Received: 16 March 2023 | Revised: 25 July 2023 | Accepted: 30 October 2023

DOI: 10.1111/sltb.13017

Suicide and Life-Threatening BEHAVIOR

ORIGINAL ARTICLE

How do explicit, implicit, and sociodemographic measures relate to concurrent suicidal ideation? A comparative machine learning approach

René Freichel MSc[1,2] | Sercan Kahveci MSc[3,4] | Brian O'Shea PhD[2,5]

[1]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

[2]Department of Psychology, Harvard University, Cambridge, Massachusetts, USA

[3]Department of Psychology, Paris-Lodron-University of Salzburg, Salzburg, Austria

[4]Centre for Cognitive Neuroscience, Paris-Lodron-University of Salzburg, Salzburg, Austria

[5]School of Psychology, University of Nottingham, Nottingham, England

Correspondence
René Freichel, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129, 1018 WS Amsterdam, Amsterdam, The Netherlands.
Email: r.freichel@uva.nl

Abstract

Introduction: Suicide is a leading cause of death, and decades of research have identified a range of risk factors, including demographics, past self-injury and suicide attempts, and explicit suicide cognitions. More recently, implicit self-harm and suicide cognitions have been proposed as risk factors for the prospective prediction of suicidal behavior. However, most studies have examined these implicit and explicit risk factors in isolation, and little is known about their combined effects and interactions in the prediction of concurrent suicidal ideation.

Methods: In an online community sample of 6855 participants, we used different machine learning techniques to evaluate the utility of measuring implicit self-harm and suicide cognitions to predict concurrent desire to self-harm or die.

Results: Desire to self-harm was best predicted using gradient boosting, achieving 83% accuracy. However, the most important predictors were mood, explicit associations, and past suicidal thoughts and behaviors; implicit measures provided little to no gain in predictive accuracy.

Conclusion: Considering our focus on the concurrent prediction of explicit suicidal ideation, we discuss the need for future studies to assess the utility of implicit suicide cognitions in the prospective prediction of suicidal behavior using machine learning approaches.

KEYWORDS

explicit suicide cognitions, implicit suicide cognitions, machine learning, predictive utility, self-harm, suicidal ideation

INTRODUCTION

Suicide is a major public health challenge and the 10th leading cause of death in North America (Fazel & Runeson, 2020). Decades of research have identified several key risk factors for suicide, including past psychiatric disorders, such as depression (Chesney et al., 2014), past self-harm behavior, suicide cognitions (Chan et al., 2016), and different sociodemographic factors, such as being middle-aged (Berkelmans et al., 2021). Research has also identified more distal

René Freichel and Sercan Kahveci contributed equally to this work.

population-wide risk factors associated with increased risk for suicidal behavior, including economic turmoil (Turecki & Brent, 2016), and seasonality (Freichel & O'Shea, 2023), with increases in deaths by suicide in spring (Christodoulou et al., 2012). In particular, the field has gained a better understanding of explicit cognitions (e.g., negative affect, Gee et al., 2020) and psychophysiological processes (e.g., sleep, Brüdern et al., 2022) in their prediction of daily self-harm behavior and suicide cognitions, thanks to the modern-day abundance of smartphones and the associated popularity of ecological momentary assessment in suicide research (Kleiman et al., 2017).

Implicit self-harm and suicide cognitions represent promising new predictors of self-harm and suicide. These assess the degree to which individuals implicitly associate themselves with constructs such as self-harm, suicide, and death, in contrast to life. To assess these implicit associations, a class of reaction time (RT)-based computerized tasks has been developed. One such task is the self-harm and suicide-related implicit association test (IAT) (Nock & Banaji, 2007a,b). Implicit measures of suicide cognition have shown promise in measuring suicidality, as they may capture automatic biases that are assumed to be more difficult to fake or conceal than regular self-report measures (Greenwald et al., 2009). Accordingly, they were shown to be robust predictors of self-harm (Randall et al., 2013), suicidal ideation, and suicide (Glenn et al., 2019; Nock et al., 2010). Scores on a death IAT were also able to discriminate between individuals with a history of suicide attempts and those without (Sohn et al., 2021), and they correlated with specific types of self-harm behavior (Glenn et al., 2017). A recent systematic review highlighted that the suicide/death IAT has reasonable retrospective and prospective criterion validity, with significant associations with past and future suicidal thoughts and behaviors (Moreno et al., 2022). Brief versions of the death IAT have been developed for use in clinical settings and they were shown to possess good psychometric properties (Millner et al., 2018). Despite these promising findings, a range of studies has questioned the predictive utility of suicide/death IATs and has found no evidence that these tasks are able to distinguish between suicide attempters and non-attempters (Rath et al., 2021; Tello et al., 2020).

Despite the multitude of promising findings, many studies examining the predictive utility of the suicide IAT do not assess whether the implicit measure adds extra information to the prediction of clinical outcomes over and beyond the wide range of commonly accepted explicit and sociodemographic risk factors. For instance, a number of studies included implicit suicide/death associations as the only predictor (Chiurliza et al., 2018), or examined their bivariate associations with suicide-related outcomes (Moreno et al., 2020) without a comparative analysis that also includes explicit associations. This is part of a broader trend where risk factors for suicide have mostly been studied in isolation (Franklin et al., 2017) and it aligns with the lack of focus on incremental validity and utility in clinical psychological science (Hunsley, 2003). Among the studies that do assess the incremental predictive value of the suicide IAT (Nock & Banaji, 2007b; Tello et al., 2020), many used traditional statistical modeling approaches that are prone to overfitting, and may thus overestimate how well these predictors predict suicide when applied to new data. As a solution for these problems, it was proposed that machine learning approaches can be used to generate predictions of suicidal behavior that capitalize on the potential existence of a large number of risk factors, which may share complex (nonlinear) interactions with each other (Fazel & O'Reilly, 2020).

Existing machine learning studies for suicide prediction have generally revealed that diagnostic indicators assessed over a longer period before suicide (e.g., 48 months) were better predictors of suicide compared to indicators over shorter (e.g., 6 months) periods (Gradus et al., 2020). Different sociodemographic factors (including male sex) and prior psychiatric history were among the top predictors of post-hospitalization suicide risk among soldiers hospitalized with psychiatric disorders (Kessler et al., 2015). Another study (Wang et al., 2021) showed that predictors derived from rapid fluctuations in momentary suicide cognitions emerged as strong predictors of future suicide attempts.

