Published in final edited form as:

Mol Psychiatry. 2020 January ; 25(1): 168–179. doi:10.1038/s41380-019-0531-0.

Suicide prediction models: a critical review of recent research with recommendations for the way forward

Ronald C. Kessler¹, Robert M. Bossarte^{2,3}, Alex Luedtke^{4,5}, Alan M. Zaslavsky¹, Jose R. Zubizarreta^{1,6}

¹Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

²West Virginia University Injury Control Research Center and Department of Behavioral Medicine and Psychiatry, West Virginia University School of Medicine, Morgantown, WV, USA

³West Virginia and VISN 2 Center of Excellence for Suicide Prevention, Canandaigua, NY, USA

⁴Department of Statistics, University of Washington, Seattle, WA, USA

⁵Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

⁶Department of Statistics, Harvard University, Cambridge, MA, USA

Abstract

Suicide is a leading cause of death. A substantial proportion of the people who die by suicide come into contact with the health care system in the year before their death. This observation has resulted in the development of numerous suicide prediction tools to help target patients for preventive interventions. However, low sensitivity and low positive predictive value have led critics to argue that these tools have no clinical value. We review these tools and critiques here. We conclude that existing tools are suboptimal and that improvements, if they can be made, will require developers to work with more comprehensive predictor sets, staged screening designs, and advanced statistical analysis methods. We also conclude that although existing suicide prediction tools currently have little clinical value, and in some cases might do more harm than good, an even-handed assessment of the potential value of refined tools of this sort cannot currently be made because such an assessment would depend on evidence that currently does not exist about the effectiveness of preventive interventions. We argue that the only way to resolve this uncertainty

[✉] Kessler@hcp.med.harvard.edu.

NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States [Title 17, United States Code] governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that use may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgement, fulfillment of the order would involve violation of copyright law. No further reproduction and distribution of this copy is permitted by transmission or any other means.

Compliance with ethical standards

Conflict of interest In the past 3 years, RCK was a consultant for Johnson & Johnson Wellness and Prevention, Sage, Shire, and Takeda and served on an advisory board for the Johnson & Johnson Services Inc. Lake Nona Life Project. Kessler is a co-owner of DataStat, Inc., a market research firm that carries out health care research. In the past 3 years, JRZ was a consultant for Johnson & Johnson. The remaining authors declare that they have no conflict of interest.

is to link future efforts to develop or evaluate suicide prediction tools with concrete questions about specific clinical decisions aimed at reducing suicides and to evaluate the clinical value of these tools in terms of net benefit rather than sensitivity or positive predictive value. We also argue for a focus on the development of individualized treatment rules to help select the right suicide-focused treatments for the right patients at the right times. Challenges will exist in doing this because of the rarity of suicide even among patients considered high-risk, but we offer practical suggestions for how these challenges can be addressed.

Introduction

Suicide is a leading cause of death [1]. Many of the people who die by suicide are known to the mental health care system well before death. In the US, two-thirds of suicide decedents are in contact with the mental health care system in the year before their death, 30% are either hospitalized for a psychiatric disorder or make an emergency department visit for a psychiatric problem during that year, and one-third of suicide decedents are in outpatient treatment for a mental disorder in the month before their death [2–5]. The value of systematically quantifying patient suicide risk at these and other contact points has been debated for over half a century [6, 7] and continues to this day [8, 9]. Although the value of classifying such patients in terms of suicide risk might seem obvious, critics argue that these classifications have no real clinical value because of the rarity of suicide, the weakness of existing suicide prediction tools, and the lack of evidence that targeted clinical interventions are effective in preventing suicides [8]. This debate has not advanced meaningfully in many years despite progress in developing suicide prediction tools and the methodology of risk prediction [10]. In this viewpoint, we revisit the core issues in this debate and propose a way forward for suicide risk assessment.

Basic suicide statistics

The global age-standardized suicide death rate in 2016 was estimated to be 11.1/100,000 person-years. The US suicide rate that year was estimated to be 12.9/100,000 person-years [11]. These estimates are probably conservative [12]. Although the suicide rate decreased worldwide over the past 15 years, it increased in the US by an average of 1.5% per year, with half the US states experiencing suicide increases >30% between 1999 and 2016 [13]. Based on this trend, suicide is now the tenth leading cause of death in the US overall, the second leading cause of death among children ages 10–15, and the third leading cause of death among adolescents and young adults [14].

Multivariate analyses to develop suicide prediction tools among high-risk patients

Most efforts to develop suicide prediction tools focus on one of three partially overlapping high-risk patient populations: patients in emergency departments with suicidal intent or after a suicide attempt, psychiatric inpatients during hospitalization, and psychiatric inpatients after discharge. Meta-analyses suggest that the suicide rate in the first of these groups is about 1,600/100,000 within 1 year of the emergency department visit [15], that the suicide

rate in the second group is about 150/100,000 inpatient-years (including suicide deaths while on temporary leave from hospitalization) [16], and that the suicide rate after psychiatric hospital discharge is between 3000/100,000 person-years in the first week after discharge and 650/100,000 person-years 4–12 months after discharge [17, 18]. As noted in the introduction, the roughly 2% of the US population who are in one or more of these three high-risk groups in a given year account for about one-third of all US suicides [5].

Numerous studies have developed suicide prediction tools for these high-risk groups [19–28] and several recent studies developed comparable prediction tools for entire health care systems [29–33]. These tools generate individual-level predicted probabilities of suicide or suicide-related behaviors (SRBs; nonfatal suicide attempts, aborted suicide attempts, and/or imminent risk resulting in an emergency department visit) over a specified time horizon (most often the next 12 months). The tools for high-risk patients are designed to be used either at intake or discharge to help guide treatment planning. The tools for total populations are designed either to target patients with high predicted probabilities for outreach or to create flags in electronic medical records to remind primary care providers to enquire about suicidality during annual physical examinations. No agreement has emerged on how best to set a high-risk threshold from the continuous probability scores generated by these prediction tools [34].

Some early suicide prediction tools used only single scales as predictors, such as self-report scales of hopelessness, depression, overall psychopathological severity, suicide intent, or attitudes toward suicide [35–38]. Other early studies and most recent studies began with a predictor set consisting of clinical and socio-demographic data extracted from medical records and sometimes included a series of patient self-report and clinician rating scales. A multivariate model of these variables is then used in these studies to predict subsequent suicide or SRBs and to generate a composite risk score made by weighting and summing the predictors in the model. Analysis methods used to develop these models ranged from trial and error cross-tabulations in early studies [39] to logistic regression [40] and various types of machine learning (ML) methods [41] in more recent studies.

All of these suicide prediction tools are suboptimal in that they failed to include at least some predictors who were found to be significant in other studies. In addition, the vast majority of these tools were developed using suboptimal statistical methods, which could lead either to overestimation due to overfitting or to underestimation due to failure to detect nonlinearities or interactions. Within the context of these limitations, the strength of most published suicide prediction tools can be characterized as *moderate*. For example, a recent meta-analysis of 17 multivariate analyses to predict inpatient suicides estimated that sensitivity (SN; the proportion of suicides that occurred among patients classified as being above the risk threshold) was about 0.70 when specificity (SP; the proportion of patients not dying by suicide that were classified correctly as being below the risk threshold) was set at 0.80 (i.e., SN was about 3.5 times as high as expected by chance at the threshold, where this ratio can be defined as $SN/[1-SP]$) and area under the receiver operating characteristic curve (AUC; the average of SN and SP) at that threshold was about 0.75 [23]. SN in this same meta-analysis was about 0.50 when SP was set at 0.90 (i.e., about 5.0 times as high as expected by chance) and AUC at that threshold was about 0.70. Given the relatively short

duration of most hospitalizations, positive predictive value (PPV; the incidence of suicide over a prespecified time horizon among patients classified as high-risk) for these studies of inpatient suicide averaged only about 0.004, but this was roughly ten times as high as the suicide rate among patients below the predicted high-risk threshold.

