The Role of Big Data Analytics in Predicting Suicide

5

Ronald C. Kessler, Samantha L. Bernecker, Robert M. Bossarte, Alex R. Luedtke, John F. McCarthy, Matthew K. Nock, Wilfred R. Pigeon, Maria V. Petukhova, Ekaterina Sadikova, Tyler J. VanderWeele, Kelly L. Zuromski, and Alan M. Zaslavsky

This chapter reviews the long history of using electronic medical records and other types of big data to predict suicide. Although a number of the most recent of these studies used machine learning (ML) methods, these studies were all suboptimal both in the features used as predictors and in the analytic approaches used to develop the prediction models. We review these limitations and describe opportunities for making improvements in future applications. We also review the controversy among clinical experts about using structured suicide risk assessment tools (be they based on ML or older prediction methods) versus in-depth clinical evaluations of needs for treatment planning. Rather than seeing them as competitors, we propose integrating

R. C. Kessler (✉) · M. V. Petukhova · E. Sadikova · A. M. Zaslavsky
Department of Health Care Policy, Harvard Medical School, Boston, MA, USA
e-mail: Kessler@hcp.med.harvard.edu

S. L. Bernecker · M. K. Nock · K. L. Zuromski
Department of Psychology, Harvard University, Cambridge, MA, USA

R. M. Bossarte · W. R. Pigeon
Departments of Behavioral Medicine and Psychiatry, West Virginia University School of Medicine, Morgantown, WV, USA

U.S. Department of Veterans Affairs Center of Excellence for Suicide Prevention, Canandaigua, NY, USA

A. R. Luedtke
Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

J. F. McCarthy
Serious Mental Illness Treatment Resource and Evaluation Center, Office of Mental Health Operations, VA Center for Clinical Management Research, Ann Arbor, MI, USA

T. J. VanderWeele
Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

I. C. Passos et al., *Personalized Psychiatry*,
https://doi.org/10.1007/978-3-030-03553-2_5

these different approaches to capitalize on their complementary strengths. We also emphasize the distinction between two types of ML analyses: those aimed at predicting which patients are at highest suicide risk, and those aimed at predicting the treatment options that will be best for individual patients. We explain why both are needed to optimize the value of big data ML methods in addressing the suicide problem.

5.1 Introduction

Suicide is the 17th leading cause of death in the world (approximately 800,000 suicides per year) and the second leading cause of death among 15–29 year olds (World Health Organization [WHO] 2018a). The actual number of suicides is likely to be higher, as some suicides are misclassified as accidental deaths (Katz et al. 2016). Psychological autopsy studies find that up to 90% of people who died by suicide in Western countries met criteria for a mental disorder (Joiner et al. 2017). In addition, up to 90% of suicide decedents in Western countries came into contact with the healthcare system in the year before death, up to two-thirds had a mental health treatment contact during that year, up to 30% had a psychiatric hospitalization or emergency department visit for a psychiatric problem during that year, and up to one-third were in mental health treatment in the month before death (Ahmedani et al. 2014; Luoma et al. 2002; Pearson et al. 2009; Schaffer et al. 2016). This high level of contact with the healthcare system represents a major opportunity to improve detection of suicide risk in health care settings and target interventions that substantially reduce suicides (Berman and Silverman 2014).

The value of systematically quantifying suicide risk has been debated for over 60 years. In 1954, Rosen argued that the low incidence of suicide poses a substantial barrier, "for in the attempt to predict suicide or any other infrequent event, a large number of false positives are obtained," which means that "such an index would have no practical value, for it would be impossible to treat as potential suicides the prodigious number of false positives" and treating only those at highest risk as potential suicides would miss the majority of true positives. Murphy (1972) countered that the practicality of suicide risk prediction depends on "what is considered appropriate treatment for persons at increased risk of suicide." This debate has continued since these early commentaries at the same time that empirical research has been carried out to improve prediction models and address the problems of false positives and false negatives. Recent studies have used machine learning (ML) methods to develop these models. We begin our review of the literature with a consideration of earlier studies on risk factors for suicide among hospital inpatients and other high-risk patients. We then discuss the ongoing controversy about using structured suicide risk assessment tools. We then review recent studies that used ML methods to predict suicide risk. Finally, we close with recommendations for future studies.

5.2 Earlier Multivariate Analyses Predicting Suicide Among Inpatients

Due to the rarity and short duration of most psychiatric hospitalizations, the proportion of all suicides that occurs among psychiatric inpatients is estimated to be no more than about 5% (Madsen et al. 2017). However, conditional suicide risk among psychiatric inpatients is nonetheless high, especially during the times they are out on temporary leave, with a recent meta-analysis estimating this rate to be 147/100,000 inpatient-years (Walsh et al. 2015) compared to a global population-wide age-standardized suicide rate of 10.7/100,000 person-years (WHO 2018b). Another recent meta-analysis reviewed the 17 studies published between 1998 and 2016 that carried out multivariate analyses of clinical risk factors to predict inpatient suicides (Large et al. 2017a). These studies all used either a cohort design or a retrospective case-control design and focused on predictors extracted from medical records, although one research group also obtained data from a retrospective questionnaire sent to treating psychiatrists. A total of 191,944 inpatients were included in these pooled studies, 1718 (0.9%) of whom died by suicide while hospitalized. The mean number of predictors considered in the studies was 78.6 (14–272 range) and the mean number in the final models was 6.1.

The methods used in developing these models likely resulted in over-fitting, as in the majority of cases univariate logistic regression analysis was used to select a subset of predictors for subsequent multivariate logistic analysis and a liberal p value was often used in selecting predictors for multivariate analysis. The multivariate analysis typically used backward stepwise selection to arrive at a parsimonious final model. No cross-validation was used to adjust for over-fitting. Recursive partitioning was used in a few studies to search for interactions, but again with no cross-validation, and the analyses otherwise assumed additivity. The focus of all the studies was on identifying "high-risk" patients by defining a threshold, typically on the individual-level predicted probability scale based on the final model, although in some cases the threshold was based on a count of dichotomously-scored predictors with positive values. We were unable to discover a principled basis for selecting thresholds in any of these studies even after a careful review of the original reports, such as to maximize sensitivity (SN; the proportion of suicides that occurred among patients classified as being above the risk threshold) for a fixed specificity (SP; the proportion of patients not dying by suicide that were classified correctly as being below the risk threshold), to equalize SN and SP, to equalize the number of false positives and the number of false negatives, or to equalize the number of false positives and r times the number of false negatives (where r = the pre-specified relative importance of false positives versus false negatives).

Although the great variety of predictors and thresholds used in these studies makes it impossible to draw firm conclusions about prediction accuracy, the authors of the meta-analysis used a random-effects model to generate a meta-analytic ROC curve across studies. SN was estimated to be about 0.70 when SP was set at 0.80 and about 0.50 when SP was set at 0.90. Given the relatively short duration of most

hospitalizations, positive predictive value (PPV; the incidence of suicide among patients classified as high-risk) was only about 0.004, but this was roughly 10 times as high as the suicide rate among patients classified below the threshold. The authors of the meta-analysis concluded from these results that risk assessment based on multivariate prediction models "is not useful as a basis of clinical decisions." Two observations were made to support this conclusion: first, that the low PPV meant that special interventions for high-risk patients would "subject many patients, who will never suicide, to excessive intrusion or coercion"; and second, that the low SN meant that patients classified as being low-risk account for a substantial proportion of inpatient suicides.

This rejection of standardized suicide risk prediction tools is consistent with the recommendations made in a number of other recent systematic reviews, meta-analyses, and commentaries (Bolton 2015; Bolton et al. 2015; Carter et al. 2017; Chan et al. 2016; Katz et al. 2017; Larkin et al. 2014; Mulder et al. 2016; Owens and Kelley 2017; Quinlivan et al. 2016; Runeson et al. 2017). This might seem to be inconsistent with clinical practice guidelines that call for mental health professionals always to make suicide risk evaluations of psychiatric inpatients and patients presenting with psychiatric crises in emergency departments (Bernert et al. 2014; Silverman et al. 2015). However, these guidelines typically advise against using structured risk prediction tools for this purpose and instead recommend that clinicians "initiate a therapeutic relationship" to make "an integrated and comprehensive psychosocial assessment" of needs and risks (National Institute for Health and Care Excellence [NICE] 2011; O'Connor et al. 2013). The notion here is that the low SN of structured suicide risk tools requires clinicians to consider all inpatients and patients in psychiatric crisis to be at risk of suicide and to focus on treatment needs rather than attempt to distinguish levels of risk.

