

Research paper

Predicting future suicidal behaviour in young adults, with different machine learning techniques: A population-based longitudinal study



Kasper van Mens^a, CWM de Schepper^b, Ben Wijnen^c, Saskia J Koldijk^b, Hugo Schnack^b, Peter de Looff^d, Joran Lokkerbol^c, Karen Wetherall^e, Seonaid Cleare^e, Rory C O'Connor^e, Derek de Beurs^{c,*}

^a Altrecht mental health Centre, Utrecht, the Netherlands

^b University medical center Utrecht, department of psychiatry, Utrecht, the Netherlands

^c Trimbos-Institute, Netherlands Institute for Mental Health and Addiction, Utrecht, the Netherlands

^d National expert Centre Specialized and Forensic Care, Den Dolder, The Netherlands

^e Glasgow university, suicidal behaviour research laboratory, Glasgow, United Kingdom

ABSTRACT

Background: The predictive accuracy of suicidal behaviour has not improved over the last decades. We aimed to explore the potential of machine learning to predict future suicidal behaviour using population-based longitudinal data.

Method: Baseline risk data assessed within the Scottish wellbeing study, in which 3508 young adults (18–34 years) completed a battery of psychological measures, were used to predict both suicide ideation and suicide attempts at one-year follow-up. The performance of the following algorithms was compared: regular logistic regression, K-nearest neighbors, classification tree, random forests, gradient boosting and support vector machine.

Results: At one year follow up, 2428 respondents (71%) finished the second assessment. 336 respondents (14%) reported suicide ideation between baseline and follow up, and 50 (2%) reported a suicide attempt. All performance metrics were highly similar across methods. The random forest algorithm was the best algorithm to predict suicide ideation (AUC 0.83, PPV 0.52, BA 0.74) and the gradient boosting to predict suicide attempt (AUC 0.80, PPV 0.10, BA 0.69).

Limitations: The number of respondents with suicidal behaviour at follow up was small. We only had data on psychological risk factors, limiting the potential of the more complex machine learning algorithms to outperform regular logistical regression.

Conclusions: When applied to population-based longitudinal data containing multiple psychological measurements, machine learning techniques did not significantly improve the predictive accuracy of suicidal behaviour. Adding more detailed data on for example employment, education or previous health care uptake, might result in better performance of machine learning over regular logistical regression.

1. Introduction

Suicide and suicide attempts are major public health issues (World Health Organization, 2014). The WHO designates suicide prevention a global imperative and urges countries to develop and implement national suicide prevention strategies. The first step to improve prevention strategies is to establish which risk factors explain and predict suicidal behaviour over time. Indeed, epidemiological studies have made some progress, in this regard. For instance, mood disorders and anxiety disorders are well known risk factors associated with future suicidal behaviour (Nock et al., 2009; Sareen et al., 2005). However, as was highlighted in a recent meta-analysis such epidemiological risk factors are not specific enough to be used in daily clinical practice, as their predictive power is only slightly better than chance for all suicidal thoughts and behavioural outcomes (Franklin et al., 2016). As the suicide research field's predictive ability has not improved in the last 50

years, these authors called for a shift from a focus on traditional modelling techniques to approaches and techniques from the field of machine learning (ML). The rationale to predict suicidal behaviour with ML compared to more traditional approaches is because the latter is focused on explanation rather than prediction. Additionally, techniques which analyse the effects of significant relations, such as using isolated predictors in simple regression models, are not adequate to model multifactorial complex behaviours such as suicide (O'Connor and Nock, 2014; Walsh et al., 2017a).

1.1. Examples of ML using electronic health records

In two studies, Walsh et al. (2017, 2018) developed risk prediction models using readily available electronic health records (EHR). In both studies, they applied a machine learning technique called random forest and demonstrated more accurate prediction of suicide attempts than

* Corresponding author.

E-mail address: dbeurs@trimbos.nl (D. de Beurs).

the more traditional regression technique (For example, the area under the curve (AUC) of the ML algorithm was 0.81 CI 0.83–0.85 versus the AUC of 0.66 CI 0.58–0.75 of standard logistical regression). [Barak-Corren et al. \(2017\)](#) also developed a risk prediction model using readily available EHR data. They predicted suicide attempts or death by suicide in a large health care system and were able to identify nearly half of all suicides and suicidal behaviours with 90% specificity. Moreover, they demonstrated that models with fewer predictors (i.e., based on commonly used risk factors only) performed worse (sensitivity of 33% vs. 45%) than those with more predictors. [Kessler et al. \(2015\)](#), in their study, collected extensive data in addition to EHR and in their best-fitted model, 53% of post-hospitalization suicides occurred within the 5% of hospitalized patients with highest predicted suicide risk ([Kessler et al., 2015](#)). These examples demonstrate the advantages of ML over traditional approaches when predicting suicidal behaviour using EHR.

1.2. ML and population-based longitudinal studies

Population-based studies are used to estimate the prevalence of suicidal behaviour in the general population (e.g. [\(Nock et al., 2010; Ten Have, Van Dorselaer, and De Graaf, 2013\)](#)). Early analyses from such studies found a significant relationship between any mental disorder and suicidal behaviour (e.g. [\(Harris and Barraclough, 1997\)](#)). However, when novel statistical techniques were used to control for comorbidity, the relation between disorders and suicidal behaviour appeared to be less strong, and for some disorders even non-significant ([Nock et al., 2009](#)). This example shows how the application of different statistical techniques can result in different conclusions.

In the current study we aim to explore the potential of machine learning when applied to population-based longitudinal data that includes 15 different psychological risk and protective factors for suicidal behaviour assessed with 211 separate items. The assessed risk factors incorporated key aspects of the integrated motivational-volitional (IMV) model of suicidal behaviour ([O'Connor and Kirtley, 2018](#)).

In brief, the IMV model proposes that suicidal behaviour results from a complex interplay of motivational and volitional phase factors. Factors within the motivational phase of the model explain how suicidal thoughts emerge in some people but not in others. Factors within the motivational phase include defeat, entrapment, and (lack of) social support. Volitional phase factors, on the other hand, are those factors that govern the transition from suicidal thinking (ideation/intent) to suicidal behaviour; they include exposure to suicide, fearlessness about death and impulsivity.