The primary goal of our study was to assess whether the IAT adds to the prediction of concurrent self-reported suicidality and desire to self-harm, over and above more easily collected measures, such as sociodemographic factors, self-reported history of self-harm and suicide, and explicit momentary self-harm and suicide cognitions. To the best of our knowledge, this is the first study to critically evaluate the predictive utility of implicit suicide cognitions using machine learning.

MATERIALS AND METHODS

Data source and procedure

We used data from Project Implicit Health (PIH), an online study platform on which respondents can voluntarily complete self-harm and suicide IATs and relevant questionnaires after providing informed consent. We included responses from individuals between April 2012 and November 2018. We only included responses from participants with a residence in the United States (US) due to the potential influence of national differences on the role of different predictor variables, which cannot be quantified in the current study due to the small sample sizes for other countries. We excluded participants with missing values in any of the predictors. From 2012 until 2014, respondents on the website were randomly assigned to either the cutting, suicide, or death IAT. Starting in 2015, the brief death IAT was added and the cutting IAT was paused, so respondents from that point onward were randomly assigned to either the suicide, death, or brief death IAT. Besides the explicit association questions, all other explicit measures and sociodemographic questions were identical across IATs and were presented in random order. The final sample size was 6855.

Measures

Current desire to self-harm or die

From the available data on self-harm and suicide, we selected two explicit items that were administered to participants before and after completing the IAT to monitor potential iatrogenic effects (Cha et al., 2016). The two questions assess participants' desire to self-harm ("How much do you want to hurt yourself right now?"), and their desire to die ("How much do you want to die right now?"). Respondents were instructed to respond to these questions on a five-point scale (0 = not at all, 1 = slightly, 2 = moderately, 3 = strongly, 4 = extremely). The responses before and after the IAT were averaged to increase the sensitivity of the two explicit measures. There were no significant differences between the pre- and post-measurements, implying no iatrogenic effects could be detected. We dichotomized these variables such that individuals were classified on whether they had any desire to self-harm and to die, be it before or after the task. This dichotomization was considered appropriate given our interest in group differences between low-risk and at-risk individuals, our goal to classify for the presence and not the severity of the desire to self-harm or to die, and our emphasis on the prediction of an extreme group (see DeCoster et al., 2009).

Explicit association and sociodemographic information

Mood ("How would you rate your mood right now?") was assessed on a 7-point Likert scale (−3 = extremely positive, 3 = extremely negative). Individuals' explicit associations between self/others and cutting/suicide/death were measured with two single-item measures, one for self and one for others, on a 9-point Likert scale (−4 = extremely strong association of self/others with cutting/suicide/death, 4 = extremely strong association self/others with life); whether the cutting, suicide, or death items were administered depended on which IAT the participant was performing. In addition, we included sociodemographic information (e.g., age, gender, race, ethnicity, educational attainment, and US citizenship) and generated three dummy variables representing the season during which the participant took the test (winter, summer, autumn, with spring as baseline). Due to a large imbalance in the numbers representing the different racial groups, we dichotomized the race variable such that it represented whether participants are White or non-White.

Self-harm and suicide history

Using responses from the abbreviated version of the Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock et al., 2007), we assessed participants' lifetime history of self-injurious thoughts and behaviors. Four items from the SITBI were used, to assess lifetime non-suicidal self-injury ("Have you ever done anything to purposely hurt yourself without wanting to die (for example cutting or burning your skin)?"), lifetime suicidal ideation ("Have you ever had thoughts of killing yourself?"), lifetime suicide plans ("Have you ever actually made a plan to kill yourself?"), and lifetime suicide attempts ("Have you ever made an actual suicide attempt, where you wanted to kill yourself, even just a little?"). Participants responded to those questions on a dichotomous scale ("Yes"/"No").

Implicit associations

We used data from four IATs (Cutting, Suicide, Death, Death brief IAT [BIAT]) to obtain estimates of the implicit association (Greenwald et al., 1998) between the self on the one hand, and cutting, suicide, or death on the other hand, respectively. Displayed one by one at the center of a screen were words relating to the self or others, images related to cutting (i.e., images of forearms cut), or words relating to suicide and death, as well as control images or words unrelated to these topics (i.e., life, alive, thrive, breathing). The cutting IAT used pictorial stimuli for self-harm-related and unrelated stimuli, whereas the suicide and death IATs featured only words. Participants were instructed to correctly classify the displayed stimuli into the appropriate categories by pressing a left or right keyboard key as fast and accurately as possible. During two blocks, the categorization rules were such that stimuli relating to cutting, suicide, or death were to be classified on the same side as self-related words, while the non-cutting or life stimuli were classified on the other side alongside words relating to others. In another set of two blocks, these contingencies were reversed, such that images/words relating to cutting, suicide, or death were to be classified on the same side as words relating to others, and on the opposite side, non-cutting or life stimuli were classified together with self-stimuli.

When target images or words were incorrectly classified, a red "X" appeared on the screen. The brief death IAT (Sriram & Greenwald, 2009) was a shortened version of the death IAT, in which participants classified words relating to death and life, and words relating to the self, but without words describing others or not-self. Various studies have supported the validity of the suicide IAT for predicting suicide attempts (Nock et al., 2010; Tello et al., 2020).

For all IATs, we calculated a standardized difference score (D-score) per participant (Greenwald et al., 2003), which represents the association between "Me" and cutting/suicide/death versus non-cutting/life, while also accounting for the corresponding "Not Me" associations. Positive values on this score imply an association between cutting/suicide/death and the self, whereas negative values on this score represented an association between non-cutting/life and the self. In addition, we computed four D-scores representing subcomponents of the aforementioned overall implicit association D-score. These scores were based on errors and RTs, respectively, both for the association between the self and cutting/suicide/death ("Not Me" associations removed), and the association between other people and cutting/suicide/death ("Me" associations removed) (O'Shea, Glenn, et al., 2020). Following the criteria described in Glenn et al. (2017), we excluded individuals with a high error rate (overall: >30% trial errors; critical blocks: >40% trial errors) and overly fast responses (overall: >10% of RT faster than 300ms; critical blocks: >25% of RT faster than 300ms).

Data analysis

We first constructed six different predictor sets (see Figure 1). All predictor sets included demographics, mood, as well as self-harm and suicide history. The six predictor sets were based on two conditions with three types of predictor sets each. In the first three predictor sets, we either excluded any IAT scores ("no IAT"), included only the overall IAT D-score ("With D-score"), or included the four decomposed D-scores ("With D-score components"). The next three predictor sets were the same as the previous three, but also included the two explicit association variables (i.e., explicit association of self and others with cutting/suicide/death). This resulted in a total of six (3×2) predictor sets. The use of these six predictor sets allowed us to examine the relative contribution of implicit and explicit associations in the prediction of concurrent desire to self-harm or die.

