Using Administrative Data to Predict Suicide After Psychiatric Hospitalization in the Veterans Health Administration System

Ronald C. Kessler[1]*, Mark S. Bauer[2,3], Todd M. Bishop[4], Olga V. Demler[5,6], Steven K. Dobscha[7,8], Sarah M. Gildea[1], Joseph L. Goulet[9,10], Elizabeth Karras[4], Julie Kreyenbuhl[11,12], Sara J. Landes[13,14], Howard Liu[1,4], Alex R. Luedtke[15,16], Patrick Mair[17], William H. B. McAuliffe[1], Matthew Nock[17], Maria Petukhova[1], Wilfred R. Pigeon[4,18], Nancy A. Sampson[1], Jordan W. Smoller[19], Lauren M. Weinstock[20] and Robert M. Bossarte[4,21]

[1] Deparment of Health Care Policy, Harvard Medical School, Boston, MA, United States, [2] Department of Psychiatry, Harvard Medical School, Boston, MA, United States, [3] Center for Healthcare Organization & Implementation Research, VA Boston Healthcare System, Boston, MA, United States, [4] Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center, Canandaigua, NY, United States, [5] Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, United States, [6] Department of Medicine, Harvard Medical School, Boston, MA, United States, [7] VA Center to Improve Veteran Involvement in Care, VA Portland Health Care System, Portland, OR, United States, [8] Department of Psychiatry, Oregon Health & Science University, Portland, OR, United States, [9] Pain, Research, Informatics, Multimorbidities & Education Center, VA Connecticut Healthcare System, West Haven, CT, United States, [10] Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, United States, [11] VA Capitol Healthcare Network (VISN 5), Mental Illness Research, Education, and Clinical Center (MIRECC), Baltimore, MD, United States, [12] Department of Psychiatry, Division of Psychiatric Services Research, University of Maryland School of Medicine, Baltimore, MD, United States, [13] South Central Mental Illness Research Education Clinical Center (MIRECC), Central Arkansas Veterans Healthcare System, North Little Rock, AR, United States, [14] Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, AR, United States, [15] Department of Statistics, University of Washington, Seattle, WA, United States, [16] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, United States, [17] Department of Psychology, Harvard University, Cambridge, MA, United States, [18] Department of Psychiatry, University of Rochester Medical Center, Rochester, NY, United States, [19] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States, [20] Department of Psychiatry & Human Behavior, Alpert Medical School of Brown University, Providence, RI, United States, [21] West Virginia University Injury Control Research Center and Department of Behavioral Medicine and Psychiatry, West Virginia University School of Medicine, Morgantown, WV, United States

There is a very high suicide rate in the year after psychiatric hospital discharge. Intensive postdischarge case management programs can address this problem but are not cost-effective for all patients. This issue can be addressed by developing a risk model to predict which inpatients might need such a program. We developed such a model for the 391,018 short-term psychiatric hospital admissions of US veterans in Veterans Health Administration (VHA) hospitals 2010–2013. Records were linked with the National Death Index to determine suicide within 12 months of hospital discharge (n=771). The Super Learner ensemble machine learning method was used to predict these suicides for time horizon between 1 week and 12 months after discharge in a 70% training sample. Accuracy was validated in the remaining 30% holdout sample. Predictors included VHA administrative variables and small area geocode data linked to patient home addresses. The models had AUC=.79–.82 for time horizons between 1 week and 6 months and

AUC=.74 for 12 months. An analysis of operating characteristics showed that 22.4%–32.2% of patients who died by suicide would have been reached if intensive case management was provided to the 5% of patients with highest predicted suicide risk. Positive predictive value (PPV) at this higher threshold ranged from 1.2% over 12 months to 3.8% per case manager year over 1 week. Focusing on the low end of the risk spectrum, the 40% of patients classified as having lowest risk account for 0%–9.7% of suicides across time horizons. Variable importance analysis shows that 51.1% of model performance is due to psychopathological risk factors accounted, 26.2% to social determinants of health, 14.8% to prior history of suicidal behaviors, and 6.6% to physical disorders. The paper closes with a discussion of next steps in refining the model and prospects for developing a parallel precision treatment model.

INTRODUCTION

Suicide is the 10th leading cause of death in the US (1). The suicide rate has increased steadily since 1999 (2) and especially among veterans (3). Transitions in care, especially psychiatric hospital discharge, are periods of particularly high suicide risk in the general population, including veterans (4, 5). Indeed, the approximately 1% of Veteran Health Administration (VHA) patients who are hospitalized for psychiatric disorders each year account for nearly 12% of all VHA suicides over the subsequent 12 months (6). Programs to reduce suicides after psychiatric hospital discharge are urgently needed. Beginning in 2008, VHA implemented a series of suicide prevention recommendations that addressed this need by requiring each VHA treatment facility to appoint a suicide prevention coordinator (7) and, more recently, to require inpatient clinicians to develop a suicide safety plan with each inpatient before discharge (8). These changes were associated with a stabilization of the previously rising postdischarge suicide rate, although the rate is still very high compared to others (6).

Although several VHA outpatient treatment programs exist, none is designed specifically for patients at high suicide risk after psychiatric hospital discharge. Intensive postdischarge case management programs, which are not used in VHA, have been shown elsewhere to be effective in reducing suicides after psychiatric hospital discharge (9–16), leading to recommendations to add such programs to existing postdischarge suicide preventive interventions (17). However, these programs can be labor-intensive, requiring frequent outpatient contacts, assertive outreach for missed outpatient appointments, and intensive community support to engage reluctant patients. It would be difficult to justify implementing such a program for all VHA patients given the rarity of postdischarge suicides (about 3/1,000 hospitalizations) (6) and the scarcity of the specially trained staff needed to implement this type of intervention in each of the nearly 100 VHA psychiatric inpatient units and over 1,000 outpatient clinics to which inpatients are discharged around the country (9).

Such a program would be more scalable, though, if it focused on recently discharged patients at high suicide risk and was implemented remotely by centralized program staff to increase efficiency. We are in the process of piloting a promising program of this sort known as the Coping Long Term with Active Suicide Program, a telephone-based adjunctive intensive case management program that has been shown to have significant aggregate effects on suicide-related behaviors (SRBs) after psychiatric hospital discharge (9, 18, 19). A first step in implementing this kind of targeted intervention would be to develop a predictive analytic model that pinpoints the inpatients with high risk of subsequent suicide. A model of this sort was developed for hospitalized US Army soldiers as part of the Army STARRS research program (20). More than 50% of all suicides in the 12 months after hospital discharge in that study occurred among the 10% of soldiers defined as being at highest risk. However, a much more extensive set of predictor variables was available to build this model in the integrated Department of Defense administrative data system than exists to build a VHA model, making it unclear if a comparable model could be developed in VHA. We present the results of a preliminary effort to investigate this question in the current report.

MATERIALS AND METHODS

Sample

We focus on the 391,018 short-term psychiatric hospital admissions of veterans in any VHA hospital in the US or its territories (i.e., American Samoa, Guam, Northern Mariana Islands, Puerto Rico, and US Virgin Islands) during calendar years 2010–2013. Records of these visits were linked with the National Death Index (21) to determine which patients died within 12 months of hospital discharge. The analysis sample was drawn from a master sample we developed that was a variation of the case-control approach used in our previous research (20). The cases in this master sample were 100% of the VHA patients who died in calendar years 2010–2014 and were last seen in the VHA system within 2 years (24 months) of their death and were classified in the National Death Index as dying either by suicide, by any opioid-related cause, or by drug overdose. This case definition excludes deaths by other external cause (i.e., murder or other accidents), some of which might have been misclassified suicides.

The controls were a stratified probability sample of other patients ever seen in the VHA system in calendar years 2008–2013 (i.e., within 2 years of any month in the time period 2010–2013). The stratification scheme for controls was hierarchical and included (i) patients that made a suicide attempt recorded in the VHA administrative records; (ii) other patients (i.e., exclusive of those that made a suicide attempt) that had a psychiatric hospitalization; (iii) other patients that had any outpatient psychiatric treatment; and (iv) all other patients. Sampling fractions within strata were set to generate a sample of controls either four times the number of cases in the stratum (in stratum i), three times the number of cases (in strata ii-iii), or two times the number of cases (in stratum iv). Sampling of controls was carried out using secondary stratification for discharge date and a variety of socio-demographic and clinical factors. This differential sampling was designed to increase power in the segments of the population where suicide rates are highest under a financial constraint on total number of controls because we purchased some of the predictor data from a commercial data aggregation firm.

For the analysis of suicides after psychiatric hospital discharge, the unweighted case-control ratio was about 1:11.5 at the person level. However, as we included all hospitalizations for all patients in the case-control sample, the case-control ratio at the level of the hospitalization was somewhat different (1:17). The person-level data were weighted by the inverse of their probabilities of selection for purposes of analysis and population projection. The 70% of case hospitalizations with the earliest discharge dates were combined with all control hospitalizations up to the same discharge date to create a training sample in which the prediction model was developed. The remaining 30% of cases and controls were held out to validate the model. The study protocol was approved by Research Ethics Committee of the Veterans Administration Center of Excellence for Suicide Prevention and Harvard Medical School with a waiver of informed consent based on the data being deidentified.

Predictors
Overview
We turned to prior studies of data from electronic health records to determine the predictor set. Troister et al. (22) carried out a comprehensive review of published studies of risk factors for civilian postdischarge suicides as of 2006 and found five replicated classes of predictors: (i) history of prior suicidal behaviors; (ii) psychopathological disorders (the most consistent being nonaffective psychosis, mood disorders, and multiple comorbid psychiatric disorders), medications for these disorders, and interactions between specific psychopathological disorders and medications known to be especially useful in protecting against suicide among patients with these disorders [e.g., lithium among patients with bipolar disorder; (23)]; (iii) quality of care after hospital discharge (e.g., low continuity of care); (iv) time since hospital discharge (inversely related to suicide risk); and (v) socio-demographics (the most consistent being male gender and recent job loss), which more recently have been conceptualized as indicators of social determinants of

health (24, 25). Studies published after the Troister et al. review found similar predictors (20, 26–28). We included indicators of all these predictor classes in our analysis.

We also included two additional predictor sets that could be considered indicators of social determinants of health: (vi) International Classification of Diseases, Ninth Revision, Clinical Modification [**ICD-9-CM**; (29)] E and V codes for social factors known to be associated with suicide, such as sexual assault victimization (30) and financial loss (31); and (vii) small-area geocode data (e.g., neighborhood deprivation, local unemployment rate). And we looked at potentially informative interactions between patient socio-demographics and neighborhood characteristics (e.g., Black patients living in predominantly White neighborhoods). Finally, we included indicators of 3 other predictor classes found in more general studies of suicides among outpatients and health plan members: (viii) physical disorders and medications for the treatment of these disorders (32); (ix) medications that are thought to be associated with increased suicide risk (33); and (x) medical procedures associated either with increased suicide risk [e.g., amputations; (34)] or decreased suicide risk [e.g., certain types of psychotherapy; (35)].

Data Sources
Three VHA data systems were used to operationalize most of the predictors:

i. *The VHA Corporate Data Warehouse* [CDW; (36)]: An integrated system containing data on patient socio-demographics along with information on all health care encounters either in VHA or paid for by VHA in the community, classified in terms of primary and secondary ICD-9-CM diagnostic and procedure codes. The CDW also contains information on prescriptions written in VHA or otherwise paid for by VHA, classified using the VHA Drug Classification System (37). The CDW also includes a comprehensive list of test results along with E codes for external causes of injury due to accidents, suicide attempts, and other types of self-inflicted injuries and V codes for other factors influencing health status and contact with the health care system that contain information about social determinants of health (38);

ii. *The Veterans Administration Suicide Prevention Applications Network* (39): This is an administrative data system for suicide behavior tracking in VHA;

iii. *The Veterans Administration Homes Registry*: The Homes Registry is a data system maintained by the National Center on Homelessness Among Veterans (40) that includes information on all veterans known either to be homeless, at risk of homelessness, or in a VHA homelessness program. For the current analysis, though, homelessness was determined by ICD-9-CM codes, Patient Treatment File (PTF) Inpatient Codes, and outpatient stop codes.

We augmented the information obtained in the three VHA data systems with small area geocode data available from various

government databases at the levels of the Census Block Group, Census Tract, or County to characterize neighborhood socio-demographic profiles and social factors either known or suspected to be associated with suicide risk. Information on patient home address from the CDW was used to link patient records to these geocode data systems.

