Scandinavica*, 117(4), 244–252. doi: 10.1111/j.1600-0447.2008.01162.x.
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3), 249–264. doi: 10.1016/S0304-3800(02)00257-0.
- Hasin, D. S., & Grant, B. F. (2015). The national epidemiologic survey on alcohol and related conditions (NESARC) waves 1 and 2: Review and summary of findings. *Social Psychiatry and Psychiatric Epidemiology*, 50(11), 1609–1640. doi: 10.1007/s00127-015-1088-0.
- Hedegaard, H., Curtin, S. C., & Warner, M. (2018). Suicide Mortality in the United States, 1999–2017. NCHS data brief, (330), 1–8. Retrieved from <https://www.cdc.gov/nchs/data/databriefs/db330-tables-508.pdf#2>.
- Heikkinen, M., Aro, H., & Lönnqvist, J. (1992). Recent life events and their role in suicide as seen by the spouses. *Acta Psychiatrica Scandinavica*, 86(6), 489–494. doi: 10.1111/j.1600-0447.1992.tb03303.x.
- Hoertel, N., Franco, S., Wall, M. M., Oquendo, M. A., Kerridge, B. T., Limosin, F., & Blanco, C. (2015). Mental disorders and risk of suicide attempt: A national prospective study. *Molecular Psychiatry*, 20(6), 718–726. doi: 10.1038/mp.2015.19.
- Holma, K. M., Melartin, T. K., Haukka, J., Holma, I. A. K., Sokero, T. P., & Isometsä, E. T. (2010). Incidence and predictors of suicide attempts in DSM-IV major depressive disorder: A five-year prospective study. *American Journal of Psychiatry*, 167(7), 801–808. doi: 10.1176/appi.ajp.2010.09050627.
- Isometsä, E. (2014). Suicidal behaviour in mood disorders – who, when, and why? *Canadian Journal of Psychiatry*, 59(3), 120–130. doi: 10.1177/070674371405900303.
- Johnston, A. K., Pirkis, J. E., & Burgess, P. M. (2009). Suicidal thoughts and behaviours among Australian adults: Findings from the 2007 national survey of mental health and wellbeing. *Australian and New Zealand Journal of Psychiatry*, 43(7), 635–643. doi: 10.1080/00048670902970874.
- Jokinen, J., Talbäck, M., Feychting, M., Ahlbom, A., & Ljung, R. (2018). Life expectancy after the first suicide attempt. *Acta Psychiatrica Scandinavica*, 137(4), 287–295. doi: 10.1111/acps.12842.
- Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., ... Ursano, R. J. (2015). Predicting suicides after psychiatric hospitalization in US army soldiers: The Army Study to Assess Risk and Resilience in Service members (Army STARRS). *JAMA Psychiatry*, 72(1), 49–57. doi: 10.1001/jamapsychiatry.2014.1754.
- Keyes, M. A., Malone, S. M., Sharma, A., Iacono, W. G., & McGue, M. (2013). Risk of suicide attempt in adopted and nonadopted offspring. *Pediatrics*, 132(4), 639–646. doi: 10.1542/peds.2012-3251.
- Kuhn, M. (2008). Caret package. *Journal of Statistical Software*, 28(5), 1–26. Retrieved from <http://www.jstatsoft.org/v28/i05/paper>.
- Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *The Lancet*, 377, 74–84. doi: 10.1016/S0140-6736(10)61422-5.
- Morey, L. C., Lowmaster, S. E., & Hopwood, C. J. (2010). A pilot study of manual-assisted cognitive therapy with a therapeutic assessment augmentation for borderline personality disorder. *Psychiatry Research*, 178(3), 531–535. doi: 10.1016/j.psychres.2010.04.055.
- National Institute on Alcohol Abuse and Alcoholism. (2006). National epidemiologic survey on alcohol and related conditions (NESARC). *Alcohol Alert*, 70(1), 1–6. Retrieved from <http://www.niaaa.nih.gov>.
- Nock, M. K., Borges, G., Bromet, E. J., Alonso, J., Angermeyer, M., Beautrais, A., ... Williams, D. (2008). Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *British Journal of Psychiatry*, 192(2), 98–105. doi: 10.1192/bjp.bp.107.040113.
- Nock, M. K., Hwang, I., Sampson, N. A., & Kessler, R. C. (2010). Mental disorders, comorbidity and suicidal behavior: Results from the national comorbidity survey replication. *Molecular Psychiatry*, 15(8), 868–876. doi: 10.1038/mp.2009.29.Mental.
- Oquendo, M. A., Galfalvy, H., Russo, S., Ellis, S. P., Grunebaum, M. F., Burke, A., & Mann, J. J. (2004). Prospective study of clinical predictors of suicidal acts after a major depressive episode in patients with major depressive or bipolar disorder. *American Journal of Psychiatry*, 161(8), 1433–1441. doi: 10.1176/appi.ajp.161.8.1433.
- Oquendo, M. A., Perez-Rodriguez, M. M., Poh, E., Sullivan, G., Burke, A. K., Sublette, M. E., ... Galfalvy, H. (2014). Life events: A complex role in the timing of suicidal behavior among depressed patients. *Molecular Psychiatry*, 19(8), 902–909. doi: 10.1038/mp.2013.128.
- Panagiotti, M., Gooding, P. A., & Tarrier, N. (2012). A meta-analysis of the association between posttraumatic stress disorder and suicidality: The role of comorbid depression. *Comprehensive Psychiatry*, 53(7), 915–930. doi: 10.1016/j.comppsy.2012.02.009.
- Passos, I. C., & Ballester, P. (2019). Positive predictive values and potential success of suicide prediction models. *JAMA Psychiatry*, 76(8), 869. doi: 10.1001/jamapsychiatry.2019.1507.