There are several elements in the modelling of the data, that makes research in the identification of future suicidal behaviour, suitable for machine learning. First, the interplay between the 15 psychological factors in the data is hypothesized to be of importance ([D. de Beurs et al., 2018](#)). For example, the IMV model states that the co-occurrence of high levels of entrapment and perceived burdensomeness result in suicide ideation. Machine learning techniques are more flexible in modelling different types of relationships. Second, it is unlikely that the relationship between all variables will be a linear combination of parameters, as shown by recent ecological momentary studies ([Hallensleben et al., 2018; Kleiman et al., 2017](#)). Additionally, another aim of this study is to investigate whether the inclusion of the individual items can increase predictive accuracy. We expect the inclusion of separate items to improve our predictive validity because they contain more detailed information than the constructs.

To explore the potential of machine learning in this context, we will compare different algorithms on their applicability in predicting suicidal behaviour by comparing performance metrics such as the positive predictive value in the prediction of 1) suicide ideation at one year follow up and 2) suicide attempt at one year follow up. Also, we will compare which risk factors are deemed most relevant by which techniques. We are interested to determine which technique is most suited

for this type of data and especially if there is added value in more complex machine learning techniques over and above standard regression techniques. Relatedly we are interested in establishing whether the inclusion of individual items from psychological scales improves prediction compared to using only the sum scores.

2. Method

2.1. Participants and procedure

The Scottish Wellbeing Study ([O'Connor et al., 2018](#)) is the first nationally representative population-based study documenting suicidal ideation and behaviour in young adults (18–34 year olds) across Scotland. A quota sampling methodology was employed, with quotas based on age (three quota groups), sex and working status. Following written consent, participants completed an hour-long interview, carried out face-to-face in their homes. Participants completed a battery of psychological and social measures that incorporated key aspects of the integrated motivational volitional (IMV) model of suicidal behaviour (see instruments sections below). All interviewers were trained in the administration of the measures. Ethical approval was obtained from the Psychology Department's ethics committee at the University of Stirling and the US Department of Defence Human Research Protections Office. Participants received £25 in compensation for taking part. All participants were given a list of support organisations at the end of the interview. Participants who agreed were then contacted after 12 months to complete a follow-up questionnaire; this was completed by email, post or phone. Participants were then compensated a further £15 in shopping vouchers for their time, and entered into a prize draw for an iPad mini as an incentive.

2.2. Instruments

Psychological risk factors for suicidal behaviour

All risk and protective factors for suicidal behaviour assessed within the Scottish Wellbeing Study are included in our simulations ([O'Connor and Kirtley, 2018](#)). The actors are: current suicidal ideation ([Beck et al., 1988](#)), depressive symptoms ([Dozois et al., 1998](#)), stress ([Cohen et al., 1983](#)), mental wellbeing ([Tennant et al., 2007](#)), defeat ([Gilbert and Allan, 1998](#)), entrapment ([Gilbert and Allan, 1998](#)), social support ([Mitchell et al., 2003](#)) socially prescribed or social perfectionism ([Hewitt and Flett, 1991](#)), interpersonal needs ([Van Orden, Cukrowicz, Witte, and Joiner, 2012](#)) and goal activation ([Wrosch et al., 2003](#)). Optimism ([Scheier et al., 1994](#)), resilience ([Campbell-Sills and Stein, 2007](#)), acquired capability ([Van Orden, Witte, Gordon, Bender, and Joiner, 2008](#)) and impulsivity ([Patton et al., 1995](#)) were also assessed. Finally, mental imagery was assessed using eight items to establish the frequency with which participants imagine death-related imagery. More details on the study can be found in ([Wetherall et al., 2018](#)).

2.3. History of suicidal ideation and suicide attempts

were assessed with two items drawn from the Adult Psychiatric Morbidity Survey ([McManus et al., 2009](#)): "Have you ever seriously thought of taking your life, but not actually attempted to do so?" and "Have you ever made an attempt to take your life, by taking an overdose of tablets or in some other way?". Scores on the two items were combined into one dichotomous variable.

2.4. Future suicidal ideation and suicide attempts

were assessed at the follow-up with the same two items used to assess history of suicidal ideation and suicide attempts, adapted to specify the last 12 months. The variables were both dichotomous.

2.5. Statistical analysis

All analyses were done in R, a statistical programming language (Computing, 2011). The Caret package was used for the machine learning analyses. Caret is a wrapper for about 200 machine learning techniques and was used to apply different algorithms to the data (Kuhn, 2008). In this study, we have chosen six popular techniques to compare: logistic regression, k-nearest neighbors, classification tree, random forest, gradient boosting and support vector machine. The techniques are described in more detail in the appendix.

2.6. Cross-validation

Since machine learning is data-driven and often uses many variables, researchers are understandably concerned with the generalizability of the models. Techniques such as cross-validation are used to prevent over-fitting and increase generalizability. In cross-validation, the data are divided repeatedly into several subsets and algorithms are trained on a subset called the training set, and are evaluated on a different subset called the test set (Kuhn and Johnson, 2013). In this study, a random training set (70%) and a random test set (30%) was selected to train and validate the models. The data splitting was done for both the analysis on future suicide ideation as for future suicide attempts. In both cases, the prevalence of the two measures was kept the same in both training and test set. On the training data, 10-fold cross-validation was used in which the data were split again, into 10 subsets from which one was used as the test set and the other 9 were used for building a model. This procedure was repeated iteratively 10 times and the performance of the model was averaged over the iterations. In each cycle of 10 steps, a different set of hyper parameters was used to build a model. The final model was built with the parameter settings from the cycle with the highest average performance measured by the Area Under the Receiver Operating Characteristic Curve (AUC). Each final model was validated on the hold-out tessel.

2.7. Data preparation

A participant's data was included if they had completed 75% or more of a psychological scale, with this resulted in minimal missing data (< 1% on any variable; range 0.31–0.86%). These small amounts of missing data were checked against demographic characteristics and there were no significant associations. Expectation maximization (EM) was applied to replace missing items for each scale; this is an iterative method used to estimate the parameters of a statistical model, and has been shown to be suitable for this type of missing data (Tsiplaktsis, 2005). One challenge in predicting suicidal behaviour is that the dataset is imbalanced, which means that the observations with suicidal behaviour are a minority class. One technique to deal with imbalanced data is Synthetic Minority Over-sampling Technique (SMOTE), which is an algorithm that creates synthetic records of suicidal behaviour in the training data, such that the prediction algorithms have more examples of suicidal behaviour to learn from (Chawla et al., 2002). In this study, we used SMOTE on the training data of the cross-validation such that the prevalence of suicidal behaviour is 43% in each iteration of cross-validation. We started with model 1, containing 25 features, including the sum scores on the 15 different psychological constructs, age, gender and whether participants had a history of suicide ideation and or suicide attempt. Additionally, we also added theoretically potential relevant subscales such as internal and external entrapment and perceived burdensomeness and thwarted belongingness to examine the effect above and beyond the total scale. Finally, the beck scale for suicide ideation was included twice, once as a 5 item screener, and once as the full scale. In the second analysis, all the 211 separate items of the 15 scales were added to model 1, to test whether this would improve predictive accuracy (model 2).