We separately analyzed each IAT type (cutting, suicide, death, death BIAT). We used the cutting IAT dataset to predict whether the respondents currently wish to engage in self-harm, and we used the IAT datasets relating to suicide and death to predict whether or not the respondents currently wish to die. This division was based on the previously reported specificity of these self-harm IATs in Glenn et al. (2017). We divided each of these datasets into 10 folds, stratified by the associated outcome variable, such that every fold included a similar proportion of people with and without a desire to self-harm or die.

We utilized six machine learning methods (see Figure 1B) to predict the desire to self-harm or die. As predictors we used the six aforementioned predictor sets separately within each of the four IAT datasets, to see to what extent adding specific IAT-related variables improved prediction accuracy. The utilized machine learning methods (see Table 1) included both three nonlinear methods, i.e., decision trees (R-package "rpart"; Therneau et al., 2022), random forests (R-package "randomForest"; Liaw & Wiener, 2002), and gradient boosting machines (R-package "gbm"; Greenwell et al., 2022), as well as three linear methods, i.e., linear discriminant analysis (R-package "MASS", Venables & Ripley, 2002), support vector machines (R-package "kernlab", Karatzoglou

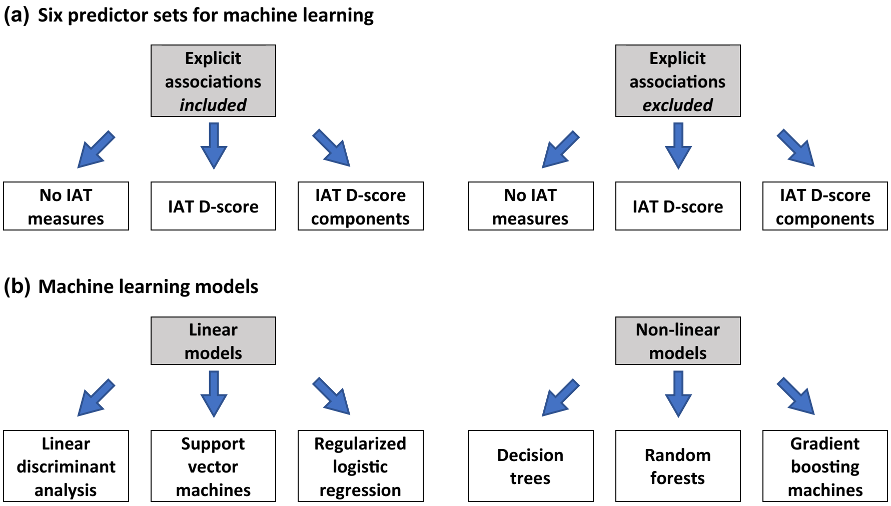
(a) Six predictor sets for machine learning
Explicit associations *included*
No IAT measures
IAT D-score
IAT D-score components
Explicit associations *excluded*
No IAT measures
IAT D-score
IAT D-score components
(b) Machine learning models
Linear models
Linear discriminant analysis
Support vector machines
Regularized logistic regression
Non-linear models
Decision trees
Random forests
Gradient boosting machines

FIGURE 1 Overview of different predictor sets (a) and machine learning models (b). Explicit association refers to the two items indicating the explicit association of self and others with cutting/suicide/death. All models shown in (a) of this figure always included demographics, mood, self-harm and suicide history as predictors.

et al., 2004), and elasticnet regularized logistic regression (R-package "glmnet"; Friedman et al., 2010). These methods were selected to form a representative range of methodologies commonly used in the prediction of health-related outcomes (e.g., Duda et al., 2016).

For every combination of dataset, predictor-set, fold, and machine learning method, we first divided the data into a training set, consisting of 9 out of 10 folds, and a test set, consisting of the remaining 1 out of 10 folds. All machine learning models were trained on the training set to predict the dichotomous outcome variable, and hyperparameters for each method were tuned using repeated 10-fold cross-validation within the training set. The explored hyperparameters and their values were the defaults provided by the *caret* R-package (2008) for each method; we listed these defaults in a document in the osf.io repository of this manuscript.

We then derived cross-validated accuracy metrics by predicting the outcome variable in the test set, on which the data was not trained (Yarkoni & Westfall, 2017). This process was repeated with different folds serving as train and test set, such that each fold served as test set once. The accuracy metrics included the accuracy, kappa, positive and negative predictive value, and Brier scores. Accuracy was defined as the percentage of correctly classified cases. Kappa was defined as the proportion of accuracy above chance, that is, (accuracy−chance accuracy)/(1−chance accuracy). Positive predictive value was defined as the percentage of actually suicidal individuals among all individuals classified as suicidal. Negative predictive value referred to the proportion of actually non-suicidal individuals among all individuals classified as not suicidal. Lastly, the Brier score is a calibration metric (Lindhiem et al., 2020) that represents the accuracy of classification probabilities, and is defined as the mean squared difference between the classification probability and the true class membership, that is, $\frac{1}{N}\sum(f-o)^2$ where f is the classification probability and o is the true class membership, being either 0 or 1 (Brier, 1950). Analyses were performed in R (R Core Team, 2022) using the package caret (Kuhn, 2008). Our analysis scripts and per-fold machine learning results are available on the Open Science Framework (link: https://osf.io/hnb8v/).

RESULTS

Sample characteristics

Respondents were predominantly young and female. They reported a large variability with respect to their current desire to self-harm or die, as well as their history of cognitions and behaviors related to self-harm and suicide (see Table 2).

TABLE 1 Brief explanations of machine learning algorithms used.

Model	Type	Interactions between variables	Description	Software (*R* package)
Decision trees	Nonlinear	Yes	Builds a single decision tree from features	"rpart" Therneau et al., 2022
Random forests	Nonlinear	Yes	Constructs a multitude of decision trees on the basis of subsets of features and data	"randomForest" Liaw & Wiener, 2002
Gradient boosting machines	Nonlinear	Yes	Sequentially generates and combines the output of decision trees that improve the classification of observations that were poorly classified by previous decision trees	"gbm" Greenwell et al., 2022
Linear discriminant analysis	Linear	No	Finds linear combination of features that best separates classes	"MASS" Venables & Ripley, 2002
Support vector machines	Linear	No	Finds the high-dimensional combination of features that best separates classes	"kernlab" Karatzoglou et al., 2004
Elastic net regularized regressions	Linear	No	Combination of LASSO and Ridge regression that suppresses weak predictors to reduce overfitting	"glmnet" Friedman et al., 2010

TABLE 2 Counts and means for all predictors and outcome variables per IAT type.