Criticisms of suicide-prediction tools

Despite their higher-than-chance SN values, commentators [42–45] conclude, with only rare exceptions [46], that suicide prediction tools are not strong enough to justify their use in clinical decision making. The argument underlying this conclusion is that all patients in high-risk settings (i.e., in EDs with mental health crises or in psychiatric hospitals) should be considered to have meaningfully elevated suicide risk and consequently should receive an in-depth clinical assessment with a focus on their service needs rather than their suicide risks, making it unnecessary for them to be administered suicide prediction scales because the information obtained in these tools will be obtained as part of the clinical assessment.

Critics go on to argue that the low PPV of suicide prediction tools at their recommended decision thresholds means that suicide prevention interventions focused on all patients classified as high-risk would “subject many patients, who will never die by suicide to excessive intrusion or coercion” [22]. In addition, critics note that the low SN of suicide prediction tools at their recommended thresholds means that suicide prevention interventions focused only on patients classified as high-risk would exclude the majority of patients who go on to take their lives [45]. In addition, critics argue that many patients perceive structured suicide prediction tools as being superficial, as they require patients to check off prespecified response boxes to survey questions that inevitably fail to capture the full complexity of patient experiences and give patients the impression that they are being treated as numbers rather than as people. Critics argue that this impersonal approach to assessment interferes with establishing the therapeutic alliance needed for more in-depth clinical risk assessments [44], a position supported by some qualitative studies [47–50]. The same criticism has also recently been made about using a computer to record patient information during the intake interview based on evidence that this predicts a weak subsequent therapeutic alliance [51]. Based on such evidence, some [52–54], but not all [55, 56], clinical practice guidelines recommend against using structured suicide prediction tools.

These tools nonetheless continue to be widely used even in settings where clinical practice guidelines say they should not be used [57]. But why? One possible reason is that structured suicide prediction tools are much less timeconsuming than the detailed clinical psychosocial assessments called for by practice guidelines and some clinicians use them to help determine which patients should subsequently receive these detailed clinical assessments. The latter assessments often require multiple sessions with the patient along with collateral interviews with loved ones about the patient’s severity of suicidal intent, precipitating problems (e.g., mental and physical health problems, life difficulties), risk factors (e.g., predispositions to suicide, warning signs for imminent risk), and protective factors (e.g., social supports, coping styles) [58]. Such assessments are carried out with only slightly more than half of all suicidal inpatients and emergency department patients even in settings where treatment guidelines stipulate that they always be carried out [57]. A comparison of the characteristics

of patients who receive these detailed clinical assessments versus those that do not suggests that clinicians use informal triage rules to decide when to administer indepth evaluations of suicide risk, although these informal rules seem to pay more attention to some relevant patient risk factors than others [59].

Clinical suicide risk evaluations versus structured assessments

An important question in evaluating criticisms of structured suicide prediction tools is whether clinical evaluations are more accurate than structured assessments in predicting subsequent suicides and SRBs. The evidence suggests not. Indeed, statistical models have long been known to be superior to unstructured clinical judgments in predicting a wide range of clinical outcomes [60], although the strength of this advantage varies over settings and decisions [61] and clinical evaluation accuracy can be improved by using structured methods to guide integration of information to arrive at judgments [10]. Consistent with this extensive literature, a meta-analysis of 13 studies examining risk factors for suicide within 12 months of psychiatric hospital discharge found that clinical judgments at discharge were not much better predictors of subsequent suicides than were each of a number of other social, historical, and clinical variables obtained from patient self-reports or extracted from administrative databases [62]. No evidence was presented in that meta-analysis to suggest that clinical judgments would have any significant incremental predictive effect in a multivariate model that included all self-report and administrative predictors. A more recent meta-analysis of seven studies found that clinical assessments were only weakly associated with subsequent suicides and SRBs among patients after hospital treatment of SRBs [28].

In-depth clinical assessments might nonetheless be important because they collect information about need for services used in developing a treatment plan based on the patient's particular circumstances. Indeed, critics of suicide risk prediction tools respond to the fact that in-depth clinical assessments do not predict future suicides better than structured suicide risk tools by emphasizing that the main goal of in-depth clinical assessment is to assess need for services rather than to assess suicide risk [19, 22, 27]. But there is a difficulty with this injunction: that clinical decisions about need for services should be informed by perceived suicide risk. For example, a number of psychotherapies exist specifically for patients with high suicide risk [63]. In addition, a number of medications exist that are recommended selectively for patients at high suicide risk [64, 65]. Accurate assessment of suicide risk is required to make informed decisions about whether to use these interventions. This fact is recognized in the strategy for suicide prevention advanced by the US National Action Alliance for Suicide Prevention [66] as well as in related guidelines for identifying risk and protective factors, assessing level of risk, and developing an intervention plan based on clinical judgments [67, 68]. However, the critics of suicide risk assessment are inexplicably silent on this point.

Improving suicide prediction models

A number of issues should be considered in charting a way forward. To begin, we need to know if suicide prediction accuracy can be improved. This, in turn, requires us to consider options for expanding the information obtained about risk factors as well as options for

analyzing this information to optimize prediction. We then need to consider how best to deal with the likelihood that PPV will continue to be relatively low even if we optimize data collection and prediction methods. Finally, we need to consider how best to address the injunction of critics of suicide risk assessment tools to treat patients based on need for services rather than on suicide risk given that an understanding of suicide risk should inform decisions about optimal treatment. We discuss these issues, in turn, in the remainder of the paper.

Expanding information about risk factors with multistage screening

We suggested above that some clinicians appear to use structured suicide prediction tools as a first-stage screen to limit the number of patients who receive more in-depth suicide risk assessments. This kind of multistage screening is routine in many other areas of medicine. For example, blood spot screening for a wide range of genetic disorders is carried out at birth for all the more than 4 million infants born in the US each year along with the infants in most other developed countries and some less developed countries [69]. Positive screens are not definitive, though, so infants whose screens are positive for one or more of the disorders in the screening battery typically receive a repeat screen or a more specific screen and subsequent clinical investigations before determining if there is a need for treatment.

We recently proposed a similar kind of three-stage screening approach to minimize the burden of administering suicide risk screening tests to all patients [70]. The first stage would use passively collected administrative data to build a ML model to predict subsequent suicides and SRBs. The key here is the term passively collected, which means the predictors would consist entirely of information abstracted from administrative records. No patients would be asked any questions. Electronic medical records are the most obvious source of such passive data, including information extracted from unstructured clinical notes using natural language processing for prior indications of suicidality [71–74] or risk factors for suicidality [75, 76]. Another useful type of passive data used in some suicide screening tests [77] is information about demographic or economic characteristics of neighborhoods or labor markets in which patients reside. This kind of information can be linked to patient addresses using aggregated data from such sources as the US Census Bureau's ongoing American Community Survey [78]. More exploratory types of passive data might also be considered in the future, such as information about suicide intent extracted from social media postings [79] or linguistic and acoustic features extracted from tape-recorded responses to open-ended questions [80].

