Empirical Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Clinical Psychological Science
2019, Vol. 7(5) 941–957

DOI: 10.1177/2167702619838464
www.psychologicalscience.org/CPS
SAGE

Predicting Imminent Suicidal Thoughts and Nonfatal Attempts: The Role of Complexity

Jessica D. Ribeiro[1], Xieyining Huang[1], Kathryn R. Fox[2], Colin G. Walsh[3], and Kathryn P. Linthicum[1]
[1]Department of Psychology, Florida State University; [2]Department of Psychology, Harvard University; and [3]Departments of Bioinformatics and Psychiatry, Vanderbilt University Medical Center

Abstract
For decades, our ability to predict suicidal thoughts and behaviors (STBs) has been at near-chance levels. The objective of this study was to advance prediction by addressing two major methodological constraints pervasive in past research: (a) the reliance on long follow-ups and (b) the application of simple conceptualizations of risk. Participants were 1,021 high-risk suicidal and/or self-injuring individuals recruited worldwide. Assessments occurred at baseline and 3, 14, and 28 days after baseline using a range of implicit and self-report measures. Retention was high across all time points (> 90%). Risk algorithms were derived and compared with univariate analyses at each follow-up. Results indicated that short-term prediction alone did not improve prediction for attempts, even using commonly cited "warning signs"; however, a small set of factors did provide fair-to-good short-term prediction of ideation. Machine learning produced considerable improvements for both outcomes across follow-ups. Results underscore the importance of complexity in the conceptualization of STBs.

Keywords
suicide, suicidal behavior, prediction, imminent risk, machine learning

Received 8/31/18; Revision accepted 2/7/19

Each year, suicide claims the lives of approximately 45,000 people in the United States and nearly 1 million people worldwide (Centers for Disease Control and Prevention, 2018; World Health Organization, 2018). Nonfatal suicide attempts are even more prevalent, with an estimated 10 to 25 attempts for every suicide death (Crosby, Gfroerer, Han, Ortega, & Parks, 2011). These figures are further dwarfed by the prevalence of suicidal thoughts, with nearly 4% of Americans endorsing suicidal thoughts in the past year (Substance Abuse and Mental Health Services Administration, 2017). Given the scope and seriousness of the problem, suicide prevention has been cited as a public health imperative. In response, there have been sweeping calls to action and suicide research has sharply increased over the past several decades (Franklin et al., 2017). Unfortunately, however, these efforts have not translated to meaningful declines in the suicide rate; instead, rates have climbed by nearly 25% in the past 2 decades (Centers for Disease Control and Prevention, 2018).

To understand this perplexing trend, Franklin and colleagues (2017) conducted a comprehensive meta-analysis of longitudinal studies examining risk factors for suicidal thoughts and behaviors (STBs). Results were sobering: Five decades of research had produced hundreds of studies and thousands of statistical tests designed to advance knowledge of suicide risk, yet our ability to predict STBs was only marginally better than chance (i.e., AUCs = 0.56–0.58). Moreover, our ability to predict had not improved since the inception of longitudinal suicide research; that is, it was at near-chance levels in 1965 and remained at that level through 2015 (Franklin et al., 2017). A series of finer-grained meta-analyses echoed these results, demonstrating that the factors most commonly cited as the strongest predictors

Corresponding Author:
Jessica D. Ribeiro, Department of Psychology, Florida State University, 1107 W. Call St., Tallahassee, FL 32306-4301
E-mail: ribeiro@psy.fsu.edu

of future STBs (e.g., prior suicidal behavior, hopelessness, psychopathology) still produced prediction on par with random guessing (Bentley et al., 2016; Chang et al., 2016; Huang et al., 2017a; Huang, Fox, Ribeiro, & Franklin, 2018; Ribeiro et al., 2016a, 2018; Smith, Velkoff, Ribeiro, & Franklin, 2018; Witte, Gauthier, Huang, Ribeiro, & Franklin, 2018).

Results of these meta-analytic efforts may seem inconsistent with prominent suicide theories and decades of clinical wisdom. However, as noted by Franklin et al. (2017), several major methodological constraints were common across studies. Given the ubiquity of these constraints across studies, methodological issues may account for the surprisingly weak prediction estimates observed in prior research.

Among the methodological constraints detected across suicide-prediction studies, two were particularly notable. First, the vast majority of studies have been focused on the prediction of eventual STBs (i.e., long-term; distal) rather than imminent STBs (i.e., short-term; acute). Follow-up periods for most studies were extremely long; the average follow-up was more than a decade. By comparison, short-term prediction studies were rare. Less than 1% of studies had follow-up periods shorter than 1 month and no studies had follow-up periods shorter than 2 weeks (Franklin et al., 2017). Although recommendations about "warning signs" based on expert clinical consensus have been proposed (e.g., Rudd et al., 2006), very few studies have tested their predictive validity over appropriate (i.e., short-term) time frames. Of the few prospective studies that have been focused on short-term prediction, the majority have involved small selected samples, potentially compromising generalizability (e.g., Britton & Conner, 2010; Busch, Fawcett, & Jacobs, 2003). As it stands, meta-analytic evidence suggests that the effects of risk factors on eventual STBs are modest at best (Franklin et al., 2017; Ribeiro et al., 2016a; Ribeiro et al., 2018). It is possible, however, that the same factors will have considerably stronger effects over shorter, more clinically useful time frames. That is, factors such as hopelessness or agitation may not be particularly strong predictors of a suicide attempt that occurs a decade later; however, they may be very strong predictors of a suicide attempt that occurs in subsequent days or weeks.

Second, most studies examined risk factors in isolation (e.g., univariate associations) or small sets of risk factors combined in fairly rudimentary ways (e.g., screeners; sum scores; two- or three-way linear interactions). The implicit assumption underlying this perspective is that the nature of suicide risk is simple; that is, the presence of a circumscribed set of necessary risk factors combined in a specific way is sufficient to predict STBs. For decades, we have relied on humans to make sense of data. As a result, our research has been focused on testing intuitive and cognitively manageable hypotheses about suicide risk. Clinically, this approach is also appealing. Clinicians are tasked with making judgments about risk in session with clients; cognitively manageable conceptualizations of risk are critical.

Yet, it is also possible that a simple conceptualization does not accurately reflect the nature of suicide risk; instead, suicide risk may be complex (Ribeiro et al., 2016b). Although not yet at the forefront of clinical psychology, the principle of complexity is ubiquitous among many other disciplines, including modern theories in biology, neuroscience, and affective science (Edelman & Gally, 2001; Barrett & Simmons, 2015). When applied to suicide, this perspective would hold that STB risk emerges from the complex combination of a multitude of factors. Reasoning from this perspective, our inability to predict suicidal phenomena accurately may be an artifact of underestimating complexity and applying overly simplified explanations. Therefore, although prediction is weak when risk factors are considered in isolation or within small theoretically informed sets (Franklin et al., 2017), our ability to predict using the same predictors may be enhanced by identifying the complex and optimized combinations among these risk factors.

In the present investigation, we addressed these two major methodological constraints. First, we focused on short-term prediction. One of the main challenges of short-term prediction studies is that STBs are rare, and suicide-prediction research has conventionally relied on in-person data-collection methods. As a result, designs have relied on geographically restricted samples and necessarily long follow-up periods to detect a sufficient number of target outcomes. Online data-collection methods can facilitate targeted, large-scale data-collection efforts focused on low base-rate events. Beyond allowing for larger data-collection efforts, online methods have additional unique advantages over traditional in-person methods, including speed of recruitment, ease of participant contact, greater likelihood of disclosing mental-health issues, improved attention to instructions, and greater geographic diversity (Casler Bickel, & Hackett, 2013; Hauser & Schwarz, 2015). Relative to in-person methods, online data collection is better suited to address many of the challenges of studying imminent STBs prospectively. In this study, we implemented a multilayer authentication approach to data collection consistent with best-practice recommendations (e.g., Bull, Vallejos, Levine, & Ortiz, 2008; Ip, Barnett, Tenerowicz, & Perry, 2010) and which we have used successfully in prior studies (see Franklin et al., 2016). Using this approach, we recruited over 1,000 individuals worldwide who were at elevated risk of imminent STBs. With follow-up assessments occurring at 3, 14, and 28 days after baseline, this study represents the largest and shortest-term prospective study of STBs among severely suicidal and self-injuring individuals to date.

Second, we used machine learning to examine the potential complexity underlying suicide risk. Broadly defined, machine learning is a branch of artificial intelligence based on the assumption that systems can learn from data, identify patterns within data, and make decisions about data with minimal human intervention. Machine-learning methods hold several distinct advantages over conventional statistical approaches typically used for the prediction of suicidal thoughts and behaviors. Chief among them, machine learning is designed to identify the most accurate and parsimonious algorithm that optimally predicts a target outcome. In doing so, machine-learning algorithms can accommodate a large number of factors and simultaneously consider highly complex combinations among those factors. Because conventional approaches rely on humans to detect meaningful patterns in data, traditional approaches are far more limited when making sense of potentially complex phenomena. Moreover, conventional statistical approaches are vulnerable to overfitting when applied to high-dimensional data sets (i.e., data sets that include a large number of potential predictors). *Overfitting* refers to spurious inflation of model performance that occurs when a model capitalizes on the idiosyncrasies (i.e., “noise”) of a given data set. An overfit model will perform well when applied to the data set in which it was derived but perform poorly once tested in new data. Protecting against overfitting is an integral feature of a machine-learning-based approach. Although conventionally derived models can produce statistically significant effects, statistical significance—especially when predicting low base-rate phenomena such as suicidal behaviors—may not correspond with predictive accuracy, robustness, or clinical significance. Machine learning, by contrast, is designed to balance predictive accuracy with generalizability.