The small amount of missing values found in this data often were nonmissing in earlier records, allowing nearest neighbor imputations. Remaining missing values and inconsistencies were reconciled using rational imputations (e.g., a patient classified as female in one record but male in both earlier and later records was recoded male). Details about missing data patterns are available in **Supplementary Tables 1** and **2**.

Predictor Classes

i. *History of prior suicidal behaviors:* ICD-9-CM E and V codes (see **Supplementary Table 3**) and the Veterans Administration Suicide Prevention Applications Network system provided information on history of suicidal ideation and attempts reported in inpatient, outpatient, or emergency department visits, including as a basis for the current hospitalization.

ii. *Psychopathological risk factors:* We created 3 variables for each of 7 retrospective time periods (past 30, 90, 180, 365, 730, 1,095 days and the veterans' entire VHA history as of January 1, 2000) for each of the 582 diagnoses or diagnostic groupings of mental disorders in the ICD-9-CM and each of the 41 mental disorder diagnoses in the Clinical Classification Software (41): yes/no for any visit with this diagnosis during the time period; a continuous count of number of days with such visits; and a stabilized 0–4 quintile transformation of the latter count. We also created a series of composite variables for common types of comorbidity among these disorders, including comorbidities thought to predict suicide that involve a combination of mental and physical disorders [e.g., (42–44)]. ICD-9-CM codes and details about each disorder are presented in **Supplementary Table 4**.

We also included information about medications used to treat the above disorders obtained from the VHA National Formulary. The latter is a three-level classification system that includes a total of 574 categories (32 major drug classes, 287 minor drug classes, 255 subclasses) to characterize the 29,290 individual pharmaceutical products available through VHA. We created count variables for prescriptions filled for the Central Nervous System class (of the 32 major drug classes) as well as for each of the minor and subclasses of this class in the 90 days and 365 days prior to the focal hospital admission. In addition, we created interaction terms to define the conjunction of two broad mental disorder diagnosis groups, schizophrenic psychoses and affective psychoses, with medications found to be associated with reduced suicide among patients with these disorders: clozapine, olanzapine, and quetiapine along with long-acting injectable antipsychotics for schizophrenic psychoses (45–47); and lithium for affective psychoses (23).

And we included a count of the number of medications used to help offset the extrapyramidal side effects of antipsychotics that can contribute to suicidality (48) in the 90 days and 365 days prior to the focal hospital admission. See **Supplementary Table 4** for the complete list of medications.

iii. *Quality of care and aftercare for psychiatric inpatients:* Recent research in the UK has found that quality of care indicators, such as extent of staff turnover and short average duration of stay, are significant predictors of postdischarge suicides (5, 14). Only superficial indicators of this sort (e.g., driving time between the patient's home and the nearest VHA treatment center) were included in the initial model-building exercise reported here, as the more comprehensive facility-level indicators we are developing were not ready at the time of analysis.

iv. *Time since hospital discharge:* As noted above in the section on analysis methods, we developed separate models for five risk time horizons: 1 week and 1, 3, 6, and 12 months since hospital discharge (22, 49).

v. *Socio-demographics:* The CDW provided information on patient age, sex, race/ethnicity, marital status, income, religion, residential characteristics (Census Division, urbanicity, homelessness), and period of service (pre-Vietnam, Vietnam era, post-Vietnam, Persian Gulf War) (**Supplementary Table 5**).

vi. *ICD-9-CM E and V codes:* The latter codes were used to count information on 14 different types of accidents, physical and sexual assaults, and perpetration of child or adult abuse (**Supplementary Table 6**). We also coded medical encounters due to housing or economic circumstances and due to other family or psychosocial circumstances (38).

vii. *Small-area geocode data:* Annual rolling 5-year average data at the levels of the Census Block Group or Census Tract were obtained from the American Community Survey (50) on a wide range of small area characteristics that previous research has shown to cluster into two dimensions associated with variation in suicide rates: neighborhood deprivation (22 indicators; e.g., low median education, high unemployment and poverty rates, percent of households receiving public assistance) and neighborhood fragmentation (5 indicators: proportions of households with single-person occupant, vacant, occupant unmarried, occupant residing in the housing unit less than 12 months, occupant owns the housing unit) (51–54). American Community Survey data were also obtained on neighborhood race/ethnicity. Based on evidence that individual-neighborhood differences sometimes predict suicides (55, 56), we created interactions of patient race/ethnicity with the percent of neighborhood residents who were of the same race/ethnicity (non-Hispanic Black, non-Hispanic White, Hispanic, Other), a ratio of patient income to median neighborhood income, and an indicator for the percent of neighborhood residents who were veterans. Based on evidence that food insufficiency is more

important than low income in predicting suicidality (57), data were obtained from the Department of Agriculture on the percent of the neighborhood that was a food desert (58). We also obtained information on the County-level suicide rate averaged over the past 2 most recently available years (59). Finally, based on evidence that economic trends are associated with trends in suicide rates (60), we obtained County-level data on the bankruptcy rate, the median debt-to-income ratio, and the unemployment rate averaged over the past 3–14 months. We also calculated changes in these statistics over the past three months or, for bankruptcy, past 2 years compared to the 3 years before that (**Supplementary Table 7**).

viii. *Physical disorders:* The Clinical Classification Software system was used to organize information about the roughly 13,000 ICD-9-CM diagnoses into a 646-variable 4-level hierarchical system. We created the same three variables at the seven retrospective time periods for each of these 646 variables as we did for the psychopathological risk factors, resulting in 13,566 variables about individual physical disorders. In addition, we created a series of composite measures for types of comorbidity reported in the literature (**Supplementary Table 8**) as potentially important predictors of suicide (61–64).

We also included information about medications used to treat physical disorders from the VHA National Formulary. As noted above, the latter is a three-level classification system that includes a total of 32 major drug classes to characterize the 29,290 individual pharmaceutical products available through VHA. In addition to the count variables noted above for all three levels of Central Nervous System drugs, we created count variables for prescriptions filled for each of the other 31 major nonpsychotropic drug classes in the 90 days and 365 days prior to the focal hospital admission.

ix. *Medications thought to cause suicide:* Literature suggests that some medications for physical disorders might predispose to suicide (65, 66). In order to investigate this possibility, we searched Food and Drug Administration approved drug labeling documents in the Food and Drug Administration Label Database (67) using the search terms *suicidality, suicidal behavior, suicidal ideation, suicide attempt, suicidal,* and *suicide* and found 49 medications that indicated suicide as an adverse reaction in the box warning section of the drug label, 112 that included suicide in the warnings and precautions section, and 79 that included suicide in the adverse reactions section. We created separate count variables for each of these three levels of possible risks to describe prescriptions in the 90 days and 365 days prior to the focal hospital admission (**Supplementary Table 9**).

x. *Medical procedures:* The CDW uses ICD-9-CM procedure codes to record inpatient procedures and the American Medical Association Current Procedural Terminology codes to recode outpatient procedures. We included measures for a mix of ranges and specific procedures within each system for each of the same seven retrospective time periods used to code diagnoses (**Supplementary Table 10**).

Analysis Methods

Numerous recent studies predicted suicide death or attempted suicide in high-risk patient populations from electronic health record data using machine learning (ML) methods. These studies focused either on psychiatric inpatients after hospital discharge (20), psychiatric outpatients after recent visits (68), or emergency department patients deemed to need a suicide risk assessment (69, 70). Results showed that ML methods have considerable promise even though all these studies were limited in a number of ways discussed elsewhere (71). We attempted to build on these prior studies by introducing five improvements:

i. Rather than choose only one or compare across a small number of alternative ML classifiers, we used the Super Learner (**SL**) ensemble ML method to combine predicted probabilities of suicide at the level of hospitalization across a large number of different ML algorithms (the "ensemble") (72). This is an important improvement over previous studies because no single ML algorithm is universally optimal. SL has a guarantee to be at least as accurate and typically has a considerably higher level of prediction accuracy than the best-performing algorithm in the ensemble. Following recent recommendations (73), we used a wide range of algorithms in the ensemble to optimize performance (**Table 1**). These included a generalized linear model with a logistic link function, a series of penalized regressions with different mixing model parameters, a series of support vector machines with different kernels, Bayesian adaptive regression trees, neural networks, random forest, and a series of gradient boosted decision trees that differed in depth and shrinkage.

ii. We used three different feature selection methods: univariate p value less than .10; and, within this set, Least Absolute Shrinkage and Selection Operator regression and random forest (see **Table 1** for descriptions) to reduce the number of potential predictors included in SL. This kind of initial feature pruning can improve out-of-sample model performance substantially (80).

iii. A number of the algorithms in our SL library require hyper-parameter tuning for optimal performance. We addressed this problem in simple cases by including a series of models for a single algorithm with different hyper-parameter values in the SL ensemble (e.g., five penalized regression classifiers that differed in values of the mixing parameter, several different support vector machines that differed in kernels). In more complex cases we used the random search method in the Classification and Regression Training package to select optimal hyper-parameter values separately for each time horizon (81).

iv. Suicide is a very rare outcome even among recently discharged psychiatric inpatients, with a case-control ratio of about 3:1,000. This kind of extreme class imbalance can pose problems for estimation because most algorithms aim to optimize overall classification accuracy and fail to adjust for the fact that false negatives may be more costly than false positives, leading the algorithms to focus on correctly

TABLE 1 | Overview of the algorithms used in the Super Learner ensemble.

Algorithm	R package	Description
Logistic regression	stats	• Traditional parametric logistic regression • Prone to overfit if independent variables are highly collinear • Optimal functional form of independent variables unknown (e.g., linear versus nonlinear)
Elastic net regularization (74)	glmnet	• Penalized regression reduces overfit due to collinear independent variables • Ridge regression shrinks coefficients for collinear independent variables toward zero, but does not fully-eliminate any independent variable • Elastic net regression allows various penalties where coefficients for collinear independent variables are shrunk toward zero (but not eliminating contributions to the predicted probability) and/or to zero (eliminating their contributions to the predicted probability) • Mixing parameter penalty (alpha) is set somewhere between .01 and .99 • Least Absolute Shrinkage and Selection Operator (LASSO) regression shrinks coefficients for collinear covariate coefficients to zero, eliminating their contributions to the predicted probability
Random forest decision trees (75)	ranger	• Decision tree methods capture interactions and non-linear associations • Independent variables are partitioned (based on values) and stacked to build decision trees and ensemble an aggregate "forest" • Random forest builds numerous trees in bootstrapped samples and generates an aggregate tree by averaging across trees (reducing overfit) • Suitable for large data sets, but may be unstable and overfitting
Bayesian additive regression trees (76)	bartMachine	• Bayesian trees are based on an underlying probability model (priors) for the structure and likelihood for data in terminal nodes • The aggregate tree is generated by averaging across tree posteriors (reducing overfit)
Extreme gradient boosting (77)	xgboost	• Extreme gradient boosting decision tree algorithm • Final predictions are formulated by models sequentially built (using gradient descent algorithm to minimize loss) to resolve residual error made by existing models
Support vector machines (78) Linear kernel Polynomial kernel Radial kernel	ksvm	• Support vector machines treats independent variables as dimensions in high dimensional space and attempts to identify the best hyperplane to separate the sample into classes (e.g., cases and noncases) • Goal is to find the hyperplane with the maximum margin between the two closest points in space • Captures linear associations, but alternate kernels can be used to capture nonlinearities (polynomial and radial basis kernels were used here)
Neural networks (79)	nnet	• Connections between predictors and the outcome are modeled as a network • Predictors affect the outcome through intermediate layers • Weights are assigned to connections • Capture interactions and non linear associations • Low interpretability

classifying the much more common noncases at the expense of misclassifying the rare cases (82). A number of strategies involving under-sampling controls, pseudo-sampling cases, and combinations have been developed to address this problem (83) and shown to improve model performance [e.g., (84, 85)]. We did this by making five copies of each case record and subsampling an equal number of control records using stratified probability sampling to create a balanced dataset for estimation. Once the model was estimated and predicted probabilities of suicide were assigned to each record, we reweighted each record in the balanced dataset by the inverse of its probability of selection to recover true unit-level predicted probabilities of suicide. We then used these weight-corrected estimates to evaluate model fit in the training sample. Five-fold cross-validation was used for internal SL cross-validation both to build optimal models with each classifier and to determine optimal weighting across classifiers in the ensemble. All five replicates of a stratified 20% of case records were included in a single five-fold cross-validation fold in order to address the problem that overfitting can occur when cases are duplicated.

v. Given that the time horizon for intervention can vary substantially depending on whether the concern is with imminent risk (e.g., suicide shortly after hospital discharge) or subsequent readjustment back into the community (e.g., suicide within 12 months of hospital discharge), we built separate models for each of 5 risk time horizons: 1 week and 1, 3, 6, and 12 months after hospital discharge.