5.3 Earlier Multivariate Analyses Predicting Suicide Among Other High-Risk Patients

Other empirical studies have been carried out for many years to predict suicide and attempted suicide in two other partly-overlapping high-risk patient populations: psychiatric inpatients after hospital discharge, and patients presenting to emergency departments after nonfatal suicide attempts (whether or not they were subsequently hospitalized). The pooled suicide rate within the first 3 months after psychiatric hospital discharge was estimated in a recent meta-analysis of these studies to be 1132/100,000 person-years, with successively lower cumulative rates in studies that followed patients 3–12 months (654/100,000 person-years), 1–5 years (494/100,000 person-years), 6–10 years (366/100,000 person-years), and more than 10 years (277/100,000 person-years) (Chung et al. 2017), although none of the individual studies that followed patients over long time periods estimated changes in conditional risk over shorter time periods. Another recent meta-analysis that focused on suicide after self-harm (whether or not the patient was hospitalized) estimated a pooled suicide incidence within 1 year of the index self-harm episode

of 1600/100,000, with higher estimates of cumulative incidence in studies that followed patients 2 years (2100/100,000), 5 years (3900/100,000), and 10 years (4200/100,000) (Carroll et al. 2014).

As detailed in several recent systematic reviews and meta-analyses (Bolton et al. 2015; Carter et al. 2017; Chan et al. 2016; Katz et al. 2017; Larkin et al. 2014; Quinlivan et al. 2016; Runeson et al. 2017), these studies were usually based on designs similar to the studies reviewed above on inpatient suicides: that is, either cohort or retrospective case-control designs, with predictors extracted from clinical records, although some studies also used patient self-report scales as predictors. The follow-up periods varied widely (6 months to 5 years). Some studies used survival analysis to study predictors over variable time periods, but no systematic effort was made in these studies to investigate change in relative importance of predictors by length of follow-up. The absence of the latter focus is a weakness because suicide risk is known to be highest shortly after clinical contact and there have been calls for increased focus on prediction during high-risk periods (Glenn and Nock 2014; Olfson et al. 2014). It was rare for risk factor analyses in these or other studies to focus on the relatively short 30-day risk window of most interest to clinicians (Franklin et al. 2017).

Some studies evaluating suicide risk prediction tools in high-risk populations were based on single scales, such as self-report scales of hopelessness (Beck and Steer 1988), depression (Beck et al. 1996), overall psychopathological severity (Lindqvist et al. 2007), suicide intent (Beck et al. 1974), and attitudes toward suicide (Koldsland et al. 2012). Other studies used multivariate prediction equations to develop composite suicide risk tools. The latter studies typically began with a predictor set, often extracted from clinical records and sometimes also including various patient self-report and clinician rating scales, used preliminary univariate analyses to select a reduced subset of significant predictors, and then formed a composite from these predictors. Trial and error cross-tabulations (e.g., Kreitman and Foster 1991) and considerations of content validity (e.g., Patterson et al. 1983) were used to construct most of the earlier tools of this sort. Logistic regression analysis or survival analysis were used to construct most of the more recently-developed empirically-derived suicide prediction tools. The predictors in some of these tools consisted entirely of socio-demographic and clinical data extracted from electronic medical records (e.g., Spittal et al. 2014), but others also included some of the patient-reported scales described above (e.g., Bilen et al. 2013; Randall et al. 2013). A few recently-developed empirically-derived tools were constructed using recursive partitioning (Cooper et al. 2006; Steeg et al. 2012; Steinberg and Phillip 1997). As in the inpatient suicide studies, single high-risk thresholds were typically specified without clear evidence of a principled basis for threshold selection, resulting in a wide range in the proportion of patients classified as being high risk. Even though the tools developed in these studies often significantly predicted subsequent suicide, reviews and meta-analyses consistently concluded, as in the inpatient studies, that operating characteristics (i.e., SN, SP, PPV) were not sufficiently strong to justify using any of these tools as a basis for clinical decision-making.

5.4 Reconsidering the Rationale for Rejecting Standardized Suicide Prediction Tools

As noted above, critics of standardized suicide risk prediction tools emphasize the fact that these tools have relatively low PPV and SN, leading clinicians to draw "false reassurance" when they use these tools in treatment planning, patients incorrectly classified as high-risk to experience needless intrusion or coercion, and patients incorrectly classified as low-risk to be denied the treatment they need. Critics also argue that patients perceive standardized risk prediction tools as superficial and that this perception interferes with establishing the kind of therapeutic alliance needed to carry out a more in-depth clinical risk assessment (Large et al. 2017b; Mulder et al. 2016; Owens and Kelley 2017). Qualitative studies debriefing UK patients who were administered standardized scales are said to be consistent with the latter concern (Hunter et al. 2013; Owens et al. 2016; Palmer et al. 2007; Taylor et al. 2009).

Arguments can be made against each of these criticisms. With regard to low PPV: Even though it is true that patients incorrectly classified as high-risk would experience additional burden by being treated if they were at high risk, a balance needs to be struck between increased intrusion-coercion for, say, 250 patients (1/0.004; the number of false positives for every true positive when PPV = 0.004, as in the Large et al. meta-analysis cited above) incorrectly classified as high-risk and saving one life. It is not at all obvious that a formal cost-benefit analysis would conclude that the cost-benefit ratio is >1.0 in such a case. In addition, recent studies have found that up to one-third of patients who do not die by suicide but are classified as high-risk are also at high risk of other experiences in the same spectrum, such as deaths classified as accidental or undetermined, nonfatal suicide attempts, serious nonfatal injuries classified as accidental, and psychiatric hospitalizations (Kessler et al. 2015; McCarthy et al. 2015). The potential to reduce incidence of these outcomes would increase the cost-effectiveness of interventions.

With regard to low SN: The suicide risk models reviewed above all searched for high-risk thresholds (i.e., thresholds to maximize SN for a given SP). There is no way to know from such analyses if a useful threshold could be specified for low-risk patients (i.e., a threshold to maximize SP for SN close to 1.0). Reanalysis, which would have to use the original data in each study, might find that a substantial proportion of patients could be isolated that had such a vanishingly small suicide risk that they could be spared the burden of further evaluation. Indeed, as elaborated below, we believe that this search for a practical low-risk threshold should be the main focus of a first-stage in a multi-stage ML analysis of suicide risk.

With regard to the claim that patients perceive structured suicide risk assessments as superficial: This claim implies that use of clinical suicide risk evaluations instead of standardized suicide risk prediction tools leads to increased detection of suicidality. However, we are aware of no experimental evaluation of this hypothesis. We do know, though, that one study found that clinicians asked to predict the likelihood that patients they are evaluating for suicide risk in at Emergency Departments (ED)

will make a suicide attempt over the next 6 months were no better than chance in their predictions (Nock et al. 2010). This suggests that detailed clinical evaluations might not be as helpful in this regard as implied by critics of standardized risk assessments. A recent systematic review is broadly consistent with this view in finding that clinical risk evaluations are not strong predictors of subsequent suicidal behaviors (Woodford et al. 2017).