Recent efforts support the viability of using machine learning to predict suicidal behaviors, considerably improving prediction accuracy (Barack-Corren et al., 2017; Kessler et al., 2015; Walsh, Ribeiro, & Franklin, 2017). Results of these efforts also align with a complex conceptualization of risk. Across studies, risk algorithms have required the integration of a large number of factors to achieve accurate prediction; however, this may be an artifact of the nature of the input data. To date, studies have relied on administrative data systems, most frequently electronic health records (EHRs). Although data from these sources are relevant and can produce accurate prediction, the nature of EHR and administrative data is fairly nonspecific to suicide. As a result, the complexity observed within the algorithms produced in these studies may be an artifact of the nonspecific nature of the input data. The present study addressed this issue, focusing on suicide-specific input variables assessed using explicit and implicit measures. We selected these factors on the basis of the existing evidence, leading theoretical perspectives, and published clinical consensus statements. Studies to date have also been focused exclusively on the prediction of nonfatal suicide attempts or suicide death, typically identifying events post hoc within existing administrative data sources. In this study, we focused on the prediction of nonfatal suicide attempts, as well as suicidal ideation, assessed prospectively using widely accepted clinical assessments.

Often, the importance of predicting suicidal ideation is overshadowed by that of suicidal behaviors, given the potential lethality of the latter. Even outside of machine-learning efforts, less attention has been paid to the prediction of suicidal ideation than to attempts or death (Franklin et al., 2017). Our rationale for including suicidal ideation in this study was two-fold. First, we reasoned that the prediction of suicidal ideation per se would be clinically relevant. Suicidal ideation, regardless of whether it results in suicidal behaviors, is associated with considerable morbidity. For individuals suffering from suicidal thoughts, the experience is often distressing and impairing, frequently interfering with their quality of life. Given the associated distress and impairment, the presence of suicidal ideation has important implications for treatment, regardless of whether a patient goes on to attempt suicide. As a clinician, knowing which patients may develop suicidal thoughts would be highly valuable, as the presence of ideation could influence the course of treatment (e.g., ideation may be an important treatment target). Furthermore, contrary to common clinical wisdom, the risk of suicide death associated with the presence of suicide ideation is comparable with the risk of prior suicide attempts; that is, although suicide attempts are often considered among the strongest univariate predictors of suicide death, meta-analytic data suggest that suicide ideation confers comparable levels of risk (Ribeiro et al., 2016a). Taken together, especially considering the prevalence of suicidal ideation relative to suicidal behaviors, advancing our ability to predict suicidal ideation per se has direct clinical relevance.

Second, we also reasoned that advancing our understanding of the nature of suicidal ideation and its prediction would be highly relevant for basic science. Relative to suicidal behaviors, suicidal thoughts can be less discrete, more heterogeneous phenomena and longer lasting compared with an isolated suicide attempt. These differences may have important implications for prediction and the nature of suicidal phenomena; however, this remains to be examined empirically.

In sum, the objective of this study was to advance our ability to predict STBs by addressing major methodological issues pervasive in prior prediction efforts. In this effort, we focused specifically on the prediction of imminent suicidal ideation and nonfatal suicide attempts among high-risk individuals. Across all time points, we expected that machine-learning risk algorithms would outperform univariate prediction. Moreover, we expected machine-learning risk algorithms to be consistent with a complex conceptualization of risk, requiring a large number of factors considered in complex ways to produce accurate prediction. We anticipated that this approach would produce accurate prediction across all time points, but prediction would be strongest as ideation and attempts became more imminent (e.g., at 3 or 14 days vs. 28 days).

Method

Participants

Participants were 1,021 adults recruited from online Web forums focused on mental health, suicide, and self-injury. Ages ranged from 18 to 63 years; the median age at baseline was 25 years (M = 26.54 years; SD = 7.31). The sample was predominantly female (66.4%); the remainder were male (28.5%), other (1.4%), or preferred not to disclose sex (3.6%). Most participants identified as White (78.9%); the remainder identified as Asian (6%), Black/African American (4%), Hispanic or Latino (4%), Native American (1%), or other (6%).

Procedures

All study procedures were approved by the Institutional Review Boards at Vanderbilt University and Florida State University.

Recruitment. Participants were recruited internationally from targeted online forums focused on mental health, self-injury, and suicide. Individuals who expressed interest in the study were asked to complete a brief questionnaire to determine eligibility. Inclusion criteria were as follows: (a) 18 years of age or older; (b) English fluency; and (c) recent, frequent nonsuicidal self-cutting (i.e., ≥ 2 self-cutting episodes in the past 2 weeks), a nonfatal suicide attempt in the past year, or recent, frequent, active suicidal ideation (i.e., active ideation more days than not in the past 2 weeks).

Several steps were taken to guard against threats to validity at this stage. First, to reduce the chances of individuals entering the study under false pretenses, inclusion criteria were not included in study advertisements or provided to potential participants until the process of informed consent. Second, relevant items used to determine eligibility were embedded in a much larger questionnaire that included irrelevant items about suicide, self-injury, and mental health. Third, we also included pairs of duplicate items and free-response items in the eligibility questionnaire. Before study entry, these items were reviewed by the study PI for inconsistent and/or irrelevant responding. Free-response items were reviewed to ensure English fluency. Fourth, only individuals with unique IP addresses were allowed to participate; this was implemented as a means of reducing the likelihood of participants entering into the study more than once.

Data collection. All data were collected anonymously. To ensure participant anonymity, we contacted and compensated participants using unidentifiable e-mails (i.e., e-mail aliases that did not include identifiable information).[1] Consent was indicated by selecting a box affirming informed consent and providing only an e-mail alias. Once informed consent was obtained, participants were provided with a unique, randomly generated participant identification number. Assessments occurred at baseline and 3, 14, and 28 days after baseline. Each assessment included a battery of measures, which are described in detail below. All measures were programmed to have a responsive design that automatically formatted for optimized use on a variety of devices, including smartphones, tablets, laptops, and desktops. Each assessment battery was accessible to participants for 48 hr and designed to take no longer than 1 hr to complete. Compensation was provided within 24 hr of completion by way of electronic Amazon.com gift cards using a compensation schedule designed to increase participant retention (i.e., $10 for each of the first three assessments, $20 for the final assessment, and a $20 bonus for completing all assessments).

Measures

Demographics. Information about age, sex, race, ethnicity, employment status, education history, sexual orientation, and native language were obtained using a brief self-report questionnaire.

Modified affect-misattribution paradigm. The affect-misattribution paradigm (AMP; Payne, Cheng, Govorun, & Stewart, 2005) is a brief computer-based task designed to measure implicit affect toward stimuli. On each trial of the AMP, the computer will flash an emotionally evocative stimulus for 75 ms, a blank screen for 125 ms, an ambiguous Chinese symbol for 100 ms, and finally a gray screen that remains until the participant presses a key.

Respondents are instructed to press one key (i.e., "p") if they judge the Chinese symbol to be more pleasant than the average symbol and another key (i.e., "q") if they judge it to be less pleasant than the average symbol. Respondents are asked to ignore the emotionally evocative stimuli during their judgments of the Chinese symbols. Despite this instruction, several studies have shown that evaluations are influenced by the nature of the stimuli; more pleasant stimuli generated more pleasant evaluations of subsequent Chinese symbols (see Payne et al., 2005). This occurs because individuals misattribute the affect evoked by the earlier emotional stimulus as being evoked by the later ambiguous Chinese symbol. Through this misattribution, the AMP assesses implicit affective reactions to the emotional stimuli that are flashed at the beginning of each trial. The AMP used in this study included 48 trials (one for each picture). Stimuli included death- or suicide-related images that ranged in intensity from low (e.g., body bag, noose, morgue) to extreme (e.g., corpse immediately following fatal gunshot to the head). Cronbach's αs were .94 for the overall suicide AMP, .85 for the low-intensity suicide images, .85 for the moderate-intensity suicide images, and .86 for the high-intensity suicide images. Positive images were also used to assess implicit attitudes toward positive stimuli (Cronbach's α = .80). The AMP was administered at baseline.

Explicit affective image ratings. Explicit affect toward a range of suicide stimuli and positive stimuli was assessed using a self-report survey. On each trial, an image was presented and respondents were asked to rate it on a scale from 0 (*extremely unpleasant*) to 5 (*neutral*) to 9 (*extremely pleasant*). Images were drawn from the set used for the AMP tasks described above. A total of 10 images were used to assess explicit affect toward suicide, using moderate-intensity stimuli depicting a range of suicide death methods (Cronbach's α = .90). Five images were used to assess affect toward positive stimuli (Cronbach's α = .79). Explicit ratings were collected at baseline.

Modified Self-Injurious Thoughts and Behaviors Interview. The Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock, Holmberg, Photos, & Michel, 2007) is an assessment of self-injurious thoughts and behaviors. The modified version used in this study was the self-report form, which has been used in prior studies (e.g., Franklin, Puzia, Lee, & Prinstein, 2014; Franklin et al., 2017). In this study, we used modules assessing nonsuicidal self-injury (NSSI), suicide plans, preparations for suicide, aborted suicide attempts, interrupted suicide attempts, and nonfatal suicide attempts. Respondents were asked to report on lifetime, past year, past month, and past week history at baseline. At all subsequent assessments, respondents were asked to report on events since baseline.

Acquired Capability for Suicide Scale–Fearlessness About Death. The *Acquired Capability for Suicide Scale–Fearlessness About Death* (ACSS-FAD; Ribeiro et al., 2014) is a single-factor, seven-item, self-report scale designed to assess fearlessness about death. All items are rated using a Likert-type scale ranging from 0 (*not at all like me*) to 4 (*very much like me*); larger total scores indicate greater fearlessness about death. The ACSS-FAD was administered at all assessments. Reliability was acceptable at all time points (αs = .85–.89).