Standard evaluations of model performance were used in the test sample. We began by generating the receiver operating characteristic (ROC) curves and then calculating area under the receiver operating characteristic curve (AUC) for the SL model developed to predict suicides over each time horizon. Each of the five SL models was used to predict suicides over each of the five thresholds in the holdout sample to determine if prediction accuracy for shorter thresholds than 12 months would be improved by developing models for each threshold rather than only developing a model to predict suicides over all 12 months.

We then calculated operating characteristics, including sensitivity (the proportion of suicide cases that were above a given prediction threshold), specificity (the proportion of suicide noncases that were below the same prediction threshold) and positive predictive value (PPV; the probability of suicide above the decision threshold within the time horizon) for a variety of thresholds. The latter included the 5%, 10%, 20%, and 60% of

observations with highest predicted probabilities of suicide based on the model as well as the thresholds needed to achieve fixed values of sensitivity and specificity. We then calculated a modification of PPV designed to adjust for differences across time horizons by computing the number of suicides per 100 patient-years rather than per 100 patients. Adjusted PPV is of interest because it allows us to estimate the expected number of patients who would otherwise die by suicide *over the time horizon* a clinician would work with under alternative scenarios in which the clinician either treated a larger number of patients over a shorter period of time or treated a smaller number of patients over a longer period of time. This is a useful distinction because conditional suicide risk is much higher in the first weeks and months after hospital discharge. This is difficult to see when focusing on a conventional PPV measure, as the latter increases as the time horizon increases. However, adjusted PPV decreases as the time horizon increases when conditional suicide risk decreases in this same way.

We then investigated which predictor variables were of greatest overall importance by using the Extreme Gradient Boosting ensemble decision tree algorithm to predict SL predicted probabilities of suicide for each time horizon (77). The importance of the splitting variable at each node of each tree was determined by examining the extent to which prediction performance at the node changed when the splitting variable was replaced by random noise. This importance measure for each split was then weighted by the proportion of the sample involved in the split and these predictor-specific weighted importance measures were summed across all nodes of all trees to arrive at a summary measure of "gain" in model prediction accuracy due to each predictor variable (86). The sum of gain across predictors was normed to 1.0. We then grouped predictors by the 10 broad categories of predictors described above.
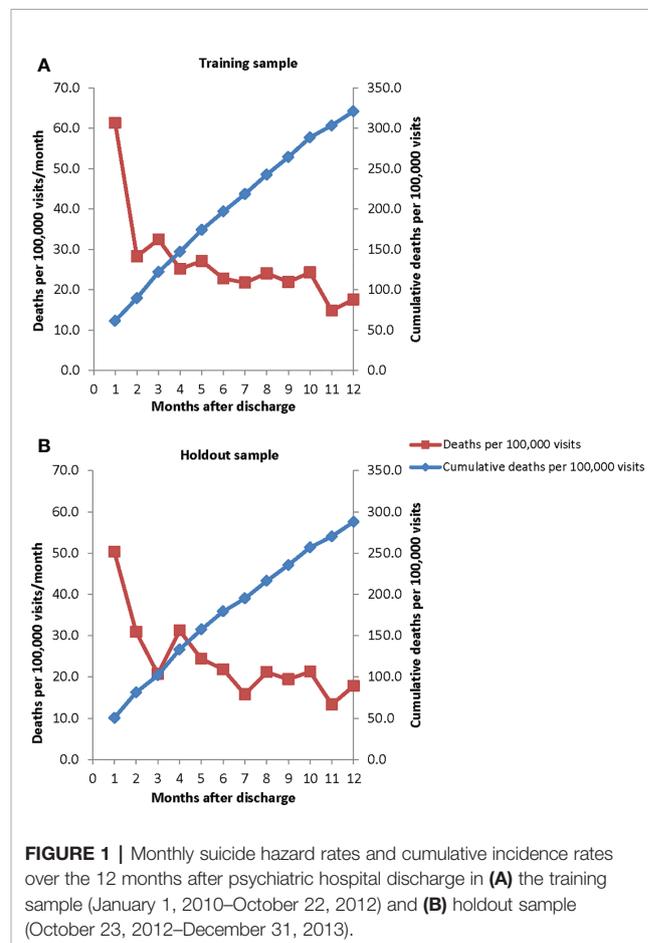
RESULTS

Outcome Distribution

There were 771 suicides among the 195,349 veterans who were hospitalized by VHA for a psychiatric problem during calendar years 2010–2013 and who died in the 365 days after discharge. These 771 veterans had 1,195 psychiatric hospitalizations in 2010–2013 out of the 391,018 such hospitalizations during this time period. As noted in the sample section, we used hospitalization as the unit of analysis, which means that we had 1,195 "cases" (i.e., hospitalizations followed within 12 months by a suicide) and 389,823 controls (i.e., other hospitalizations). The 70% of case hospitalizations with the earliest months of discharge (n=864; January 1, 2010–October 22, 2012) were combined with all control hospitalizations up through the same discharge month to create a training sample in which the prediction model was developed. The remaining 30% of case hospitalizations (n=331) and the associated controls were held out to validate the model.

The observed suicide rate at the level of hospitalization over the 12 months after hospital discharge was 315.5 per 100,000

person-years in the training sample (**Figure 1A**) and 282.5 in the holdout sample (**Figure 1B**). In both samples, the suicide rate varied significantly and inversely with time since discharge (χ^2_{11} = 115.3–51.3, p < .001), with the highest suicide rate in the first week after discharge (1,104.3–1,290.7 per 100,000 person-years), a much lower rate in the remainder of the first month after discharge (626.7–399.9), an even lower rate in the 2–3 months after discharge (363.4–309.3), and a more gradually decreasing rate over subsequent months (from 299.0–308.9 4–6 months after discharge to 246.9–215.8 7–12 months after discharge).

Stratification Variable Distributions

We noted above that the sample was stratified to match the population of all psychiatric hospital admissions over the study period on the cross-classification of diverse socio-demographic and geographic variables as well as on a measure of whether the patient had a suicide risk flag on their medical record in the month prior to their hospitalization (**Supplementary Table 11**). A discussion of VHA suicide risk flags is presented elsewhere (87). The great majority of patients in both the training sample and the holdout sample were male (92.6%–93.4%) and had a median age of 54. Consistent with this age distribution, the plurality served most recently either in the Vietnam era (37.1%–39.4%), followed by the Persian Gulf War era (29.9%–36.3%) and

FIGURE 1 | Monthly suicide hazard rates and cumulative incidence rates over the 12 months after psychiatric hospital discharge in **(A)** the training sample (January 1, 2010–October 22, 2012) and **(B)** holdout sample (October 23, 2012–December 31, 2013).

the years between the Vietnam and Persian Gulf War eras (22.9%–25.1%). The majority were non-Hispanic White (60.1%–61.3%) and others mostly non-Hispanic Black (24.3%–24.9%). The plurality was either divorced (38.4%–38.8%) and the others mostly either married (24.3%–26.0%) or never married (23.0%–23.3%). The majority lived in Census-defined metropolitan statistical areas of more than 1 million residents (51.4%–53.5%) or 250,000–1 million residents (23.1%–23.8%) and only a small proportion (8.0%–9.8%) lived in areas with populations of less than 20,000. The vast majority reported being Christian (72.8%–73.3%) and most of these either Baptists (24.1%–28.2%) or Roman Catholics (18.9%–19.8%), whereas most others (19.9%–21.1%) reported having no religion. A strikingly high 35.6%–41.1% had been homeless at some time in the 12 months before hospitalization, including 16.1%–19.0% homeless at time of admission. Finally, 15.1% in the training sample and 8.3% in the holdout sample had a high risk of suicide flag on their medical records in the month prior to the time of their hospitalization. In comparing these distributions to those of all patients making VHA visits for any reason over the same time periods weighted by number of visits, the stratification variables most strongly associated with psychiatric hospitalization were ages 20–55 (55.4%–57.7% versus 26.7%–27.6%), never married (23.0%–23.3% versus 12.4%–12.6%), separated-divorced (47.3%–61.7% versus 31.5%–31.8%), post-Vietnam era (22.9%–25.1% versus 12.0%–13.4%), Persian Gulf War era (29.9%–36.3% versus 18.0%–20.5%), currently or recently homeless (35.6%–41.1% versus 8.7%–10.3%), and having a high risk flag (8.3%–15.1% versus 0.5%–0.8%).

Stratification Variable Models

We estimated initial multivariate logistic regression models that used all the stratification variables to predict suicides over each of the five risk time horizons. Logistic regression coefficients were exponentiated to create odds-ratios (OR). As described below, these models were subsequently used as controls to screen each other potential predictor one at a time. The stratification variable models were globally significant for each time horizon ($\chi^2_{35} = 59.8$–535.6, p < .001). Only three variables were significant in the 1-week model: race/ethnicity, with significant ORs of 4.2 for non-Hispanic Whites and 5.3 for "other" race/ethnicity (not non-Hispanic White, non-Hispanic Black, and other Hispanic) compared to non-Hispanic Blacks; religion, with significantly reduced ORs of 0.00–0.17 for Black and other Baptists and 0.12 for "other" Christians (not Roman Catholic) compared to other Protestants; and having a high risk suicide flag on the medical record prior to the time of hospitalization (OR=2.5) (**Supplementary Table 12**). Most of these predictors remained significant in models for longer time horizons (OR=2.4–3.0 for non-Hispanic White; OR=3.4–4.5 for "other" race/ethnicity; OR=0.26–0.5 for Baptists, OR=1.8–2.6 high risk flag), a pattern also found for most of the predictors that became significant only in models with longer time horizons. But the OR for "other" Christians was no longer significant over longer time horizons (OR=0.7–0.8) and the ORs for Roman Catholics, non-Christians, and veterans with no religion became significant (OR=0.5–0.6

for Roman Catholics in the 3- through 12-month models; OR=0.5–0.6 for non-Christians in the 3-month and 12-month models; OR=0.8–0.7 for no religion in the 6-month and 12-month models).

Four additional variables became significant in the 1-month model: male gender (OR=3.3, decreasing from 2.2 to 1.3 over longer time horizons); age (OR=0.4–0.5 for the two youngest quintiles, increasing to 0.6–0.8 over longer time horizons; and nonsignificant OR=0.6–0.9 for the two oldest quintiles, decreasing to significant OR=0.7 in the 12-month model); marital status (OR=2.1 for never married increasing to OR=2.5–2.9 in the 6- and 12-month models; and nonsignificant OR=1.1–1.3 for currently married in the 1- and 3-month models increasing to significant OR=1.6–1.7 in the 6- and 12-month models compared to currently separated); most recent era of active duty service (OR=2.6 for Persian Gulf War era in the 1-month model and OR=2.0–2.3 in models for longer time horizons; a nonsignificant OR=1.0 for the Pre-Vietnam era that became significant OR=1.9 in the 12-month model; and a nonsignificant OR=0.9 for the Vietnam era that became significant ORs=1.5–1.8 in the 6- and 12-month models compared to the post-Vietnam era). Homelessness became significant in the 3-month model (OR=0.7 for currently homeless remaining significant OR=0.6–0.07 in the 6- and 12-month models; OR=0.7 for recently homeless remaining significant in the 6-month model OR=0.7 but not in the 12-month model OR=0.9). Census Region and patient income became significant in the 6-month model (OR=1.8–1.6 for Midwest, OR=1.7–1.6 for South, and OR=1.4–1.4 for West in the 6- and 12-month models compared to the Northeast; OR=1.5 in the 12-month model for no income, OR=1.5–1.4 in the 6- and 12-month models for low income, and OR=1.3–1.7 for high-average and high incomes in the 6- and 12-month models compared to low-average income).

Super Learner Results
Feature Selection

As noted above in the *Analysis Methods* section, we used three different feature selection methods to prune the more than 89,000 potential predictors included in the dataset. This process resulted in the selection of 1,221 features for the 1-week model, 2,411 for the 1-month model, 4,074 for the 3-month model, 5,675 for the 6-month model, and 8,071 for the 12-month model.

Classifier Weighting

As noted above in the *Analysis Methods* section, SL generates a cross-validated weight that defines the relative importance of the different classifiers in the ensemble. Neural network was the best classifier for the 1-week model and random forest was best for the other models and second most important for the 1-week model (**Supplementary Table 13**). Extreme gradient boosting was one of the top 5 classifiers in all models, support vector machines (with varying kernels) in 4 of the 5, generalized linear models in 4 of the 5, Bayesian additive regression trees in 1 of the 5, and elastic net in 1 of the 5.