2.8. Evaluation of techniques

The fields of medicine and biostatistics often deal with classification problems. Health professionals need to decide whether a spot on a patient's arm is a tumour or not, whether a patient is at risk for heart failure etc. Similarly, in the field of psychiatry, mental health professionals classify a patient as having a high or a low risk of future suicidal behaviour. In classification terms, a person can be a case (i.e. at risk for future suicidal behaviour) or a control (i.e. not at risk for future suicidal behaviour). The quality of the classification is then determined by the percentage of true cases who are indeed recognized as cases (true positives), the percentage of true cases who are wrongly classified as controls (false negatives), the percentage of true controls who are rightfully classified as controls (true negatives) and the percentage of true controls who are wrongly classified as cases (false positives). These four outcomes are contained in a so-called confusion matrix. As the controls (thankfully) outnumber the cases in clinical practice and in the general population, there is specific interest in the positive predictive value (TP/ TP + FP) (Walsh et al., 2017a). The overall accuracy of the models is measured with the balanced accuracy, which describes the average proportion of correct classifications and is more suited for classification problems with imbalanced data.

2.9. Variable importance

The Caret package allows the researcher to order variables within a technique according to their importance. Although each model uses its own statistical methods to determine the most important variables, conceptually, variable importance can be understood as an ordering of the variables according to their relative contribution to the model. We will present the top five most important psychological constructs in the paper. The order of all variables per model can be found in the supplementary material. As not all models will select the same variables, a rule of thumb is that variables that occur in most models are indeed most important.

3. Results

3508 respondents finished baseline assessments. At one year follow up, 2420 respondents finished the second assessment. 333 respondents (14%) reported suicide ideation between baseline and follow up, and 50 (2%) reported a suicide attempt.

3.1. Prediction of future suicidal ideation

Table 1 shows the results of the two models on the training and the test dataset (model 1: sum scores, subscales and background variables, model 2: model 1 + all separate items). The results on the test set are for most model quite similar to the training set, which indicates that the results are robust. On the test set, for model 1, all metrics are highly similar across analytic methods. The random forest of model 1 had the highest PPV but the difference with for example generalized linear model seems neglectable. The confusion matrix of the random forest in model 1 and all the receiver operating curves are shown in **Table 2** and **Fig. 1**. The random forest and the gradient boosting algorithm of model 2 improved mildly on PPV when compared to model 1. random forest of model 2 was highly similar to model 1. Interestingly, the other algorithms did slightly worse when all items were added.

3.2. Prediction of future suicide attempt

Table 2 presents the metrics of the different techniques predicting suicide attempt at follow up (2% prevalence) on the training and test data. For most models, the test set results are similar to the training set results which indicates that the models are robust. On the test data, for model 1, the AUC and PPV was highest for the random forest algorithm.

Table 1

comparison of ML techniques predicting suicidal ideation at follow up: Model 1: predictors = sum scores, subscales and background variables: Model 2: model 1 + all separate items.

| Training set | | | | | |
|------------------------|----------------------|-------------|-------------|---------------------------|-------------------|
| Model 1 | area under the curve | sensitivity | specificity | positive predictive value | balanced accuracy |
| Logistic regression | 0.81 | 0.655 | 0.85 | 0.37 | 0.75 |
| K-nearest neighbour | 0.8 | 0.69 | 0.77 | 0.29 | 0.73 |
| Classification tree | 0.74 | 0.6 | 0.74 | 0.34 | 0.72 |
| Random forest | 0.85 | 0.63 | 0.87 | 0.4 | 0.75 |
| Gradient boosting | 0.81 | 0.71 | 0.8 | 0.43 | 0.76 |
| Support vector machine | 0.8 | 0.65 | 0.85 | 0.37 | 0.75 |
| Model 2 | area under the curve | sensitivity | specificity | positive predictive value | balanced accuracy |
| Logistic regression | 0.64 | 0.54 | 0.74 | 0.22 | 0.64 |
| K-nearest neighbour | 0.78 | 0.84 | 0.54 | 0.2 | 0.69 |
| Classification tree | 0.73 | 0.58 | 0.85 | 0.34 | 0.72 |
| Random forest | 0.83 | 0.55 | 0.9 | 0.42 | 0.72 |
| Gradient boosting | 0.79 | 0.53 | 0.91 | 0.44 | 0.72 |
| Support vector machine | 0.69 | 0.48 | 0.81 | 0.26 | 0.65 |
| Test set | | | | | |
| Model 1 | area under the curve | sensitivity | specificity | positive predictive value | balanced accuracy |
| Logistic regression | 0.81 | 0.65 | 0.86 | 0.5 | 0.76 |
| K-nearest neighbour | 0.82 | 0.63 | 0.85 | 0.47 | 0.74 |
| Classification tree | 0.74 | 0.62 | 0.87 | 0.41 | 0.75 |
| Random forest | 0.82 | 0.53 | 0.9 | 0.53 | 0.71 |
| Gradient boosting | 0.84 | 0.71 | 0.8 | 0.43 | 0.76 |
| Support vector machine | 0.78 | 0.59 | 0.85 | 0.45 | 0.72 |
| Model 2 | area under the curve | sensitivity | specificity | positive predictive value | balanced accuracy |
| Logistic regression | 0.65 | 0.49 | 0.81 | 0.35 | 0.65 |
| K-nearest neighbour | 0.83 | 0.89 | 0.57 | 0.31 | 0.73 |
| Classification tree | 0.79 | 0.59 | 0.85 | 0.46 | 0.72 |
| Random forest | 0.83 | 0.6 | 0.88 | 0.52 | 0.74 |
| Gradient boosting | 0.82 | 0.53 | 0.9 | 0.53 | 0.71 |
| Support vector machine | 0.64 | 0.33 | 0.85 | 0.32 | 0.59 |

Table 2

Confusion matrix of the random forest algorithm from model 1 on the test set.

| | Future suicide ideation | No future suicide ideation |
|---|-------------------------|----------------------------|
| Future suicide ideation (prediction) | 70 | 63 |
| No future suicide ideation (prediction) | 63 | 563 |

Again, adding all items (model 2) did not improve the metrics when compared to model 1. Logistic regression and k-nearest neighbors were even worse than when only including the sum scores.