Affective States Questionnaire. The Affective States Questionnaire (ASQ; Hendin, Maltsberger, & Szanto, 2007) is a self-report questionnaire assessing the presence, severity, and duration of nine different intense, negatively valenced affective states. These states include anxiety, rage, desperation, loneliness, hopelessness, abandonment, self-hatred, guilt, and humiliation. The ASQ was administered at all time points. Internal consistency of the scale at all assessments was good (αs = .88–.93).

Beck Hopelessness Scale. The Beck Hopelessness Scale (BHS; Beck & Steer, 1988) is a 20-item self-report measure designed to assess attitudes (positive and negative) about the future. Respondents are asked to identify whether each statement is true (1) or false (0) for them. Higher total scores on the BHS are indicative of greater negative attitudes about the future. Respondents completed the BHS at all assessments. The scale's internal consistency was strong (αs = .91–.93).

Beck Scale for Suicide Ideation. The Beck Scale for Suicide Ideation (BSS; Beck & Steer, 1991) is a self-report instrument assessing suicidal desire and resolved plans and preparation. The scale consists of 21 items rated from 0 to 2. Higher total scores correspond to greater severity of suicidal symptoms. In this study, the BSS was administered at all time points. In machine-learning models examining suicide ideation as an outcome, only the suicidal desire subscale was used, as the items derived from the resolved plans and preparations subscale included questions related to suicidal behaviors (e.g., past suicide attempts, suicide plans, preparations for suicide, etc.). Note that these subscales were empirically derived and have demonstrated good psychometric properties in past research (Joiner, Rudd, & Rajab, 1997). As measured by Cronbach's α, the internal consistency of the total scale and suicidal desire subscale were acceptable (total range = .87–.88; desire subscale range = .86–.89).

Brief Agitation Measure. The Brief Agitation Measure (BAM; Ribeiro, Bender, Selby, Hames, & Joiner, 2011) is a three-item self-report measure used to assess agitation in the past week. Items are rated on a 7-point Likert-type scale. Total scores range from 3 to 21; higher scores reflect greater agitation severity. Agitation was assessed using the BAM at all time points (αs = .84–.89).

Brief Symptom Index-18. The Brief Symptom Index-18 (BSI-18; Derogatis, 2000) measures psychological symptoms experienced in the past 7 days. The scale includes 18 items that are rated on a 5-point scale from 0 (*not at all*) to 4 (*extremely*). The total score reflects a global severity index of distress; higher scores indicate higher levels of psychological distress. Cronbach's α indicated strong reliability at all assessments (.92–.94).

Disgust With Life Scale. The Disgust With Life Scale (DWLS; Chu, Bodell, Ribeiro, & Joiner, 2015; Ribeiro, Bodell, & Joiner, 2012) is a self-report inventory designed to assess disgust with the self, others, and the world. Using a 5-point Likert scale, respondents are instructed to indicate agreement with a given item. Greater levels of disgust with life are indicated by higher total scores. The DWLS demonstrated strong internal consistency at all time points (Cronbach's αs = .91–.92).

Insomnia Severity Index. The Insomnia Severity Index (ISI; Bastien, Vallières, & Morin, 2001) is a seven-item self-report questionnaire designed to evaluate insomnia symptoms. Items assess the severity of initial (sleep onset), middle (sleep maintenance), and terminal (early morning awakening) problems as well as satisfaction, impairment, and distress associated with existing sleep problems. Each item is rated on a 5-point Likert scale that ranges from 0 (*not at all*) to 4 (*extremely*). Higher scores indicate greater insomnia symptom severity. Internal consistency of the ISI was acceptable at all time points (αs = .86–.90).

Social Functioning Scale. The Social Functioning Scale (SFS; Birchwood, Smith, Cochrane, Wetton, & Copestake, 1990) is a self-report questionnaire designed to evaluate domains of basic social functioning and self-care. Items from this scale were used to assess behaviors consistent with social withdrawal. These items were administered at all time points. As this scale was not designed to be a unitary factor, Cronbach's alpha is not appropriate. For data analyses, items were considered independently.

Statistical modeling

All statistical analyses were performed in the R software environment (R Core Team, 2014). Data were preprocessed using IBM SPSS (Version 24), R via *glm* in base R, the R packages *randomForest* (Liaw & Wiener, 2018) and *pROC* (Robin et al., 2018).

Missing data. Missing data, which were minimal (< 1%), were handled using multiple imputation. Data were not imputed for outcome variables; participants who did not have outcome data available for a particular follow-up time point were excluded from analyses at that time point only.

Outcome variables. The main outcomes of interest for this study were suicide ideation and nonfatal suicide attempts. For ideation analyses, we were primarily interested in developing models to predict the presence (vs. absence) of ideation. Accordingly, BSS-Desire scores were dichotomized such that a score of 0 corresponded to the absence of ideation and scores greater than 0 corresponded to the presence of ideation. We reasoned that presence versus absence of suicide ideation would be most clinically useful because (a) episodes of suicide ideation can be fairly long lasting; therefore, predicting changes in ideation over short-term intervals among individuals already experiencing ideation may not be that useful; and (b) small changes in BSS-Desire scores may not always represent clinically meaningful changes, but shifts from a score of 0 to one greater than 0 would be clinically meaningful.[2] As a predictor, BSS-Desire was retained as a continuous score. For attempt analyses, the SITBI item assessing presence versus absence of nonfatal suicide attempt was used.

Univariate analyses. As discussed above, one possible explanation for the poor prediction performance observed in prior research (see Franklin et al., 2017) is the focus on distal or long-term rather than proximal or short-term prediction. To evaluate this possibility, we also examined univariate effects of our predictor variables on suicide ideation (as measured by BSS-Desire scores) and nonfatal suicide attempts (as measured by the SITBI) at 3, 14, and 28 days after baseline using nonregularized univariate logistic regression. Bootstrap optimism corrections were also applied to these analyses, as described below.

Machine-learning analyses. Outcome variables were the same as those used in univariate analyses (i.e., BSS-Desire; SITBI nonfatal suicide attempt). For each model, only baseline variables were used as predictors. Predictors were identical across outcomes, except that suicidal desire at baseline (i.e., BSS-Desire) was excluded from predictive models of ideation. Random forest was selected as the primary predictive modeling framework in this effort. Though we considered alternate models (e.g., least absolute shrinkage and selection operator, or LASSO) that achieved comparable performance, we ultimately chose random forest because it holds several advantages particularly relevant to the current effort. First, random forest is flexible and robust to several data-analytic issues (e.g., outliers, nonnormal distributions). Given the nonnormal distributions of several features in our data set (e.g., suicide attempts), this was particularly useful. Second, relative to other predictive modeling frameworks, the process underlying random forest is fairly straightforward, resulting in models that are often more readily interpretable.

In brief, random forest is a nonparametric ensemble learning method that consists of a set of decision trees developed through recursive sampling of bootstrapped subsets of predictor data. A tree-like data structure is derived by randomly combining predictors to form nodes (i.e., *parent*) and subnodes (i.e., *child*). Splits in the decision trees, which represent binary (yes/no) questions phrased in terms of predictors, divide the predictor space. The number of predictors at each node can vary; in this study, this parameter was set at the square root of the total number of predictors. Predictor selection at each node was determined by using the error minimization of the mean square error (*MSE*) approach (James, Witten, Hastie, & Tibshirani, 2013). This approach minimized *MSE* across all other randomly selected predictors at a given node. A single decision tree is the result of iterating until a terminal node (i.e., a node that does not have a child node) is achieved. This process is then repeated a set number of times to produce a multitude of decision trees; in this study, this parameter was set at 500. Results are aggregated across decision trees via majority voting from the 500 trees to determine risk estimates.

Bootstrapping was used to estimate and adjust for model optimism (i.e., overfitting). Using this approach, the predictive model is first trained using all available data and bootstrap replicates are then derived from the original data set. The models derived from the bootstrapped replicates are subsequently applied to the original data, which results in an estimate of "out-of-bag" performance. Differences between bootstrapped and out-of-bag performance are calculated for each bootstrap sample. The average of these differences reflects the degree of optimism in the original model performance. The original model performance is then corrected by subtracting this estimate. Common alternatives to this approach include cross-validation and holdout sets; however, evidence suggests that bootstrapping may provide more conservative estimates of model performance and lower absolute and mean-squared errors, particularly in smaller samples (Hastie, Tibshirani, & Friedman, 2001; Smith, Seaman, Wood, Royston, & White, 2014). Given the size of the present sample, we elected to apply the bootstrap. In this study, we used 100 bootstrap replicates.

Variable importance. Relative importance of predictors within each model was also quantified. Values for each predictor are randomly permuted and classification accuracy is reassessed using the newly permuted values. Accuracy estimates using permuted data are subsequently subtracted from the original performance estimates that were based on unpermuted data. Differences in performance are then averaged over all trees and then standardized using the standard deviation of the differences. Taken together, this provides an estimate of the relative importance of predictors with respect to overall model accuracy.

Performance metrics. Model performance was evaluated using area under the receiver operating characteristic curve (AUC), precision, and recall. AUC is a common index of discrimination performance; that is, the accuracy of identifying case subjects from control subjects. An AUC of 1.0 represents perfect discrimination, whereas 0.50 represents chance discrimination. In general, the following breakdown is used to summarize performance: 0.90 to 1.0 = excellent; 0.80 to 0.90 = good; 0.70 to 0.80 = fair; 0.60 to 0.70 = poor; 0.50 to 0.60 = fail. Although a commonly used metric, AUC can produce overly optimistic estimates of performance when there is an imbalance in the number of case subjects versus control subjects.

In instances of case imbalance, precision and recall are suggested alternatives. *Precision* (positive predictive value) refers to the proportion of true positives over all case subjects identified as positive (i.e., the number of individuals who attempted suicide over the number of individuals who were predicted to attempt suicide). *Recall* (sensitivity) refers to the proportion of actual positives identified as such (i.e., the proportion of suicide attempters identified as suicide attempters). Precision and recall metrics can be considered using the same 0.50-to-1.0 descriptors noted above for AUC.