Area Under the Receiver Operating Characteristic Curve

The AUCs of the five SL models when applied to the holdout sample were in the range .67–.74. Models for the shortest time horizons had the lowest AUCs (.67 for 1 week, .68 for 1 month). Models for the longer time horizons had higher AUCs (.73 for 3 months, .74 for 6 months, .74 for 12 months). These results are shown in the diagonals entries in **Table 2**. But a comparison of the performance of each model predicting suicides over each time horizon (i.e., comparing all entries in a single column of **Table 2**) found an unexpected result: that the model built to predict suicides over the 12-month time horizon also outperformed all other models predicting suicides over each shorter time horizon. AUC = .79 versus .67–.77 to predict suicides within the first week of hospital discharge, AUC = .82 versus .63–.77 to predict suicides within the first month of discharge, AUC = .78 versus .61–.73 to predict suicides within 3 months of discharge, AUC = .80 versus .61–.74 to predict suicides within 6 months of discharge, and AUC = .74 versus .60–.71 to predict suicides within 12 months of discharge. With the exception of a single inversion in predicting suicides within the first week of discharge (between the SL models designed to predict suicides within 3 and 6 months after discharge), a consistently monotonic association was found within each time horizon for AUC to increase as the time horizon for model development increased. Based on this result, we focused subsequent analyses on the SL model developed to predict suicides within 12 months of discharge. Consistent with the guarantee that SL outperforms the best classifiers in the ensemble, the AUCs of this best SL model averaged 0.02 higher than the best individual classifier in the ensemble (random forest) across all time horizons (**Supplementary Figure 1**).

Operating Characteristics

Inspection of the ROC curves for the best-fitting SL model (i.e., the model developed to predict suicides over the 12-month time horizon) in the holdout sample showed that the slope was steepest for 1-specificity in the range 0–0.05, which corresponds roughly to the 5% of patients with highest predicted suicide risk in the model (**Figure 2**). The sensitivities
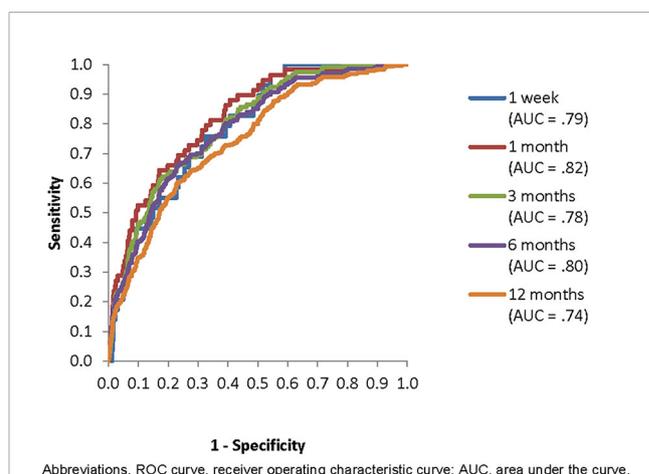
at this threshold show that these patients accounted for 24.1% of all suicides in the holdout sample that occurred in the 1 week after hospital discharge, 32.2% in 1 month, 26.9% in 3 months, 26.4% in 6 months, and 22.4% in 12 months (**Table 3**). This means that an intensive postdischarge case management program that was delivered only to the 5% of hospitalized patients with highest predicted suicide risk would capture 22.4%–32.2% of the patients who would otherwise die by suicide by one or more of the time horizons. Other consistent inflection points in the slope were at 1-specificity of about 0.2 and 0.6. A case management program delivered to the 20% or 60% of patients with highest predicted risk would capture 55.2%–66.1% (20% decision threshold) and 90.3%–100% (60% decision rule) of the patients who would otherwise die by suicide at one or more of the time horizons.

The proportion of patients receiving the intervention who would otherwise go on to die by suicide (i.e., PPV) is an important consideration in determining the potential value of any targeted suicide prevention intervention. As noted above in the section on analysis methods, PPV increases as the number of patients above the decision threshold decreases and as the time horizon increases. The highest PPV for our model is 1.2% for the .05 threshold over a 12-month time horizon. In other words, this is the proportion of patients above that threshold who would be expected to die by suicide in the 12 months after hospital discharge in the absence of any interventions beyond those currently provided by VHA. PPV decreases to 0.4% at the most liberal threshold considered (.60) over the same time horizon. By far the lowest PPVs are for the 1-week time horizon, where values are in the range 0.12%–0.04% across thresholds. That is, we would expect 0.12%–0.04% of patients above the .60 and .05 thresholds, respectively, to die by suicide in the first week after hospital discharge.

The adjusted PPVs, in comparison, decrease rather than increase as time horizons increase. The highest adjusted PPV is 6.1% for the .05 threshold over a 1-week time horizon. In other

TABLE 2 | Area under the receiver operating characteristic curve (AUC) of the super learner model's developer in the training sample for each time horizon to predict suicides in the holdout sample over each of the five time horizons.

	Time horizon for prediction in the holdout sample				
	1 week	1 month	3 months	6 months	12 months
Time horizon for model development					
1-week		.67	.63	.61	.62	.60
1-month		.71	.68	.70	.70	.67
3-month		.77	.76	.73	.72	.69
6-month		.75	.77	.73	.74	.71
12-month		.79	.82	.78	.80	.74

Abbreviations. ROC curve, receiver operating characteristic curve; AUC, area under the curve.

FIGURE 2 | Receiver operating characteristic (ROC) curve for the best Super Learner model (to predict suicides within 12 months of hospital discharge) applied in the holdout sample to predict suicides over each of the five time horizons.

TABLE 3 | Operating characteristics at a range of thresholds of the best Super Learner model (developed to predict suicides within 12 months of hospital discharge) applied in the holdout sample to predict suicides over each of the five time horizons.

Threshold P	SN		SP		PPV		Adjusted PPV	
	%	(SE)	%	(SE)	S/100k	(SE)	%	(SE)
Suicides within 1 week of hospital discharge								
.05	24.1	(0.1)	94.9	(0.1)	0.12	(0.01)	6.1	(0.5)
.10	44.8	(0.2)	89.9	(0.1)	0.11	(0.01)	5.7	(0.5)
.20	55.2	(0.2)	79.9	(0.1)	0.07	(0.01)	3.5	(0.4)
.60	100.0	(0.0)	39.9	(0.1)	0.04	(0.01)	2.1	(0.3)
Suicides within 1 month of hospital discharge								
.05	32.2	(0.1)	94.9	(0.1)	0.3	(0.0)	3.8	(0.2)
.10	52.5	(0.2)	89.9	(0.1)	0.3	(0.0)	3.1	(0.2)
.20	66.1	(0.1)	79.9	(0.1)	0.2	(0.0)	2.0	(0.1)
.60	98.3	(0.0)	39.9	(0.1)	0.1	(0.0)	1.0	(0.1)
Suicides within 3 months of hospital discharge								
.05	26.9	(0.1)	94.9	(0.1)	0.5	(0.0)	2.1	(0.1)
.10	46.2	(0.2)	89.9	(0.1)	0.5	(0.0)	1.9.	(0.1)
.20	63.0	(0.1)	79.9	(0.1)	0.3	(0.0)	1.3	(0.1)
.60	95.8	(0.1)	39.9	(0.1)	0.2	(0.0)	0.6	(0.0)
Suicides within 6 months of hospital discharge								
.05	26.4	(0.1)	94.9	(0.1)	0.9	(0.0)	1.8	(0.1)
.10	40.4	(0.1)	89.9	(0.1)	0.7	(0.0)	1.4	(0.0)
.20	61.5	(0.1)	79.9	(0.1)	0.5	(0.0)	1.1	(0.0)
.60	93.8	(0.1)	39.9	(0.1)	0.3	(0.0)	0.6	(0.0)
Suicides within 12 months of hospital discharge								
.05	22.4	(0.1)	94.9	(0.1)	1.2	(0.0)	1.2	(0.0)
.10	35.0	(0.1)	90.0	(0.1)	1.0	(0.0)	1.0	(0.0)
.20	55.3	(0.2)	80.0	(0.1)	0.8	(0.0)	0.8	(0.0)
.60	90.3	(0.1)	40.0	(0.1)	0.4	(0.0)	0.4	(0.0)

Threshold P, the proportion of patients classified at above the clinical threshold based on their predicted probabilities of suicide; SN, Sensitivity; SP, specificity; PPV, positive predictive value stated in terms of suicides expected per 100 patients over the time horizon; S, suicides; Adjusted PPV, PPV adjusted for the length of the time horizon to reflect the expected ratio of suicides to number of person-years of intervention over that time horizon; SE, standard error.

words, we would expect 6.1 deaths per year during the first week after hospital discharge for every 100 patients discharged per week over that time period (i.e., a total of 5,200 patients, each considered to be at risk for only the first week after discharge). This comparatively high adjusted PPV speaks to the potential value of special interventions focused on the high rates of imminent risk among patients shortly after discharge. Although actual benefit will depend on effectiveness, we see here that the potential benefit for a fixed level of clinical effort of longer preventive interventions with the small proportion of patients at high risk is greater than the benefit of shorter interventions for larger proportions of patients. For example, if the alternative interventions were equally effective and limited to the time horizons considered, our results suggest that more lives would be saved for a fixed level of effort by intervening with the 10% of patients at highest risk for 6 months (Adj PPV = 1.4%) than with the 20% of patients at highest risk for 3 months (Adj PPV = 1.3%) or the 60% of patients at highest risk for 1 month (Adj PPV = 1.0%).

In considering optimal allocation of intervention resources in this way, it is important to note that patients with high suicide risk also have significantly elevated risk of other negative outcomes, such as other types of death by external cause, suicide attempts, severe and permanently disabling injuries, and repeat psychiatric hospitalizations (20, 88). These other outcomes might be reduced by intensive postdischarge interventions designed to reduce suicide death. A formal analysis of intervention net benefit would be needed to take these added benefits into consideration. This would require us to make estimates of intervention effects among patients at different levels of risk across interventions that vary in duration. An added complication is that intervention effectiveness might vary depending on patient severity related to differential suicide risk. We discuss the logic of such an analysis below in the subsection on *Criticisms of Suicide Prediction Models* as part of the description of the concept of intervention net benefit.

Predictor Importance

The top 10 predictors accounted for 43.7% of overall model performance based on the gain metric in the Extreme Gradient Boosting algorithm (77), the top 25 predictors for 58.1%, the top 50 predictors for 70.3%, and the top 100 predictors for 81.4% of model performance. 3,521 predictors were needed to account for 100% of model performance. Sorting these predictors into categories shows that psychopathological risk factors were most important (accounting for 51.1% of overall model performance) followed by social determinants of health (26.2% of overall model performance, including 11.8% for socio-demographics, 2.9% for V codes, and 11.5% for small area geocode data) and history of suicidal behaviors (14.8%) (**Table 4**). The other categories of predictors were much less important (physical disorders 6.6%; medications classified by the US Food and Drug Administration as potential risk factors for suicide 0.7%).

DISCUSSION

Comparisons With Other Suicide Prediction Models

Most prior efforts to develop suicide prediction tools focused 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. These models are designed for use either at intake or discharge to help guide treatment planning. Meta-analyses suggest that the suicide rate among emergency department patients presenting with suicidal intent or after a suicide attempt is about 1,600/100,000 within 1 year of the emergency department visit (89), that the suicide rate among inpatients is about 150/100,000 inpatient-years (90), and

TABLE 4 | Predictor variable importance[1] overall by category and for the predictors in the top 10, 11–25, and 26–50 in the best Super Learner model (to predict suicides within 12 months of hospital discharge)[2].