Variable	IAT type			
	Cutting	Suicide	Death	Death BIAT
Number of participants	964	2322	2278	1291
Outcome variables				
Current desire to self-harm (%)	28.53	31.22	27.92	27.96
Current desire to die (%)	29.15	34.45	33.98	33.54
Demographic predictors				
Age	26.49	25.9	25.79	26.01
Educational level	3.3	3.22	3.26	3.24
Citizenship (%)	87.97	89.28	89.03	89.08
Hispanic (%)	10.48	10.29	10.89	11.39
Gender (%)	68.67	71.79	72.39	70.57
Non-white (%)	21.47	25.32	24.54	24.79
Seasonality predictors				
Tested in summer (%)	22.51	19.51	22.39	14.56
Tested in autumn (%)	21.99	32.52	30.03	24.17
Tested in winter (%)	26.45	21.45	20.85	38.11
Self-harm and suicide history predictors				
Ever engaged in self-harm (%)	60.89	61.24	63.08	64.06
Ever thought of suicide (%)	82.26	84.32	84.64	84.97
Ever planned suicide (%)	36.72	41.82	40.08	43.22
Ever attempted suicide (%)	27.8	33.72	31.61	33.77
Mood predictor				
Mood	0.33	0.29	0.25	0.28
Explicit association predictors				
Explicit association of others with cutting/suicide/death	−1.27	−1.87	−0.98	−1.08
Explicit association of self with cutting/suicide/death	−0.52	−0.64	−0.01	0.09
D-score predictor				
Overall D-score	−0.12	−0.36	−0.36	−0.16
D-score component predictors				
RT-based Self-reference D-score	−0.11	−0.29	−0.32	−0.12
Error-based Self-reference D-score	−0.04	−0.02	−0.09	−0.16
RT-based Other-reference D-score	−0.15	−0.5	−0.47	−0.23
Error-based Other-reference D-score	−0.06	−0.12	−0.08	−0.08

Model prediction for desire to self-harm and die

Comparing machine learning methods, we found that regularized logistic regression, gradient boosting machines, linear discriminant analysis, and support vector machines performed more or less on par with each other, as displayed in Figure 2. Random forests, and especially decision trees, had lower accuracies than the aforementioned four methods, and will not be further interpreted.

As displayed in Figure 2, adding explicit associations to the predictor set significantly improved prediction accuracy; this was especially true for explicit associations between the self and self-harm or suicide (rather than death). The simple IAT D-score tended to improve accuracy when added to a predictor set without explicit associations, especially in the case of the self-harm and death IATs, and to a smaller extent in the death BIAT and suicide IATs. Adding IAT D-score subcomponents instead of the full D-score slightly improved prediction accuracy in

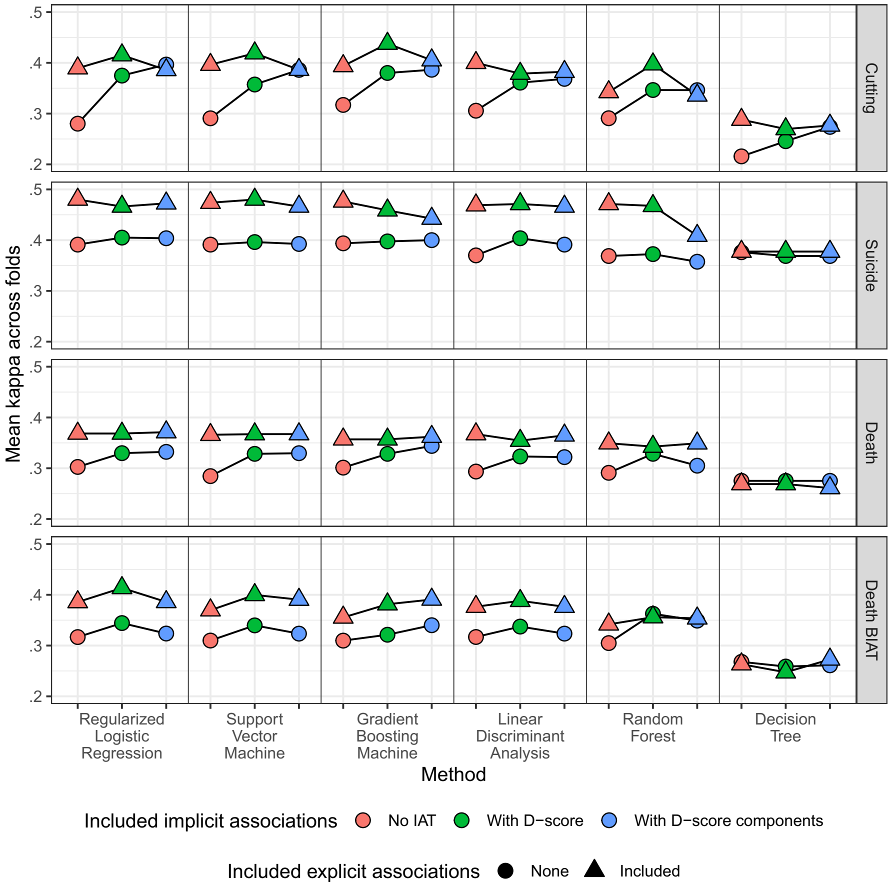

FIGURE 2 Machine learning model comparisons. The figure shows accuracy above chance (Kappa) for all different folds during the cross-validation procedure. Higher kappa indicates a better classification accuracy. Kappa was computed as: (accuracy–chance accuracy)/(1–chance accuracy). All models shown in this figure always included demographics, mood, self-harm and suicide history as predictors. Explicit association refers to the two items assessing the explicit association of self and others with cutting/suicide/death.