Model calibration was also assessed. Predictive models are designed to output probabilities of an outcome across individuals. Model calibration represents a model's ability to yield predictions that reflect underlying outcome probabilities in the target population. More simply stated, *calibration* refers to the extent to which an algorithm accurately predicts risk on an individual level. In this study, we evaluated model calibration using Brier scores. Brier scores range from 0 to 1, and 0 represents perfect calibration.

Results

Baseline descriptives

At baseline, nearly 90% of the sample endorsed engaging in nonsuicidal self-injury at least once in their lifetime. The most frequent method reported was nonsuicidal self-cutting; approximately 80% of the sample reported at least one self-cutting episode in their lifetime. Among individuals with a lifetime history, many were still actively engaging in the behavior: Roughly 23% reported a self-cutting episode in the past week, 41% in the past month, and more than 66% in the past year. About 12% of participants with a history of NSSI reported no desire to stop engaging in NSSI, and more than 91% reported that it was likely that they would engage in NSSI in the future.

Planning and preparing for suicide were also common. Most participants (96%) endorsed having a suicide plan at least once in their lifetime and just over half of the sample (51%) reportedly thought about their suicide plan at least once in the week before baseline. Of individuals who had a suicide plan, 52% reported their plan was highly specific, involving a specific method used in a specific place and at a specific time. Most individuals also expressed strong intent to act on the suicide plan; less than 2% reported "absolutely no intent" and more than 34% reported "extremely high intent" to act on the plan or plans. More than 76% of the sample also endorsed having made preparations to attempt suicide in the past, such as acquiring means and writing a suicide note.

Lifetime prevalence of suicidal behaviors was high. The majority of the sample endorsed a history of nonfatal suicide attempts (62.3%). Among individuals who endorsed a suicide attempt history, most had attempted more than once (75.6%). The number of lifetime attempts ranged from 0 to 126 (M = 3.92, SD = 11.08, median = 1.0). Approximately 28.6% of the sample reported attempting suicide at least once in the past year, 8.2% in the past month, and 2.7% in the week before baseline. Aborted suicide attempts were also common, occurring in 57.7% of the sample. On average, participants reported a mean of 6.29 aborted suicide attempts in their lifetimes (SD = 26.94, median = 1.00). About 37.3% of the sample reported an aborted suicide attempt in the year before baseline (M = 1.21, SD = 3.08, median = 0) and 13.4% in the month prior (M = 0.22; SD = 0.71, median = 0). Interrupted attempts were less frequent, as expected; 28.6% reported a lifetime history. The number of interrupted attempts in the year before baseline ranged from 0 to 26 (M = 0.39, SD = 2.17, median = 0).

Retention rates

The majority of the participants (97.3%) completed at least one follow-up assessment after baseline. At 3, 14, and 28 days after baseline, retention rates were 95.5%, 93.1%, and 90.8%, respectively.

Outcome prevalence

Suicidal ideation was present among 79.3% (n = 772), 79.7% (n = 757), and 80.1% (n = 742) of the sample that completed assessments at 3, 14, and 28 days after baseline, respectively. Prevalence of nonfatal attempts over the same time periods was 2.5% (n = 24), 5.3% (n = 50), and 8.2% (n = 76).

Univariate analyses

For full results, refer to Tables 1 and 2.

Suicide ideation. Univariate prediction of suicide ideation was weak across all metrics for most predictors. The average AUC of all predictors was approximately 0.61 across time points; estimates ranged from 0.47 to 0.84. Precision ranged from 0.76 to 0.96, with an average of about 0.88 across time points. Recall, on average, fell in the poor range, with an average estimate of 0.52 across time points. Individual-predictor recall estimates ranged from 0.05 to 0.97. A fairly wide range was also observed for Brier scores, with estimates between 0.14 and 0.76 and an average of 0.44 across time points.

Several univariate predictors were particularly notable. BHS, self-predicted likelihood of future attempt, and prior-month planning evidenced fair-to-good prediction as indexed by AUC, precision, and recall metrics; Brier scores approached 0.30 (see Table 1). Prediction using baseline suicidal-ideation scores (BSS-Desire) was good at 3 and 14 days, with improved calibration (Brier scores = 0.15); at 28 days, prediction remained fair, but calibration suffered slightly (Brier score = 0.26).

Suicide attempt. Univariate analyses predicting nonfatal suicide attempt produced AUCs that ranged between 0.43 and 0.77 in the prediction of suicide attempts, with an average of approximately 0.61. Precision was weak across all time points, ranging from 0.02 to 0.41 with an average of 0.11 across all time points and predictors. The average recall estimate across all predictors and time points was 0.57, though estimates spanned from weak to excellent. Calibration was poor for most predictors across time points, with an average of 0.36 across all predictors and time points. No univariate predictor emerged as particularly strong, especially with respect to precision.

Machine-learning analyses

Model performance. A total of 51 variables were included in building the random forest models for suicide attempts; for suicide ideation models, baseline suicide ideation was omitted to result in a total of 50 baseline variables. Given that the duration of suicide ideation may last for days or weeks, we reasoned that omitting baseline ideation would ensure more conservative prediction models.[3] Full model-performance metrics, including confusion matrices for both outcomes across follow-up, are presented in Tables 3 and 4.

Table 1. Univariate Analyses Predicting Suicidal Desire Across Follow-Up

	3 days					14 days					28 days				
Variable	AUC	95% CI	Precision	Recall	Brier	AUC	95% CI	Precision	Recall	Brier	AUC	95% CI	Precision	Recall	Brier
Aborted suicide attempt															
Past month	0.56	[0.54, 0.58]	0.95	0.15	0.68	0.57	[0.55, 0.58]	0.96	0.16	0.67	0.57	[0.55, 0.58]	0.96	0.16	0.66
Lifetime	0.61	[0.57, 0.65]	0.86	0.61	0.39	0.60	[0.56, 0.63]	0.85	0.62	0.39	0.61	[0.57, 0.65]	0.85	0.62	0.39
Past year	0.63	[0.60, 0.66]	0.91	0.42	0.50	0.63	[0.60, 0.66]	0.91	0.42	0.49	0.63	[0.60, 0.66]	0.90	0.42	0.49
Fearlessness about death (ACSS-FAD)	0.62	[0.59, 0.66]	0.86	0.63	0.37	0.62	[0.58, 0.66]	0.86	0.63	0.37	0.63	[0.59, 0.67]	0.86	0.63	0.37
AMP-suicide															
High intensity	0.49	[0.46, 0.53]	0.79	0.52	0.49	0.52	[0.48, 0.56]	0.81	0.48	0.51	0.49	[0.46, 0.53]	0.77	0.52	0.49
Low intensity	0.52	[0.48, 0.56]	0.80	0.53	0.48	0.53	[0.49, 0.57]	0.81	0.54	0.47	0.47	[0.43, 0.51]	0.76	0.45	0.54
Moderate intensity	0.49	[0.45, 0.53]	0.78	0.45	0.54	0.52	[0.48, 0.56]	0.81	0.47	0.51	0.52	[0.48, 0.56]	0.78	0.54	0.48
Likelihood of future attempt	0.74	[0.71, 0.77]	0.92	0.72	0.27	0.74	[0.70, 0.77]	0.92	0.72	0.27	0.74	[0.71, 0.77]	0.91	0.73	0.27
Nonfatal suicide attempt															
Past month	0.53	[0.52, 0.55]	0.93	0.08	0.74	0.54	[0.53, 0.55]	0.96	0.09	0.73	0.54	[0.53, 0.55]	0.96	0.09	0.71
Lifetime	0.60	[0.56, 0.63]	0.84	0.65	0.37	0.59	[0.55, 0.63]	0.84	0.66	0.37	0.59	[0.55, 0.63]	0.83	0.65	0.38
Past year	0.60	[0.57, 0.63]	0.91	0.32	0.57	0.60	[0.57, 0.63]	0.91	0.32	0.57	0.60	[0.57, 0.63]	0.91	0.32	0.56
Agitation (BAM)	0.63	[0.60, 0.67]	0.86	0.68	0.34	0.62	[0.58, 0.66]	0.86	0.68	0.35	0.59	[0.55, 0.62]	0.83	0.67	0.37
Hopelessness (BHS)	0.76	[0.73, 0.79]	0.93	0.74	0.25	0.75	[0.71, 0.78]	0.91	0.79	0.23	0.71	[0.68, 0.75]	0.90	0.73	0.28
Psychopathology (BSI)	0.70	[0.66, 0.73]	0.90	0.68	0.31	0.69	[0.66, 0.73]	0.90	0.68	0.32	0.65	[0.61, 0.68]	0.87	0.64	0.36
Suicidal desire (BSS)	0.84	[0.81, 0.87]	0.95	0.86	0.15	0.84	[0.81, 0.87]	0.95	0.85	0.15	0.79	[0.76, 0.82]	0.95	0.70	0.26
NSSI (self-cutting)															
Past month	0.60	[0.57, 0.62]	0.92	0.30	0.58	0.60	[0.57, 0.63]	0.92	0.29	0.58	0.58	[0.56, 0.61]	0.90	0.29	0.58
Lifetime	0.53	[0.50, 0.56]	0.80	0.83	0.30	0.54	[0.50, 0.57]	0.81	0.83	0.29	0.55	[0.52, 0.58]	0.80	0.84	0.29
Past year	0.60	[0.58, 0.63]	0.92	0.30	0.57	0.58	[0.56, 0.61]	0.92	0.25	0.61	0.57	[0.54, 0.59]	0.89	0.25	0.61
Self-disgust (DWLS-Self)	0.63	[0.59, 0.67]	0.86	0.65	0.36	0.63	[0.59, 0.67]	0.87	0.64	0.36	0.62	[0.58, 0.66]	0.85	0.64	0.37
Disgust with others (DWLS-Other)	0.73	[0.69, 0.76]	0.91	0.71	0.28	0.72	[0.69, 0.76]	0.91	0.71	0.28	0.70	[0.67, 0.74]	0.90	0.70	0.30
Explicit ratings of suicide images	0.53	[0.52, 0.54]	0.97	0.07	0.74	0.53	[0.51, 0.54]	0.94	0.06	0.75	0.53	[0.52, 0.54]	0.96	0.06	0.74
Interrupted suicide attempt															
Past month	0.52	[0.51, 0.53]	0.93	0.05	0.76	0.53	[0.52, 0.54]	0.97	0.05	0.76	0.52	[0.51, 0.53]	0.95	0.05	0.75
Lifetime	0.53	[0.49, 0.56]	0.82	0.29	0.61	0.53	[0.50, 0.57]	0.83	0.30	0.61	0.53	[0.50, 0.57]	0.82	0.29	0.60
Past year	0.53	[0.50, 0.55]	0.86	0.13	0.71	0.53	[0.51, 0.55]	0.88	0.13	0.71	0.53	[0.51, 0.56]	0.88	0.13	0.69
Insomnia (ISI)	0.63	[0.59, 0.66]	0.87	0.62	0.38	0.62	[0.58, 0.66]	0.87	0.61	0.38	0.59	[0.55, 0.63]	0.84	0.61	0.40
Suicide plan															
Past month	0.75	[0.72, 0.79]	0.93	0.69	0.28	0.75	[0.72, 0.78]	0.93	0.69	0.28	0.73	[0.70, 0.76]	0.92	0.68	0.30
Lifetime	0.53	[0.51, 0.55]	0.80	0.97	0.22	0.53	[0.51, 0.55]	0.80	0.97	0.21	0.52	[0.50, 0.54]	0.78	0.97	0.23
Past year	0.71	[0.68, 0.74]	0.94	0.56	0.38	0.71	[0.68, 0.74]	0.94	0.56	0.38	0.70	[0.67, 0.73]	0.93	0.56	0.38
Preparations for suicide	0.63	[0.60, 0.67]	0.85	0.81	0.26	0.61	[0.58, 0.65]	0.84	0.81	0.27	0.62	[0.59, 0.66]	0.84	0.82	0.27
Suicidal intent	0.63	[0.60, 0.67]	0.86	0.71	0.32	0.64	[0.60, 0.68]	0.86	0.72	0.31	0.63	[0.59, 0.67]	0.85	0.71	0.32