I. History of prior suicidal behaviors: 14.8% total importance

    4 in the top 10: 1Y suicide attempt (D); intake suicide attempt (D); 30D and 2Y suicide attempt (D)

    2 in the top 11–25: 3Y suicide attempt (D); High risk flag in month prior to admission (D)

II. Psychopathological risk factors: 51.1% total importance

    3 in the top 10: LT psychiatric hospitalizations (C); LT outpx cocaine dependence (D); LT outpx drug dependence (C)

    8 in the top 11–25: LT outpx neurotic, personality, and other nonpsychotic disorders (C); LT inpx cocaine dependence (D); 1Y nonopioid analgesic medication (C); LT outpx unspecified schizophrenia (C); LT inpx cocaine dependence (Q); LT outpx unspecified schizophrenia (D); LT ED neurotic, personality and other nonpsychotic disorders (C); Current mood disorder (D)

    9 in the top 26–50: LT outpx specialty MH neurotic, personality, and other nonpsychotic disorders (C); 90D sedative/hypnotic medication (C); 90D outpx recurrent MDD (C); Current episodic mood disorder (D); 3Y outpx nondependent drug abuse (C); LT outpx specialty MH drug dependence (C); LT PCP outpx alcohol dependence (C); LT inpx neurotic, personality, and other nonpsychotic disorders (C); 2Y outpx recurrent MDD (C)

III. Quality of care: Inadequately assessed

IV. Time since hospital discharge: Not included in the model[1]

V-VII. Social determinants of health: 26.2%

    V. Socio-demographics: 11.8% total importance

        2 in the top 10: Non-Hispanic Black (D); Age (C)

        2 in the top 26–50: Sex (D); Non-Hispanic White (D)

    VI. ICD-9-CM E and V codes: 2.9% total importance

        1 in the top 11–25: 2Y housing problem (C)

        2 in the top 26–50: Current other psychological or physical stress (D); 2Y multiple housing problems (C)

    VII. Small-area geocode data: 11.5% total importance

        1 in the top 10: BG % Non-Hispanic White x Px Non-Hispanic White (C)

        3 in the top 11–25: County social capital (% voted in presidential election) (C); BG % Non-Hispanic Black x Px Non-Hispanic Black (C); County bankruptcy rate change, past 2Y vs. 3–5Y (C)

        8 in the top 26–50: BG % Non-Hispanic Black (C); BG % Hispanic (C); County social capital (composite) (C); County social capital (charity rate) (C); County suicide rate past 3Y (C); County violent crime rate past 1Y (C); County unemployment rate past 1Y (C); County debt-to-income ratio change past 2Y vs. 3–5Y (C)

VIII. Physical disorders: 6.6% total importance

    1 in the top 11–25: LT pain diagnosis (D)

    4 in the top 26–50: 1Y vitamin medications (C); LT heart disease (D); LT high-risk physical disorders (C)[3]; 3Y chronic obstructive pulmonary disease (D)

IX. Medications classified by FDA as increased risk of suicide: 0.7% total importance

    0 in the top 50

X. Medical procedures: 0.2% total importance

    0 in the top 50

ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; FDA, US Food and Drug Administration.

[1]*Variable importance was defined in terms of the Extreme Gradient Boosting (XGBoost) "gain" measure (77). We did not consider time since hospital discharge, which was category (iv) in our list of predictor categories, because the model was designed to predict suicide at any time in the 12 months after hospital discharge.*

[2]***Geocode:*** *BG = The small-area geocode variable was defined over the Block Group of the patient's residence; County = The small-area geocode variable was defined over the County of the patient's residence;* ***Variable metric:*** *D = A yes-no dichotomous variable; C = A continuous count variable; Q = A truncated continuous variable transformed to quintiles of the count distribution;* ***Time:*** *30D = The predictor was assessed over the 30 days before hospital admission; 1Y = The predictor was assessed over the 1 year (365 days) before hospital admission; 2Y = The predictor was assessed over the 2 years (730 days) before hospital admission; 3Y = The predictor was assessed over the 3 years (1,095 days) before hospital admission; LT = The predictor was assessed over the lifetime of the patient's contact with the VHA system beginning January 1, 2000;* ***Treatment sector:*** *Intake = At the time of the focal hospitalization; ED = Emergency department; Inpx = Psychiatric inpatient; Outpx = Any outpatient treatment; Specialty outpx = Outpatient treatment by a mental health treatment provider; PCP outpx = Outpatient treatment by anyone other than a mental health treatment provider; No mention of treatment sector = Aggregation of the diagnosis or treatment across all sectors.*

[3]*High-risk physical disorders were defined based on Ahmedani et al. (32) as asthma, back pain, cancer (esophageal; head and neck; Hodgkin lymphoma; lung; mesothelioma; pancreatic; prostate; stomach; testicular), congestive heart failure, chronic obstructive pulmonary disease, diabetes mellitus, epilepsy, fibromyalgia, heart disease, HIV/AIDS, hypertension, migraine, multiple sclerosis, osteoporosis, Parkinson's disease, psychogenic pain, renal disease, sleep disorders, and traumatic brain injury.*

that the suicide rate after psychiatric hospital discharge is between 3,000/100,000 person-years in the first week after discharge and 650/100,000 person-years 4–12 months after discharge (49, 91). Although only about 2% of the US population in a given year either visit an emergency department with suicide intent, visit an emergency department after a suicide attempt, or are hospitalized for a psychiatric problem, such individuals account for nearly one-third of all US suicides (92).

As reviewed elsewhere, ML methods with a single classifier were used in most recent studies aimed at building suicide prediction models in these high-risk patient populations using predictors that included information collected from patient self-report scales, clinical rating scales, and administrative data (71). Prediction accuracy in our model was generally comparable to that in these earlier models even though we did not use any patient self-report data or clinician ratings data in building our model. Sensitivity was typically 0.6–0.7 in these earlier models when specificity was set at 0.8 (compared to sensitivity of 0.55–0.66 in our model) and 0.4–0.5 when specificity was set at 0.9 (compared to sensitivity of 0.35–0.52 in our model). In other words, the performance of our model is roughly comparable to that of previous ML models designed to predict suicide in high-risk patient samples, although the vast majority of previous

studies focused on samples of emergency department patients or patients who made suicide attempts rather than on hospitalized patients. As noted below, we anticipate that ongoing refinements will improve the prediction accuracy of our model, perhaps substantially, but it is likely that an optimal version of such a model will have operating characteristics not dramatically higher than those found here unless a breakthrough occurs in the discovery of a critical biomarker.

In terms of predictors, our finding that psychopathological risk factors and prior suicidality were important is not surprising. It is somewhat surprising, though, that bipolar disorder did not figure more prominently than it did among the psychopathological risk factors. It is noteworthy in this regard that bipolar disorder was a powerful predictor when considered alone in the initial screening models, but other predictors more efficiently captured the variance due to bipolar disorder in the SL analysis. In particular, patients with bipolar disorder had a stronger history of suicidality and more risk factors involving social determinants of health than other patients. It is important to recognize in this regard, though, that predictor importance was defined in terms of prediction and not intervention. This means that interventions focused on improving bipolar disorder treatment among high-risk patients might very well be useful in reducing suicides even though a diagnosis of bipolar disorder was not one of the most important variables in making up the composite risk index.

Another surprising finding was that social determinants of health were quite important in making up the index. This might mean that we need to think in terms of upstream interventions to address the suicide problem, a possibility of increasing interest in many areas of medicine (93–95). However, as an opposite side of the coin in the above discussion of bipolar disorder, the fact that social determinants of health indicators were important *predictors* does not necessarily mean that they would be useful intervention targets, as they might be risk markers rather than causal risk factors (96). The association of homelessness with reduced probability of suicide in the model predicting SL scores is a case in point. The gross association of homelessness with suicide is positive, but the OR for homelessness in the multivariate model suggests that homelessness is protective. This is due to the fact that the high suicide rate among homeless patients is considerably lower than expected based on the fact that these patients experience a wide array of other risk factors for suicide that are assessed in the model. This kind of subadditive multivariate interaction should not be taken as evidence that homelessness is somehow protective against suicide, but as an indication that homelessness is a strong marker of the existence of this subadditive interaction.

A final surprising result involved our finding that the model developed for the 12-month time horizon outperformed the models developed for shorter time horizons in predicting suicides across those shorter time horizons. This is an important finding given that most prior ML models to predict suicides from administrative data used much shorter time horizons than 12 months. In particular, the VHA Reach Vet model (88), which is used to target preventive interventions to

the roughly 35,000 VHA patients each year considered to be at highest suicide risk, is based on a 30-day time horizon, which we showed clearly to be suboptimal in the current application to inpatients. Our finding might reflect the fact that we had more statistical power to detect meaningful associations in the 12-month model because of the larger number of cases than in the models for shorter time horizons. This might explain why the number of predictors that passed our feature screening step increased substantially as the time horizon increased. Another possibility is that suicides became more predictable after the early weeks and months after hospital discharge, but this possibility is inconsistent with the finding in an ancillary analysis not reported here that SL models estimated for *conditional* suicide risk between the 1st week and 1st month after hospital discharge and between 2 and 3, 4 and 6, and 7 and 12 months after discharge did not vary substantially in their AUCs. Based on this result, it is possible that models designed specifically to predict imminent suicide risk in the first week after discharge or for longer time horizons less than 12 months (e.g., 1, 3, or 6 months) might improve on the model developed to predict 12-month suicides if a sufficiently large sample was available for training. Larger samples are available for other segments of the VHA population (e.g., outpatients with common mental disorders who report suicidality). As a result, the issue of the optimal time horizon for model development needs to be revisited anew each time a new population segment is targeted for model building.

Another observation related to model optimality involves the fact that we developed our model specifically for a particular segment of the patient population (i.e., psychiatric inpatients) at a particular time when clinical decision-making occurs (i.e., at time of discharge, when a discharge plan needs to be formulated). Other recently developed suicide risk ML models have this same characteristic, such as a model for suicide risk in the months after an outpatient primary care visit designed to provide decision support for clinicians in making specialty referrals [e.g., (97)]. It is possible that models of this sort perform better than models based on total populations of health plans [e.g., (88, 98)]. An interesting point of comparison is the VHA Reach Vet model, which was developed by analyzing data available from all veterans seen in the VHA system regardless of whether they carried diagnoses of or received treatment for mental disorders (88). The Reach Vet model attempted to isolate the 0.1% of all VHA users with highest suicide risk. These veterans were found to account for approximately 2% of all suicide deaths in VHA. Our model, in comparison, focuses on the roughly 1% of VHA patients who are hospitalized for a psychiatric disorder in a year. We found that 10% of this 1% of patients (i.e., the same 0.1% of all VHA patients defined as being above the risk threshold in the Reach Vet model) account for 35% of all the suicides that occur among inpatients in the 12 months after discharge. As noted in the introduction, psychiatric inpatients account for 12% of all VHA suicides over the 12 months after discharge, which means that the 10% of these ex-inpatients with highest risk account for approximately 4% of all VHA suicides (i.e., 35% of 12%), which is roughly twice as high a proportion as in the Reach Vet model. This might reflect the more sophisticated modeling procedures

or expanded set of predictors in our model compared to the Reach Vet model, but the targeted focus on one segment of the patient population might also have played a part. We plan to investigate this issue in ongoing analyses by developing parallel models for other segments of the VHA patient population and determining if these models improve on the performance of a model for the overall VHA population.

Criticisms of Suicide Prediction Models

Many critics have argued that prediction models like the one we presented here are not strong enough to justify use for clinical decision-making even if the costs of generating, updating, and making model results available to clinicians for decision support are de minimus (99–102). Two reasons are typically given for this conclusion: first, that the low PPV of the models at the decision thresholds would mean that interventions focused on patients classified high-risk would "subject many patients, who will never die by suicide, to excessive intrusion or coercion" (103); and second, that the low sensitivity of the models at these thresholds would mean that only a minority of suicides occurred among patients classified high-risk. Clinicians unaware of this low sensitivity might draw "false reassurance" from negative predictions and deny needed treatment to patients who have a meaningful risk of suicide but are incorrectly classified low-risk (102). These criticisms have become institutionalized in clinical practice guidelines that advise clinicians against using structured suicide prediction models and instead recommend that clinicians implement "an integrated and comprehensive psychosocial assessment" (104) of needs and risks with all psychiatric inpatients, psychiatric emergency department patients, and other patients considered to be at elevated suicide risk (105, 106).

But are clinical evaluations any more accurate than structured assessments in predicting subsequent suicides and SRBs? The evidence suggests not. Statistical models have long been known to be superior to unstructured clinical judgments in predicting a wide range of clinical outcomes (107), although varying over settings and decisions (108). Consistent with this evidence, 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 stronger predictors of subsequent suicides than were several other social, historical, and clinical variables assessed by patient self-report or extracted from administrative databases (26). A more recent meta-analysis of seven studies found that clinical assessments were only weakly associated with subsequent suicides among patients after hospital treatment of SRBs (109). Based on this evidence, review authors conclude that clinicians should focus on need for services rather than on suicide risk in assessing suicidal patients. But this recommendation overlooks the fact that clinical decisions about need for services should be informed by perceived suicide risk. This fact is recognized in the strategy for suicide prevention advanced by the US National Action Alliance for Suicide Prevention (110) 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 (111, 112). However, given the greater

accuracy of prediction models than clinical judgments about patient suicide risk, it makes sense for clinicians to have access to the results of these models as input in developing intervention plans.