A number of suicide prediction models using passively collected administrative data have been developed and are reviewed elsewhere [70]. Importantly, these models are much better at predicting low risk than high risk. That is, they do a good job of screening out the vast majority of patients who have lower than average suicide risk. For example, it is not uncommon for these models to identify up to three-fourths of patients as having suicide risks less than half those of the total patient population. However, they are not nearly as good at defining a small number of patients with extremely high risk. Models such as these would be ideal to screen out a high proportion of patients from second-stage screening.

Second-stage screening would then involve administering a thoughtfully assembled short battery of self-report scales to patients who are not screened out in the first stage. The goal would be to pinpoint the even smaller set of patients who might benefit from a third-stage comprehensive clinical assessment of suicide risk. A question can be raised about the cost-benefit ratio of the second-stage screen given the criticism noted above that structured suicide risk scales are sometimes perceived by patients as being superficial. Although this criticism is doubtlessly accurate for some patients, experimental research suggests that it is far from the norm given that people are more likely to report embarrassing and stigmatizing thoughts and behaviors, including those involving suicidality, when they are given more confidential modes of reporting, such as self-administered questionnaires, rather than face-to-face interviews [81–84]. Consistent with this observation, a number of novel self-report scales have been developed recently to address the problem of patient under reporting of suicidality that could be used as part of second-stage screening batteries. Included here are performance-based neurocognitive tests of suicide-related implicit cognitions [85], self-report scales about suicide-related beliefs [86], and scales that assess volitional factors shown to predict suicide such as fearlessness of death, impulsivity, and exposure to past suicidal behaviors [87]. Although no attempt has been made to determine what an optimal set of such scales might be or to estimate the prediction strength of a comprehensive battery of these scales, encouraging results exist in studies finding that accuracy of predicting subsequent SRBs increases substantially when a number of these scales are combined into multivariate models rather than considered one at a time [40, 88].

An interesting study that illustrates the potential value of such a staged approach to suicide risk screening was carried out in the US Army Study to Assess Risk and Resilience in Servicemembers (STARRS) [89]. The STARRS investigators developed a first-stage ML model using passively collected administrative data for the entire Army in 2004–2009 to predict SRBs and then applied this model to an independent 2011–2014 sample that consisted of soldiers who participated in a STARRS survey about risk factors for suicide. These survey respondents were then followed 24–48 months after their survey through administrative records and these records were used to determine which respondents experienced an SRB [90]. Results showed that the 70% of survey respondents with the lowest predicted risk scores based on the administrative model had a SRB rate <50% that of the total sample, with the remaining 30% of the sample accounting for 81% of all subsequent SRBs. A second-stage model was then developed in this 30% subsample, focusing on soldiers who denied a history of suicidality in the survey [90]. Although this second-stage model was much less refined than if the survey had been designed explicitly to develop such a model, the 10% of second-stage respondents at highest cross-validated risk of SRB based on the model (i.e., 3% [10% of 30%] of all soldiers in the total sample) accounted for 45% of suicide attempts in the total sample. It might reasonably be argued that third-stage in-depth clinical psychosocial risk assessments should be limited to that 3% of soldiers given that they were the only ones with an observed SRB rate above the mean in the total sample.

Improving data analysis methods

We noted above that recent attempts to develop suicide prediction tools used ML methods. These methods offer a number of advantages over classical methods, including improved ability to handle nonlinear and interactive associations and to work with databases in which either the outcome is rare, the number of predictors is large relative to the number of patients with the outcome, and/or the predictors are highly correlated [91]. It is important to recognize, though, that these advantages are limited to the goal of prediction rather than to understanding underlying causal processes or unique associations associated with particular predictors. Indeed, ML methods sacrifice accuracy of coefficient estimation in the service of increasing overall model prediction accuracy. In addition, some of the advantages of ML can be achieved as well with thoughtful use of classical analysis methods [92]. There are also some cases in which classical methods either perform as well as or better than some ML methods depending on the structure of the data under investigation [93]. This makes it important to weigh the benefits of improved ML model prediction accuracy against the costs of decreased intuitiveness of results, although methods exist to address the latter problem by using methods to assess predictor variable importance [94].

Most studies using ML methods to develop suicide prediction tools compare their results with those obtained with logistic regression and show that the ML results are superior (see the review by Kessler et al. [70]). However, improvements to existing ML model results are nonetheless possible in at least four ways. First, most ML-based suicide prediction tools failed to use crossvalidation, thereby risking overfitting. Second, the vast majority of these tools used only one ML algorithm or compared performance across a small number of different ML algorithms, whereas prediction accuracy could be improved by using ensemble ML methods that combine individual-level predictions across many algorithms [95]. Third, automated machine learning [96, 97] makes it increasingly possible to refine feature transformation and pruning, algorithm selection, and hyperparameter tuning well beyond the default values used in most studies that used ML to develop suicide prediction tools [98]. Fourth, few existing studies to develop suicide prediction tools addressed the problem of extreme imbalance in the outcome. A number of toolkits exist to evaluate the relative effectiveness of different imbalance correction methods [99]. Future efforts to develop ML models for suicide prediction should use these best-practice methods.

Improving decision-making based on the results of prediction models

The problem of low PPV

We noted above that a main criticism of existing suicide prediction tools is that their low PPV undermines their practical value. Even though the three-stage screening approach described above coupled with use of an expanded predictor set and optimal analysis methods would almost certainly result in substantial improvements in suicide prediction, PPV would still inevitably be low in optimal ML suicide prediction models because of the rarity of SRBs. As we noted above, some critics consider this low PPV unacceptable based on the fact that false positives, who would vastly outnumber true positives even in an optimal prediction tool, would be subjected to unnecessary intrusion or coercion [8].

However, a balance needs to be struck between saving lives and increasing intrusion-coercion for some larger number of patients. In the meta-analysis cited above of studies to predict inpatient suicides, for example, a PPV of 0.004 was asserted to be too low [22]. But is increased intrusion and coercion for 249 patients ($1/0.004 = 1$) too high a price to save one life? It is not obvious that a formal cost-benefit analysis would draw such a conclusion. If it did, then the critics would be correct in this particular case. But as this example shows, the facile assertion that PPV is too low to have practical value cannot be assumed to be true without a good deal more thought than has been given to the matter in published critiques. This is especially true given that recent studies have found that up to one-third of patients who do not die by suicide but are classified as at high suicide risk are also at high risk for other extreme negative outcomes, such as deaths classified as accidental or undetermined, nonfatal suicide attempts, serious nonfatal injuries, and psychiatric hospitalizations [100, 101]. The potential to reduce incidence of these outcomes would increase the cost-effectiveness of high-risk interventions.

This perspective on evaluating the benefits relative to costs of interventions for rare outcomes is captured at a more formal level by the concept of Net Benefit (NB). NB is the standardized difference between the number of true positives at or above each potential decision threshold on a continuous PPV scale and the discounted number of false positives at or above that threshold, where the discount rate is defined as $PPV/[1-PPV]$ at the threshold (not at or above the threshold) [102]. NB provides more intuitive and clinically useful information than SN, SP, or PPV because it explicitly evaluates the relative value of intervening with a true positive (i.e., someone who would die by suicide in the absence of intervention) and not intervening with a false positive and thereby helps determine an optimal decision threshold (i.e., the threshold with the highest NB) across the full range of the predicted outcome scale [103]. This means that NB, unlike SN, SP, or PPV, can be used to decide whether an intervention should be used in clinical practice and, if so, where the threshold is on the prediction scale above which it should be used. Importantly, indirect costs (e.g., the costs of intrusion, coercion, and other adverse implications noted by critics) need to be taken into consideration along with direct costs (e.g., the value of clinician time, the costs of diagnostic tests) when calculating NB, as the indirect costs of suicide risk screening and labeling can be very real [8]. Ethical and legal issues arise here that, although not addressed in this paper, also need to be taken into consideration in determining indirect costs [104].