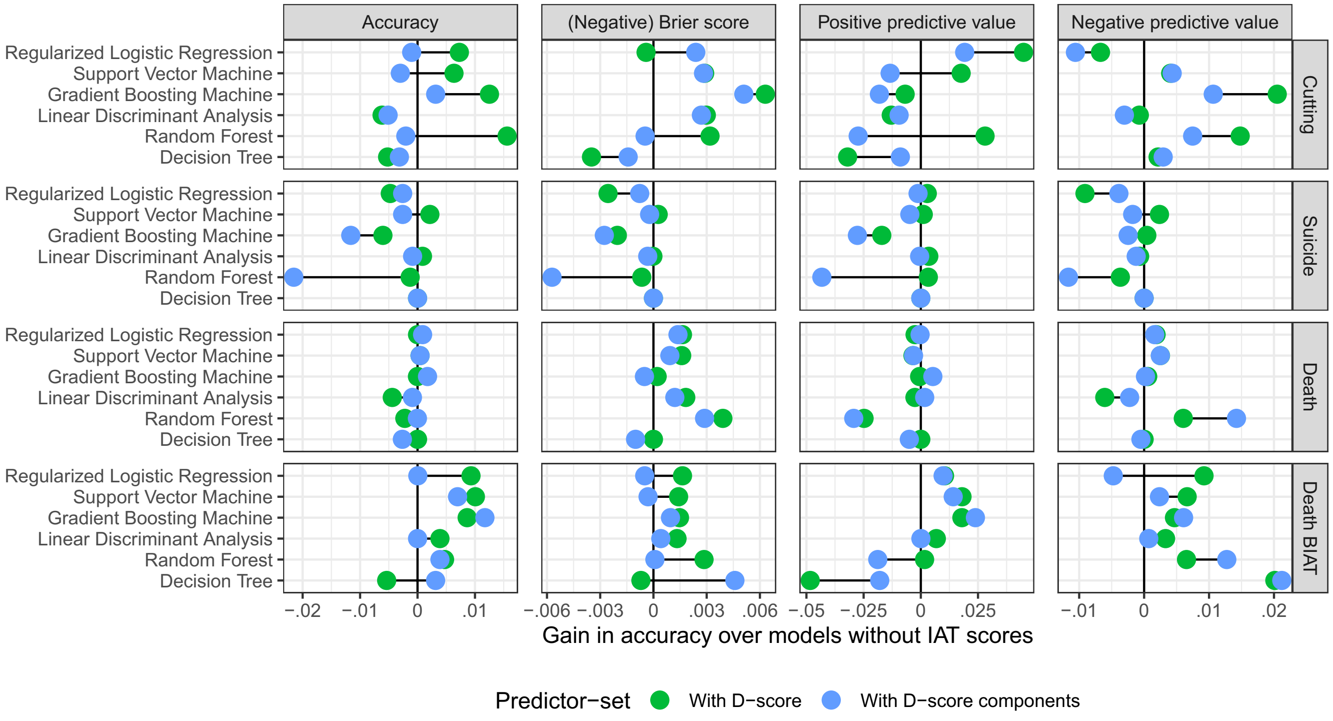

FIGURE 3 Gain in accuracy for different machine learning models (in models with explicit associations included).

case of the cutting and death IATs, and it tended to reduce or leave unaffected the prediction accuracy in case of the brief death IAT and suicide IAT.

Looking at the models that included explicit associations as predictors, the four best-performing machine-learning methods achieved cross-validated kappas

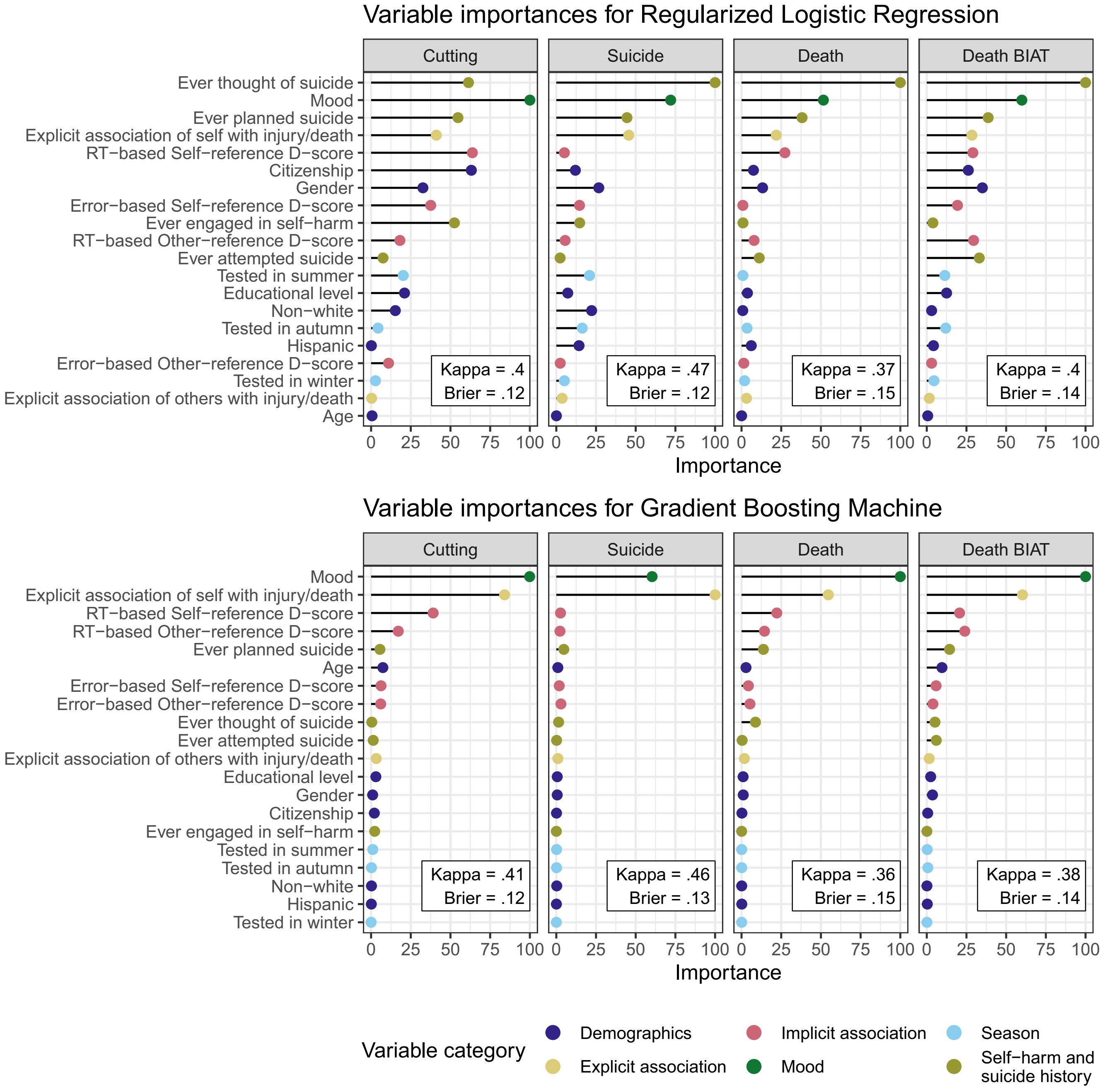

FIGURE 4 Predictor variable importance comparison of regularized logistic regression and gradient boosting machine (in models with explicit associations included).