There is a third criticism that could be raised here involving our suggestion that a prediction model such as the one we present could be used to target a new kind of intensive postdischarge case management intervention. It is important to note that we have no way to know if such an intervention would be effective in reducing postdischarge suicides and, if so, if it would be cost-effective to do so compared to existing interventions. Our assumption at the onset was that this kind of intervention would be too expensive to provide to all patients being discharged from psychiatric hospitalizations and that any hope of it being practical would require targeting. But it needs to be said that the existence of a prediction model would not make it practical to implement the intervention if it was not cost-effective relative to other uses of the equivalent clinical resources.

Assuming that such an intervention could be cost-effective if it was targeted correctly, where should the decision threshold be set using a model of the sort we developed? No agreement has emerged on this question (113). The thresholds we considered were based on observed inflection points in the ROC curves, but PPV will inevitably be very low for a rare outcome such as suicide for any decision threshold that includes more than a very small proportion of patients. And this reintroduces the criticism that prediction models with low PPV are not clinically useful. However, this argument is incorrect. As discussed in more detail elsewhere (70), Net Benefit (NB), not PPV, should be considered the key operating characteristic in evaluating the value of new interventions. NB is the standardized difference between the number of true positives at or above a decision threshold and the discounted number of false positives at or above that threshold, where the discount rate explicitly evaluates the value of intervening with a true positive (i.e., someone who would die by suicide in the absence of intervention) relative to the costs (both direct and indirect) of intervening with a false positive. Once the cost-benefit ratio is determined, an optimal decision threshold can be calculated, noting that the optimal decision threshold might be 0; that is, it is shown empirically that the intervention is not cost-effective for any patient. It is important to note that NB can be positive even when PPV is low if the costs of intervening with false positives are low relative to the benefits of intervening with true positives. That is why it is considered cost-effective to prescribe statins to adults aged 40–75 with mildly elevated total cholesterol even though annual PPV is only .0075 (which is lower than the PPVs rejected by critics of suicide prediction models as too low for targeting interventions) and statin treatment requires nearly 500 person-years of treatment to prevent one case of atherosclerotic cardiovascular disease (114).

Future Directions
Improving Model Performance

We are continuing to make refinements by expanding the predictor set in several important ways. First, as noted in the introduction, recent UK research found suggestive evidence that

a number of inpatient unit characteristics, such as staff turnover and average length of stay, were significant predictors of postdischarge suicide rates (14, 15). We are in the process of assembling a unit-level time series dataset for all the roughly 100 psychiatric inpatient units in VHA to assess these and other unit characteristics as potential indicators of treatment quality relevant to suicide risk. The same UK studies found that a number of policies for treating high-risk outpatients, such as the use of community outreach teams, were important predictors of geographic variation in suicide rates. We are assembling a time series dataset with an expanded set of such indicators for each VHA outpatient clinic where VHA psychiatric inpatients are transferred after hospital discharge.

Second, we are trying to expand the indicators of social determinants of health by using natural language processing (NLP) methods to elicit additional information from clinical notes. A growing number of methodological studies have shown that NLP can be used to generate such measures from clinical notes (115–118) and to elicit information about a wide range of social determinants of health relevant to suicide beyond the information captured in ICD-9 V codes (119–123).

Third, based on the unexpectedly strong influence of social determinants of health in our model, we are expanding the assessment of this domain in our next phase of model-building by adding the 450 variables in the LexisNexis Social Determinants of Health database to our predictor set (124). These variables assess various aspects of employment, finances, marital status, parenting status, and involvement with the criminal justice system for close to 300 million Americans and their neighborhoods. In addition to using these individual-level and neighborhood-level variables additively, we will also create more complex multivariate profiles to characterize mismatch between patients and their neighborhoods on a wide range of characteristics.

Precision Treatment

The criticism noted above that suicide prediction models have low sensitivity speaks to an important issue not addressed by these models: that the patients at highest suicide risk are not necessarily the patients most likely to be helped by existing interventions. As it happens, though, another class of models can be developed to help predict which available intervention is most likely to help a specific patient and the extent of that help (71). The estimates of predicted risk based on models of the sort presented in this paper can be used as input to such precision treatment models. Or the intervention could be limited to patients with meaningfully elevated predicted risk based on an initial model of this type. Or the predicted values from a model like the one developed here could be provided to clinicians as input to their clinical decision-making. But the critical distinction between the type of model developed in the current paper and precision treatment models is that the latter focus on interactions between patient characteristics and specific treatment alternatives with the goal of developing an individualized treatment rule (ITR) that predicts which treatment option is likely to be best for which patients (70,

125, 126). We plan to develop such a model as part of a pragmatic trial for intensive case management after psychiatric hospital discharge focused on the 60% of patients with meaningfully elevated risk of postdischarge suicide.

It is also possible to develop preliminary precision treatment models from the kinds of observational study designs used in the current report (70). However, this can be done only for intervention that is already being used in practice. If promising ITRs are developed in such studies, rigorous evaluation is needed in pragmatic trials. This is a much more feasible order of operations in some cases than beginning with a controlled trial sufficiently large to support the development of an ITR. As an example, we are involved in an investigation of this sort to study the circumstances under which patients should versus should not be hospitalized after nonfatal suicide attempts (SAs). As detailed elsewhere (71), it is unclear whether hospitalization (which occurs after 50% of VHA outpatient SAs) reduces subsequent SRBs (either suicide or subsequent SAs). A recent study carried out by UK investigators attempted to address this issue in a prospective observational study that used propensity score methods to adjust for baseline differences among acutely suicidal patients managed in four different ways (psychiatric hospitalization, general hospital admission as a psychiatric inpatient, psychiatric outpatient treatment, specialist evaluation without referral for treatment). Differences in subsequent 12-month suicidal behaviors across the four groups were largely nonsignificant after this risk adjustment (127, 128), implying either that whether a patient was hospitalized had no effect or that hospitalization was beneficial for some patients and harmful for a roughly equal number of patients. We are investigating the latter possibility.

Although theorizing exists about the patients most likely to be helped and those most likely to be hurt by hospitalization (129), little empirical research exists on these hypotheses (90). We are attempting to develop an ITR to provide guidance in making this decision in the immediate aftermath of an outpatient suicide attempt. It is infeasible to use controlled treatment trials to study the aggregate effects of these decisions given the large samples required (130) and rarity of SRBs other than among high-risk patients for whom randomization would be unethical. However, modern statistical methods applied to large electronic health record databases to adjust for ("balance") baseline differences in patients across types of treatment can be used to estimate aggregate treatment effects (131, 132). Such methods often yield results similar to those in controlled treatment trials (133, 134). Extensions exist to develop ITRs using ML methods (135–137). This is, in fact, what we are attempting to do: to see if ensemble ML methods can be used to develop an ITR that allows us to estimate which specific patients should be hospitalized and which ones not after SAs, with the goal of reducing subsequent SRBs. If our observational analysis suggests that a useful ITR can be developed, we will implement a pragmatic trial to determine the validity of that conclusion. Such an ITR could have considerable value for clinical practice in providing empirical guidance in making this critical treatment decision.

Implementation Beyond VHA

Some of the key predictors in our model are unavailable in health systems other than the VHA, making it impossible to apply our model outside of VHA. The fact that most VHA patients are males is another unique characteristic of our sample. However, several innovations implemented here could be used in building models to predict suicides in other health systems. The most notable of these are expansions of the predictor set, investigation of diverse time horizons, and use of ensemble ML methods. With regard to the predictor set, we made more use than previous ML suicide studies of small-area geocode data and E-V codes to generate additional information about social determinants of health. We constructed composite condition indices that cut across the ICD hierarchy using information from external sources regarding such organizing constructs as pain severity and Food and Drug Administration drug warnings. And we used unsupervised analysis methods to develop multimorbidity profiles that we considered along with composites based on the ICD hierarchy. Extensive detail on predictor variable construction is provided in **Supplementary Tables**.

Conclusions

We found that a model could be developed using only administrative data available while patients are still hospitalized to target inpatients with high risk of postdischarge suicide for intensive postdischarge case management. If an effective intervention of this or another sort could be provided to the 60% of patients classified by our model as having highest suicide risk prior to hospital discharge, we estimate that 90.3%–100% of the patients who would otherwise go on to die by suicide would be reached. If additional intervention could be provided only to the 20% of patients classified by our model as having highest suicide risk, we estimate that 55.3%–66.1% of the patients who would otherwise go on to die by suicide would be reached. It is important to note, though, that we provided no evidence to suggest that an additional intervention would be effective in this way. It might not be. This remains to be seen.

However, we noted at the onset that intensive postdischarge case management programs, which are not used in VHA, have been shown elsewhere to be effective in reducing suicides after psychiatric hospital discharge (9–16), leading to recommendations to add such programs to existing postdischarge suicide preventive interventions (17). The motivation for our model development exercise was the assumption that these labor-intensive programs could not be implemented cost-effectively given the rarity of postdischarge suicide unless they were targeted to recently discharged patients at high suicide risk. Our aim was to determine whether a prediction model could be developed that had a sufficiently high concentration of risk to make the consideration of targeting plausible. We have done that, showing, for example, that nearly one-third of all suicides occurring in the first month after hospital discharge occur among the 5% of patients classified by our model as having highest risk. Whether this is the optimal decision threshold for implementing a new postdischarge intervention or, indeed, if such an intervention would be cost-effective at any threshold in VHA is beyond the scope of this report. A cost-effectiveness analysis based on simulated data using information about the effects of existing interventions and our PPV estimates would be the next logical step in deciding whether the evidence is sufficiently strong to justify implementing a pragmatic trial [e.g., (138)].

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. The data used in this report were obtained from VHA clinical records based on a VA IRB-approved protocol. We are prohibited from sharing the data outside of those listed on the IRB-approved protocol. Requests to access these datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of the VA Center of Excellence for Suicide Prevention and Harvard Medical School with a waiver of informed consent based on the fact that the data were deidentified. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

RK, TB, SD, EK, MN, WP, LW, and RB conceptualized and designed the study. SG, HL, and NS organized the database. MB, SD, SG, JG, EK, JK, SL, WM, WP, NS, and JS contributed to the selection of predictors. HL and RB performed the statistical analysis with direct supervision from MP and NS. OD, AL, and PM worked with RK to develop the analysis plan. RK wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2020.00390/full#supplementary-material

REFERENCES

1. Xu J, Murphy SL, Kochanek KD, Bastian BA. Deaths: final data for 2013. *Natl Vital Stat Rep* (2016) 64:1–119.

2. Case A, Deaton A. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proc Natl Acad Sci USA* (2015) 112:15078–83. doi: 10.1073/pnas.1518393112

3. Walser RD, Garvert DW, Karlin BE, Trockel M, Ryu DM, Taylor CB. Effectiveness of Acceptance and Commitment Therapy in treating depression and suicidal ideation in Veterans. *Behav Res Ther* (2015) 74:25–31. doi: 10.1016/j.brat.2015.08.012

4. US Department of Veterans Affairs. *National Strategy for Prevention Veteran Suicide, 2018-2028*. Washington, DC: US Department of Veterans Affairs Office of Mental Health and Suicide Prevention (2018).

5. Hunt IM, Kapur N, Webb R, Robinson J, Burns J, Shaw J, et al. Suicide in recently discharged psychiatric patients: a case-control study. *Psychol Med* (2009) 39:443–9. doi: 10.1017/S0033291708003644

6. Britton PC, Bohnert KM, Ilgen MA, Kane C, Stephens B, Pigeon WR. Suicide mortality among male veterans discharged from Veterans Health Administration acute psychiatric units from 2005 to 2010. *Soc Psychiatry Psychiatr Epidemiol* (2017) 52:1081–7. doi: 10.1007/s00127-017-1377-x

7. US Department of Veterans Affairs. *Uniform mental health services in VA medical centers and clinics (VHA Handbook 1160.01)*. Washington, D.C: Veterans Health Administration (2008).

8. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. *Cog Behav Pract* (2012) 19:256–64. doi: 10.1016/j.cbpra.2011.01.001

9. Miller IW, Gaudiano BA, Weinstock LM. The Coping Long Term with Active Suicide Program: description and pilot data. *Suicide Life Threat Behav* (2016) 46:752–61. doi: 10.1111/sltb.12247

10. Knox KL, Stanley B, Currier GW, Brenner L, Ghahramanlou-Holloway M, Brown G. An emergency department-based brief intervention for veterans at risk for suicide (SAFE VET). *Am J Public Health* (2012) 102:S33–S7. doi: 10.2105/AJPH.2011.300501

11. Matarazzo BB, Farro SA, Billera M, Forster JE, Kemp JE, Brenner LA. Connecting veterans at risk for suicide to care through the HOME Program. *Suicide Life Threat Behav* (2017) 47:709–17. doi: 10.1111/sltb.12334

12. Stanley B, Brown GK, Brenner LA, Galfalvy HC, Currier GW, Knox KL, et al. Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. *JAMA Psychiatry* (2018) 75:894–900. doi: 10.1001/jamapsychiatry.2018.1776

13. Anderson M, Jenkins R. The national suicide prevention strategy for England: the reality of a national strategy for the nursing profession. *J Psychiatr Ment Health Nurs* (2006) 13:641–50. doi: 10.1111/j.1365-2850.2006.01011.x

14. Kapur N, Ibrahim S, While D, Baird A, Rodway C, Hunt IM, et al. Mental health service changes, organisational factors, and patient suicide in England in 1997-2012: a before-and-after study. *Lancet Psychiatry* (2016) 3:526–34. doi: 10.1016/S2215-0366(16)00063-8

15. While D, Bickley H, Roscoe A, Windfuhr K, Rahman S, Shaw J, et al. Implementation of mental health service recommendations in England and Wales and suicide rates, 1997-2006: a cross-sectional and before-and-after observational study. *Lancet* (2012) 379:1005–12. doi: 10.1016/S0140-6736(11)61712-1

16. De Leo D, Heller T. Intensive case management in suicide attempters following discharge from psychiatric care. *Aust J Prim Health* (2007) 13:49–58. doi: 10.1071/PY07038

17. Olfson M, Marcus SC, Bridge JA. Focusing suicide prevention on periods of high risk. *JAMA* (2014) 311:1107–8. doi: 10.1001/jama.2014.501

18. Miller IW, Gaudiano BA, Weinstock LM. Introduction to the Coping Long-Term with Active Suicide Program (CLASP). In: *Mini workshop presented at the annual meeting of the Association for Behavioral and Cognitive Therapies*. Washington, DC: Association for Behavioral and Cognitive Therapies (2018).

19. Weinstock LM, Primack J. The Coping Long-Term with Active Suicide Program (CLASP) across vulnerable transitions in care: treatment description, recent data, and future directions. In: . *VA Suicide Prevention Cyberseminar Series (SoSP) Webinar*. Washington, DC: Veterans' Health Administration (2018).

20. 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. doi: 10.1001/jamapsychiatry.2014.1754

21. Centers for Disease Control and Prevention. (2015). *National Death Index*. Centers for Disease Control and Prevention, Department of Health & Human Services. [Accessed August 1, 2016].

22. Troister T, Links PS, Cutcliffe J. Review of predictors of suicide within 1 year of discharge from a psychiatric hospital. *Curr Psychiatry Rep* (2008) 10:60–5. doi: 10.1007/s11920-008-0011-8

23. Caley CF, Perriello E, Golden J. Antiepileptic drugs and suicide-related outcomes in bipolar disorder: a descriptive review of published data. *Ment Health Clin* (2018) 8:138–47. doi: 10.9740/mhc.2018.05.138

24. Adler NE, Glymour MM, Fielding J. Addressing social determinants of health and health inequalities. *JAMA* (2016) 316:1641–2. doi: 10.1001/jama.2016.14058

25. Fitzpatrick SJ. Reshaping the ethics of suicide prevention: responsibility, inequality and action on the social determinants of suicide. *Public Health Ethics* (2018) 11:179–90. doi: 10.1093/phe/phx022

26. 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. doi: 10.3109/00048674.2011.590465

27. Bickley H, Hunt IM, Windfuhr K, Shaw J, Appleby L, Kapur N. Suicide within two weeks of discharge from psychiatric inpatient care: a case-control study. *Psychiatr Serv* (2013) 64:653–9. doi: 10.1176/appi.ps.201200026

28. Park S, Choi JW, Kyoung Yi K, Hong JP. Suicide mortality and risk factors in the 12 months after discharge from psychiatric inpatient care in Korea: 1989-2006. *Psychiatry Res* (2013) 208:145–50. doi: 10.1016/j.psychres.2012.09.039

29. Centers for Disease Control and Prevention. (2013). *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. Centers for Disease Control and Prevention, Department of Health & Human Services. [Accessed: December 3, 2019].

30. Dworkin ER, Menon SV, Bystrynski J, Allen NE. Sexual assault victimization and psychopathology: a review and meta-analysis. *Clin Psychol Rev* (2017) 56:65–81. doi: 10.1016/j.cpr.2017.06.002

31. Chen T, Roberts K. Negative Life Events and Suicide in the National Violent Death Reporting System. *Arch Suicide Res* (2019). 1–15 doi: 10.1080/13811118.2019.1677275

32. Ahmedani BK, Peterson EL, Hu Y, Rossom RC, Lynch F, Lu CY, et al. Major physical health conditions and risk of suicide. *Am J Prev Med* (2017) 53:308–15. doi: 10.1016/j.amepre.2017.04.001

33. Robertson HT, Allison DB. Drugs associated with more suicidal ideations are also associated with more suicide attempts. *PloS One* (2009) 4:e7312. doi: 10.1371/journal.pone.0007312

34. Arias Vázquez PI, Castillo Avila RG, Dominguez Zentella MDC, Hernández-Díaz Y, González-Castro TB, Tovilla-Zárate CA, et al. Prevalence and correlations between suicide attempt, depression, substance use, and functionality among patients with limb amputations. *Int J Rehabil Res* (2018) 41:52–6. doi: 10.1097/MRR.0000000000000259

35. Méndez-Bustos P, Calati R, Rubio-Ramírez F, Olié E, Courtet P, Lopez-Castroman J. Effectiveness of psychotherapy on suicidal risk: a systematic review of observational studies. *Front Psychol* (2019) 10:277. doi: 10.3389/fpsyg.2019.00277

36. US Department of Veterans Affairs. (2019). *Corporate Data Warehouse (CDW)*. US Department of Veterans Affairs Office of Information & Technology. [Accessed December 6, 2019].

37. US Department of Veterans Affairs. (2019). *VA National Formulary - Pharmacy Benefits Management Services*. US Department of Veterans affairs Pharmacy Benefits Management Strategic Health Group. https://www.pbm.va.gov/nationalformulary.asp [Accessed October 1, 2019].

38. Torres JM, Lawlor J, Colvin JD, Sills MR, Bettenhausen JL, Davidson A, et al. ICD social codes: an underutilized resource for tracking social needs. *Med Care* (2017) 55:810–6. doi: 10.1097/MLR.0000000000000764

39. 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. doi: 10.1177/0033354916670133

40. US Department of Veterans Affairs. (2019). *Homeless veterans: VA National Center on Homelessness Among Veterans*. VA Center on Homelessness Among Veterans. [Accessed January 2, 2019].

41. Elixhauser A, Steiner C, Palmer L. (2015). *Clinical Classifications Software (CCS). US Agency for Healthcare Research and Quality*. Agency for Healthcare Research and Quality. [Accessed December 3, 2019].

42. Finley EP, Bollinger M, Noël PH, Amuan ME, Copeland LA, Pugh JA, et al. A national cohort study of the association between the polytrauma clinical triad and suicide-related behavior among US Veterans who served in Iraq and Afghanistan. *Am J Public Health* (2015) 105:380–7. doi: 10.2105/AJPH.2014.301957

43. Owen-Smith AA, Ahmedani BK, Peterson E, Simon GE, Rossom RC, Lynch FL, et al. The mediating effect of sleep disturbance on the relationship between nonmalignant chronic pain and suicide death. *Pain Pract* (2019) 19:382–9. doi: 10.1111/papr.12750

44. Pugh MJV, Finley EP, Copeland LA, Wang C-P, Noel PH, Amuan ME, et al. Complex comorbidity clusters in OEF/OIF veterans: the polytrauma clinical triad and beyond. *Med Care* (2014) 52:172–81. doi: 10.1097/MLR.0000000000000059

45. Meltzer HY, Alphs L, Green AI, Altamura AC, Anand R, Bertoldi A, et al. Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Arch Gen Psychiatry* (2003) 60:82–91. doi: 10.1001/archpsyc.60.1.82

46. Pompili M, Baldessarini RJ, Forte A, Erbuto D, Serafini G, Fiorillo A, et al. Do atypical antipsychotics have antisuicidal effects? a hypothesis-generating overview. *Int J Mol Sci* (2016) 17:1700. doi: 10.3390/ijms17101700

47. Pompili M, Orsolini L, Lamis DA, Goldsmith DR, Nardella A, Falcone G, et al. Suicide prevention in schizophrenia: do long-acting injectable antipsychotics (LAIs) have a role? *CNS Neurol Disord Drug Targets* (2017) 16:454–62. doi: 10.2174/1871527316666170223163629

48. Stroup TS, Gray N. Management of common adverse effects of antipsychotic medications. *World Psychiatry* (2018) 17:341–56. doi: 10.1002/wps.20567

49. 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. doi: 10.1001/jamapsychiatry.2017.1044

50. US Census Bureau. (2011). *Detailed characteristics, 2006-2010 and 2007-2011 American Community Survey 5-year estimates*. United States Census Bureau. [Accessed December 1, 2019].

51. Evans J, Middleton N, Gunnell D. Social fragmentation, severe mental illness and suicide. *Soc Psychiatry Psychiatr Epidemiol* (2004) 39:165–70. doi: 10.1007/s00127-004-0733-9

52. Steelesmith DL, Fontanella CA, Campo JV, Bridge JA, Warren KL, Root ED. Contextual factors associated with county-level suicide rates in the United States, 1999 to 2016. *JAMA Netw Open* (2019) 2:e1910936. doi: 10.1001/jamanetworkopen.2019.10936

53. Whitley E, Gunnell D, Dorling D, Smith GD. Ecological study of social fragmentation, poverty, and suicide. *BMJ* (1999) 319:1034–7. doi: 10.1136/bmj.319.7216.1034

54. Fontanella CA, Saman DM, Campo JV, Hiance-Steelesmith DL, Bridge JA, Sweeney HA, et al. Mapping suicide mortality in Ohio: a spatial epidemiological analysis of suicide clusters and area level correlates. *Prev Med* (2018) 106:177–84. doi: 10.1016/j.ypmed.2017.10.033

55. Wetherall K, Daly M, Robb KA, Wood AM, O'Connor RC. Explaining the income and suicidality relationship: income rank is more strongly associated with suicidal thoughts and attempts than income. *Soc Psychiatry Psychiatr Epidemiol* (2015) 50:929–37. doi: 10.1007/s00127-015-1050-1

56. Liu K. To Compare is to despair? a population-wide study of neighborhood composition and suicide in Stockholm. *Soc Probl* (2017) 64:532–57. doi: 10.1093/socpro/spw044

57. Alaimo K, Olson CM, Frongillo EA. Family food insufficiency, but not low family income, is positively associated with dysthymia and suicide symptoms in adolescents. *J Nutr* (2002) 132:719–25. doi: 10.1093/jn/132.4.719

58. Ver Ploeg M, Breneman V, Farrigan T, Hamrick K, Hopkins D, Kaufman P, et al. *Access to affordable and nutritious food: measuring and understanding food deserts and their consequences: report to congress*. Washington, DC: United States Department of Agriculture Economic Research Service (2009). https://www.ers.usda.gov/webdocs/publications/42711/12716_ap036_1_.pdf [Accessed November 7, 2019].

59. Institute for Health Metrics and Evaluation (IHME). *United States mortality rates by county 1980-2014*. Seattle, WA: Institute for Health Metrics and Evaluation (2016). Data retrieved from: http://ghdx.healthdata.org/us-data [Accessed November 16, 2019].