Table 3 and 4 and Fig. 2

3.3. Variable importance

Fig. 3 presents the order of the variables of the algorithm with the highest ppv (random forest) for future suicidal ideation. The first five variables are internal entrapment, defeat, entrapment, mental imagery, and interpersonal needs.

Table 5 presents the top 5 of all machine learning models both for the outcome future suicide ideation and future suicide attempt. For future suicide ideation, the most selected variables are: internal entrapment and perceived burdensomeness. With regard to suicide attempt at follow up, defeat, optimism, internal entrapment and depressive symptoms were amongst the most important variables.

4. Discussion

In this study we compared several popular ML techniques and compared them on their applicability in predicting suicidal ideation

and behaviour within a population-based longitudinal sample.

4.1. Predicting future suicidal ideation

When using the sum scores as predictors, the more regular algorithm logistic regression was amongst the best predictive algorithms. For example, on the test set, when predicting future suicidal ideation, the PPV of the logistic regression was 0.5 (AUC 0.81, BA 0.76) and for the best performing algorithm, random forest, it was 0.53 (AUC 0.83, BA 0.74). As shown in the results, no large differences in any of the metrics were found between the different techniques. Logistic regression, random forests and gradient boosting seemed to do slightly better when compared to the other algorithms, but the difference is unlikely to be clinically relevant.

4.2. Separate items as predictors

When including all items, none of the techniques improved on the AUC. Less complex techniques such as logistic regression and k-nearest neighbors showed a reduction in PPV. The PPV of the random forest on the test set was 0.52 (AUC 0.83, BA 0.74) compared to 0.35 (AUC 0.65, BA 0.65) of the logistic regression. This is in line with Walsh et al's findings (Walsh et al., 2017a). They reported that the random forest performed much better than the multiple logistic regression, (AUC 0.68(0.66–0.67) versus (AUC 0.80(0.80–0.81)) although they did not report on the positive predictive value of the logistic regression making it difficult to compare the metrics. Interestingly, when including all items, the predictive metrics of the complex techniques were highly comparable with the metrics of logistic regression with only the sum scores. Overall, the highest positive predictive value was 53%, indicating that in about 2 or 3 out of 5 patients we could predict the

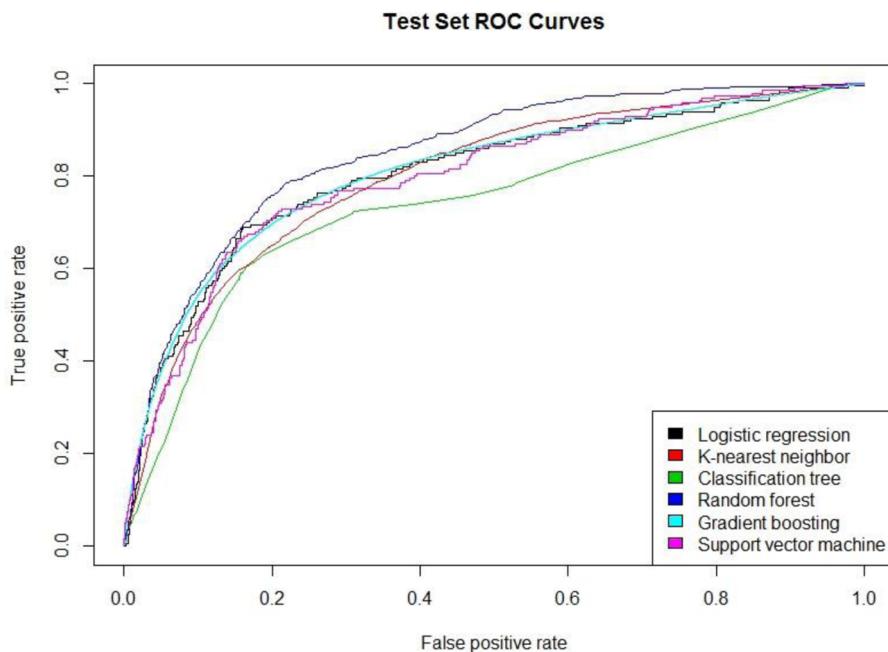


Fig. 1. Receiver operating curves of all algorithms in model 1.

presence of suicide ideation at one year follow up.

4.3. Predicting future suicide attempts

Future suicide attempt at follow-up is clinically the most relevant

variable. The best positive predictive value on the test set was 0.10. This indicates that when we assess all risk factors, we could predict 1 in 10 suicide attempts. As this was a population-based sample, it is important to realize that the base rate was very low (2%). Even after applying oversampling techniques, the prediction of such relatively rare

Table 3

comparison of ML techniques predicting suicide attempt at follow up: Model 1: predictors = sum scores, subscales and background variables: Model 2: model 1 + all separate items.

| Training set | | area under the curve | sensitivity | specificity | positive predictive value | balanced accuracy |
|------------------------|---------|----------------------|-------------|-------------|---------------------------|-------------------|
| Model 1 | Model 2 | | | | | |
| Logistic regression | 0.67 | 0.49 | 0.83 | 0.06 | 0.66 | |
| K-nearest neighbour | 0.76 | 0.65 | 0.8 | 0.06 | 0.72 | |
| Classification tree | 0.7 | 0.57 | 0.8 | 0.06 | 0.69 | |
| Random forest | 0.73 | 0.5 | 0.9 | 0.09 | 0.7 | |
| Gradient boosting | 0.74 | 0.48 | 0.87 | 0.07 | 0.68 | |
| Support vector machine | 0.74 | 0.6 | 0.85 | 0.08 | 0.72 | |
| Test set | | area under the curve | Sensitivity | specificity | positive predictive value | balanced accuracy |
| Model 1 | Model 2 | | | | | |
| Logistic regression | 0.56 | 0.57 | 0.55 | 0.03 | 0.56 | |
| K-nearest neighbour | 0.76 | 0.76 | 0.59 | 0.04 | 0.67 | |
| Classification tree | 0.64 | 0.34 | 0.83 | 0.04 | 0.58 | |
| Random forest | 0.74 | 0.4 | 0.9 | 0.08 | 0.65 | |
| Gradient boosting | 0.75 | 0.39 | 0.91 | 0.09 | 0.65 | |
| Support vector machine | 0.65 | 0.29 | 0.93 | 0.08 | 0.61 | |