In addition, there is evidence that in some cases a structured suicide risk assessment yields better predictions than a clinical evaluation. In an early study on the use of computerized screening for suicide risk, patients in a crisis intervention clinic were asked to complete a computerized assessment of suicidality and then asked whether they would have preferred to have given this information directly to a doctor or to the computer (Greist et al. 1973). The majority of patients said they preferred to provide the information to the computer. A subsequent study building on this finding used a series of computerized self-report questions to assess hospitalized patients who had been admitted because of suicide attempts and then had a psychiatrist carry out an independent face-to-face evaluation blinded to patient reports on the computerized assessment (Levine et al. 1989). Retrospective comparisons showed that patients who subsequently engaged in suicidal behaviors were more likely to admit sensitive symptoms to the computer than to the psychiatrist. This finding is consistent with a good deal of experimental research showing that the likelihood of reporting embarrassing or stigmatizing thoughts and behaviors increases when respondents are randomized to more confidential modes of reporting (Brown et al. 2013; Gnambs and Kaspar 2015). Based on the above results, a computerized version of the self-report Columbia Suicide Severity Rating Scale (CSSRS; Posner et al. 2011) was developed and administered to 6760 patients with psychiatric disorders and 2077 patients with physical disorders who participated in 33 different prospective clinical research studies (Greist et al. 2014). The vast majority (89.9%) of subsequent suicidal behaviors were predicted accurately by the CSSRS.

These results are important given that detailed clinical suicide risk evaluations are carried out only with slightly more than half of all psychiatric inpatients and ED patients in psychiatric crises even when official policies call for these evaluations to be carried out (Cooper et al. 2013). Furthermore, structured suicide risk assessment tools continue to be widely used even when clinical practice guidelines explicitly suggest that they not be used (Quinlivan et al. 2014). Why? One possibility is that the time-consuming nature of detailed clinical suicide risk evaluations leads them to be used only selectively. Gold-standard clinical evaluations of this sort are very time-consuming, often requiring multiple sessions (Rudd 2014) to assess needs (e.g., mental and physical health problems, life difficulties, reasons for recent self-harm and for possible future self-harm, and needs for diverse interventions) and risks (e.g., the nature of the patient's suicidal thinking and behaviors, predispositions to suicide, previous suicide attempts, hopelessness, impulsivity/self-control, suicide warning signs for imminent risk, and protective factors).

How is the decision made to carry out these detailed evaluations with some patients but not others? We are aware of no discussion of this question in the literature. One possibility worth considering is that standardized suicide prediction

tools might be useful in helping clinicians make this decision. Not enough research has been focused on this possibility to know how helpful existing tools could be in this respect, but, as noted below, the small amount of existing evidence suggests that this might be a fruitful direction for future research. The goal would be to define a *low-risk* (not high-risk) threshold for patients who would not be subjected to a more in-depth clinical risk evaluation because of the low proportion of actual suicides that occurs among such patients. If a ML-based decision support tool based on a structured assessment battery could be developed of this sort, one that yielded a meaningful SP for a SN near 1.0, it would almost certainly improve substantially on whatever current decision rules clinicians are using in deciding which patients to evaluate and which not.

It is clear from the results of recent prospective studies that any such assessment battery would have to go beyond patient self-reports of suicidality. These studies have shown that a substantial proportion of the patients who went on to die by suicide shortly after making healthcare visits denied being suicidal during those visits when asked explicitly about suicidality (Louzon et al. 2016; Simon et al. 2013). However, a number of novel structured self-report suicide risk assessment tools developed recently have been shown to have higher predictive validity than previously-developed tools and to be predictive among patients who deny being suicidal. These new tools include: performance-based neurocognitive tests of suicide-related implicit cognitions (Nock et al. 2010); self-reports of suicide-related beliefs (Bryan et al. 2014) and volitional factors such as fearlessness of death, impulsivity, and exposure to past suicidal behaviors (Dhingra et al. 2015); and tools based on linguistic and acoustic features extracted from tape-recorded responses to open-ended questions that do not ask about suicidality (Pestian et al. 2017). It is also worthwhile remembering that previously-developed structured suicide prediction tools measure many of the same dimensions that guidelines call for including in detailed clinical suicide risk evaluations and that these structured tools have been shown to be significant predictors of subsequent suicidal behaviors even though they are not sufficiently strong predictors when considered one at a time to guide clinical decision-making (Bolton et al. 2015; Carter et al. 2017). It is plausible to think that a comprehensive computerized battery that includes all these measures along with the detailed EMR data used in the recent ML prediction models reviewed below would be able to define a low-risk segment of the patient population that had a sufficiently low predicted risk of suicide not to receive a subsequent in-depth clinical evaluation.

Although we are aware of no attempt to develop a comprehensive structured predictor battery of this sort, encouraging results have been found in studies that administered a small number of structured suicide risk tools and found that prediction accuracy is improved significantly by combining them rather than considering them one at a time (Randall et al. 2013; Stefansson et al. 2015). It would not be difficult to expand this line of investigation with existing data. For example, Quinlivan et al. (2017) administered seven commonly-used structured suicide risk assessment tools to a sample of patients who were referred to liaison psychiatry following suicide attempts and followed those patients for 6 months to evaluate

the predictive validity of each tool for repeat suicide attempts or suicide deaths. Four of the eight tools had statistically significant odds-ratios (ORs = 3.9–8.7). Yet the researchers nonetheless concluded that "risk scales on their own have little role in the management of suicidal behavior" (Reutfors et al. 2010). This conclusion was drawn even though no attempt was made to combine the significant scales into a multivariate composite that might have had better prediction accuracy than the individual scales considered one at a time. This negative conclusion is also curious in that the same researchers noted that defining a low-risk threshold might be useful by stating that "(t)he use of risk scales is dependent on clinical context. For example, clinicians may prefer scales with high sensitivity for screening or ruling out a risk of a condition, or scales high in specificity for later stages of assessment or ruling in patients for treatment." Yet the thresholds used in their analysis were for the most part high-risk thresholds, making it impossible to draw any conclusions about the value of the tools reviewed in defining a low-risk patient subgroup.

5.5 Machine Learning Analyses Predicting Suicide Among High-Risk Patients

A number of recent studies have extended the approaches taken in the high-risk multivariate predictor studies reviewed above by using ML methods instead of logistic regression. Results show that ML methods have a great deal of promise in predicting suicide even though all the studies carried out so far have limitations that we review later in the chapter. These studies focused on suicides among psychiatric inpatients in the 12 months after hospital discharge (Kessler et al. 2015), suicides among psychiatric outpatients in the 12 months after visits (Kessler et al. 2017b), and suicide attempts in the 12 months after receiving a formal suicide risk assessment among patients in a psychiatric hospital or ED who were deemed to be at sufficiently high risk to receive such an assessment (Tran et al. 2014). The sample sizes ranged from a low of 68 post-hospitalization suicides among 53,760 hospitalized patients (Kessler et al. 2015) to a high of 1562 serious suicide attempts among 7399 patients who received suicide risk assessments (Tran et al. 2014).

All these studies used electronic medical record (EMR) data as predictors, defined a clear retrospective data capture time period for feature aggregation (2–5 years before baseline), allowed for strength of associations to vary by length of retrospective time period and time-since-baseline, used a multi-step process of feature transformation and pruning based on cross-validation in a training sample followed by evaluation in a separate validation sample, and used standard over-sampling or up-weighting of cases (He and Garcia 2009) in the training sample to deal with the problem of extreme class imbalance. Two of the studies used preliminary bootstrap recursive partitioning to search for interactions, and all the studies used some form of penalized logistic regression (either lasso or elastic net) to estimate the final model. All of the studies evaluated model performance by examining SN and PPV at predefined levels of SN and focused on high-risk prediction. One of the studies compared the prediction accuracy of the ML model

with that of a structured suicide risk assessment and found that prediction based on the former was substantially better than prediction based on the latter (Tran et al. 2014).

Several of the studies suggested that their results had clinical implications. One found that more than 50% of the suicides in the year after psychiatric hospitalization among US Army personnel occurred among the 5% of inpatients classified by ML at the time of hospital discharge as being at highest suicide risk (Kessler et al. 2015). Although PPV was only 3.8%, more than one-third of these highest-risk patients experienced at least one other extreme negative outcome, such as death judged to be accidental or unclassifiable, serious nonfatal injury, attempted suicide, or repeat psychiatric hospitalization, leading the authors to suggest that it might be cost-effective to target patients defined by the ML classifier as being highest-risk for the type of intensive post-hospital case management program that is recommended but not mandated by the US Department of Defense (VA Office of Inspector General 2007). Another US Army study found that an ML model was able to isolate a small number of soldiers (about 500 out of an Army of 500,000) that accounted for a very high proportion of all suicides in the five-week high-risk period after index psychiatric outpatient visits (1047.1/100,000 person-years), leading to a recommendation to target these highest-risk outpatients to receive one of the evidence-based psychotherapies that have been developed specifically to treat suicidality (Jobes et al. 2015).