between 0.35 and 0.48, corresponding to accuracies between 78% and 84%. Predictions were more accurate in the suicide IAT dataset (mean kappa = 0.44 compared to 0.35 in other datasets), though not because of the inclusion of the suicide IAT; there were only little gains, and sometimes losses in accuracy, from adding the IAT D-score as a predictor when explicit associations were already included as a predictor. Rather, the differences in accuracy between IAT types may have been caused by either different sample characteristics, or the different explicit predictors that were used with the different IATs. There were no gains in accuracy for the suicide IAT, death IAT, and death BIAT from adding D-scores as predictors when explicit associations were already included as predictors; the self-harm IAT dataset did see improvements in prediction accuracy from adding D-scores when explicit associations were already included, though these gains were much reduced in comparison to when these explicit associations were not included. Looking at overall accuracy rather than kappa, desire to self-harm was best predicted using gradient boosting, achieving 84% accuracy with all explicit measures and cutting IAT D-scores as predictors; desire to die was best predicted using elasticnet-regularized logistic regression, achieving 80% accuracy with all explicit measures and brief death IAT D-scores as predictors.

Predictive utility of implicit self-harm cognitions

In Figure 3, we illustrated the changes in prediction accuracy that are caused by adding implicit measures, either as a single D-score or as a set of four decomposed indicators, to the full explicit predictor set. Adding the D-score improved prediction accuracy in the cutting IAT and brief death IAT, and it had no effect on the death IAT and sometimes decreased accuracy in the suicide IAT. However, changes in accuracy were minimal when they did occur, ranging from −2.15% to +1.56%. Linear discriminant analysis and decision trees showed no substantial added benefit from adding D-scores or their components as predictors. Besides these effects, there were no consistent patterns in the data with respect to changes in accuracy caused by adding the implicit measures.

Importance of different predictor variables

For two of the most successful machine learning methods, gradient boosting and regularized logistic regression, we further examined the importance of all predictors. This analysis is depicted in Figure 4. Gradient boosting depended primarily on mood and explicit associations with cutting/suicide/death, with the RT-based self-referenced D-score being the third most important variable in three out of four IATs. Elasticnet-regularized logistic regression depended on a broader range of variables, including past suicidal thoughts and behaviors. These variables were especially important in the prediction of suicidality, while mood was the most important variable for the prediction of self-harm. Both methods consistently showed that the suicide IAT was not an important predictor of suicidality.

DISCUSSION

The primary goal of our study was to evaluate the predictive utility of implicit self-harm cognitions in predicting current desire to self-harm and die.

Low predictive utility of implicit associations

Our findings indicate that reasonable accuracy (84% in the best-performing model) can be achieved for the prediction of concurrent desire to self-harm or die. However, the self-harm, suicide, and death IATs offered little (<2%) and often no predictive value on top of explicit measures that are much easier to collect, such as explicit associations, lifetime indicators of suicide or self-harm behavior, and current affective state (e.g., mood). In some cases, the inclusion of the IAT even led to a decrease in accuracy. The death and suicide IATs (but not the brief death IAT) consistently showed no predictive value over and above explicit measures. In comparison, the brief death IAT did improve the prediction of suicidality, which supports prior findings showing that this shorter death IAT does not perform worse than the standard version (Millner et al., 2018). Overall, the limited gain in prediction accuracy across IATs above explicit measures is in line with prior evidence from inpatient samples showing that IAT scores were unrelated to important suicide-related outcome measures (Rath et al., 2021) and did not differ between individuals with and without a recent suicide attempt (Barnes et al., 2017; Tello et al., 2020).

Key risk factors predict desire to self-harm and die

Despite the limited incremental gains in prediction accuracy from utilizing the IAT, our findings do point to a handful of other predictors that allowed us to predict desires to self-harm and die well above chance: mood, history of suicidal behavior, and explicit associations between the self and self-harm, suicide, and death. Previous machine learning studies similarly point to the top predictors of desire to die found in this study, namely explicit associations between the self and death or suicide, past suicide or self-harm behavior and ideation, and current affective state (Kessler et al., 2015). These findings suggest that it may thus be fruitful to understand suicide as a complex classification problem (Ribeiro et al., 2016).

Limitations

Several limitations with respect to the data source and study design should be considered. First, we used data from an online study platform that allows users to voluntarily complete a self-harm or suicide-related IAT. Thus, the self-selection of participants in PIH studies led to an opt-in sample that is disproportionately young, female, and shows an unusually high base rate of self-harm and suicide-related past behaviors. Various studies have shown, however, that patterns found in Project Implicit Health samples have been associated with meaningful regional and objective outcomes (Giasson & Chopik, 2020; O'Shea, Watson, et al., 2020). While the higher rate of suicidality and self-harm in the current sample limited the generalizability of our findings, it did allow us more power to detect potential gains in prediction accuracy from adding predictors. This makes it all the more striking that the IAT failed to improve prediction accuracy beyond 1.6% at most. Second, we used dichotomized averages of single-item measures to assess the presence of suicidal ideation and the desire to self-injure. Although both items showed a high internal test–retest reliability (i.e., correlation between pre-, and post-IAT responses), there is some evidence (Millner et al., 2015) pointing to the risk of misclassification when using single-item measures to assess the presence of suicidal behaviors.

Finally, we believe that our finding of limited predictive value in implicit IATs is at least in part because the outcome variables were explicit self-report measures whose variance was easily explained with similarly explicit self-report measures of known risk factors, with little added value in an implicit measure. One important reason we use implicit measures is their purported ability to bypass the respondent's biases and intentionally untruthful responses. This benefit, if present, will not come to light when we use the IAT to predict participants' self-reported desire to self-harm or die; instead, it may lead to a misclassification when a truly suicidal patient reports not wanting to die but nevertheless displays suicide-endorsing attitudes on the IAT. Conversely, when a participant is correctly classified as self-reporting no suicidality on the basis of self-report measures, it is unclear whether that is because of a true self-reported absence of a desire to die, or because the participant is denying their elevated suicidality on both predictor and outcome measures. Hence, it remains possible that implicit measures of associations between self and death offer incremental utility in the prospective prediction of self-harm and suicidal *behavior*, as reported by Sohn et al. (2021). Thus, future research could use similar machine learning techniques using clinical samples to obtain a thorough judgment of the IAT's contribution to the prospective prediction of self-harm and suicide behavior.