| Training set | | area under the curve | Sensitivity | specificity | positive predictive value | balanced accuracy |
|------------------------|---------|----------------------|-------------|-------------|---------------------------|-------------------|
| Model 1 | Model 2 | | | | | |
| Logistic regression | 0.72 | 0.53 | 0.83 | 0.06 | 0.68 | |
| K-nearest neighbour | 0.72 | 0.53 | 0.78 | 0.05 | 0.66 | |
| Classification tree | 0.69 | 0.53 | 0.84 | 0.07 | 0.69 | |
| Random forest | 0.8 | 0.4 | 0.92 | 0.09 | 0.66 | |
| Gradient boosting | 0.8 | 0.47 | 0.88 | 0.08 | 0.67 | |
| Support vector machine | 0.79 | 0.53 | 0.87 | 0.08 | 0.7 | |
| Test set | | area under the curve | Sensitivity | specificity | positive predictive value | balanced accuracy |
| Model 1 | Model 2 | | | | | |
| Logistic regression | 0.53 | 0.47 | 0.55 | 0.02 | 0.51 | |
| K-nearest neighbour | 0.66 | 0.533 | 0.74 | 0.04 | 0.64 | |
| Classification tree | 0.77 | 0.6 | 0.83 | 0.07 | 0.71 | |
| Random forest | 0.8 | 0.4 | 0.92 | 0.09 | 0.66 | |
| Gradient boosting | 0.8 | 0.47 | 0.91 | 0.1 | 0.69 | |
| Support vector machine | 0.63 | 0.4 | 0.9 | 0.08 | 0.65 | |

Table 4

Confusion matrix of the Gradient boosting algorithm from model 2 on the test set.

| | Future suicide attempt | No future suicide attempt |
|--|------------------------|---------------------------|
| Future suicide attempt (prediction) | 7 | 63 |
| No future suicide attempt (prediction) | 8 | 647 |

and complex behaviour is extremely difficult. When training a model to predict cases using matched controls, or a relative small control group, as done in other studies, the prediction accuracy greatly improves (Barak-Corren et al., 2017; Walsh et al., 2017b). However, as soon as the low prevalence is taken into account, the chance of being a case even after a positive test result remains really low. Since suicidal behaviour is such a rare event, even the slightest prediction errors result in high false positive rates, meaning that a person will be incorrectly classified as suicidal. As with the prediction of suicidal ideation, the metrics did not differ so much between algorithms. Still, as suicide attempts are so devastating, an increase in PPV from 6% to 9% may still be relevant from a suicide prevention perspective. Therefore, when implementing machine learning models in a clinical practice, the models should not be seen as substitution of a professional, but as complementary step in the identification and stratification of people at risk for future suicidal behaviour. Although the PPVs are not great, we still argue that the machine learning approach of model building and validation offers hope for suicide prevention. This approach includes multiple checks to see whether a statistical model actually fits the data via techniques such as cross-validation and bootstrapping. This renders the results more robust and reproducible. When adhering to the machine learning approach of model building, logistic regression itself is essentially a form of machine learning, as it is an algorithm to learn from observations.

4.4. Best predictors

The risk factors that were most important to predict future suicidal ideation were internal entrapment, defeat and perceived

burdenomeness. This resonates with the integrated motivational volitional model that states that suicidal thoughts develop in response to feelings of internal entrapment, and that perceived burdenomeness moderates the relation between internal entrapment and suicide ideation (O'Connor and Kirtley, 2018). Although novel ecological momentary assessment studies have shown that the transition from defeat to entrapment to suicide ideation via perceived burdenomeness may be more complicated than suggested by the IMV (Hallensleben et al., 2018; Kleiman et al., 2017), the present findings do indicate the importance of these central elements of the IMV model.

Internal entrapment and depressive symptoms were also the most important variables in the prediction of future suicide attempts. Depressive feelings are not explicitly included in the IMV model but are likely to influence many factors involved in the development of suicidal ideation. The relationship between depression and suicide ideation is well established (Turecki & Brent, 2016). EMA studies have also found that changes in depressive feelings are associated with changes in suicidal ideation, although the direction of this relationship is unknown. Additionally, optimism and resilience were most often selected as important protective variables. Interestingly, the volitional variables, i.e. variables that govern the transition from thoughts to action such as acquired capability were not often included as most important. Also, a history of suicidal attempt only entered any of the models twice, although it is one of the most important known risk factors for future suicidal behaviour (Franklin et al., 2016). These unexpected findings may reflect the fact that the number of suicide attempts was small, that they included a mixture of first time attempts and repeat attempts, and that the overwhelming majority of participants who attempted suicide would also have reported suicide ideation. As a consequence, the study may simply not have been powered to distinguish suicide attempts from suicidal ideation in the models. Future research should also assess the full range of volitional variables (O'Connor and Kirtley, 2018).

4.5. Comparing different techniques

So why did the performance metrics not differ highly between different techniques or when including more information? Several reasons may explain this (Hand, 2006). For one, the sum scores are, of course, based on the separate items, but overall they have less measurement

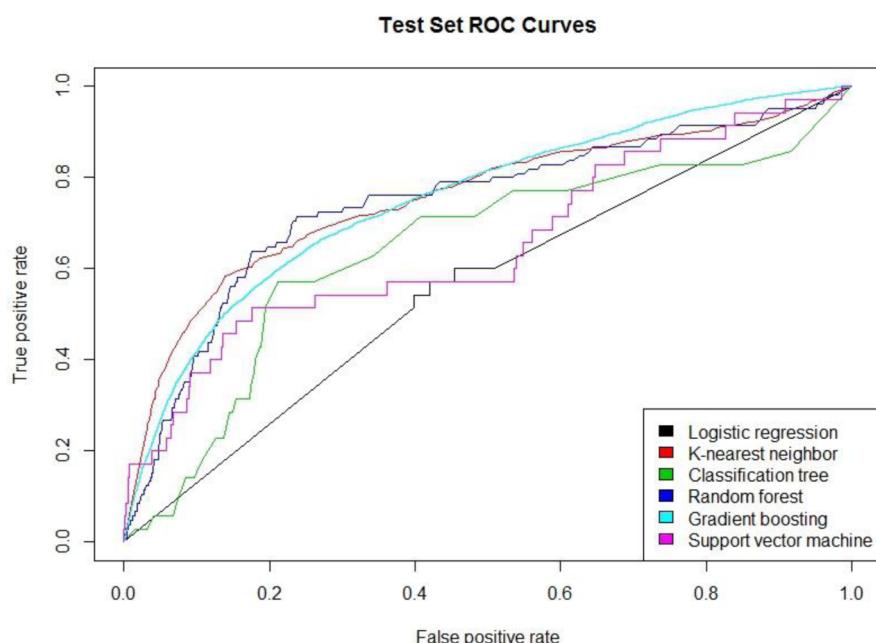


Fig. 2. Receiver operating curves of all algorithms from model 2 on the test set.