5.6 Machine Learning Analyses Predicting Suicide in Total Patient Populations

Other ML studies have attempted to predict future suicides or suicide attempts among all patients in a healthcare system (Barak-Corren et al. 2017; Ben-Ari and Hammond 2015; Choi et al. 2018; Kessler et al. 2017a; Walsh et al. 2017). Samples in these studies were typically quite large. Barak-Corren et al. (2017), for example, developed a ML model to predict future suicide attempts ($n = 20{,}246$) in a commercial health system based on an analysis of 1.7 M patients followed for up to 15 years (9.0 M person-years). Kessler et al. (2017a) developed a ML model to predict suicide deaths among patients in the US Veterans Affairs health system, the Veterans Health Administration (VHA), in 2009–2011 using a person-month data array that included information at the month before death for all 6360 VHA suicide decedents and a 1% time-matched person-month probability sample of 2,112,008 VHA service users alive at the end of an index control month over those years. This analysis built on an earlier proof-of-concept model (McCarthy et al. 2015).

As with the high-risk studies reviewed in the previous subsection, the total-population studies used structured EMR data as predictors. One also used natural language processing (NLP) methods to define features based on information extracted from clinical notes (Ben-Ari and Hammond 2015). All studies defined a clear retrospective data capture time period for feature aggregation (2–5 years), and most, but not all, cases allowed for strength of associations to vary by length

of retrospective time frame and time-since-baseline. They all defined a clear risk time horizon (between 30 days and 15 years). They all used a multi-step process of feature transformation and pruning based on cross-validation in a training sample followed by testing in a separate validation sample. Most of the studies used over-sampling or up-weighting of cases in the training sample to deal with the problem of extreme class imbalance. Although analyses were consistently based on a single algorithm (artificial neural networks, naïve Bayes, penalized regression, random forests), some studies compared results across different classifiers before selecting a best one defined in terms of mean-squared error (e.g., adaptive splines, Bayesian additive regression trees, generalized boosting, support vector machines). Most, but not all, studies evaluated model performance by examining SN and PPV at predefined levels of SN, and all studies focused on high-risk assessment aimed at targeting preventive interventions rather than on low-risk assessment aimed at limiting the number of patients who would receive more in-depth clinical evaluations.

For the most part, lift (i.e., incidence of the outcome among patients classified as high-risk versus in the total patient population) was relatively high at the upper ends of the prediction scales in these studies, with SN at a fixed SP of 0.95 equal to 0.28 in the VHA suicide study (Kessler et al. 2017a) and in the range 0.28–0.50 (Barak-Corren et al. 2017; Ben-Ari and Hammond 2015) in the studies predicting suicide attempts. PPV, of course, was quite low at these thresholds due to the rarity of the outcomes. Despite the models not focusing on low-risk prediction, the 25% of patients with the lowest predicted risk in a number of these studies (Barak-Corren et al. 2017; Ben-Ari and Hammond 2015) accounted for very low (3–7%) proportions of suicidal outcomes.

5.7 Other Machine Learning Studies Aimed at Predicting Suicidality

Another group of ML studies attempted to predict either current or past patient self-reported suicidality from information obtained in administrative records and/or patient self-report scales (e.g., Barros et al. 2017; Hettige et al. 2017; Ilgen et al. 2009; Jordan et al. 2018; Oh et al. 2017; Passos et al. 2016). The rationale for these efforts was that model predictions might help unobtrusively to detect "unseen" cases of suicidality when applied in other samples. A related series of studies applied ML methods to complex feature sets made up of various biomarkers in order to predict current self-reported suicidality, using such predictors as immune markers (Dickerson et al. 2017) and altered fMRI neural signatures in response to life- and death-related words (Just et al. 2017). Other related studies used text analysis to extract predictive information from clinical notes (McCoy et al. 2016; Poulin et al. 2014) or new technologies, such as smartphones and wearable sensors that might allow passive monitoring of suicidality (Braithwaite et al. 2016; Cook et al. 2016). Samples in all these studies were small because of the high expense of the biomarkers and/or new technologies. The analyses typically used only a single ML

classifier rather than an ensemble, although some studies compared results across different classifiers. Relatively simple feature selection methods were used in most of these applications. Little was said in most of them about the methods used for hyper-parameter tuning or dealing with the problem of class imbalance. Most applications used internal cross-validation but did not divide their small samples into separate training and validation sets. Practical prediction accuracy (i.e., estimates of SN or PPV for fixed high values of SN) was seldom emphasized, although overall prediction strength (AUC) was typically moderate, suggesting that these methods would be most useful if combined with administrative data to create a rich multivariate feature set.

5.8 Future Directions in Using ML for Suicide Risk Prediction

Although the studies reviewed above suggest that ML methods have considerable promise in predicting suicide, the field has as yet not fully realized that promise. A number of changes would likely improve prediction accuracy and clinical value. First, as illustrated in the last section, the feature sets used in the ML analyses of suicide carried out until now could be expanded beyond the structured EMR data that have so far been the mainstay of these analyses. In addition to the methods described in the last section, information on residential zip code could be used to extract small area geocode data from public sources on a number of important predictors of suicide such as local unemployment rates (Nordt et al. 2015) and neighborhood social capital (Holtkamp and Weaver 2018). Data from commercial search engines could be used to obtain more detailed socio-demographic information than the information on age, sex, and marital status typically available in EMRs and to extract information from public records on individual-level legal, financial, and criminal justice experiences that predict suicide (e.g., Accurint 2018).

Second, prediction accuracy could be improved by using ensemble ML methods combining individual-level predictions across algorithms. The Super Learner ensemble method, for example, has been shown to yield considerably higher levels of prediction accuracy than the best-performing algorithm in the ensemble (Polley et al. 2016). Automated machine learning (AutoML; Feurer et al. 2015; Olson et al. 2017) is also making it increasingly possible to refine feature transformation-pruning, algorithm selection, and hyperparameter tuning (Urbanowicz et al. 2017). AutoML can also be used to address the extreme imbalance problem by automatically implementing toolkits to evaluate the relative effectiveness of different imbalance correction methods (e.g., Chawla 2010).

Third, greater consideration is needed of the clinical value of different outcome time horizons in light of the fact that several studies have shown that optimal model features and coefficients differ depending on time horizon. In the ideal case, the time horizon would be chosen in light of the intervention the model is being designed to guide. This does not always occur. For example, the ML analysis described earlier predicting suicide among users of the VHA system was designed to facilitate

VHA implementation of their Recovery Engagement And Coordination for Health-Veterans Enhanced Treatment (REACH VET) program (VA Office of Public and Intergovernmental Affairs 2017) among highest-risk VHA users. However, the ML model had a 30-day time horizon even though it often takes more than 30 days to make initial contact with the targeted Veteran and the program continues for at least 90 days. This raises the question whether the REACH VET ML model should have had a longer (e.g., 180-day) time horizon and, if so, the extent to which different Veterans would have been selected for intervention if this had been done.

Fourth, ML modeling efforts need to be better coordinated with the clinical interventions they are designed to support in ways other than time horizon. Most notably, ML model development up to now has focused on high-risk prediction even though a good argument could be made that models based on the feature sets considered up to now are likely to be more useful in low-risk prediction. If that is the case, then, as suggested earlier in the chapter, a first-stage ML model based on structured predictors could be used to help select which patients should receive more intensive clinical suicide risk evaluations.

Fifth, more work needs to be done to determine the extent to which high-risk predictions based on ML models could be improved by adding information from subsequently-administered structured and/or clinical risk evaluations. Tran et al. (2014) had an opportunity to do something along these lines by virtue of the fact that their sample consisted exclusively of patients who had been the subjects of in-depth clinical suicide risk assessments, but the authors focused instead on the extent to which predictions based on ML outperformed predictions based on clinical evaluations rather than seeing how much overall prediction improved by combining the two sets of predictors.