In cases where there is uncertainty or variation across treatment providers in the relative valuations of detecting true positives and excluding true negatives, a *decision curve* can be plotted to provide information on the NB of intervening at different decision thresholds (Fig. 1). The X axis is the minimum PPV the clinician would require to designate a patient as warranting intervention and the Y axis is the NB of the screening tool at that threshold given the SN and SP of the tool [105]. Clinicians can use these curves to select decision thresholds that match their own views about the relative values of saving a life (i.e., intervening with a true positive) and increasing clinician and patient burden (i.e., intervening with false positives). It is also possible to compare alternative screening scales in the same figure to determine which one has the highest NB at the decision threshold selected by a particular clinician.

We illustrate this approach with decision curves for hypothetical prediction tools for repeat SRBs among patients recently discharged from a psychiatric hospital after an SRB. Twelve-month incidence of this outcome is about 8% in the US Veterans Administration health care system [106]. Intensive case management is thought to be one way to reduce this incidence [107], but this would be a laborintensive intervention. A prediction tool used prior to hospital discharge to predict which patients are at high risk of this outcome might be used to offer this intervention only to high-risk patients [100]. Figure 1 illustrates how decision curves would be constructed for a tool with perfect prediction (i.e., $SN = SP = 1.0$), for decision rules to offer the intervention either to all patients or to no patients, and for dichotomous tools with various combinations of SN and SP. A tool that classified patients perfectly would have an NB equal to incidence (i.e., 0.08) at all values of the decision threshold (i.e., the X axis), as such a test would result in treatment of all true positives but no false positives. A decision rule not to offer the intervention to any patients, in comparison, would have an NB of 0 at all values of the decision threshold, whereas a decision rule to offer the intervention to all patients would have an NB of 0.08 at a decision threshold of 0 (i.e., where the cost of treating false positives would be assumed to be 0) and linearly decreasing NB as the decision threshold increased (including an NB of 0 when the decision threshold equaled the incidence). The NB of imperfect classification tools, finally, would have Y intercepts equal to $0.08 \times \text{incidence}$, as the costs of false positives would be 0 when the decision threshold was 0, and NB would decrease nonlinearly as the decision threshold increased. The speed of the decrease in NB with increase in the decision threshold would be inversely related to SP for a fixed SN. In cases where SN is higher for one tool in a pair and SP higher for the other (as in the $SN = 0.95, SP = 0.95$ and $SN = 0.50, SP = 0.99$ curves in the figure), the test with the higher SN would have comparatively higher NB at low decision thresholds and the test with the higher SP comparatively higher NB at higher decision thresholds.

The fact that critics of suicide prediction tools have rejected these tools based on their low PPVs means that these critics implicitly assumed a high discount rate (i.e., a high decision threshold) in their implicit calculations of NB. But this assumption is unwarranted in the absence of explicit justifications that, with rare exceptions [28, 57], are not provided by these critics. As noted elsewhere [9], NB can be positive even when PPV is low, as when statin therapy is considered to have a positive NB for patients ages 40–75 with mildly elevated total cholesterol even though nearly 500 person-years of treatment are needed to prevent one case of atherosclerotic cardiovascular disease [108]. A roughly equivalent intervention designed to prevent suicides or SRBs (i.e., one with low unit cost that would need to be provided to a large number of patients to prevent one suicide or SRB) might be one that sent brief caring text messages to suicidal patients. Such interventions have been shown to reduce SRBs among suicide patients [109]. The cost of this intervention is very low both in terms of clinician time and iatrogenic effects, which means that NB of the intervention might be optimal at a very low PPV using the standard approach of federal agencies in the US to calculate the Value of a Statistical Life [110].

We are not asserting here that NB analyses will demonstrate that existing or future suicide prediction tools have value in guiding any clinical decisions that are currently being made on a wide-scale basis. Even in the case of the example presented here, where one could easily

imagine that intensive case management would be cost-effective even if it only rarely saved a life, an argument could be made that all patients should receive intensive case management given the high value of a life in relation to the costs of the intervention. If so, then none of the hypothetical suicide prediction tools considered here or others that could be imagined would have value, as the decision to treat all will always be close to optimal when the decision threshold is close to 0. However, as shown in the figure, suicide prediction tools can have higher NB than the decision rule to treat all in situations where intervention costs (direct and indirect) are high and the decision threshold is consequently meaningfully >0 . It is important not to dismiss this possibility out of hand in the way critics of suicide prediction tools have done up to now.

The problem of low SN

We noted above that the second main criticism of existing suicide prediction tools is their low SN, which means that many of the patients who go on to die by suicide or experience an SRB are not classified by these tools as having high risk. Critics argue based on this fact that need for services, as determined by in-depth evaluation of risk and protective factors, should be the focus on clinical attention rather than risk of suicide. As we noted earlier in this paper, though, clinical decisions about need for services should be influenced by assessments of suicide risk. At the same time, critics of suicide prediction tools are correct in noting that information about the magnitude of a patient's suicide risk may not in itself tell us a great deal about optimal treatment other than in the extreme case of involuntary commitment. But the conclusion from this fact that prediction tools should be abandoned in favor of clinical judgment is unlikely to be correct. As we know that clinical judgment about suicide risk is poor, why would we expect clinical judgment to be any better in synthesizing information about risk and protective factors to determine an optimal treatment plan?

It is important to recognize in this regard that suicide-focused clinical interventions generally have relatively weak effects [111] aside from a few widely accepted universal interventions [67]. One possible reason for this is that heterogeneity of treatment effects exists in these interventions. This possibility is likely for two reasons. First, suicides and SRBs occur in conjunction with manifold mental disorders. Despite many similar issues in managing suicidality across these disorders, meaningful differences exist in causal risk factors across these disorders. Second, SRB-focused interventions target intermediate outcomes, such as increased perceptions of belongingness in caring text interventions [109] or "suicide drivers" in Collaborative Assessment and Management of Suicidality therapy [112]. But each of these is an issue for only a subset of patients, leading inevitably to weak aggregate effects even if large effects exist among the subset of patients for whom the intervention focus is most relevant. It is consequently of considerable importance to make thoughtful decisions in selecting among alternative interventions with roughly comparable aggregate effects for particular patients. This would not resolve the problem of low SN, but would reframe the problem to one of developing a model to predict the type of treatment needed by a given patient rather than a model to predict the patient's suicide risk.

The need for individualized treatment rules

We currently have no way of predicting optimal SRB-focused treatments at the individual level because virtually no systematic research exists on heterogeneity of treatment effects for SRB-focused interventions. Instead, existing suicide prediction models are designed to target the patients at highest suicide risk and to leave it to the discretion of clinicians to decide on the best treatments for their patients without the models giving any guidance about the latter decisions [113]. But important questions exist about these decisions. For example, although several outpatient therapies have been shown in replicated trials to reduce SRBs with modest aggregate effect sizes [63], none of these trials was designed to investigate which patients benefit from these therapies or carried out comparative effectiveness analyses to determine which specific SRB-focused therapies are optimal for which patients. Clinicians consequently must rely on clinical judgment informed by their past personal experience in deciding whether to use one of these SRB-focused treatments rather than more general treatment and, if using an SRB-focused treatment, which one to use.