CONCLUSION

Using a dataset with 6855 responses collected over 7 years, our findings speak to the value of machine learning in utilizing the complexity of known risk factors for self-harm and suicide in the prediction of the desire to self-harm and die. Our results challenge the notion that implicit self-harm and suicide IATs aid in the prediction of concurrent self-reported desires to self-harm and die over and beyond known risk factors. Further research is needed to determine the optimal temporal resolution (retrospective, concurrent, prospective; Freichel et al., 2023) at which implicit self-harm and suicide cognitions predict suicidal thoughts and actual behaviors over and beyond known explicit risk factors.

CONFLICT OF INTEREST STATEMENT

The author(s) declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

DATA AVAILABILITY STATEMENT

This study involved an analysis of existing data rather than new data collection. We report how we determined our sample size in the study. Our data analysis scripts and per-fold machine learning results are available on the Open Science Framework (link: https://osf.io/hnb8v/).

ETHICS STATEMENT

We are using existing data collected from the Project Implicit Health (PIH) database. Human-subject ethics approval was granted to PIH by the University of Virginia.

ORCID

René Freichel https://orcid.org/0000-0002-9478-0575
Sercan Kahveci https://orcid.org/0000-0002-4139-5710
Brian O'Shea https://orcid.org/0000-0001-9736-238X

REFERENCES

Barnes, S. M., Bahraini, N. H., Forster, J. E., Stearns-Yoder, K. A., Hostetter, T. A., Smith, G., Nagamoto, H. T., & Nock, M. K. (2017). Moving beyond self-report: Implicit associations about death/life prospectively predict suicidal behavior among veterans. *Suicide and Life-Threatening Behavior*, *47*(1), 67–77. https://doi.org/10.1111/sltb.12265

Berkelmans, G., van der Mei, R., Bhulai, S., & Gilissen, R. (2021). Identifying socio-demographic risk factors for suicide using data on an individual level. *BMC Public Health*, *21*(1), 1702. https://doi.org/10.1186/s12889-021-11743-3

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Brüdern, J., Hallensleben, N., Höller, I., Spangenberg, L., Forkmann, T., Rath, D., Strauß, M., Kersting, A., & Glaesmer, H. (2022). Sleep disturbances predict active suicidal ideation the next day: An ecological momentary assessment study. *BMC Psychiatry*, *22*(1), 65. https://doi.org/10.1186/s12888-022-03716-6

Cha, C. B., Glenn, J. J., Deming, C. A., D'Angelo, E. J., Hooley, J. M., Teachman, B. A., & Nock, M. K. (2016). Examining potential iatrogenic effects of viewing suicide and self-injury stimuli. *Psychological Assessment*, *28*(11), 1510–1515. https://doi.org/10.1037/pas0000280

Chan, M. K. Y., Bhatti, H., Meader, N., Stockton, S., Evans, J., O'Connor, R. C., Kapur, N., & Kendall, T. (2016). Predicting suicide following self-harm: Systematic review of risk factors and risk scales. *The British Journal of Psychiatry*, *209*(4), 277–283. https://doi.org/10.1192/bjp.bp.115.170050

Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: A meta-review. *World Psychiatry*, *13*(2), 153–160. https://doi.org/10.1002/wps.20128

Chiurliza, B., Hagan, C. R., Rogers, M. L., Podlogar, M. C., Hom, M. A., Stanley, I. H., & Joiner, T. E. (2018). Implicit measures of suicide risk in a military sample. *Assessment*, *25*(5), 667–676. https://doi.org/10.1177/1073191116676363

Christodoulou, C., Douzenis, A., Papadopoulos, F. C., Papadopoulou, A., Bouras, G., Gournellis, R., & Lykouras, L. (2012). Suicide and seasonality. *Acta Psychiatrica Scandinavica*, *125*(2), 127–146. https://doi.org/10.1111/j.1600-0447.2011.01750.x

DeCoster, J., Iselin, A.-M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, *14*, 349–366. https://doi.org/10.1037/a0016956

Duda, M., Ma, R., Haber, N., & Wall, D. P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Translational Psychiatry*, *6*(2), Article 2. https://doi.org/10.1038/tp.2015.221

Fazel, S., & O'Reilly, L. (2020). Machine learning for suicide research–can it improve risk factor identification? *JAMA Psychiatry*, *77*(1), 13–14. https://doi.org/10.1001/jamapsychiatry.2019.2896

Fazel, S., & Runeson, B. (2020). Suicide. *New England Journal of Medicine*, *382*(3), 266–274. https://doi.org/10.1056/NEJMra1902944

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143*(2), 187–232. https://doi.org/10.1037/bul0000084

Freichel, R., & O'Shea, B. A. (2023). Suicidality and mood: The impact of trends, seasons, day of the week, and time of day on explicit and implicit cognitions among an online community sample. *Translational Psychiatry*, *13*(1), 1–9. https://doi.org/10.1038/s41398-023-02434-1

Freichel, R., Wiers, R., O'Shea, B., McNally, R. J., & de Beurs, D. (2023). Between the group and the individual: The need for within-person panel study approaches in suicide research. *Psychiatry Research*, *330*, 115549. https://doi.org/10.1016/j.psychres.2023.115549

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Gee, B. L., Han, J., Benassi, H., & Batterham, P. J. (2020). Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. *Digital Health*, *6*, 2055207620963958. https://doi.org/10.1177/2055207620963958

Giasson, H. L., & Chopik, W. J. (2020). Geographic patterns of implicit age bias and associations with state-level health outcomes across the United States. *European Journal of Social Psychology*, *50*(6), 1173–1190. https://doi.org/10.1002/ejsp.2707

Glenn, C. R., Millner, A. J., Esposito, E. C., Porter, A. C., & Nock, M. K. (2019). Implicit identification with death predicts suicidal thoughts and behaviors in adolescents. *Journal of Clinical Child & Adolescent Psychology*, *48*(2), 263–272. https://doi.org/10.1080/15374416.2018.1528548

Glenn, J. J., Werntz, A. J., Slama, S. J. K., Steinman, S. A., Teachman, B. A., & Nock, M. K. (2017). Suicide and self-injury-related implicit cognition: A large-scale examination and replication. *Journal of Abnormal Psychology*, *126*(2), 199–211. https://doi.org/10.1037/abn0000230

Gradus, J. L., Rosellini, A. J., Horváth-Puhó, E., Street, A. E., Galatzer-Levy, I., Jiang, T., Lash, T. L., & Sørensen, H. T. (2020). Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. *JAMA Psychiatry*, *77*(1), 25–34. https://doi.org/10.1001/jamapsychiatry.2019.2905