Note: AUC = area under the receiver-operating-characteristic curve; CI = confidence interval; ACSS-FAD = Acquired Capability for Suicide Scale–Fearlessness About Death; AMP = Affect Misattribution Procedure; BAM = Brief Agitation Measure; BHS = Beck Hopelessness Scale; BSI-18 = Brief Symptom Inventory–18; BSS = Beck Scale for Suicide Ideation; NSSI = nonsuicidal self-injury; DWLS = Disgust With Life Scale; ISI = Insomnia Severity Index.

Table 2. Univariate Analyses Predicting Nonfatal Suicide Attempts Across Follow-Up

	3 days					14 days					28 days				
Variable	AUC	95% CI	Precision	Recall	Brier	AUC	95% CI	Precision	Recall	Brier	AUC	95% CI	Precision	Recall	Brier
Aborted attempt															
Past month	0.69	[0.58, 0.79]	0.10	0.48	0.13	0.67	[0.60, 0.74]	0.18	0.46	0.14	0.62	[0.57, 0.68]	0.23	0.35	0.15
Lifetime	0.56	[0.47, 0.66]	0.03	0.65	0.56	0.59	[0.53, 0.65]	0.07	0.75	0.55	0.59	[0.54, 0.64]	0.11	0.74	0.53
Past year	0.64	[0.54, 0.74]	0.08	0.38	0.14	0.64	[0.57, 0.71]	0.16	0.40	0.15	0.65	[0.60, 0.71]	0.19	0.49	0.22
Fearlessness about death (ACSS-FAD)	0.49	[0.38, 0.59]	0.02	0.39	0.54	0.52	[0.45, 0.59]	0.06	0.51	0.49	0.53	[0.47, 0.59]	0.09	0.53	0.48
AMP-Suicide															
High intensity	0.48	[0.38, 0.59]	0.02	0.39	0.48	0.50	[0.42, 0.57]	0.04	0.44	0.53	0.51	[0.45, 0.56]	0.08	0.46	0.48
Low intensity	0.48	[0.38, 0.59]	0.02	0.45	0.53	0.53	[0.46, 0.60]	0.06	0.48	0.47	0.50	[0.44, 0.56]	0.08	0.51	0.53
Moderate intensity	0.43	[0.33, 0.54]	0.02	0.27	0.47	0.51	[0.44, 0.58]	0.05	0.50	0.53	0.53	[0.48, 0.59]	0.09	0.51	0.46
Likelihood of future attempt	0.74	[0.65, 0.82]	0.06	0.77	0.31	0.74	[0.68, 0.80]	0.13	0.79	0.30	0.70	[0.66, 0.75]	0.16	0.79	0.37
Nonfatal suicide attempt															
Past month	0.78	[0.68, 0.88]	0.20	0.62	0.07	0.77	[0.70, 0.84]	0.39	0.60	0.07	0.67	[0.61, 0.72]	0.41	0.39	0.10
Lifetime	0.70	[0.68, 0.71]	0.04	0.99	0.59	0.69	[0.66, 0.72]	0.08	1.00	0.57	0.68	[0.64, 0.71]	0.13	0.94	0.54
Past year	0.82	[0.73, 0.90]	0.15	0.70	0.12	0.73	[0.66, 0.80]	0.25	0.56	0.12	0.68	[0.63, 0.74]	0.30	0.46	0.14
Agitation (BAM)	0.66	[0.57, 0.75]	0.04	0.73	0.43	0.63	[0.57, 0.70]	0.09	0.70	0.42	0.54	[0.48, 0.60]	0.09	0.60	0.52
Hopelessness (BHS)	0.57	[0.47, 0.67]	0.03	0.56	0.46	0.63	[0.56, 0.69]	0.08	0.70	0.44	0.64	[0.58, 0.69]	0.13	0.71	0.42
Psychopathology (BSI)	0.77	[0.68, 0.85]	0.07	0.77	0.25	0.70	[0.63, 0.76]	0.12	0.68	0.28	0.62	[0.57, 0.68]	0.13	0.62	0.38
Suicidal desire (BSS)	0.73	[0.65, 0.81]	0.06	0.79	0.34	0.75	[0.70, 0.80]	0.13	0.84	0.32	0.65	[0.59, 0.71]	0.15	0.62	0.33
NSSI (self-cutting)															
Past month	0.70	[0.60, 0.81]	0.08	0.54	0.18	0.68	[0.61, 0.75]	0.16	0.51	0.18	0.63	[0.58, 0.69]	0.20	0.41	0.19
Lifetime	0.57	[0.52, 0.62]	0.03	0.89	0.80	0.55	[0.50, 0.59]	0.06	0.88	0.78	0.55	[0.51, 0.58]	0.09	0.88	0.76
Past year	0.64	[0.54, 0.74]	0.08	0.35	0.14	0.66	[0.59, 0.73]	0.18	0.42	0.14	0.63	[0.58, 0.69]	0.23	0.37	0.16
Disgust with others (DWLS-Other)	0.57	[0.47, 0.68]	0.03	0.56	0.46	0.62	[0.55, 0.68]	0.08	0.68	0.44	0.52	[0.46, 0.58]	0.09	0.51	0.49
Self-disgust (DWLS-Self)	0.60	[0.50, 0.70]	0.04	0.60	0.41	0.68	[0.62, 0.75]	0.10	0.74	0.36	0.61	[0.56, 0.67]	0.12	0.66	0.43
Explicit ratings of suicide images	0.55	[0.47, 0.62]	0.06	0.10	0.08	0.57	[0.52, 0.63]	0.18	0.18	0.09	0.53	[0.49, 0.56]	0.17	0.08	0.12
Interrupted suicide attempt															
Past month	0.62	[0.53, 0.72]	0.17	0.25	0.06	0.61	[0.55, 0.67]	0.31	0.24	0.07	0.59	[0.54, 0.63]	0.37	0.20	0.10
Lifetime	0.50	[0.41, 0.60]	0.02	0.23	0.30	0.56	[0.49, 0.63]	0.07	0.39	0.30	0.54	[0.48, 0.59]	0.10	0.34	0.31
Past year	0.60	[0.51, 0.70]	0.07	0.25	0.14	0.58	[0.52, 0.64]	0.12	0.24	0.16	0.60	[0.54, 0.65]	0.21	0.28	0.16
Insomnia (ISI)	0.66	[0.56, 0.76]	0.05	0.63	0.35	0.65	[0.58, 0.71]	0.10	0.63	0.35	0.62	[0.57, 0.68]	0.13	0.65	0.40
Suicide plan															
Past month	0.70	[0.60, 0.79]	0.06	0.65	0.27	0.69	[0.62, 0.75]	0.12	0.64	0.26	0.66	[0.60, 0.71]	0.15	0.66	0.34
Lifetime	0.49	[0.45, 0.53]	0.02	0.05	0.07	0.52	[0.51, 0.53]	0.05	0.95	0.91	0.50	[0.47, 0.52]	0.06	0.92	0.89
Past year	0.61	[0.51, 0.71]	0.04	0.56	0.36	0.65	[0.58, 0.71]	0.10	0.65	0.35	0.62	[0.56, 0.68]	0.13	0.57	0.34
Preparations for suicide	0.58	[0.51, 0.64]	0.03	0.85	0.74	0.59	[0.55, 0.63]	0.06	0.92	0.72	0.59	[0.55, 0.62]	0.10	0.92	0.69
Suicidal intent	0.52	[0.42, 0.62]	0.02	0.56	0.56	0.63	[0.56, 0.69]	0.08	0.69	0.42	0.59	[0.54, 0.65]	0.11	0.73	0.53

Note: AUC = area under the receiver-operating-characteristic curve; CI = confidence interval; ACSS-FAD = Acquired Capability for Suicide Scale–Fearlessness About Death; AMP = Affect Misattribution Procedure; BAM = Brief Agitation Measure; BHS = Beck Hopelessness Scale; BSI-18 = Brief Symptom Inventory–18; BSS = Beck Scale for Suicide Ideation; NSSI = nonsuicidal self-injury; DWLS = Disgust With Life Scale; ISI = Insomnia Severity Index.