The development of comprehensive individualized treatment rules (ITRs) using precision medicine analysis methods would be of enormous value in addressing the above problems by providing guidance on which treatments are likely to be best for which patients and which times. ITRs are different from risk prediction tools because the patients at highest suicide risk are not necessarily the patients most likely to be helped by available interventions. This means that the models to develop ITRs can be thought of as evaluating interactions between predictors of treatment response and treatment type in comparative effectiveness studies, ideally evaluated in controlled treatment effectiveness trials that have real-world validity [114]. When significant interactions of this sort exist, the predictors are referred to as *prescriptive* predictors; that is, predictors of greater response to one type of treatment than another. A difficulty arises, though, when the number of prescriptive predictors is large and/or when the functional forms of the interactions are complex, in which case conventional estimation methods break down. ML methods can be used in these cases [115] and, in fact, can be applied even when treatment is not randomly assigned by using double-robust estimation methods [41, 116] so long as all confounders that are a common cause of treatment assignment and suicide within strata of treatment are measured. Minor violations of the latter assumption are not expected to induce major biases in the analysis if, as in the case of suicide, loss to follow-up outcome assessment is low [117].

ITRs have the potential to provide decision support for a great many clinical decisions that are currently being made based on clinical judgment. Consider, for example, the decision to hospitalize a patient after a suicide attempt. Despite hospitalization being the standard of care, some experts believe that hospitalization can itself be a traumatic experience and can cause negative societal reactions after discharge that contribute to the high suicide rate of recently discharged psychiatric inpatients [70]. Consistent with this concern, recent non-experimental comparative effectiveness studies using propensity score adjustment methods to compare patients who were hospitalized versus treated as outpatients after a nonfatal SRB found that the hospitalized patients were no less likely to die by suicide [118] and were significantly more likely to experience a repeat SRB [119] over the subsequent 12 months than the patients treated as outpatients. This was true, importantly, after adjusting

statistically for baseline differences in the severity indicators that differentiated patients who were hospitalized from patients who were treated as outpatients. This might be because hospitalization has no effect in reducing repeat SRB. Or it might be that some patients are helped and others hurt by hospitalization after an SRB. If the latter is the case, as many experts believe it is, careful weighing of risks and benefits is needed to determine which patients are likely to be helped and which ones harmed by hospitalization [120]. No guidance exists currently on how this weighing should be done. It is plausible to think that an ITR could be developed to decrease the current suicide and SRB rates after nonfatal SRBs by improving decisions about which patients to hospitalize. ML methods exist that would allow this to be done if relevant predictors were assessed [121].

Conclusions

Three broad conclusions can be drawn from the above review. First, existing suicide prediction tools have little current clinical value. But this is not because of the low SN and low PPV that have been the focus of critics of these tools, as prediction tools can have clinical value despite these low values. The lack of clinical value is instead due to the fact that we lack the kind of information about the effectiveness of targeted suicide preventive interventions that is needed to calculate NB. And NB, rather than SN or PPV, needs to be recognized as the key operating characteristic of screening scales for determining clinical relevance. Evaluation of NB requires concrete consideration of specific clinical decision options and serious weighing of the expected risks and benefits of each. Principled decision thresholds can be set so long as these expectations are clear, NB is specified, and models are developed that produce predicted probabilities of the suicide or SRBs for each patient. In the absence of such information, it is impossible to use screening scales to support clinical decision-making or to know if these tools might have clinical value. The use of structured suicide risk tools in the absence of such information has the potential to do more harm than good by increasing patient burden, disrupting therapeutic alliances, and leading to misguided clinical decisions.

Second, in the course of developing new suicide prediction tools keyed to specific clinical decisions aimed at reducing suicides, it will be important to address a number of technical issues to optimize model performance. These include expansion of the predictors used in the models, which would be facilitated by adopting a multistage screening approach, and using improved analysis methods.

Third, future research is needed to develop ITRs for patients at elevated suicide risk. Existing suicide prediction models, which attempt to pinpoint the patients at highest suicide risk, do not tell us which suicide-focused treatments are right for which patients. This is the key clinical question critics of suicide risk screening have in mind when they say that in-depth clinical assessments should focus on need for treatment rather than risk of suicide. As it is now, though, treatment guidelines offer no practical guidance on how to select optimal individualized treatments. The technology exists for addressing this question by reorienting suicide prediction models to focus on prescriptive predictors and develop ITRs. This needs to be a central focus of future suicide prediction research.

Acknowledgements

This work was supported, in part, by the Department of Veterans Affairs Center of Excellence for Suicide Prevention and the Precision Treatment of Mental Disorders Initiative. The contents are solely the responsibility of the authors and do not necessarily represent the views of the Veteran's Health administration. The authors appreciate the helpful comments of Matthew K. Nock, Vicki Shahly, Murray B. Stein, and Robert J. Ursano on an earlier version of the paper.

References

1. World Health Organization (WHO). Mental health: suicide data. 2018 http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ Accessed 30 April 2019.
2. Ahmedani B, Simon G, Stewart C, Beck A, Waitzfelder B, Rossom R, et al. Health care contacts in the year before suicide death. *J Gen Intern Med*. 2014;29:870–7.
3. Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. *Am J Psychiatry*. 2002;159:909–16. [PubMed: 12042175]
4. Pearson A, Saini P, Da Cruz D, Miles C, While D, Swinson N, et al. Primary care contact prior to suicide in individuals with mental illness. *Br J Gen Pr*. 2009;59:825.
5. Schaffer A, Sinyor M, Kurdyak P, Vigod S, Sareen J, Reis C, et al. Population-based analysis of health care contacts among suicide decedents: identifying opportunities for more targeted suicide prevention strategies. *World Psychiatry*. 2016;15:135–45. [PubMed: 27265704]
6. Rosen A Detection of suicidal patients: an example of some limitations in the prediction of infrequent events. *J Consult Psychol*. 1954;18:397–403. [PubMed: 13221698]
7. Murphy GE. Clinical identification of suicidal risk. *Arch Gen Psychiatry*. 1972;27:356–9. [PubMed: 5051624]
8. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. *JAMA Psychiatry*. 2019;76:642–51. [PubMed: 30865249]
9. Kessler RC. Clinical epidemiological research on suicide-related behaviors: where we are and where we need to go. *JAMA Psychiatry* 2019;76:777–8.
10. Garb HN, Wood JM. Methodological advances in statistical prediction. *Psychol Assess*. 2019 10.1037/pas0000673.
11. Naghavi M. Global, regional, and national burden of suicide mortality 1990–2016: systematic analysis for the Global Burden of Disease Study 2016. *BMJ*. 2019;364:194. [PubMed: 31339847]
12. Katz C, Bolton J, Sareen J. The prevalence rates of suicide are likely underestimated worldwide: why it matters. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51:125–7. [PubMed: 26590856]
13. Stone D, Simon T, Fowler K, Kegler S, Yuan K, Holland K et al. Vital signs: trends in state suicide rates - United States, 1999–2016 and circumstances contributing to suicide - 27 States, 2015. *MMWR Morb Mortal Wkly Rep*. 2018; 67: 617–24. [PubMed: 29879094]
14. Heron M Deaths: leading causes for 2016. National Vital Statistics Reports. 67. Hyattsville, MD: National Center for Health Statistics 2018 https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_06.pdf.
15. Carroll R, Metcalfe C, Gunnell D. Hospital presenting self-harm and risk of fatal and non-fatal repetition: systematic review and meta-analysis. *PLoS ONE*. 2014;9:e89944. [PubMed: 24587141]
16. Walsh G, Sara G, Ryan CJ, Large M. Meta-analysis of suicide rates among psychiatric in-patients. *Acta Psychiatr Scand*. 2015;131:174–84. [PubMed: 25559375]
17. Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, Large MM. Suicide rates after discharge from psychiatric facilities: a systematic review and meta-analysis. *JAMA Psychiatry*. 2017;74:694–702. [PubMed: 28564699]
18. Chung D, Hadzi-Pavlovic D, Wang M, Swaraj S, Olfson M, Large M. Meta-analysis of suicide rates in the first week and the first month after psychiatric hospitalisation. *BMJ Open*. 2019;9: e023883.