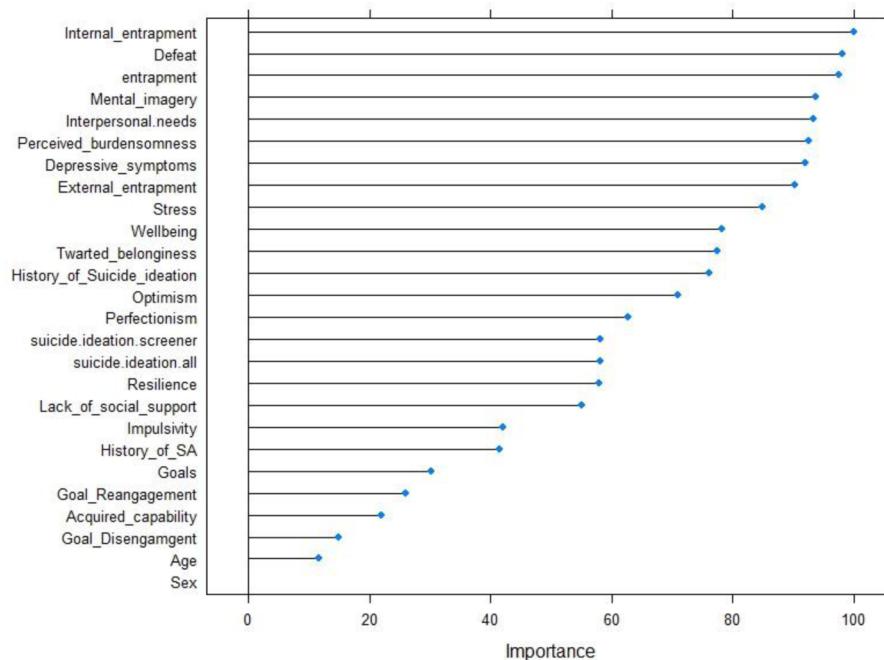


Fig. 3. Ranking of the importance of each variable according to random forest model to predict future suicide ideation.

error. Model fitting is a sequential process that starts with the largest aspects of the data and then progressively moves to uncover smaller aspects. In our data it seems that the initial step taken by relative simple models at the start already explains so much variance that the extra accuracy of more sophisticated models is fairly small. Although it is hypothesized that the complex interactions between different constructs underlies suicidal behaviour, it may be that our data simply did not capture the necessary constructs accurately enough to model their interaction. When adding more detailed information from other domains such as employment status, education and previous health care uptake, the complex non-linear interaction between the different predictors from different domains might result in superior performance of complex machine learning techniques over regular logistical regression.

Another argument relates to the error in class labels. All algorithms are based on the assumption that there are no errors in the classification, or in the assessment of psychological constructs. Even with salient behaviour such as a suicide attempt, we know that people answer inconsistently over time (Eikelenboom et al., 2014). Respondents might not be willing to disclose that they attempted suicide, or even that they still have suicidal ideation. It might be that respondents did take an overdose of pills, but afterwards reasoned that it was not a “real”

attempt. These potential errors in classification make it difficult to improve prediction. Finally, it is intrinsically difficult to predict future human behaviour: many factors can influence the outcome during the time interval between baseline and follow-up.

4.6. Feature selection

In this study, many scales were assessed during a face to face interview session which lasted one hour. This is of course not feasible in daily practice. However, it is highly likely that not all scales or all items are needed. For example, De Beurs et al. (2014) showed that when using computer adaptive testing techniques, the 19 item Beck Scale for Suicide Ideation could be reduced to 4 items without losing discriminative ability (D. P. De Beurs, de Vries, de Groot, de Keijser, and Kerkhof, 2014). We are currently working on a feature selection study that will allow us to select only the most informative single items of all scales, thereby hoping to reduce the cost of assessment.

In sum we found that the differences in accuracy between the algorithms was negligible. Therefore, the decision about which algorithm to implement in clinical practice will depend on other factors. For example, less complex algorithms are often easier to interpret and easier

Table 5
Top 5 most important predictors for future suicidal ideation and future suicide attempt.

| Suicide ideation | Variable 1 | Variable 2 | Variable 3 | Variable 4 | Variable 5 |
|------------------------|-----------------------------|-----------------------------|----------------------------|-----------------------------|-------------------------|
| Logistic regression | History of suicide ideation | Age | History of suicide attempt | Mental imagery | Acquired capability |
| K-nearest neighbour | Internal entrapment | Defeat | Entrapment | Mental imagery | Interpersonal needs |
| Classification tree | History of suicide ideation | Interpersonal needs | Defeat | Perceived burdensomness | Mental imagery |
| Random forest | Defeat | Internal entrapment | Mental imagery | Perceived burdensomness | Depressive symptoms |
| Gradient boosting | Internal entrapment | History of suicide ideation | Depressive symptoms | Suicide ideation full scale | Mental imagery |
| Support vector machine | Internal entrapment | Defeat | Entrapment | Interpersonal needs | Perceived burdensomness |

| Suicide attempt | Variable 1 | Variable 2 | Variable 3 | Variable 4 | Variable 5 |
|------------------------|----------------------------|-----------------------------|---------------------------|-----------------------------|---------------------------|
| Logistic regression | Acquired capability | Defeat | Lack of social support | Mental imagery | Stress |
| K-nearest neighbour | Defeat | Resilience | Internal entrapment | Entrapment | External Entrapment |
| Classification tree | Depressive symptoms | Defeat | Internal entrapment | History of suicide ideation | Suicide ideation screener |
| Random forest | Depressive symptoms | Suicide ideation full scale | Suicide ideation screener | Acquired capability | Perfectionism |
| Gradient boosting | History of suicide attempt | Depressive symptoms | Resilience | Acquired capability | Age |
| Support vector machine | Defeat | Internal entrapment | Resilience | Entrapment | Optimism |

to implement on a computational device. The field of machine learning offers some promising approaches and techniques, however which technique to apply will depend on the situation and the available data. Although complex machine learning techniques did not show added value in the prediction of suicidal behaviour when applied a dataset containing multiple psychological measurements, the addition of data from other domains, such as on employment, education, and previous healthcare uptake might result in better performance of machine learning over regular logistical regression.