5.9 Machine Learning Models for Clinical Decision Support in Treatment Planning

We noted above that critics of structured suicide risk prediction tools argue that all psychiatric inpatients and ED patients should be considered at risk of suicide and should receive in-depth clinical evaluations rather than structured suicide risk assessments. But this raises the question how the information about needs should be applied to formulate a treatment plan. A number of special types of psychotherapy exist for patients at high suicide risk (e.g., Jobes et al. 2017; Linehan et al. 2015; Rudd et al. 2015) that have been shown to improve on usual care in reducing suicidal behavior (Jobes et al. 2015). However, these interventions are more labor-intensive than usual care and require special clinical training, making it important to have some principled basis for knowing which patients need these interventions. The same could be said for the decision to offer combined pharmacotherapy and psychotherapy (versus only one), which is known to be of value for some but not all patients (Kessler 2018), and the use of ketamine as a pharmacologic treatment for patients at imminent suicide risk (Wilkinson and Sanacora 2016). How do clinicians make decisions about what suicidal patients need after carrying out in-depth suicide

needs assessments? Critics of structured suicide risk prediction tools are silent on this question.

ML has the potential to provide clinical decision support in making these decisions, but in doing so it needs to be recognized that the patients at highest suicide risk are not necessarily the patients most likely to be helped by available interventions. This means that different ML modeling strategies need to be used to predict differential treatment response than to predict differential risk. Speaking in general terms, the models for differential treatment response can be thought of as evaluating interactions between *prescriptive* predictors of treatment response (i.e., predictors of greater response to some types of treatment than others) and treatment type, ideally evaluated in controlled treatment effectiveness trials that have real-world validity (Cohen and DeRubeis 2018). 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 (VanderWeele et al. 2018). ML methods can be applied even when treatment is not randomly assigned by using double-robust estimation methods (Vermeulen and Vansteelandt 2015), so long as either strong predictors exist of nonrandom treatment assignment or if, as in the case of suicide, loss to follow-up outcome assessment is low (Luedtke and van der Laan 2016).

To illustrate the potential value of this approach, consider the VHA's REACH VET initiative. This initiative was implemented in 2016 based on the results of an ML model that used 2008–2011 data. A separate prescriptive ML model to evaluate differential response to the REACH VET intervention could be estimated by predicting suicide deaths among high-risk VHA patients in the 12 months after selection by the initial ML intervention targeting model in 2014 (2 years before the intervention was initiated, which means that none of these high-risk patients received the intervention) and in 2016 (the year the intervention was initiated, when all the high-risk patients were "randomized" to the intervention). An expanded set of features that included not only structured EMR data, but also NLP data extracted from clinical notes, geocode data linked to zip codes, and individual-level public records data extracted from commercial sources, could be used as predictors in the analysis. Difference-in-difference before-after comparison analysis could be used by combining patients above the intervention threshold with an equal or greater number of patients just slightly below the threshold in order to adjust for possible time trends. To the extent that prescriptive ML analysis shows that some high-risk VHA patients do not profit from the current REACH VET intervention, more intensive interventions could be targeted to patients with this profile in future implementations. It might even be possible to use a group-randomized (by treatment center) design (Treweek and Zwarenstein 2009) to assign the high-risk VHA patients predicted not to be helped by the current REACH VET intervention to different high-intensity evidence-based interventions designed specifically to treat suicidal patients, such as Dialectical Behavior Therapy, Cognitive Therapy for Suicide Prevention, or Collaborative Assessment and Management of Suicidality. This design would allow a more refined prescriptive ML analysis subsequently to be

carried out to create a clinical decision support tool that helped clinicians implement precision treatment planning for high-risk VHA patients.

5.10 Conclusions

Improvements are needed in both the big data and the ML methods used to analyze these data if the full potential of ML is to be realized in addressing the suicide problem. It is likely that the prediction accuracy of the ML models reviewed here could be improved, perhaps substantially so, at low cost by more nuanced EMR feature transformation and by expanding the features to include information extracted from clinical notes using NLP and, in the US, from public data sources using zip code links (small area geocode data) and from commercially aggregated individual-level public records. Even better prediction is likely in health plans that routinely screen patients with self-reports of various sort (e.g., periodic completion of a self-report depression scale; Louzon et al. 2016; Simon et al. 2013). The ML analysis methods used in existing suicide prediction studies could also be improved substantially by using recently-developed ensemble and AutoML methods that optimize feature transformation-pruning, hyperparameter tuning, and adjustments for extreme imbalance in the outcome. Further work is needed to determine sample sizes at which such ML approaches are effective, especially for outcomes as rare as suicide.

We have no way of knowing how much suicide prediction accuracy would be improved by implementing all these feature expansions and ML analysis improvements, but it is almost certain that prediction accuracy would be insufficient to allow treatment planning to be based on such a model. Rather than use this fact, as critics have, to reject structured suicide risk assessment out of hand, it makes much more sense to see this phase of ML analysis as a useful first step in a multi-step process of need and risk evaluation. It is not inconceivable that SP in such an improved total-population first-stage ML model would be very close to 1.0 below a threshold that included a substantial proportion of patients. If so, it might be practical to ask all patients above that low-risk threshold to complete a structured self-report suicide risk assessment that included the full range of scales and performance-based neurocognitive tests that have been found to predict suicidal behavior in previous studies. A second-stage ML analysis in that subsample could then be carried out that used the predictors from the prior total-population analysis and the self-report measures obtained in the structured risk assessment to target the subset of patients who would receive an in-depth clinical suicide risk evaluation. The information in the self-report battery could be used as a starting point for this evaluation in the service of developing a treatment plan. A third-stage ML clinical decision support model based on input from all three predictor sets (i.e., the EMR data and other passive data available in the total-population, the structured patient self-report data available in the subsamples defined by the first ML model, and the clinical data collected in the smaller subgroup targeted by the second ML model) could then be developed to provide clinical decision support for this

treatment planning. Part of the treatment process might then involve the use of new technologies supported by additional ML analyses, such as pharmacogenomics screening to select optimal medications (El-Mallakh et al. 2016) and use of new technologies to monitor ongoing treatment response as well as imminent suicide risk (Vahabzadeh et al. 2016). This kind of nested use of successively more refined ML models in which structured data are combined with clinical evaluations is likely to hold the key to maximizing the value of big data ML analysis in improving detection and treatment of suicidal patients.

References

Accurint (2018) http://www.accurint.com. Accessed 20 Feb 2018

Ahmedani BK, Simon GE, Stewart C, Beck A, Waitzfelder BE, Rossom R et al (2014) Health care contacts in the year before suicide death. J Gen Intern Med 29(6):870–877. https://doi.org/10.1007/s11606-014-2767-3

Barak-Corren Y, Castro VM, Javitt S, Hoffnagle AG, Dai Y, Perlis RH et al (2017) Predicting suicidal behavior from longitudinal electronic health records. Am J Psychiatry 174(2):154–162. https://doi.org/10.1176/appi.ajp.2016.16010077

Barros J, Morales S, Echávarri O, García A, Ortega J, Asahi T et al (2017) Suicide detection in Chile: proposing a predictive model for suicide risk in a clinical sample of patients with mood disorders. Rev Bras Psiquiatr 39:1–11

Beck AT, Steer RA (1988) BHS, Beck Hopelessness Scale: manual. Psychological Corporation, San Antonio

Beck AT, Schuyler D, Herman I (1974) Development of suicidal intent scales. In: Beck AT, Lettieri DJ, HLP R, National Institute of Mental Health, Center for Studies of Suicide Prevention, University of Pennsylvania Department of Psychiatry (eds) The prediction of suicide. Charles Press, Bowie, pp 45–56

Beck AT, Steer RA, Brown G (1996) Manual for the Beck Depression Inventory-II. Psychological Corporation, San Antonio

Ben-Ari A, Hammond K (2015) Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 Persian gulf war. In: Paper presented at the 2015 48th Hawaii international conference on system sciences, 5–8 Jan 2015, pp 3168–3175. https://doi.org/10.1109/hicss.2015.382