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037//0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers. (2022). *gbm: Generalized Boosted Regression Models* (2.1.8.1). https://github.com/gbm-developers, https://CRAN.R-project.org/package=gbm

Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment*, *15*, 443–445. https://doi.org/10.1037/1040-3590.15.4.443

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab—An S4 package for kernel methods in R. *Journal of Statistical Software*, *11*, 1–20. https://doi.org/10.18637/jss.v011.i09

Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., Brown, M., Cai, T., Colpe, L. J., & Cox, K. L. (2015). Predicting suicides after psychiatric hospitalization in US Army soldiers: The Army study to assess risk and resilience in servicemembers (Army STARRS). *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. *Journal of Abnormal Psychology*, *126*(6), 726–738. https://doi.org/10.1037/abn0000273

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lindhiem, O., Petersen, I. T., Mentch, L. K., & Youngstrom, E. A. (2020). The importance of calibration in clinical psychology. *Assessment*, *27*(4), 840–854. https://doi.org/10.1177/1073191117752055

Millner, A. J., Coppersmith, D. D. L., Teachman, B. A., & Nock, M. K. (2018). The brief death implicit association test: Scoring recommendations, reliability, validity, and comparisons with the death implicit association test. *Psychological Assessment*, *30*(10), 1356–1366. https://doi.org/10.1037/pas0000580

Millner, A. J., Lee, M. D., & Nock, M. K. (2015). Single-item measurement of suicidal behaviors: Validity and consequences of misclassification. *PLoS One*, *10*(10), e0141606. https://doi.org/10.1371/journal.pone.0141606

Moreno, M., Gutiérrez-Rojas, L., & Porras-Segovia, A. (2022). Implicit cognition tests for the assessment of suicide risk: A systematic review. *Current Psychiatry Reports*, *24*(2), 141–159. https://doi.org/10.1007/s11920-022-01316-5

Moreno, M., Porras-Segovia, A., Lopez-Castroman, J., Peñuelas-Calvo, I., Díaz-Oliván, I., Barrigón, M. L., & Baca-García, E. (2020). Validation of the Spanish version of the death/suicide implicit association test for the assessment of suicidal behavior. *Journal of Affective Disorders Reports*, *1*, 100012. https://doi.org/10.1016/j.jadr.2020.100012

Nock, M. K., & Banaji, M. R. (2007a). Assessment of self-injurious thoughts using a behavioral test. *American Journal of Psychiatry*, *164*(5), 820–823. https://doi.org/10.1176/ajp.2007.164.5.820

Nock, M. K., & Banaji, M. R. (2007b). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Consulting and Clinical Psychology*, *75*(5), 707–715. https://doi.org/10.1037/0022-006X.75.5.707

Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-injurious thoughts and behaviors interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment*, *19*(3), 309–317. https://doi.org/10.1037/1040-3590.19.3.309

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science*, *21*(4), 511–517. https://doi.org/10.1177/0956797610364762

O'Shea, B. A., Glenn, J. J., Millner, A. J., Teachman, B. A., & Nock, M. K. (2020). Decomposing implicit associations about life and death improves our understanding of suicidal behavior. *Suicide and Life-Threatening Behavior*, *50*(5), 1065–1074. https://doi.org/10.1111/sltb.12652

O'Shea, B. A., Watson, D. G., Brown, G. D. A., & Fincher, C. L. (2020). Infectious disease prevalence, not race exposure, predicts both implicit and explicit racial prejudice across the United States. *Social Psychological and Personality Science*, *11*(3), 345–355. https://doi.org/10.1177/1948550619862319

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Randall, J. R., Rowe, B. H., Dong, K. A., Nock, M. K., & Colman, I. (2013). Assessment of self-harm risk using implicit thoughts. *Psychological Assessment*, *25*(3), 714–721. https://doi.org/10.1037/a0032391

Rath, D., Teismann, T., Schmitz, F., Glaesmer, H., Hallensleben, N., Paashaus, L., Spangenberg, L., Schönfelder, A., Juckel, G., & Forkmann, T. (2021). Predicting suicidal behavior by implicit associations with death? Examination of the death IAT in two inpatient samples of differing suicide risk. *Psychological Assessment*, *33*(4), 287–299. https://doi.org/10.1037/pas0000980

Ribeiro, J. D., Franklin, J. C., Fox, K. R., Bentley, K. H., Kleiman, E. M., Chang, B. P., & Nock, M. K. (2016). Letter to the editor: Suicide as a complex classification problem: Machine learning and related techniques can advance suicide prediction–a reply to Roaldset (2016). *Psychological Medicine*, *46*(9), 2009–2010. https://doi.org/10.1017/S0033291716000611

Sohn, M. N., McMorris, C. A., Bray, S., & McGirr, A. (2021). The death-implicit association test and suicide attempts: A systematic review and meta-analysis of discriminative and prospective utility. *Psychological Medicine*, *51*(11), 1789–1798. https://doi.org/10.1017/S0033291721002117

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, *56*(4), 283–294. https://doi.org/10.1027/1618-3169.56.4.283

Tello, N., Harika-Germaneau, G., Serra, W., Jaafari, N., & Chatard, A. (2020). Forecasting a fatal decision: Direct replication of the predictive validity of the suicide–implicit association test. *Psychological Science*, *31*(1), 65–74. https://doi.org/10.1177/0956797619893062

Therneau, T., Atkinson, B., & port, B. R. (producer of the initial R., & maintainer 1999–2017. (2022). *rpart: Recursive Partitioning and Regression Trees* (4.1.16). https://CRAN.R-project.org/package=rpart

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

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Wang, S. B., Coppersmith, D. D., Kleiman, E. M., Bentley, K. H., Millner, A. J., Fortgang, R., Mair, P., Dempsey, W., Huffman, J. C., & Nock, M. K. (2021). A pilot study using frequent inpatient assessments of suicidal thinking to predict short-term postdischarge suicidal behavior. *JAMA Network Open*, *4*(3), e210591. https://doi.org/10.1001/jamanetworkopen.2021.0591

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: a Journal of the Association for Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

How to cite this article: Freichel, R., Kahveci, S., & O'Shea, B. (2023). How do explicit, implicit, and sociodemographic measures relate to concurrent suicidal ideation? A comparative machine learning approach. *Suicide and Life-Threatening Behavior*, *00*, 1–12. https://doi.org/10.1111/sltb.13017