- Passos, I. C., Ballester, P. L., Barros, R. C., Librenza-Garcia, D., Mwangi, B., Birmaher, B., ... Kapczinski, F. (2019). Machine learning and big data analytics in bipolar disorder: A position paper from the International Society For Bipolar Disorders Big data task force. *Bipolar Disorders*, 21(7), 582–594. doi: 10.1111/bdi.12828.
- Passos, I. C., Mwangi, B., Cao, B., Hamilton, J. E., Wu, M.-J., Zhang, X. Y., ... Soares, J. C. (2016). Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. *Journal of Affective Disorders*, 193, 109–116. doi: 10.1016/j.jad.2015.12.066.
- Perera, S., Eisen, R. B., Dennis, B. B., Bawor, M., Bhatt, M., Bhatnagar, N., ... Samaan, Z. (2016). Body mass index is an important predictor for suicide: Results from a systematic review and meta-analysis. *Suicide and Life-Threatening Behavior*, 46(6), 697–736. doi: 10.1111/sltb.12244.
- Sareen, J., Cox, B. J., Stein, M. B., Afifi, T. O., Fleet, C., & Asmundson, G. J. G. (2007). Physical and mental comorbidity, disability, and suicidal behavior associated with posttraumatic stress disorder in a large community sample. *Psychosomatic Medicine*, 69(3), 242–248. doi: 10.1097/PSY.0b013e31803146d8.
- Schaffer, A., Isometsä, E. T., Tondo, L., H. Moreno, D., Turecki, G., Reis...C., ... Yatham, L. N. (2014). International society for bipolar disorders task force on suicide: Meta-analyses and meta-regression of correlates of suicide attempts and suicide deaths in bipolar disorder. *Bipolar Disorders*, 17(1), 1–16. doi: 10.1111/bdi.12271.
- Slap, G., Goodman, E., & Huang, B. (2001). Adoption as a risk factor for attempted suicide during adolescence. *Pediatrics*, 108(2), e30. doi: 10.1542/peds.108.2.e30.
- Soloff, P. H., & Chiappetta, L. (2017). Suicidal behavior and psychosocial outcome in borderline personality disorder at 8-year follow-up. *Journal of Personality Disorders*, 31(6), 774–789. doi: 10.1521/pedi\_2017\_31\_280.
- Temes, C. M., Frankenburg, F. R., Fitzmaurice, G. M., & Zanarini, M. C. (2019). Deaths by suicide and other causes among patients with borderline

- personality disorder and personality-disordered comparison subjects over 24 years of prospective follow-up. *Journal of Clinical Psychiatry*, 80(1), 30–36. doi: 10.4088/JCP.18m12436.
- Tondo, L., Lepri, B., & Baldessarini, R. J. (2007). Suicidal risks among 2826 Sardinian major affective disorder patients. *Acta Psychiatrica Scandinavica*, 116(6), 419–428. doi: 10.1111/j.1600-0447.2007.01066.x.
- Torous, J., & Walker, R. (2019). Leveraging digital health and machine learning toward reducing suicide – from Panacea to practical tool. *JAMA Psychiatry*, 76(10), 999–1000. doi: 10.1001/jamapsychiatry.2019.1231.
- Turecki, G. (2014). The molecular bases of the suicidal brain. *Nature Reviews Neuroscience*, 15(12), 802–816. doi: 10.1038/nrn3839.
- Turecki, G., & Brent, D. A. (2015). Suicide and suicidal behaviour. *The Lancet*, 6736, 15. doi: 10.1016/S0140-6736(15)00234-2.
- Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *The Lancet*, 385(9969), 717–726. doi: 10.1016/S0140-6736(14)61995-4.
- Van Heeringen, K., & Mann, J. J. (2014). The neurobiology of suicide. *The Lancet Psychiatry*, 1(1), 63–72. doi: 10.1016/S2215-0366(14)70220-2.
- Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3), 457–469. doi: 10.1177/2167702617691560.
- WHO. (2014). *Preventing suicide: A global imperative*. Geneva: World Health Organization. Retrieved from [https://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779\\_eng.pdf?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779_eng.pdf?sequence=1).
- WHO. (2018). *National suicide prevention strategies: Progress, examples and indicators*. Geneva: World Health Organization. Retrieved from <https://apps.who.int/iris/bitstream/handle/10665/279765/9789241515016-eng.pdf?ua=1>.
- Wilcox, H. C., Storr, C. L., & Breslau, N. (2009). Posttraumatic stress disorder and suicide attempts in a community sample of urban American young adults. *Archives of General Psychiatry*, 66(3), 305–311. doi: 10.1001/archgenpsychiatry.2008.557.
- Zalsman, G., Hawton, K., Wasserman, D., van Heeringen, K., Arensman, E., Sarchiapone, M., ... Zohar, J. (2016). Suicide prevention strategies revisited: 10-year systematic review. *The Lancet Psychiatry*, 3(7), 646–659. doi: 10.1016/S2215-0366(16)30030-X.
- Zanarini, M. C., Frankenburg, F. R., Reich, D. B., Fitzmaurice, G., Weinberg, I., & Gunderson, J. G. (2008). The 10-year course of physically self-destructive acts reported by borderline patients and axis II comparison subjects. *Acta Psychiatrica Scandinavica*, 117(3), 177–184. doi: 10.1111/j.1600-0447.2008.01155.x.
- Zhang, J., Yan, F., Li, Y., & McKeown, R. E. (2013). Body mass index and suicidal behaviors: A critical review of epidemiological evidence. *Journal of Affective Disorders*, 148(2–3), 147–160. doi: 10.1016/j.jad.2012.05.048.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 67(2), 301–320. doi: 10.1111/j.1467-9868.2005.00503.x.