Berman AL, Silverman MM (2014) Suicide risk assessment and risk formulation part II: suicide risk formulation and the determination of levels of risk. Suicide Life Threat Behav 44(4):432–443. https://doi.org/10.1111/sltb.12067

Bernert RA, Hom MA, Roberts LW (2014) 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 38(5):585–592. https://doi.org/10.1007/s40596-014-0180-1

Bilen K, Ponzer S, Ottosson C, Castren M, Pettersson H (2013) Deliberate self-harm patients in the emergency department: who will repeat and who will not? Validation and development of clinical decision rules. Emerg Med J 30(8):650–656. https://doi.org/10.1136/emermed-2012-201235

Bolton JM (2015) Suicide risk assessment in the emergency department: out of the darkness. Depress Anxiety 32(2):73–75. https://doi.org/10.1002/da.22320

Bolton JM, Gunnell D, Turecki G (2015) Suicide risk assessment and intervention in people with mental illness. BMJ 351:h4978

Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL (2016) Validating machine learning algorithms for twitter data against established measures of suicidality. JMIR Ment Health 3(2):e21. https://doi.org/10.2196/mental.4822

Brown JL, Swartzendruber A, DiClemente RJ (2013) Application of audio computer-assisted self-interviews to collect self-reported health data: an overview. Caries Res 47(Suppl 1):40–45. https://doi.org/10.1159/000351827

Bryan CJ, Rudd DM, Wertenberger E, Etienne N, Ray-Sannerud BN, Morrow CE et al (2014) 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 159:15–22. https://doi.org/10.1016/j.jad.2014.02.021

Carroll R, Metcalfe C, Gunnell D (2014) Hospital presenting self-harm and risk of fatal and non-fatal repetition: systematic review and meta-analysis. PLoS One 9(2):e89944. https://doi.org/10.1371/journal.pone.0089944

Carter G, Milner A, McGill K, Pirkis J, Kapur N, Spittal MJ (2017) Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. Br J Psychiatry 210(6):387–395. https://doi.org/10.1192/bjp.bp.116.182717

Chawla N (2010) Data mining for imbalanced datasets: an overview. In: Maimon O, Rokach L (eds) Data mining and knowledge discovery handbook, 2nd edn. Springer, Berlin, pp 875–886

Chan MK, Bhatti H, Meader N, Stockton S, Evans J, O'Connor RC et al (2016) Predicting suicide following self-harm: systematic review of risk factors and risk scales. Br J Psychiatry 209(4):277–283. https://doi.org/10.1192/bjp.bp.115.170050

Choi SB, Lee W, Yoon JH, Won JU, Kim DW (2018) Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea. J Affect Disord 231:8–14. https://doi.org/10.1016/j.jad.2018.01.019

Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, Large MM (2017) Suicide rates after discharge from psychiatric facilities: a systematic review and meta-analysis. JAMA Psychiat 74(7):694–702. https://doi.org/10.1001/jamapsychiatry.2017.1044

Cohen ZD, DeRubeis RJ (2018) Treatment selection in depression. Annu Rev Clin Psychol. https://doi.org/10.1146/annurev-clinpsy-050817-084746

Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E (2016) Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Comput Math Methods Med 2016:8708434. https://doi.org/10.1155/2016/8708434

Cooper J, Kapur N, Dunning J, Guthrie E, Appleby L, Mackway-Jones K (2006) A clinical tool for assessing risk after self-harm. Ann Emerg Med 48(4):459–466. https://doi.org/10.1016/j.annemergmed.2006.07.944

Cooper J, Steeg S, Bennewith O, Lowe M, Gunnell D, House A et al (2013) Are hospital services for self-harm getting better? An observational study examining management, service provision and temporal trends in England. BMJ Open 3(11):e003444. https://doi.org/10.1136/bmjopen-2013-003444

Dhingra K, Boduszek D, O'Connor RC (2015) Differentiating suicide attempters from suicide ideators using the integrated motivational-volitional model of suicidal behaviour. J Affect Disord 186:211–218. https://doi.org/10.1016/j.jad.2015.07.007

Dickerson F, Adamos M, Katsafanas E, Khushalani S, Origoni A, Savage C et al (2017) The association between immune markers and recent suicide attempts in patients with serious mental illness: a pilot study. Psychiatry Res 255:8–12. https://doi.org/10.1016/j.psychres.2017.05.005

El-Mallakh RS, Roberts RJ, El-Mallakh PL, Findlay LJ, Reynolds KK (2016) Pharmacogenomics in psychiatric practice. Clin Lab Med 36(3):507–523. https://doi.org/10.1016/j.cll.2016.05.001

Feurer M, Klein A, Eggensperger K, Springenberg JT, Blum M, Hutter F (2015) Efficient and robust automated machine learning. Paper presented at the proceedings of the 28th International Conference on Neural Information Processing Systems - volume 2, Montreal, Canada, 2015

Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X et al (2017) Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 143(2):187–232. https://doi.org/10.1037/bul0000084

Glenn CR, Nock MK (2014) Improving the short-term prediction of suicidal behavior. Am J Prev Med 47(3 Suppl 2):S176–S180. https://doi.org/10.1016/j.amepre.2014.06.004

Gnambs T, Kaspar K (2015) Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behav Res Methods 47(4):1237–1259. https://doi.org/10.3758/s13428-014-0533-4

Greist JH, Gustafson DH, Stauss FF, Rowse GL, Laughren TP, Chiles JA (1973) A computer interview for suicide-risk prediction. Am J Psychiatry 130(12):1327–1332. https://doi.org/10.1176/ajp.130.12.1327

Greist JH, Mundt JC, Gwaltney CJ, Jefferson JW, Posner K (2014) Predictive value of baseline electronic Columbia-Suicide Severity Rating Scale (eC-SSRS) assessments for identifying risk of prospective reports of suicidal behavior during research participation. Innov Clin Neurosci 11(9–10):23–31

He H, Garcia EA (2009) Learning from imbalanced data. IEEE Trans Knowl Data Eng 21(9):1263–1284. https://doi.org/10.1109/tkde.2008.239

Hettige NC, Nguyen TB, Yuan C, Rajakulendran T, Baddour J, Bhagwat N et al (2017) Classification of suicide attempters in schizophrenia using sociocultural and clinical features: a machine learning approach. Gen Hosp Psychiatry 47:20–28. https://doi.org/10.1016/j.genhosppsych.2017.03.001

Holtkamp CR, Weaver RC (2018) Quantifying the relationship between social capital and economic conditions in Appalachia. Appl Geogr 90:175–186. https://doi.org/10.1016/j.apgeog.2017.12.010

Hunter C, Chantler K, Kapur N, Cooper J (2013) Service user perspectives on psychosocial assessment following self-harm and its impact on further help-seeking: a qualitative study. J Affect Disord 145(3):315–323. https://doi.org/10.1016/j.jad.2012.08.009

Ilgen MA, Downing K, Zivin K, Hoggatt KJ, Kim HM, Ganoczy D et al (2009) Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. J Clin Psychiatry 70(11):1495–1500. https://doi.org/10.4088/JCP.08m04795

Jobes DA, Au JS, Siegelman A (2015) Psychological approaches to suicide treatment and prevention. Curr Treat Options Psychiatry 2(4):363–370. https://doi.org/10.1007/s40501-015-0064-3

Jobes DA, Comtois KA, Gutierrez PM, Brenner LA, Huh D, Chalker SA et al (2017) A randomized controlled trial of the collaborative assessment and management of suicidality versus enhanced care as usual with suicidal soldiers. Psychiatry 80(4):339–356. https://doi.org/10.1080/00332747.2017.1354607

Joiner TE Jr, Buchman-Schmitt JM, Chu C (2017) Do undiagnosed suicide decedents have symptoms of a mental disorder? J Clin Psychol 73(12):1744–1752. https://doi.org/10.1002/jclp.22498

Jordan P, Shedden-Mora MC, Lowe B (2018) Predicting suicidal ideation in primary care: an approach to identify easily assessable key variables. Gen Hosp Psychiatry 51:106–111. https://doi.org/10.1016/j.genhosppsych.2018.02.002