Authors' contribution

DdB developed the study concept. In hackaton sessions, DdB, KM, SK, CS, JL, RW, PL, HS and BW worked on the analysis as a group. KM and DdB drafted the initial version. All authors offered feedback on several drafts. RC, KW and SC offered specific input on the method section of the paper, and the interpretation of the findings. All authors approved the final version of the paper for submission.

Funding

This study was funded, in part, by a grant from US Department of defense (W81XWH-12-1-0007). The writing and analysis of this study was funded by the fellowship GGZ scholarship of DdB, awarded by the Netherlands organization for health research and development (ZONMW). Project number is 636,320,002. The funders had no role in study design, data collection, data analysis, data interpretation or writing of this article. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the funders.

Declaration of Competing Interest

All authors declare to have no conflict of interest.

Acknowledgement

DdB developed the study concept. In hackaton sessions, DdB, KM, SK, CS, JL, RW, PL, HS and BW worked on the analysis as a group. KM and DdB drafted the initial version. All authors offered feedback on several drafts. RC, KW and SC offered specific input on the method section of the paper, and the interpretation of the findings. All authors approved the final version of the paper for submission.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.jad.2020.03.081](https://doi.org/10.1016/j.jad.2020.03.081).

References

Barak-Corren, Y., Castro, V.M., Javitt, S., Hoffnagle, A.G., Dai, Y., Perlis, R.H., ..., Reis, B.Y., 2017. Predicting suicidal behaviour from longitudinal electronic health records. *Am. J. Psychiatry* 174 (2), 154–162. <https://doi.org/10.1176/appi.ajp.2016.16010077>.

Beck, A.T., Steer, R.A., Ranieri, W.F., 1988. Scale for suicide ideation: psychometric properties of a self-report version. *J. Clin. Psychol.* 44 (4), 499–505. [https://doi.org/10.1002/1097-4679\(198807\)44:4>499::AID-JCLP2270440404<3.0.CO;2-6](https://doi.org/10.1002/1097-4679(198807)44:4>499::AID-JCLP2270440404<3.0.CO;2-6).

Campbell-Sills, L., Stein, M.B., 2007. Psychometric analysis and refinement of the Connor-Davidson resilience scale (CD-RISC): validation of a 10-item measure of resilience. *J. Trauma. Stress* 20 (6), 1019–1028. <https://doi.org/10.1002/jts.20271>.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *J. Artificial Intelligence Res.* 16, 321–357. <https://doi.org/10.1613/jair.953>.

Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. *J. Health Soc. Behav.* 24 (24), 386–396. <https://doi.org/10.2307/2136404>.

Computing, R.F., 2011. R development core team. R: A Language and Environment for Statistical Computing 55, 275–286.

de Beurs, D., Fried, E., Wetheral, K., Cleare, S., O' Connor, D., Ferguson, E., ..., O' Connor, R., 2018. Exploring the psychology of suicidal ideation: a theory driven network analysis. *Preprint* <https://osf.io/>.

De Beurs, D.P., de Vries, A.L., de Groot, M.H., de Keijser, J., Kerkhof, A.J., 2014. Applying computer adaptive testing to optimize online assessment of suicidal behaviour: a simulation study. *J. Med. Internet Res.* 16 (9), e207. <https://doi.org/10.2196/jmir.3511>.

Dozois, D.J.A., Dobson, K.S., Ahnberg, J.L., 1998. A psychometric evaluation of the beck depression inventory-ii. *Psychol. Assess.* 10 (2), 83–89. <https://doi.org/10.1037/1040-3590.10.2.83>.

Eikelenboom, M., Smit, J.H., Beekman, A.T.F., Kerkhof, A.J.F.M., Penninx, B.W.J.H., 2014. Reporting suicide attempts: consistency and its determinants in a large mental health study. *Int. J. Methods Psychiatr. Res.* 23 (2), 257–266. <https://doi.org/10.1002/mpr.1423>.

Franklin, J.C., Ribeiro, J.D., Fox, K.R., Bentley, K.H., Kleiman, E.M., Huang, X., ..., Nock, M.K., 2016. Risk factors for suicidal thoughts and behaviours: a meta-analysis of 50 years of research. *Psychol. Bull.* <https://doi.org/10.1037/bul0000084>.

Gilbert, P., Allan, S., 1998. The role of defeat and entrapment (arrested flight) in depression: an exploration of an evolutionary view. *Psychol. Med.* 28 (3), 585–598. <https://doi.org/10.1017/S0033291798006710>.

Hallensleben, N., Spangenberg, L., Forkmann, T., Rath, D., Hegerl, U., Kersting, A., ..., Glaesmer, H., 2018. Investigating the dynamics of suicidal ideation: preliminary findings from a study using ecological momentary assessments in psychiatric inpatients. *Crisis*. <https://doi.org/10.1027/0227-5910/a000464>.

Hand, D.J., 2006. Rejoinder: classifier technology and the illusion of progress. *Statistical Science* 21 (1), 30–34. <https://doi.org/10.1214/088342306000000079>.

Harris, E.C., Barracough, B., 1997. Suicide as an outcome for mental disorders. a meta-analysis. *Br. J. Psychiatry* 170 (MAR.), 205–228. <https://doi.org/10.1192/bj.p.170.3.205>.

Hewitt, P.L., Flett, G.L., 1991. Perfectionism in the self and social contexts: conceptualization, assessment, and association with psychopathology. *J. Pers. Soc. Psychol.* 60 (3), 456–470. <https://doi.org/10.1037/0022-3514.60.3.456>.

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 servicemembers (Army STARSS). *JAMA Psychiatry* 72 (1), 49–57. <https://doi.org/10.1001/jamapsychiatry.2014.1754>.