Table 3. Classification Table Across Follow-Up

Category	Suicide ideation			Nonfatal suicide attempt		
	T2	T3	T4	T2	T3	T4
True positive	757	744	706	15	34	51
False positive	42	44	48	1	1	1
False negative	15	13	17	9	16	25
True negative	160	149	155	950	899	849

Note: T2 = Time 2; T3 = Time 3; T4 = Time 4.

Suicidal ideation. Discrimination performance, as measured by AUC, was good across all time points with tight confidence intervals. Prediction at 3 days (AUC = 0.89; 95% confidence interval, or CI = [0.86, 0.91]) was fairly comparable with prediction at 14 (AUC = 0.88; 95% CI = [0.85, 0.91]) and 28 days after baseline (AUC = 0.87; 95% CI = [0.84, 0.90]). Precision was excellent, ranging between 0.94 and 0.95. Sensitivity was also good, with estimates of 0.98 across all time points. Brier scores were between 0.06 and 0.07, indicating good calibration.

Suicide attempts. Prediction, as indexed by AUC, was good at 3 days (AUC = 0.82; 95% CI = [0.73, 0.92]), 14 days (AUC = 0.83; 95% CI = [0.77, 0.90]), and 28 days (AUC = 0.83; 95% CI = [0.78, 0.89]). Precision was strong across all time points; values approached 1.00. Recall ranged from 0.64 to 0.67. All models were well calibrated, and Brier scores ranged from 0.01 to 0.03.

Variable importance. Variable importance as quantified by mean decrease in model accuracy for ideation and attempt are displayed in Table S1 in the Supplemental Material available online. Note that although variable importance analyses can be informative, the effects of individual predictors are conditional on the larger models; consequently, inferences should not be made about predictors in isolation. That is, considering the relative importance for predictors in isolation would not be equivalent with respect to predictive power as considering the model as a whole.

Suicide ideation. Across time points, several factors consistently emerged among the top 10 most important to overall model performance. These included self-predicted likelihood of future suicide attempt, self-predicted likelihood of a future suicide plan, suicide plan in the past month, psychopathology severity (BSI), disgust with self (DWLS-Self), fearlessness about death (ACSS-FAD), and acute hopelessness (ASQ-Hopelessness).

Suicide attempts. Nonfatal suicide attempt in the past month and multiple attempter status emerged within the top 10 predictors with respect to relative importance across all time points. Feelings of self-hatred (ASQ) evidenced the strongest relative effect on model performance at 3 days and notably weakened in its importance as a suicide attempt became less imminent, whereas suicidal ideation (BSS-Desire), nonfatal suicide attempt in the past month, suicide plan in the past month, multiple attempter status, self-perceived lethality of all prior suicide attempts, and the most severe suicide attempt notably strengthened in their importance over time. Psychopathology severity (BSI) emerged among the strongest predictors at 14 days, though less prominent at 3- and 28-day follow-ups.

Discussion

For decades, our ability to predict suicidal thoughts and behaviors has been weak (Franklin et al., 2017). This general finding—now replicated across multiple meta-analytic efforts (see Bentley et al., 2016; Carter et al., 2017; Chang et al., 2016; Huang, Ribeiro, Musacchio, & Franklin, 2017; Huang et al., 2018; Ribeiro et al., 2016a, 2018; Smith et al., 2018; Witte et al., 2018)—may help

Table 4. Machine-Learning Risk Algorithm Performance Across Follow-Up

Variable	AUC	95% CI	Recall	Precision	Calibration
Suicide ideation					
3 days	0.89	[0.86, 0.91]	0.98	0.95	0.06
14 days	0.88	[0.85, 0.91]	0.98	0.94	0.06
28 days	0.87	[0.84, 0.90]	0.98	0.94	0.07
Suicide attempt					
3 days	0.82	[0.73, 0.92]	0.64	0.99	0.01
14 days	0.84	[0.77, 0.90]	0.67	0.98	0.02
28 days	0.83	[0.78, 0.89]	0.67	0.98	0.03

Note: AUC = area under the receiver-operating-characteristic curve; CI = confidence interval.

to explain why rates of STBs have not abated despite considerable increases in research. Critically, however, these meta-analyses also revealed major methodological constraints present across much of the existing literature, and the ubiquity of these issues precludes definitive conclusions about conventional approaches to prediction. In this study, we addressed two critical methodological issues pervasive in prior research: (a) the reliance on extremely long follow-up periods and (b) the application of simple conceptualizations of risk. We posited that prediction accuracy would improve by addressing these issues. Results supported the following conclusions.

First, shorter follow-up periods alone generally did not improve prediction accuracy barring a few exceptions in the prediction of suicide ideation. One potential explanation of the weak prediction estimates observed in recent meta-analyses was the reliance on long follow-up periods. Although meta-analyses revealed that certain risk factors (e.g., hopelessness, agitation, insomnia, etc.) were not strong predictors of STBs years later, it was possible that these factors might be much stronger predictors in the days and weeks preceding STBs. Meta-analyses indicated that AUCs in the long term, on average, ranged between 0.56 and 0.62 (see Bentley et al., 2016; Franklin et al., 2017; Ribeiro et al., 2016a), indicating weak prediction. Most univariate estimates observed in the present study remained weak, despite examining these factors over shorter and more clinically meaningful follow-up windows, particularly in the prediction of nonfatal suicide attempts. Critically, although many predictors did not demonstrate strong performance in the short term, this does not mean that these factors have no relevance to STBs; rather, these findings suggest that these factors, when considered in isolation, may be insufficient to produce accurate short-term prediction.

Although machine-learning algorithms demonstrated superior performance, as we discuss below, several predictors of suicide ideation did evidence fair-to-good performance in univariate analyses. Baseline suicide ideation, not surprisingly, was the strongest univariate predictor. The remaining factors that performed well in univariate analyses were constructs closely related to suicidal ideation, including hopelessness and planning for suicide over the past month. This suggests that, in the case of suicide ideation, the accuracy of univariate prediction can be considerably improved by examining the phenomenon over shorter time frames. Note that the performance associated with these predictors weakened slightly over time. Taken together, this pattern of findings may reflect the temporal stability of suicidal ideation over the course of a few days or weeks. In contrast to suicide attempts, which are discrete behaviors typically lasting only minutes or hours, episodes of suicide ideation can persist for weeks or months. In part, the improved performance over shorter time frames observed in this study may be an artifact of the nature of suicide ideation. As prediction windows extend outside the typical duration of ideation, it is possible that more complex algorithms are necessary for accurate prediction. Research exploring how these predictors may perform over slightly longer yet still clinically meaningful time periods (e.g., a few months) would be valuable.

Second, complex models of risk can considerably improve our ability to predict imminent suicidal thoughts and nonfatal suicide attempts relative to simple models of risk. In this study, machine-learning risk algorithms that considered complex combinations among many predictors produced stronger prediction than univariate models. These results are consistent with a growing body of evidence supporting the utility of machine learning for suicide risk prediction (see Kessler et al., 2015, 2016; Ribeiro et al., 2016a; Walsh, Ribeiro, & Franklin, 2017, 2018). In this study, we were able to obtain comparable performance with fewer predictors (i.e., ~50) relative to prior machine-learning efforts using administrative data sources, some of which have required several hundred predictors. One concern that might be raised is whether the performance observed in the machine-learning algorithms is simply a statistical artifact due to the number of predictors included in the model. We have explored this possibility in prior efforts (see Walsh et al., 2017, 2018). Although including more predictors does result in stronger performance relative to univariate analyses, it fails to optimize prediction, resulting in algorithms that continue to weakly predict suicide outcomes. This suggests that the strength of machine learning lies not only in its ability to integrate a large number of predictors, but also in its ability to combine those predictors in an optimized fashion.

To examine algorithms more granularly, we conducted analyses to evaluate variable importance. In these analyses, we examined mean decrease in model accuracy to estimate the relative importance of predictors. Results of these analyses highlight the variable nature of predictor importance for suicide attempts and relative stability of predictor importance for suicide ideation, at least within the short time frames (i.e., days, weeks) examined in this study. Although it may seem reasonable to draw conclusions about the absolute importance of individual risk factors on prediction on the basis of these results, the importance of factors within machine-learning models are conditional to the context (i.e., the broader algorithm) in which those factors were modeled. Algorithmic performance depends heavily on the complex interactions modeled among factors; consequently, algorithms are not reducible to their component factors. We therefore strongly advise against drawing inferences about the importance of factors in isolation.

Third, our ability to predict suicidal ideation was slightly stronger than our ability to predict nonfatal suicide attempts. Although AUC estimates were fairly comparable across outcomes, tighter 95% CIs suggested more robust prediction estimates of suicide ideation. Precision was strong for both outcomes; however, sensitivity could be improved for suicide-attempt prediction. Two issues may contribute to these findings. First, as discussed above, suicide ideation, particularly over shorter time periods, may be an easier problem to predict. It is an episodic phenomenon of longer duration than a discrete suicidal behavior. Within such circumscribed time frames, accurate prediction may not be as challenging, particularly when leveraging information from closely related constructs such as hopelessness. Second, suicidal behavior occurred much less frequently than suicide ideation, though rates of both outcomes were far higher than what would be expected in the general population (Centers for Disease Control and Prevention, 2018). Accordingly, we would expect suicide-ideation models to be more robust, given that there were substantially more instances of the target outcome on which to build the model.