19. Carter G, Milner A, McGill K, Pirkis J, Kapur N, Spittal MJ. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *Br J Psychiatry*. 2017;210:387–95. [PubMed: 28302700]
20. Chan KCG, Yam SCP, Zhang Z. Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *J R Stat Soc Ser B Stat Methodol*. 2016;78:673–700.
21. Katz C, Randall JR, Sareen J, Chateau D, Walld R, Leslie WD, et al. Predicting suicide with the SAD PERSONS scale. *Depress Anxiety*. 2017;34:809–16. [PubMed: 28471534]
22. Large M, Myles N, Myles H, Corderoy A, Weiser M, Davidson M, et al. Suicide risk assessment among psychiatric inpatients: a systematic review and meta-analysis of high-risk categories. *Psychol Med*. 2018;48:1119–27. [PubMed: 28874218]
23. Large M, Kaneson M, Myles N, Myles H, Gunaratne P, Ryan C. Meta-analysis of longitudinal cohort studies of suicide risk assessment among psychiatric patients: heterogeneity in results and lack of improvement over time. *PLoS ONE*. 2016;11:e0156322. [PubMed: 27285387]
24. Larkin C, Di Blasi Z, Arensman E. Risk factors for repetition of self-harm: a systematic review of prospective hospital-based studies. *PLoS ONE*. 2014;9:e84282. [PubMed: 24465400]
25. Quinlivan L, Cooper J, Davies L, Hawton K, Gunnell D, Kapur N. Which are the most useful scales for predicting repeat self-harm? A systematic review evaluating risk scales using measures of diagnostic accuracy. *BMJ Open*. 2016;6:e009297.
26. Runeson B, Odeberg J, Pettersson A, Edbom T, Jildevik Adamsson I, et al. Instruments for the assessment of suicide risk: a systematic review evaluating the certainty of the evidence. *PLoS ONE*. 2017;12:e0180292. [PubMed: 28723978]
27. Bolton JM, Gunnell D, Turecki G. Suicide risk assessment and intervention in people with mental illness. *Brit Med J*. 2015;351:h4978. [PubMed: 26552947]
28. Woodford R, Spittal MJ, Milner A, McGill K, Kapur N, Pirkis J, et al. Accuracy of clinician predictions of future self-harm: a systematic review and meta-analysis of predictive studies. *Suicide Life Threat Behav*. 2019;49:23–40. [PubMed: 28972271]
29. Barak-Corren Y, Castro VM, Javitt S, Hoffnagle AG, Dai Y, Perlis RH, et al. Predicting suicidal behavior from longitudinal electronic health records. *Am J Psychiatry*. 2017;174:154–62. [PubMed: 27609239]
30. Ben-Ari A, Hammond K. Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 Persian gulf war. 2015 48th Hawaii International Conference on System Sciences Kauai, HI; 2015;3168–75. 10.1109/HICSS.2015.
31. Choi SB, Lee W, Yoon J-H, Won J-U, Kim DW. Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea. *J Affect Disord*. 2018;231:8–14. [PubMed: 29408160]
32. Kessler RC, Hwang I, Hoffmire CA, McCarthy JF, Petukhova MV, Rosellini AJ, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration. *Int J Methods Psychiatr Res*. 2017;26 10.1002/mpr.1575
33. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci*. 2017;5:457–69.
34. Simon R Improving suicide risk assessment with evidence-based psychiatry In: Pompili M, Taterelli R editors. Evidence-based practice in suicidology: a sourcebook. Cambridge MA: Hogrefe Publishing; 2011, p. 45–54.
35. Beck A, Steer R. BHS, Beck Hopelessness Scale: manual. San Antonio TX: Psychological Corporation; 1988.
36. Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II. San Antonio TX: Psychological corporation; 1996.
37. Beck A, Schuyler D, Herman I. Development of suicidal intent scales In: Beck A, Lettieri D, Resnik H editors. The prediction of suicide. Bowie, MD: Charles Press; 1974, p. 45–56.
38. Koldslund BO, Mehlum L, Mellestad LS, Walby FA, Diep LM. The suicide assessment scale: psychometric properties of a Norwegian language version. *BMC Res Notes*. 2012;5:417 10.1186/1756-0500-5-417 [PubMed: 22870950]

39. Kreitman N, Foster J. The construction and selection of predictive scales, with special reference to parasuicide. *Br J Psychiatry*. 1991;159:185–92. [PubMed: 1773234]
40. Randall JR, Rowe BH, Dong KA, Nock MK, Colman I. Assessment of self-harm risk using implicit thoughts. *Psychol Assess*. 2013;25:714–21. [PubMed: 23647043]
41. Robins JM, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *J Am Stat Assoc*. 2015;90:122–9.
42. Bolton JM. Suicide risk assessment in the emergency department: out of the darkness. *Depress Anxiety*. 2015;32:73–75. [PubMed: 25421638]
43. Hoge CW. Suicide reduction and research efforts in service members and veterans-sobering realities. *JAMA Psychiatry*. 2019 10.1001/jamapsychiatry.2018.4564.
44. Mulder R, Newton-Howes G, Coid JW. The futility of risk prediction in psychiatry. *Br J Psychiatry*. 2016;209:271–2. [PubMed: 27698212]
45. Owens D, Kelley R. Predictive properties of risk assessment instruments following self-harm. *Br J Psychiatry*. 2017;210:384–6. [PubMed: 28572432]
46. Wortzel HS, Nazem S, Bahraini NH, Matarazzo BB. Why suicide risk assessment still matters. *J Psychiatr Pr*. 2017; 23:436–40.
47. Hunter C, Chantler K, Kapur N, Cooper J. Service user perspectives on psychosocial assessment following self-harm and its impact on further help-seeking: a qualitative study. *J Affect Disord*. 2013;145:315–23. [PubMed: 22925352]
48. Owens C, Hansford L, Sharkey S, Ford T. Needs and fears of young people presenting at accident and emergency department following an act of self-harm: secondary analysis of qualitative data. *Br J Psychiatry*. 2016;208:286–91. [PubMed: 26450583]
49. Taylor TL, Hawton K, Fortune S, Kapur N. Attitudes towards clinical services among people who self-harm: systematic review. *Br J Psychiatry*. 2009;194:104–10. [PubMed: 19182168]
50. Palmer L, Blackwell H, Strevens P. Service users' experience of emergency services following self harm: a national survey of 509 patients. College Centre for Quality Improvement, Royal College of Psychiatrists 2007 <https://www.rcpsych.ac.uk/> Accessed 20 Feb 2018
51. Rosen DC, Nakash O, Alegria M. The impact of computer use on therapeutic alliance and continuance in care during the mental health intake. *Psychother (Chic)*. 2016;53:117–23.
52. Self-harm in over 8s: long-term management. National Institute for Health and Care Excellence (NICE) 2011 Accessed 30 April 2019
53. O'Connor E, Gaynes BN, Burda BU, Soh C, Whitlock EP. Screening for and treatment of suicide risk relevant to primary care: a systematic review for the US Preventive Services Task Force. *Ann Intern Med*. 2013;158:741–54. [PubMed: 23609101]
54. National Institute for Health and Care Excellence. Preventing suicide in community and custodial settings. 2018 <https://www.nspa.org.uk/wp-content/uploads/2018/09/preventing-suicide-in-community-and-custodial-settings-pdf-66141539632069.pdf>.
55. Bernert R, Hom M, Roberts L. A review of multidisciplinary clinical practice guidelines in suicide prevention: toward an emerging standard in suicide risk assessment and management, training and practice. *Acad Psychiatry*. 2014;38:585–92. [PubMed: 25142247]
56. Silverman JJ, Galanter M, Jackson-Triche M, Jacobs DG, Lomax JW, Riba MB, et al. The American Psychiatric Association practice guidelines for the psychiatric evaluation of adults. *Am J Psychiatry*. 2015;172:798–802. [PubMed: 26234607]
57. Quinlivan L, Cooper J, Steeg S, Davies L, Hawton K, Gunnell D, et al. Scales for predicting risk following self-harm: an observational study in 32 hospitals in England. *BMJ Open*. 2014;4:e004732.
58. Rudd M Core competencies, warning signs, and a framework for suicide risk assessment in clinical practice In: Nock M editor. *The Oxford handbook of suicide and self-injury*. 1st ed New York: Oxford University Press; 2014 p. 323–36.
59. Cooper J, Steeg S, Bennewith O, Lowe M, Gunnell D, House A, et al. Are hospital services for self-harm getting better? An observational study examining management, service provision and temporal trends in England. *BMJ Open*. 2013;3:e003444.
60. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989;243:1668–74. [PubMed: 2648573]