Just MA, Pan L, Cherkassky VL, McMakin D, Cha C, Nock MK et al (2017) Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. Nat Hum Behav 1:911–919. https://doi.org/10.1038/s41562-017-0234-y

Katz C, Bolton J, Sareen J (2016) The prevalence rates of suicide are likely underestimated worldwide: why it matters. Soc Psychiatry Psychiatr Epidemiol 51(1):125–127. https://doi.org/10.1007/s00127-015-1158-3

Katz C, Randall JR, Sareen J, Chateau D, Walld R, Leslie WD et al (2017) Predicting suicide with the SAD PERSONS scale. Depress Anxiety 34(9):809–816. https://doi.org/10.1002/da.22632

Kessler RC (2018) The potential of predictive analytics to provide clinical decision support in depression treatment planning. Curr Opin Psychiatry 31(1):32–39. https://doi.org/10.1097/yco.0000000000000377

Kessler RC, Warner CH, Ivany C, Petukhova MV, Rose S, Bromet EJ et al (2015) Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study to Assess Risk

and Resilience in Servicemembers (Army STARRS). JAMA Psychiat 72(1):49–57. https://doi.org/10.1001/jamapsychiatry.2014.1754

Kessler RC, Hwang I, Hoffmire CA, McCarthy JF, Petukhova MV, Rosellini AJ et al (2017a) Developing a practical suicide risk prediction model for targeting high-risk patients in the veterans health administration. Int J Methods Psychiatr Res 26(3). https://doi.org/10.1002/mpr.1575

Kessler RC, Stein MB, Petukhova MV, Bliese P, Bossarte RM, Bromet EJ et al (2017b) Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). Mol Psychiatry 22(4):544–551. https://doi.org/10.1038/mp.2016.110

Koldsland BO, Mehlum L, Mellesdal LS, Walby FA, Diep LM (2012) The suicide assessment scale: psychometric properties of a Norwegian language version. BMC Res Notes 5:417. https://doi.org/10.1186/1756-0500-5-417

Kreitman N, Foster J (1991) The construction and selection of predictive scales, with special reference to parasuicide. Br J Psychiatry 159:185–192

Large M, Myles N, Myles H, Corderoy A, Weiser M, Davidson M et al (2017a) Suicide risk assessment among psychiatric inpatients: a systematic review and meta-analysis of high-risk categories. Psychol Med 48(7):1119–1127. https://doi.org/10.1017/s0033291717002537

Large MM, Ryan CJ, Carter G, Kapur N (2017b) Can we usefully stratify patients according to suicide risk? BMJ 359:j4627

Larkin C, Di Blasi Z, Arensman E (2014) Risk factors for repetition of self-harm: a systematic review of prospective hospital-based studies. PLoS One 9(1):e84282. https://doi.org/10.1371/journal.pone.0084282

Levine S, Ancill RJ, Roberts AP (1989) Assessment of suicide risk by computer-delivered self-rating questionnaire: preliminary findings. Acta Psychiatr Scand 80(3):216–220

Lindqvist D, Nimeus A, Traskman-Bendz L (2007) Suicidal intent and psychiatric symptoms among inpatient suicide attempters. Nord J Psychiatry 61(1):27–32. https://doi.org/10.1080/08039480601122064

Linehan MM, Korslund KE, Harned MS, Gallop RJ, Lungu A, Neacsiu AD et al (2015) Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: a randomized clinical trial and component analysis. JAMA Psychiat 72(5):475–482. https://doi.org/10.1001/jamapsychiatry.2014.3039

Louzon SA, Bossarte R, McCarthy JF, Katz IR (2016) Does suicidal ideation as measured by the PHQ-9 predict suicide among VA patients? Psychiatr Serv 67(5):517–522. https://doi.org/10.1176/appi.ps.201500149

Luedtke AR, van der Laan MJ (2016) Optimal individualized treatments in resource-limited settings. Int J Biostat 12(1):283–303. https://doi.org/10.1515/ijb-2015-0007

Luoma JB, Martin CE, Pearson JL (2002) Contact with mental health and primary care providers before suicide: a review of the evidence. Am J Psychiatry 159(6):909–916. https://doi.org/10.1176/appi.ajp.159.6.909

Madsen T, Erlangsen A, Nordentoft M (2017) Risk estimates and risk factors related to psychiatric inpatient suicide-an overview. Int J Environ Res Public Health 14(3). https://doi.org/10.3390/ijerph14030253

McCarthy JF, Bossarte R, Katz IR, Thompson C, Kemp J, Hannemann C et al (2015) Predictive modeling and concentration of the risk of suicide: implications for preventive interventions in the US Department of Veterans Affairs. Am J Pub Health 105(9):1935–1942. https://doi.org/10.2105/AJPH.2015.302737

McCoy TH Jr, Castro VM, Roberson AM, Snapper LA, Perlis RH (2016) Improving prediction of suicide and accidental death after discharge from general hospitals with natural language processing. JAMA Psychiatry 73(10):1064–1071. https://doi.org/10.1001/jamapsychiatry.2016.2172

Mulder R, Newton-Howes G, Coid JW (2016) The futility of risk prediction in psychiatry. Br J Psychiatry 209(4):271–272. https://doi.org/10.1192/bjp.bp.116.184960

Murphy GE (1972) Clinical identification of suicidal risk. Arch Gen Psychiatry 27:356–359

National Institute for Health and Care Excellence (NICE) (2011) Self-harm in over 8s: long-term management. //www.nice.org.uk/guidance/cg133. Accessed 5 Jan 2018

Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR (2010) Measuring the suicidal mind: implicit cognition predicts suicidal behavior. Psychol Sci 21(4):511–517. https://doi.org/10.1177/0956797610364762

Nordt C, Warnke I, Seifritz E, Kawohl W (2015) Modelling suicide and unemployment: a longitudinal analysis covering 63 countries, 2000-11. Lancet Psychiatry 2(3):239–245. https://doi.org/10.1016/s2215-0366(14)00118-7

O'Connor E, Gaynes BN, Burda BU, Soh C, Whitlock EP (2013) Screening for and treatment of suicide risk relevant to primary care: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 158(10):741–754. https://doi.org/10.7326/0003-4819-158-10-201305210-00642

Oh J, Yun K, Hwang JH, Chae JH (2017) Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. Front Psych 8:192. https://doi.org/10.3389/fpsyt.2017.00192

Olfson M, Marcus SC, Bridge JA (2014) Focusing suicide prevention on periods of high risk. JAMA 311(11):1107–1108. https://doi.org/10.1001/jama.2014.501

Olson RS, Sipper M, La Cava W, Tartarone S, Vitale S, Fu W et al. (2017) A system for accessible artificial intelligence. arXiv.org. arXiv:1705.00594v2

Owens D, Kelley R (2017) Predictive properties of risk assessment instruments following self-harm. Br J Psychiatry 210(6):384–386. https://doi.org/10.1192/bjp.bp.116.196253

Owens C, Hansford L, Sharkey S, Ford T (2016) 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 208(3):286–291. https://doi.org/10.1192/bjp.bp.113.141242

Palmer L, Blackwell H, Strevens P (2007) Service users' experience of emergency services following self harm: a national survey of 509 patients. College Centre for Quality Improvement, Royal College of Psychiatrists. https://www.rcpsych.ac.uk/pdf/National%20SU%20Survey%20Final%20Self%20Harm%20Project.pdf. Accessed 20 Feb 2018

Passos IC, Mwangi B, Cao B, Hamilton JE, Wu MJ, Zhang XY et al (2016) Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach. J Affect Disord 193:109–116. https://doi.org/10.1016/j.jad.2015.12.066

Patterson WM, Dohn HH, Bird J, Patterson GA (1983) Evaluation of suicidal patients: the SAD PERSONS scale. Psychosomatics 24(4):343–345, 348–349. https://doi.org/10.1016/s0033-3182(83)73213-5

Pearson A, Saini P, Da Cruz D, Miles C, While D, Swinson N et al (2009) Primary care contact prior to suicide in individuals with mental illness. Br J Gen Pract 59(568):825–832. https://doi.org/10.3399/bjgp09X472881