Taken together, these findings have several important theoretical implications. For decades since its inception, suicide science has concentrated on identifying circumscribed sets of risk factors that could be used to identify individuals at risk. Recent evidence, including the results of this study, indicates that a shift in focus is necessary to advance suicide prediction. We believe that the integration of complexity into the conceptualization of STBs will be critical to this shift. Presently, the dominant perspective in the suicide field holds that STBs are the product of a discrete, determinate set of a small number of factors combined in rudimentary ways. Alternatively, a complex conceptualization would hold that suicidal thoughts and behaviors emerge from complex combinations of a large number of biopsychosocial factors. As emergent phenomena, STBs are posited to be irreducible to the component factors from which they emerge. Although simple conceptualizations of STBs allow for more cognitively manageable explanations, they are largely incompatible with the findings of this study, as well as with recent meta-analytic and machine-learning evidence.

Indeterminacy, a corollary of complexity, also merits discussion. Briefly described, *indeterminacy* (sometimes referred to as *degeneracy* or *equifinality*) refers to the idea that different mechanisms or pathways can produce identical outcomes. Our study provides preliminary evidence for the role of indeterminacy in the emergence of STBs. The design of the present study deviates from prior machine-learning STB risk-prediction efforts in several important ways. One critical distinction is the nature of the input data. Before this study, efforts designed to develop machine-learning-based risk algorithms largely relied on existing administrative data sources, such as EHRs. As a result, psychological constructs historically hypothesized as highly relevant to STBs (e.g., hopelessness, suicidal capability, etc.) were not directly represented. This study, in contrast, was able to directly assess a wide range of psychological constructs. Despite having unique data sources and using different input variables, the present study achieved comparable predictive performance. One possible explanation for this finding is that STB risk is indeterminate; that is, there may be many different pathways, involving different sets of biopsychosocial factors, from which STBs may emerge. Correspondingly, we expect that many diverse risk algorithms can produce accurate prediction of STBs, though future research is needed to test this hypothesis more stringently.

Relatedly, one issue that may be raised with respect to this and other machine-learning prediction studies concerns the potential ramifications of omitting potentially important predictors. In this study, variables were selected on the basis of a review of published clinical consensus statements regarding short-term prediction (e.g., Rudd et al., 2006), as well as meta-analytic reviews of prospective suicide-prediction studies (e.g., Franklin et al., 2017). This review resulted in a combination of more static, presumably distal predictors (e.g., history of suicidal behavior, demographics, etc.) and more dynamic, presumably proximal predictors (e.g., changes in affect and arousal, etc.). In addressing this concern, we believe two fundamental points are important to highlight. First, machine learning is designed to map the optimal function between a given set of input (i.e., predictor) variables and the target output (i.e., outcome). If a given variable is removed or omitted from a model, algorithms can still adjust to optimize prediction of the target outcome. Particularly when predicting fairly complex outcomes such as suicidal phenomena, accurate prediction may be achieved via a multitude of different risk algorithms that consider a wide range of factors. Therefore, even the removal or omission of a highly relevant predictor variable (e.g., prior suicide attempts) may not impede the development of accurate prediction (see Walsh et al., 2017).

Second, we believe that the concern of potentially omitting an important variable is due in part to a common misunderstanding that risk factors are isomorphic with causes. Although we used a longitudinal design that can speak to risk, the resultant risk algorithms and risk factors within these algorithms are not necessarily causal risk factors. In other words, the objective of building risk algorithms is to optimize prediction, which may or may not map on to the causal mechanisms underlying STBs. As we discuss above, machine-learning prediction efforts are not designed to find a singular "best" algorithm, but instead optimize prediction of an outcome with a given set of relevant predictors available, which may or may not be causal. Identifying causal risk factors requires controlled experiments (Kraemer et al., 1997).

Beyond implications for theory, the results of this study can also inform clinical practice. A major shortcoming of prior literature has been that research has overwhelmingly been focused on eventual death by suicide (i.e., prediction over years or decades; Franklin et al., 2017). Yet, clinicians are tasked with assessing risk of imminent suicidal behavior (i.e., prediction over days or weeks). Alarmingly, however, less than 1% of all prospective STB studies have follow-up periods of 1 month or less, suggesting that very little is known about imminent risk prediction (Franklin et al., 2017). This investigation was, in part, an effort to address this critical gap in knowledge. We hypothesized that many risk factors, particularly those that have been suggested through clinical consensus to be warning signs of imminent suicidal behavior (e.g., agitation, insomnia; Rudd et al., 2006), would have greater clinical utility and predictive validity in the short term. Our results do not support that hypothesis; rather, our results support the potential utility of integrating machine-learning-based risk algorithms into clinical decision making. It may be tempting to examine predictor-importance values and reduce algorithms into more streamlined, cognitively manageable sets of predictors; however, we believe this would be a misstep. Doing so would jeopardize the potential predictive power of the algorithms, as the strength of this approach lies not only in the predictors used but, critically, the complex combinations among them. As such, pursuing research to develop scalable, user-centered clinical decision-support tools or systems for providers is critical. Ideally, these tools would provide accurate information about risk and attendant management strategies to providers while simultaneously being minimally disruptive to existing clinical workflows. The results of this study would represent an initial step toward that end; future research is needed to translate risk algorithms into scalable, user-centered clinical decision tools.

The present effort is not without limitations and results should be considered in that context. First, the sample used in this study was predominantly young, female, and White; evaluating the generalizability of our findings for demographically diverse samples would be useful. Relatedly, examining whether risk functions differently for certain suicidal thoughts and behaviors across demographic groups in which the prevalence of suicidal thoughts and behaviors varies would be prudent. Given the well-documented differences in the prevalence of certain suicidal thoughts and behaviors among men and women, examining whether risk functions differently across these groups may be particularly valuable. The smaller size of the present sample precluded such analyses, as the risk of overfitting would be high; however, future studies sufficiently powered to investigate these issues are needed. Second, risk algorithms developed in this study may not be ideal to predict initial onset of STBs (i.e., first-onset ideation; initial suicide attempt). Given the rarity of STBs, especially over narrow prediction windows, we recruited individuals on the basis of recency and severity of self-injurious thoughts and behaviors. Accurate prediction of incident or first-onset STBs may require different risk algorithms than those developed in this effort.

Third, because data collection was anonymous, we cannot ascertain whether participant dropout was due to suicide death or extended hospitalization as a result of nonfatal suicide attempt or severe ideation. Retention rates were high; therefore, the probability of this issue significantly influencing results is likely low. Fourth, external validation of the algorithms derived in this study is critical to ensure generalizability of findings. This issue is not unique to the present study, but rather one that is critical to all existing machine-learning efforts. Although we implemented techniques to decrease the likelihood of overfitting, internal validation strategies, such as the bootstrap optimism correction technique applied in this study, cannot replace external validation efforts. Given the rarity of prediction efforts in the short term, we were unable to provide evidence of external validity in this effort; however, this represents a critical next step. Though slightly more computationally demanding than alternative internal-validation strategies (e.g., holdout sets), we reasoned that the evidence demonstrating less-biased performance using the bootstrap to correct for optimism in smaller data sets justified the trade-off. Nevertheless, without evidence of external validity, the risk of overfit models remains.

Limitations notwithstanding, this study also has notable strengths. First, it represents among the largest efforts to examine short-term prediction among severely suicidal and self-injuring individuals. Second, because the sample severity was so high, outcome prevalence was also considerably higher than expected in the general or most clinical populations, allowing for more reliable estimates of model performance. Third, sample severity also resulted in stringent comparisons. In other words, our models were able to accurately predict who would think about and attempt suicide among individuals who evidenced many of the strongest univariate predictors of doing so (e.g., STB history; psychopathology; Franklin et al., 2017). Fourth, this is the first study applying machine learning to predict suicidal ideation. Fifth, we used gold-standard psychological instruments to assess our outcomes of interest (i.e., BSS; SITBI). This is particularly notable, as most prior machine-learning suicide-prediction studies have relied on ICD codes within EHRs, which have been shown to be vulnerable to misclassification of STB outcomes (Walsh et al., 2017). Sixth, variables were focused on psychological constructs with strong empirical or theoretical rationale for their relation to STB prediction, which has not been feasible in prior machine-learning suicide-prediction efforts. Seventh, we assessed these constructs using a range of explicit and implicit measures, thereby assessing factors using a combination of objective and subjective measures.

In sum, the present study was designed to advance our ability to predict STBs. To this end, we addressed two fundamental methodological constraints of prior literature that were viable explanations for the poor prediction accuracy observed in recent meta-analytic studies (see Franklin et al., 2017). Taken together, our findings suggest that focusing on short-term prediction alone will not substantially improve our ability to predict using our conventional approaches to prediction; however, conceptualizing risk as complex and using methods such as machine learning to model that complexity can. We look forward to studies that build on the present effort, particularly those focused on effectively translating risk algorithms into clinical practice.

Action Editor

John J. Curtin served as action editor for this article.

Author Contributions

J. D. Ribeiro oversaw all aspects of the study. J. D. Ribeiro also conceptualized the study idea, designed the study, oversaw data collection, conducted data analysis and interpretation, and wrote the manuscript. X. Huang and C. G. Walsh assisted with data analysis and interpretation and provided critical feedback on the manuscript. X. Huang, K. R. Fox, and K. P. Linthicum assisted in data collection and provided substantive feedback on the manuscript. All of the authors approved the final manuscript for submission

Acknowledgments

Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Military Suicide Research Consortium, the Department of Veterans Affairs, or the Department of Defense.