61. ġisdottir S, White MJ, Spengler PM, Maugherman AS, Anderson LA, Cook RS, et al. The meta-analysis of clinical judgment project: fifty-six years of accumulated research on clinical versus statistical prediction. *Couns Psychol.* 2006;34:341–82.
62. Large M, Sharma S, Cannon E, Ryan C, Nielssen O. Risk factors for suicide within a year of discharge from psychiatric hospital: a systematic meta-analysis. *Aust N Z J Psychiatry.* 2011;45:619–28. [PubMed: 21740345]
63. Jobes D, Au J, Siegelman A. Psychological approaches to suicide treatment and prevention. *Curr Treat Options Psychiatry.* 2015;2:363–70.
64. Smith KA, Cipriani A. Lithium and suicide in mood disorders: updated meta-review of the scientific literature. *Bipolar Disord.* 2017;19:575–86. [PubMed: 28895269]
65. Vermeulen JM, van Rooijen G, van de Kerkhof MPJ, Sutherland AL, Correll CU, de Haan L. Clozapine and long-term mortality risk in patients with schizophrenia: a systematic review and meta-analysis of Studies Lasting 1.1–12.5 Years. *Schizophr Bull.* 2019;45:315–29. [PubMed: 29697804]
66. United States. Public Health Service. Office of the Surgeon General. 2012 National strategy for suicide prevention: goals and objectives for action. National Action Alliance for Suicide Prevention. Washington, DC: National Action Alliance for Suicide Prevention; 2012 <https://www.surgeongeneral.gov/library/reports/national-strategy-suicide-prevention/full-report.pdf>.
67. Brodsky BS, Spruch-Feiner A, Stanley B. The zero suicide model: applying evidence-based suicide prevention practices to clinical care. *Front Psychiatry.* 2018;9:33. [PubMed: 29527178]
68. Jacobs DG. Suicide Assessment Five-step Evaluation and Triage for mental health professionals (SAFE-T). 2009 https://www.integration.samhsa.gov/images/res/SAFE_T.pdf.
69. Fernhoff PM. Newborn screening for genetic disorders. *Pediatr Clin North Am.* 2009;56:505–13. [PubMed: 19501689]
70. Kessler R, Bernecker S, Bossarte R, Luedtke A, McCarthy JF, Nock MK, et al. The role of big data analytics in predicting suicide In: Passos I, Mwangi B, Kapczinski F, editors. *Personalized Psychiatry—big data analytics in mental health.* Springer Nature, 2019 p. 77–98.
71. Hammond KW, Laundry RJ, O’Leary TM, Jones WP. Use of text search to effectively identify lifetime prevalence of suicide attempts among Veterans. 2013 46th Hawaii International Conference on System Sciences; Wailea, Maui, HI; 2013 p. 2676–83.
72. Hammond KW, Laundry RJ. Application of a hybrid text mining approach to the study of suicidal behavior in a large population. 2014 47th Hawaii International Conference on System Science; Waikoloa, HI; 2014 p. 2555–61.
73. Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. *Sci Rep.* 2018;8:7426. [PubMed: 29743531]
74. Carson NJ, Mullin B, Sanchez MJ, Lu F, Yang K, Menezes M, et al. Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records. *PLoS ONE.* 2019;14:e0211116. [PubMed: 30779800]
75. Hammond KW, Ben-Ari AY, Laundry RJ, Boyko EJ, Samore MH. The feasibility of using large-scale text mining to detect adverse childhood experiences in a VA-treated population. *J Trauma Stress.* 2015;28:505–14. [PubMed: 26579624]
76. McCoy TH, Pellegrini AM, Perlis RH. Research domain criteria scores estimated through natural language processing are associated with risk for suicide and accidental death. *Depress Anxiety.* 2019;36:392–9. [PubMed: 30710497]
77. Simon GE, Johnson E, Lawrence JM, Rossom RC, Ahmedani B, Lynch FL, et al. Predicting Suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry.* 2018;175:951–60. [PubMed: 29792051]
78. American Community Survey (ACS). United States Census Bureau. 2018 <https://www.census.gov/programs-surveys/acs/about.html> Accessed 8 Aug 2019.
79. Lopez-Castroman J, Moulahi B, Aze J, Bringay S, Deninotti J, Guillaume S, et al. Mining social networks to improve suicide prevention: a scoping review. *J Neurosci Res.* 2019 10.1002/jnr24404.