Pestian JP, Sorter M, Connolly B, Cohen KB, McCullumsmith C, Gee JT et al (2017) A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. Suicide Life Threat Behav 47(1):112–121. https://doi.org/10.1111/sltb.12312

Polley E, LeDell E, van der Laan M (2016) SuperLearner: Super learner prediction [computer program]. R package version 2.0–21: The Comprehensive R Archive Network

Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA et al (2011) The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 168(12):1266–1277. https://doi.org/10.1176/appi.ajp.2011.10111704

Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B et al (2014) Predicting the risk of suicide by analyzing the text of clinical notes. PLoS One 9(1):e85733. https://doi.org/10.1371/journal.pone.0085733

Quinlivan L, Cooper J, Steeg S, Davies L, Hawton K, Gunnell D et al (2014) Scales for predicting risk following self-harm: an observational study in 32 hospitals in England. BMJ Open 4(5):e004732. https://doi.org/10.1136/bmjopen-2013-004732

Quinlivan L, Cooper J, Davies L, Hawton K, Gunnell D, Kapur N (2016) Which are the most useful scales for predicting repeat self-harm? A systematic review evaluating risk scales using measures of diagnostic accuracy. BMJ Open 6(2):e009297. https://doi.org/10.1136/bmjopen-2015-009297

Quinlivan L, Cooper J, Meehan D, Longson D, Potokar J, Hulme T et al (2017) Predictive accuracy of risk scales following self-harm: multicentre, prospective cohort study. Br J Psychiatry 210(6):429–436. https://doi.org/10.1192/bjp.bp.116.189993

Randall JR, Rowe BH, Dong KA, Nock MK, Colman I (2013) Assessment of self-harm risk using implicit thoughts. Psychol Assess 25(3):714–721. https://doi.org/10.1037/a0032391

Reutfors J, Brandt L, Ekbom A, Isacsson G, Sparen P, Osby U (2010) Suicide and hospitalization for mental disorders in Sweden: a population-based case-control study. J Psychiatr Res 44(12):741–747. https://doi.org/10.1016/j.jpsychires.2010.02.003

Rosen A (1954) Detection of suicidal patients: an example of some limitations in the prediction of infrequent events. J Consult Psychol 18(6):397–403

Rudd MD (2014) Core competencies, warning signs, and a framework for suicide risk assessment in clinical practice. In: Nock MK (ed) The Oxford handbook of suicide and self-injury, 1st edn. Oxford University Press, Cary, pp 323–336. https://doi.org/10.1093/oxfordhb/9780195388565.013.0018

Rudd MD, Bryan CJ, Wertenberger EG, Peterson AL, Young-McCaughan S, Mintz J et al (2015) Brief cognitive-behavioral therapy effects on post-treatment suicide attempts in a military sample: results of a randomized clinical trial with 2-year follow-up. Am J Psychiatry 172(5):441–449. https://doi.org/10.1176/appi.ajp.2014.14070843

Runeson B, Odeberg J, Pettersson A, Edbom T, Jildevik Adamsson I, Waern M (2017) Instruments for the assessment of suicide risk: a systematic review evaluating the certainty of the evidence. PLoS One 12(7):e0180292. https://doi.org/10.1371/journal.pone.0180292

Schaffer A, Sinyor M, Kurdyak P, Vigod S, Sareen J, Reis C et al (2016) Population-based analysis of health care contacts among suicide decedents: identifying opportunities for more targeted suicide prevention strategies. World Psychiatry 15(2):135–145. https://doi.org/10.1002/wps.20321

Silverman JJ, Galanter M, Jackson-Triche M, Jacobs DG, Lomax JW, Riba MB et al (2015) The American Psychiatric Association practice guidelines for the psychiatric evaluation of adults. Am J Psychiatry 172(8):798–802. https://doi.org/10.1176/appi.ajp.2015.1720501

Simon GE, Rutter CM, Peterson D, Oliver M, Whiteside U, Operskalski B et al (2013) Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? Psychiatr Serv 64(12):1195–1202. https://doi.org/10.1176/appi.ps.201200587

Spittal MJ, Pirkis J, Miller M, Carter G, Studdert DM (2014) The Repeated Episodes of Self-Harm (RESH) score: a tool for predicting risk of future episodes of self-harm by hospital patients. J Affect Disord 161:36–42. https://doi.org/10.1016/j.jad.2014.02.032

Steeg S, Kapur N, Webb R, Applegate E, Stewart SL, Hawton K et al (2012) The development of a population-level clinical screening tool for self-harm repetition and suicide: the ReACT self-harm rule. Psychol Med 42(11):2383–2394. https://doi.org/10.1017/s0033291712000347

Stefansson J, Nordstrom P, Runeson B, Asberg M, Jokinen J (2015) Combining the Suicide Intent Scale and the Karolinska Interactive Violence Scale in suicide risk assessments. BMC Psychiatry 15:226. https://doi.org/10.1186/s12888-015-0607-6

Steinberg D, Phillip C (1997) CART – classification and regresstion trees. Salford Systems, San Diego

Taylor TL, Hawton K, Fortune S, Kapur N (2009) Attitudes towards clinical services among people who self-harm: systematic review. Br J Psychiatry 194(2):104–110. https://doi.org/10.1192/bjp.bp.107.046425

Tran T, Luo W, Phung D, Harvey R, Berk M, Kennedy RL et al (2014) Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. BMC Psychiatry 14:76. https://doi.org/10.1186/1471-244x-14-76

Treweek S, Zwarenstein M (2009) Making trials matter: pragmatic and explanatory trials and the problem of applicability. Trials 10:37. https://doi.org/10.1186/1745-6215-10-37

Urbanowicz RJ, Meeker M, Lacava W, Olson RS, Moore JH (2017) Relief based feature selection: introduction and review. arXiv.org. arXiv:1711.08421

VA Office of Inspector General (2007) Health care inspection: impementing VHA's mental health strategic plan initiatives for suicide prevention. https://www.va.gov/oig/54/reports/VAOIG-06-03706-126.pdf

VA Office of Public and Intergovernmental Affairs (2017) VA REACH VET initiative helps save veterans lives: program signals when more help is needed for at-risk veterans. U.S. Department of Veterans Affairs. https://www.va.gov/opa/pressrel/pressrelease.cfm?id=2878. Accessed 12 May 2017

Vahabzadeh A, Sahin N, Kalali A (2016) Digital suicide prevention: can technology become a game-changer? Innov Clin Neurosci 13(5–6):16–20

VanderWeele TJ, Leudtke AR, van der Laan MJ, Kessler RC (2018) Selecting optimal subgroups for treatment using many covariates. arXiv.org. arXiv:1802.09642

Vermeulen K, Vansteelandt S (2015) Bias-reduced doubly robust estimation. J Am Stat Assoc 110(511):1024–1036. https://doi.org/10.1080/01621459.2014.958155

Walsh G, Sara G, Ryan CJ, Large M (2015) Meta-analysis of suicide rates among psychiatric in-patients. Acta Psychiatr Scand 131(3):174–184. https://doi.org/10.1111/acps.12383

Walsh CG, Ribeiro JD, Franklin JC (2017) Predicting risk of suicide attempts over time through machine learning. Clin Psychol Sci 5(3):457–469. https://doi.org/10.1177/2167702617691560

Wilkinson ST, Sanacora G (2016) Ketamine: a potential rapid-acting antisuicidal agent? Depress Anxiety 33(8):711–717. https://doi.org/10.1002/da.22498

Woodford R, Spittal MJ, Milner A, McGill K, Kapur N, Pirkis J et al (2017) Accuracy of clinician predictions of future self-harm: a systematic review and meta-analysis of predictive studies. Suicide Life Threat Behav. https://doi.org/10.1111/sltb.12395

World Health Organization (WHO) (2018a) Mental health: suicide data. http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/. Accessed 20 Feb 2018

World Health Organization (WHO) (2018b) Age-standardized suicide rates (per 100 000 population), 2015. Global Health Observatory (GHO) data. http://www.who.int/gho/mental_health/suicide_rates/en/. Accessed 03 Mar 2018