Declaration of Conflicting Interests

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

Funding

This work was supported in part by funding from the Military Suicide Research Consortium (MSRC), an effort supported by the Office of the Assistant Secretary of Defense for Health Affairs (Award No. W81XWH-10-2-0181).

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/2167702619838464

Notes

1. Note that because the validity of e-mail handles cannot be verified, we considered any e-mail to be an alias and unidentifiable. Nevertheless, we encouraged participants to use e-mail handles that did not include any potentially identifiable information (e.g., name, location, date of birth, etc.). We also provided participants with a document describing potentially identifiable information and asked them to verify that their e-mail did not include any potentially identifiable information.
2. Results of analyses retaining BSS-Desire as a continuous outcome were highly consistent with those of analyses in which it was dichotomized.
3. Note that models were also run with baseline ideation as a predictor of future ideation; as anticipated, these models performed even more strongly than the models omitting baseline ideation. However, this is likely because the predictor and outcome were measuring the same phenomenon.

References

Barak-Corren, Y., Castro, V. M., Javitt, S., Hoffnagle, A. G., Dai, Y., Perlis, R. H., . . . Reis, B. Y. (2016). Predicting suicidal behavior from longitudinal electronic health records. *American Journal of Psychiatry*, *174*, 154–162.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*, 419–429.

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*, *2*, 297–307.

Beck, A. T., & Steer, R. A. (1988). *BHS, Beck Hopelessness Scale: Manual*. San Antonio, TX: Psychological Corporation.

Beck, A. T., & Steer, R. A. (1991). *BSI, Beck scale for suicide ideation: Manual*. San Antonio, TX: Psychological Corporation.

Bentley, K. H., Franklin, J. C., Ribeiro, J. D., Kleiman, E. M., Fox, K. R., & Nock, M. K. (2016). Anxiety and its disorders as risk factors for suicidal thoughts and behaviors: A meta-analytic review. *Clinical Psychology Review*, *43*, 30–46.

Birchwood, M., Smith, J. O., Cochrane, R., Wetton, S., & Copestake, S. (1990). The Social Functioning Scale. The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. *The British Journal of Psychiatry*, *157*, 853–859.

Britton, P. C., & Conner, K. R. (2010). Suicide attempts within 12 months of treatment for substance use disorders. *Suicide and Life-Threatening Behavior*, *40*, 14–21.

Bull, S. S., Vallejos, D., Levine, D., & Ortiz, C. (2008). Improving recruitment and retention for an online randomized controlled trial: Experience from the Youthnet study. *AIDS care*, *20*, 887–893.

Busch, K., Fawcett, J., & Jacobs, D. (2003). Clinical correlates of inpatient suicide. *Journal of Clinical Psychiatry*, *64*, 14–19.

Carter, G., Milner, A., McGill, K., Pirkis, J., Kapur, N., & Spittal, M. J. (2017). Predicting suicidal behaviours using clinical instruments: Systematic review and meta-analysis of positive predictive values for risk scales. *The British Journal of Psychiatry*, *210*, 387–395.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*, 2156–2160.

Centers for Disease Control and Prevention. (2018). Injury prevention & control: Data & statistics (WISQARS). https://webappa.cdc.gov/sasweb/ncipc/mortrate.html

Chang, B. P., Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., & Nock, M. K. (2016). Biological risk factors for suicidal behaviors: A meta-analysis. *Translational Psychiatry*, *6*, e887. doi:10.1038/tp.2016.165

Chu, C., Bodell, L. P., Ribeiro, J. D., & Joiner, T. E. (2015). Eating disorder symptoms and suicidal ideation: The moderating role of disgust. *European Eating Disorders Review*, *23*, 545–552.

Crosby, A., Gfroerer, J., Han, B., Ortega, L., & Parks, S. E. (2011). *Suicidal thoughts and behaviors among adults aged ≥18 years — United States, 2008-2009*. Atlanta, GA: Centers for Disease Control and Prevention.

Derogatis, L. (2000). *Brief Symptom Inventory 18*. Minneapolis, MN: NCS Pearson.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, *98*, 13763–13768.

Franklin, J. C., Fox, K. R., Franklin, C. R., Kleiman, E. M., Ribeiro, J. D., Jaroszewski, A. C., . . . Nock, M. K. (2016). A brief mobile app reduces nonsuicidal and suicidal self-injury: Evidence from three randomized controlled trials. *Journal of Consulting and Clinical Psychology*, *84*, 544–557.

Franklin, J. C., Puzia, M. E., Lee, K. M., Lee, G. E., Hanna, E. K., Spring, V. L., & Prinstein, M. J. (2013). The nature of pain offset relief in nonsuicidal self-injury: A laboratory study. *Clinical Psychological Science*, *1*, 110–119.

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., . . . Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143*, 187–232.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: Data mining, inference, and predictions*. New York, NY: Springer.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400–407.

Hendin, H., Maltsberger, J. T., & Szanto, K. (2007). The role of intense affective states in signaling a suicide crisis. *The Journal of Nervous and Mental Disease*, *195*, 363–368.

Huang, X., Fox, K. R., Ribeiro, J. D., & Franklin, J. C. (2018). Psychosis as a risk factor for suicidal thoughts and behaviors: A meta-analysis of longitudinal studies. *Psychological Medicine*, *48*, 765–776.

Huang, X., Ribeiro, J. D., Musacchio, K. M., & Franklin, J. C. (2017). Demographics as predictors of suicidal thoughts and behaviors: A meta-analysis. *PLOS ONE*, *12*(7), Article e0180793. doi:10.1371/journal.pone.0180793

Ip, E. J., Barnett, M. J., Tenerowicz, M. J., & Perry, P. J. (2010). The touro 12-step: A systematic guide to optimizing survey research with online discussion boards. *Journal of Medical Internet Research*, *12*(2), Article e16. doi:10.2196/jmir.1314

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York, NY: Springer.

Joiner, T. E., Rudd, M. D., & Rajab, M. H. (1997). The Modified Scale for Suicidal Ideation: Factors of suicidality and their relation to clinical and diagnostic variables. *Journal of Consulting and Clinical Psychology*, *106*, 260–265.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., . . . Nierenberg, A. A. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*, *21*, 1366–1371.

Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., . . . Fullerton, C. S. (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*, 49–57.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, *54*, 337–343.

Liaw, A., & Wiener, M. (2018). randomForest: Breiman and Cutler's random forests for classification and regression (Version 4.6-14) [Software]. Retrieved from https://cran.r-project.org/web/packages/randomForest/

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*, 309–317.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293.

R Core Team. (2014). R: A language and environment for statistical computing (Version 3.1.2) [Computer software]. Retrieved from https://www.r-project.org/index.html

Ribeiro, J. D., Bender, T. W., Selby, E. A., Hames, J. L., & Joiner, T. E. (2011). Development and validation of a brief self-report measure of agitation: The Brief Agitation Measure. *Journal of Personality Assessment*, *93*, 597–604.

Ribeiro, J. D., Bodell, L., & Joiner, T. E. (2012, November). *Disgust with self, others, and world in suicidality*. Poster Presented at the 46th Annual Meeting of the Association for Behavioral and Cognitive Therapies, National Harbor, MD.

Ribeiro, J. D., Franklin, J. C., Fox, K. R., Bentley, K. H., Kleiman, E. M., Chang, B. P., & Nock, M. K. (2016a). Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: A meta-analysis of longitudinal studies. *Psychological Medicine*, *46*, 225–236.

Ribeiro, J. D., Franklin, J. C., Fox, K. R., Bentley, K. H., Kleiman, E. M., Chang, B. P., & Nock, M. K. (2016b). 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*, 2009–2010.

Ribeiro, J. D., Huang, X., Fox, K. R., & Franklin, J. C. (2018). Depression and hopelessness as risk factors for suicide ideation, attempts and death: Meta-analysis of longitudinal studies. *The British Journal of Psychiatry*, *212*, 279–286.

Ribeiro, J. D., Witte, T. K., Van Orden, K. A., Selby, E. A., Gordon, K. H., Bender, T. W., & Joiner, T. E., Jr. (2014). Fearlessness about death: The psychometric properties and construct validity of the revision to the Acquired Capability for Suicide Scale. *Psychological Assessment*, *26*, 115–126.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., . . . Müller, M. (2018). pROC: Display and analyze ROC curves (Version 1.12.1) [Software]. Retrieved from https://cran.r-project.org/web/packages/pROC

Rudd, M. D., Berman, A. L., Joiner, T. E., Jr., Nock, M. K., Silverman, M. M., Mandrusiak, M., . . . Witte, T. (2006). Warning signs for suicide: Theory, research, and clinical applications. *Suicide and Life-Threatening Behavior*, *36*, 255–262.

Smith, A. R., Velkoff, E. A., Ribeiro, J. D., & Franklin, J. (2018). Are eating disorders and related symptoms risk factors for suicidal thoughts and behaviors? A meta-analysis. *Suicide and Life-Threatening Behavior*, *49*, 221–239.

Smith, G. C., Seaman, S. R., Wood, A. M., Royston, P., & White, I. R. (2014). Correcting for optimistic prediction in small data sets. *American Journal of Epidemiology*, *180*, 318–324.

Substance Abuse and Mental Health Services Administration. (2017). *Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health* (HHS Publication No. SMA 17-5044, NSDUH Series H-52). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*, 457–469.

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of Child Psychology and Psychiatry*, *59*, 1261–1270.

Witte, T. K., Gauthier, J. M., Huang, X., Ribeiro, J. D., & Franklin, J. C. (2018). Is externalizing psychopathology a robust risk factor for suicidal thoughts and behaviors? A meta-analysis of longitudinal studies. *Journal of Clinical Psychology*, *74*, 1607–1625.

World Health Organization. (2018, January 31). *Suicide*. Retrieved from http://www.who.int/mediacentre/factsheets/fs398/en/