80. Pestian JP, Sorter M, Connolly B, Bretonnel Cohen K, McCullumsmith C, Gee JT, et al. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. *Suicide Life Threat Behav.* 2017;47:112–21. [PubMed: 27813129]
81. Brown JL, Swartzendruber A, Diclemente RJ. Application of audio computer-assisted self-interviews to collect self-reported health data: an overview. *Caries Res.* 2013;47:40–5. [PubMed: 24107606]
82. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. *Behav Res Methods.* 2015;47:1237–59. [PubMed: 25410404]
83. Greist JH, Laughren TP, Gustafson DH, Stauss FF, Rowse GL, Chiles JA. A computer interview for suicide-risk prediction. *Am J Psychiatry.* 1973;130:1327–32. [PubMed: 4585280]
84. Levine S, Ancill RJ, Roberts AP. Assessment of suicide risk by computer-delivered self-rating questionnaire: preliminary findings. *Acta Psychiatr Scand.* 1989;80:216–20. [PubMed: 2683590]
85. Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR. Measuring the suicidal mind: implicit cognition predicts suicidal behavior. *Psychol Sci.* 2010;21:511–7. [PubMed: 20424092]
86. Bryan CJ, Rudd MD, Wertenberger E, Etienne N, Ray-Sannerud BN, Morrow CE, et al. Improving the detection and prediction of suicidal behavior among military personnel by measuring suicidal beliefs: an evaluation of the Suicide Cognitions Scale. *J Affect Disord.* 2014;159:15–22. [PubMed: 24679384]
87. Dhingra K, Boduszek D, O'Connor RC. Differentiating suicide attempters from suicide ideators using the Integrated Motivational-Volitional model of suicidal behaviour. *J Affect Disord.* 2015;186:211–8. [PubMed: 26247914]
88. Stefansson J, Nordstrom P, Runeson B, Asberg M, Jokinen J. Combining the Suicide Intent Scale and the Karolinska Interpersonal Violence Scale in suicide risk assessments. *BMC Psychiatry.* 2015;15:226. [PubMed: 26400812]
89. Ursano RJ, Colpe LJ, Heeringa SG, Kessler RC, Schoenbaum M, Stein MB. The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Psychiatry.* 2014;77:107–19. [PubMed: 24865195]
90. Bernecker SL, Zuromski KL, Gutierrez PM, Joiner TE, King AJ, Liu H, et al. Predicting suicide attempts among soldiers who deny suicidal ideation in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Behav Res Ther.* 2018 10.1016/j.brat.2018.11.018
91. Boulesteix AL, Schmid M. Machine learning versus statistical modeling. *Biom J.* 2014;56:588–93. [PubMed: 24615669]
92. Harrell JFE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed 2015 Cham, Switzerland: Springer International Publishing; 2015.
93. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019 10.1016/j.clinepi.2019.02.004.
94. Archer K, Kimes R. Empirical characterization of random forest variable importance measures. *Comput Stat Data Anal.* 2008;52:2249–60.
95. Polley E, LeDell E, van der Laan MJ. Super learner: super learner prediction. R package version 2.0-21: The Comprehensive R Archive Network; 2016 [Computer software]. Available at: <https://cran.rstudio.org/> Accessed 30 April 2019.
96. Feurer M, Klein A, Eggenberger K, Springenberg JT, Blum M, Hutter F. Efficient and robust automated machine learning. *Proceedings of the 28th International Conference on Neural Information Processing Systems Vol 2*; Montreal, Canada; 2015.
97. Olson RS, Sipper M, La Cava W, Tartarone S, Vitale S, Fu W et al. A system for accessible artificial intelligence. arXiv:1705.00594v2. 2017 Available from: <https://arxiv.org/abs/1705.00594>.
98. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief based feature selection: introduction and review. arXiv:1711.08421. 2018 Available from: <https://arxiv.org/abs/1711.08421>.

99. Chawla N Data mining for imbalanced datasets: an overview In: Maimon O, Rokach L editors. Data mining and knowledge discovery handbook. 2nd ed Berlin/Heidelberg, Germany: Springer; 2010 p. 875–86.
100. Kessler RC, Warner CH, Ivany C, Petukhova MV, Rose S, Bromet EJ, et al. Predicting suicides after psychiatric hospitalization in US Army Soldiers: the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2015;72:49–57. [PubMed: 25390793]
101. McCarthy JF, Bossarte RM, Katz IR, Thompson C, Kemp J, Hannemann CM, et al. Predictive modeling and concentration of the risk of suicide: implications for preventive interventions in the US Department of Veterans Affairs. *Am J Public Health*. 2015;105:1935–42. [PubMed: 26066914]
102. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *Brit Med J*. 2016;352:i6. [PubMed: 26810254]
103. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur Urol*. 2018;74:796–804. [PubMed: 30241973]
104. McKernan LC, Clayton EW, Walsh CG. Protecting life while preserving liberty: ethical recommendations for suicide prevention with artificial intelligence. *Front Psychiatry*. 2018;9:650. [PubMed: 30559686]
105. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26:565–74.
106. Hoffmire C, Stephens B, Morley S, Thompson C, Kemp J, Bossarte RM. VA Suicide Prevention Applications Network: a national health care system-based suicide event tracking system. *Public Health Rep*. 2016;131:816–21. [PubMed: 28123228]
107. Miller IW, Gaudiano BA, Weinstock LM. The coping long term with active suicide program: description and pilot. *Suicide Life Threat Behav*. 2018;46:752–61.
108. Stone JN, Robinson GJ, Lichtenstein HA, Bairey Merz NC, Blum BC, Eckel HR, et al. 2013 ACC/AHA Guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Circulation*. 2014;129:S1–45. [PubMed: 24222016]
109. Comtois KA, Kerbrat AH, DeCou CR, Atkins DC, Majeres JJ, Baker JC, et al. Effect of augmenting standard care for military personnel with brief caring text messages for suicide prevention: a randomized clinical trial. *JAMA Psychiatry*. 2019;76:474–83. [PubMed: 30758491]
110. Mortality risk valuation. United States Environmental Protection Agency; [updated February 8, 2018]; Available from: <https://www.epa.gov/environmental-economics/mortality-risk-valuation> Accessed Mar 2019.
111. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. *Lancet Psychiatry*. 2016;3: 646–59. [PubMed: 27289303]
112. Jobes DA. The Collaborative Assessment and Management of Suicidality (CAMS): an evolving evidence-based clinical approach to suicidal risk. *Suicide Life Threat Behav*. 2012;42:640–53. [PubMed: 22971238]
113. Frakt AB, Prentice JC, Pizer SD, Elwy AR, Garrido MM, Kilbourne AM, et al. Overcoming challenges to evidence-based policy development in a large, integrated delivery system. *Health Serv Res*. 2018;53:4789–807. [PubMed: 29862494]
114. Cohen ZD, DeRubeis RJ. Treatment selection in depression. *Annu Rev Clin Psychol*. 2018;14:209–36. [PubMed: 29494258]
115. VanderWeele T, Luedtke A, van der Laan MJ, Kessler RC. Selecting optimal subgroups for treatment using many covariates. arXiv:1802.09642. 2018 Available from: <https://arxiv.org/abs/1802.09642>.
116. Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. *J Am Stat Assoc*. 2015;110:910–22.

117. Luedtke AR, van der Laan MJ. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann Stat.* 2016;44:713–42. [PubMed: 30662101]
118. Steeg S, Carr M, Emsley R, Hawton K, Waters K, Bickley H, et al. Suicide and all-cause mortality following routine hospital management of self-harm: propensity score analysis using multicentre cohort data. *PLoS ONE.* 2018;13:e0204670. [PubMed: 30261030]
119. Ichimura A, Kato K, Taira T, Otsuka H, Seki T, Nakagawa Y et al. Psychiatric hospitalization after emergency treatment for deliberate self-harm is associated with repeated deliberate self-harm. *Arch Suicide Res.* 2018 10.1080/13811118.2018.1438323.
120. Large MM, Kapur N. Psychiatric hospitalisation and the risk of suicide. *Br J Psychiatry.* 2018;212:269–73. [PubMed: 30056818]
121. Luedtke AR, van Der, Laan MJ. Optimal individualized treatments in resource-limited settings. *Int J Biostat.* 2016;12:283–303. [PubMed: 27227725]

Fig. 1.
Illustrative decision curves of alternative intervention possibilities to prevent suicide.
Adapted from [105]. We assume that incidence of repeat SRBs is 0.08. See the text for background