Kleiman, E.M., Turner, B.J., Fedor, S., Beale, E.E., Huffman, J.C., Nock, M.K., 2017. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. *J. Abnorm. Psychol.* <https://doi.org/10.1037/abn000273>.

Kuhn, M., 2008. caret package. *J. Stat. Softw.* 28 (5), 1–26. Retrieved from <http://www.jstatsoft.org/v28/05/paper>.

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. *Appl. Predictive Model.* <https://doi.org/10.1007/978-1-4614-6849-3>.

McManus, S., Meltzer, H., Brugha, T.T., Bebbington, P.P., Jenkins, R., 2009. Adult psychiatric morbidity in England, 2007 results of a household survey. The Health and Social Care Information Centre, Social Care Statistics. <https://doi.org/10.13140/2.1.1563.5205>.

Mitchell, P.H., Powell, L., Blumenthal, J., Norten, J., Ironson, G., Pitula, C.R., ..., Berkman, L.F., 2003. A short social support measure for patients recovering from myocardial infarction: the enriched social support inventory. *J. Cardiopulm. Rehabil.* 23 (6), 398–403. <https://doi.org/10.1097/00008483-200311000-00001>.

Nock, M.K., Hwang, I., Sampson, N.A., Kessler, R.C., 2010. Mental disorders, comorbidity and suicidal behaviour: results from the national comorbidity survey replication. *Mol. Psychiatry* 15 (8), 868–876. <https://doi.org/10.1038/mp.2009.29>.

Nock, M.K., Hwang, I., Sampson, N., Kessler, R.C., Angermeyer, M., Beauvais, A., ..., Williams, D.R., 2009. Cross-national analysis of the associations among mental disorders and suicidal behaviour: findings from the WHO world mental health surveys. *PLoS Med.* 6, e1000123. <https://doi.org/10.1371/journal.pmed.1000123>.

O'Connor, R.C., Kirtley, O.J., 2018. The integrated motivational-volitional model of suicidal behaviour. *SUBMITTED*.

O'Connor, R.C., Nock, M.K., 2014. The psychology of suicidal behaviour. *The Lancet Psychiatry*. [https://doi.org/10.1016/S2215-0366\(14\)70222-6](https://doi.org/10.1016/S2215-0366(14)70222-6).

O'Connor, R.C., Wetherall, K., Cleare, S., Eschle, S., Drummond, J., Ferguson, E., ..., O'Carroll, R.E., 2018. Suicide attempts and non-suicidal self-harm: a national prevalence study of young adults. *SUBMITTED*.

Patton, J.H., Stanford, M.S., Barratt, E.S., 1995. Factor structure of the barratt impulsiveness scale. *J. Clin. Psychol.* 51 (6), 768–774. [https://doi.org/10.1002/1097-4679\(199511\)51:6>768::AID-JCLP2270510607<3.0.CO;2-1](https://doi.org/10.1002/1097-4679(199511)51:6>768::AID-JCLP2270510607<3.0.CO;2-1).

Sareen, J., Cox, B.J., Afifi, T.O., de Graaf, R., Asmundson, G.J.G., ten Have, M., Stein, M.B., 2005. Anxiety disorders and risk for suicidal ideation and suicide attempts. *Arch. Gen. Psychiatry* 62 (11), 1249. <https://doi.org/10.1001/archpsyc.62.11.1249>.

Scheier, M.F., Carver, C.S., Bridges, M.W., 1994. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test. *J. Pers. Soc. Psychol.* 67 (6), 1063–1078. <https://doi.org/10.1037/0022-3514.67.6.1063>.

Ten Have, M., Van Dorsselaer, S., De Graaf, R., 2013. Prevalence and risk factors for first onset of suicidal behaviours in the Netherlands mental health survey and incidence study-2. *J. Affect. Disord.* 147 (1–3), 205–211. <https://doi.org/10.1016/j.jad.2012.11.005>.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ..., Stewart-Brown, S., 2007. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health Qual. Life Outcomes* 5 (1), 63. <https://doi.org/10.1186/1477-7525-5-63>.

Tsikriktsis, N., 2005. A review of techniques for treating missing data in Om survey research. *J. Oper. Manage.* <https://doi.org/10.1016/j.jom.2005.03.001>.

Turecki, G., Brent, D.A., 2016. Suicide and suicidal behaviour 387(10024). *Lancet*, London, England, pp. 1227–1239. [https://doi.org/10.1016/S0140-6736\(15\)00234-2](https://doi.org/10.1016/S0140-6736(15)00234-2).

Van Orden, K.A., Cukrowicz, K.C., Witte, T.K., Joiner, T.E., 2012. Thwarted belongingness and perceived burdensomeness: construct validity and psychometric properties of the interpersonal needs questionnaire. *Psychol. Assess.* 24 (1), 197–215. <https://doi.org/10.1037/a0025358>.

Van Orden, K.A., Witte, T.K., Gordon, K.H., Bender, T.W., Joiner, T.E., 2008. Suicidal desire and the capability for suicide: tests of the interpersonal-psychological theory of suicidal behaviour among adults. *J. Consult. Clin. Psychol.* 76 (1), 72–83. <https://doi.org/10.1037/0022-006X.76.1.72>.

Walsh, C.G., Ribeiro, J.D., Franklin, J.C., 2017a. Predicting risk of suicide attempts over time through machine learning. *Clin. Psychol. Sci.* 5 (3), 457–469. <https://doi.org/10.1177/2167702617691560>.

Walsh, C.G., Ribeiro, J.D., Franklin, J.C., 2017b. Predicting risk of suicide attempts over time through machine learning. *Clin. Psychol. Sci.* <https://doi.org/10.1177/2167702617691560>.

Wetherall, K., Cleare, S., Eschle, S., Ferguson, E., O'Connor, D.B., O'Carroll, R.E., O'Connor, R.C., 2018. From ideation to action: differentiating between those who think about suicide and those who attempt suicide in a national study of young adults. *J. Affect. Disord.* <https://doi.org/10.1016/j.jad.2018.07.074>.

World Health Organization, 2014. Preventing suicing: A global Imperative. WHO <https://doi.org/10.1002/9780470774120>.

Wrosch, C., Scheier, M.F., Miller, G.E., Schulz, R., Carver, C.S., 2003. Adaptive self-regulation of unattainable goals: goal disengagement, goal reengagement, and subjective well-being. *Pers. Soc. Psychol. Bull.* 29 (12), 1494–1508. <https://doi.org/10.1177/0146167203256921>.