60. Reeves A, McKee M, Stuckler D. Economic suicides in the Great Recession in Europe and North America. *Br J Psychiatry* (2014) 205:246–7. doi: 10.1192/bjp.bp.114.144766

61. Goulet JL, Kerns RD, Bair M, Becker WC, Brennan P, Burgess DJ, et al. The musculoskeletal diagnosis cohort: examining pain and pain care among veterans. *Pain* (2016) 157:1696–703. doi: 10.1097/j.pain.0000000000000567

62. Ilgen MA, Kleinberg F, Ignacio RV, Bohnert ASB, Valenstein M, McCarthy JF, et al. Noncancer pain conditions and risk of suicide. *JAMA Psychiatry* (2013) 70:692–7. doi: 10.1001/jamapsychiatry.2013.908

63. Maixner W, Fillingim RB, Williams DA, Smith SB, Slade GD. Overlapping chronic pain conditions: implications for diagnosis and classification. *J Pain* (2016) 17:T93–T107. doi: 10.1016/j.jpain.2016.06.002

64. Mayhew M, DeBar LL, Deyo RA, Kerns RD, Goulet JL, Brandt CA, et al. Development and assessment of a crosswalk between ICD-9-CM and ICD-10-CM to identify patients with common pain conditions. *J Pain* (2019) 20:1429–45. doi: 10.1016/j.jpain.2019.05.006

65. Gorton HC, Webb RT, Kapur N, Ashcroft DM. Non-psychotropic medication and risk of suicide or attempted suicide: a systematic review. *BMJ Open* (2016) 6:e009074. doi: 10.1136/bmjopen-2015-009074

66. Gupta A, Chadda RK. Adverse psychiatric effects of non-psychotropic medications. *BJPsych Adv* (2016) 22:325–34. doi: 10.1192/apt.bp.115.015735

67. US Food and Drug Administration. (2019). *FDALabel: full-text search of drug labeling*. US Food and Drug Administration. [Accessed November 30, 2019].

68. Kessler RC, Stein MB, Petukhova MV, Bliese P, Bossarte RM, Bromet EJ, et al. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry* (2017) 22:544–51. doi: 10.1038/mp.2016.110

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

70. Kessler RC, Bossarte RM, Luedtke A, Zaslavsky AM, Zubizarreta JR. Machine learning methods for developing precision treatment rules with observational data. *Behav Res Ther* (2019) 120:103412. doi: 10.1016/j.brat.2019.103412

71. Kessler RC, Bernecker SL, Bossarte RM, Luedtke AR, McCarthy JF, Nock MK, et al. "The Role of Big Data Analytics in Predicting Suicide." In: Passos IC, Mwangi B, Kapczinski, editors. *Personalized Psychiatry: Big Data Analytics in Mental Health*. Cham, Switzerland: Springer Nature Switzerland (2019). p. 77–98.

72. 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. Available from: https://cran.r-project.org/web/packages/SuperLearner/index.html.

73. LeDell E, van der Laan MJ, Petersen M. AUC-maximizing ensembles through metalearning. *Int J Biostat* (2016) 12:203–18. doi: 10.1515/ijb-2015-0035

74. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* (2010) 33:1–22. doi: 10.18637/jss.v033.i01

75. Breiman L. Random forests. *Mach Learn* (2001) 45:5–32. doi: 10.1023/A:1010933404324

76. Chipman HA, McCulloch RE. (2016). *BayesTree: Bayesian Additive Regression Trees*. R package version 0.3-1.4. The Comprehensive R Archive Network. Available at: http://CRAN.R-project.org/package=BayesTree [Accessed December 9, 2019].

77. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *arXiv:* (2016) 1603:02754. doi: 10.1145/2939672.2939785

78. Steinwart I, Christmann A. *Support vector machines*. New York, NY: Springer (2008). p. 466.

79. Venables WN, Ripley BD. *Modern Applied Statistics with S. 4th*. New York, NY: Springer (2002). p. 498.

80. Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electrical Eng* (2014) 40:16–28. doi: 10.1016/j.compeleceng.2013.11.024

81. Kuhn M. (2019). *The caret package*. Max Kuhn. Available at: https://topepo.github.io/caret/index.html [Accessed December 2, 2019].

82. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowledge Data Eng* (2009) 21:1263–84. doi: 10.1109/TKDE.2008.239

83. Chawla NV. Data mining for imbalanced datasets: An overview. In: Maimon O, Rokach L, editors. *Data Mining and Knowledge Discovery Handbook, 2nd*. Berlin/Heidelberg, Germany: Springer (2010). p. 875–86.

84. Lee PH. Resampling methods improve the predictive power of modeling in class-imbalanced datasets. *Int J Environ Res Public Health* (2014) 11:9776–89. doi: 10.3390/ijerph110909776

85. Rahman MM, Davis DN. Addressing the class imbalance problem in medical datasets. *Int J Mach Learn Comput* (2013) 3:224. doi: 10.7763/IJMLC.2013.V3.307

86. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist* (2001) 29:1189–232. doi: 10.1214/aos/1013203451

87. US Department of Veterans Affairs. (2019). *Patient Record Flags (PRF) User Guide*. US Department of Veterans Affairs. [Accessed December 5, 2019].

88. 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. doi: 10.2105/AJPH.2015.302737

89. 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. doi: 10.1371/journal.pone.0089944

90. Walsh G, Sara G, Ryan CJ, Large M. Meta-analysis of suicide rates among psychiatric in-patients. *Acta Psychiatr Scand* (2015) 131:174–84. doi: 10.1111/acps.12383

91. 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. doi: 10.1136/bmjopen-2018-023883

92. 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. doi: 10.1002/wps.20321

93. Houlihan J, Leffler S. Assessing and addressing social determinants of health: a key competency for succeeding in value-based care. *Prim Care* (2019) 46:561–74. doi: 10.1016/j.pop.2019.07.013

94. Coughlin SS. Social determinants of breast cancer risk, stage and, survival. *Breast Cancer Res Treat* (2019) 177:537–48. doi: 10.1007/s10549-019-05340-7

95. Sokol R, Austin A, Chandler C, Byrum E, Bousquette J, Lancaster C, et al. Screening children for social determinants of health: a systematic review. *Pediatrics* (2019) 144:e20191622 doi: 10.1542/peds.2019-1622

96. Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensern PS, Kupfer DJ. Coming to terms with the terms of risk. *Arch Gen Psychiatry* (1997) 54:337–43. doi: 10.1001/archpsyc.1997.01830160065009

97. 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. doi: 10.1176/appi.ajp.2018.17101167

98. 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. doi: 10.1176/appi.ajp.2016.16010077

99. Bolton JM. Suicide risk assessment in the emergency department: out of the darkness. *Depress Anxiety* (2015) 32:73–5. doi: 10.1002/da.22320

100. Hoge CW. Suicide reduction and research efforts in service members and veterans-sobering realities. *JAMA Psychiatry* (2019) 76:464–6. doi: 10.1001/jamapsychiatry.2018.4564

101. Mulder R, Newton-Howes G, Coid JW. The futility of risk prediction in psychiatry. *Br J Psychiatry* (2016) 209:271–2. doi: 10.1192/bjp.bp.116.184960

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

103. 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. doi: 10.1017/S0033291717002537

104. National Institute for Health and Care Excellence (NICE). (2011). *Self-harm in over 8s: long-term management*. National Institute for Health and Care Excellence. [Accessed April 30, 2019].

105. Bernert RA, Hom MA, Roberts LW. 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. doi: 10.1007/s40596-014-0180-1

106. 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. doi: 10.1176/appi.ajp.2015.1720501

107. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science* (1989) 243:1668–74. doi: 10.1126/science.2648573

108. Ægisdóttir 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. doi: 10.1177/0011000005285875

109. 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. doi: 10.1111/sltb.12395

110. US Public Health Service. *National strategy for suicide prevention: goals and objectives for action*. Washington, DC: US Department of Health and Human Services (2012) (2012). [Accessed November 12, 2019].

111. 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. doi: 10.3389/fpsyt.2018.00033

112. Jacobs DG. (2009). *Suicide Assessment Five-step Evaluation and Triage for Mental Health Professionals*. Education Development Center Inc. [Accessed December 10, 2019].

113. Simon R. "Improving suicide risk assessment with evidence-based psychiatry." In: Pompoli M, Taterelli R, editors. *Evidence-Based Practice in Suicidology: A Sourcebook*. Cambridge, MA: Hogrefe Publishing (2011). p. 45–54.

114. 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–S45. doi: 10.1016/j.jacc.2013.11.002

115. 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. doi: 10.1002/jts.22058

116. Bejan CA, Angiolillo J, Conway D, Nash R, Shirey-Rice JK, Lipworth L, et al. Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. *J Am Med Inform Assoc* (2018) 25:61–71. doi: 10.1093/jamia/ocx059

117. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using clinical notes and natural language processing for automated HIV risk assessment. *J Acquir Immune Defic Syndr* (2018) 77:160–6. doi: 10.1097/QAI.0000000000001580

118. Ehrenfeld JM, Gottlieb KG, Beach LB, Monahan SE, Fabbri D. Development of a natural language processing algorithm to identify and evaluate transgender patients in electronic health record systems. *Ethn Dis* (2019) 29:441–50. doi: 10.18865/ed.29.S2.441

119. Hatef E, Predmore Z, Lasser EC, Kharrazi H, Nelson K, Curtis I, et al. Integrating social and behavioral determinants of health into patient care and population health at Veterans Health Administration: a conceptual framework and an assessment of available individual and population level data sources and evidence-based measurements. *AIMS Public Health* (2019) 6:209–24. doi: 10.3934/publichealth.2019.3.209

120. Hatef E, Rouhizadeh M, Tia I, Lasser E, Hill-Briggs F, Marsteller J, et al. Assessing the availability of data on social and behavioral determinants in structured and unstructured electronic health records: a retrospective analysis of a multilevel health care system. *JMIR Med Inform* (2019) 7:e13802. doi: 10.2196/13802

121. Vest JR, Grannis SJ, Haut DP, Halverson PK, Menachemi N. Using structured and unstructured data to identify patients' need for services that address the social determinants of health. *Int J Med Inform* (2017) 107:101–6. doi: 10.1016/j.ijmedinf.2017.09.008

122. Björkenstam C, Kosidou K, Björkenstam E. Childhood adversity and risk of suicide: cohort study of 548 721 adolescents and young adults in Sweden. *BMJ* (2017) 357:j1334. doi: 10.1136/bmj.j1334

123. Miranda-Mendizábal A, Castellví P, Parés-Badell O, Almenara J, Alonso I, Blasco MJ, et al. Sexual orientation and suicidal behaviour in adolescents and young adults: systematic review and meta-analysis. *Br J Psychiatry* (2017) 211:77–87. doi: 10.1192/bjp.bp.116.196345

124. LexisNexis. (2019). *LexisNexis: Solutions for professionals who shape the world*. LexisNexis. [Accessed January 2, 2019].

125. Kessler RC, Chalker SA, Luedtke AR, Sadikova E, Jobes DA. A preliminary precision treatment rule for remission of suicide ideation. *Suicide Life Threat Behav* (2019). 50:558–72. doi: 10.1111/sltb.12609

126. Wu C-S, Luedtk AR, Sadikova E, Tsai H-J, Liao S-C, Liu C-C, et al. Development and validation of a machine learning individualized treatment rule for patients with first-episode schizophrenia. *JAMA Netw Open* (2020) 3:e1921660. doi: 10.1001/jamanetworkopen.2019.21660

127. Steeg S, Emsley R, Carr M, Cooper J, Kapur N. Routine hospital management of self-harm and risk of further self-harm: propensity score analysis using record-based cohort data. *Psychol Med* (2018) 48:315–26. doi: 10.1017/S0033291717001702

128. 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. doi: 10.1371/journal.pone.0204670

129. Large MM, Kapur N. Psychiatric hospitalisation and the risk of suicide. *Br J Psychiatry* (2018) 212:269–73. doi: 10.1192/bjp.2018.22

130. Luedtke A, Sadikova E, Kessler RC. Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clin Psychol Sci* (2019) 7:445–61. doi: 10.1177/2167702618815466

131. Hirshberg DA, Zubizarreta JR. On two approaches to weighting in causal inference. *Epidemiology* (2017) 28:812–6. doi: 10.1097/EDE.0000000000000735

132. Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. *J Am Stat Assoc* (2015) 110:910–22. doi: 10.1080/01621459.2015.1023805

133. Dahabreh IJ, Sheldrick RC, Paulus JK, Chung M, Varvarigou V, Jafri H, et al. Do observational studies using propensity score methods agree with randomized trials? a systematic comparison of studies on acute coronary syndromes. *Eur Heart J* (2012) 33:1893–901. doi: 10.1093/eurheartj/ehs114

134. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* (2014) 4:MR000034. doi: 10.1002/14651858.MR000034.pub2

135. Zhu R, Zhao Y-Q, Chen G, Ma S, Zhao H. Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics* (2017) 73:391–400. doi: 10.1111/biom.12593

136. Zhou X, Mayer-Hamblett N, Khan U, Kosorok MR. Residual weighted learning for estimating individualized treatment rules. *J Am Stat Assoc* (2017) 112:169–87. doi: 10.1080/01621459.2015.1093947

137. Luedtke AR, van der Laan MJ. Evaluating the impact of treating the optimal subgroup. *Stat Methods Med Res* (2017) 26:1630–40. doi: 10.1177/0962280217708664

138. Bernecker SL, Zuromski KL, Curry JC, Kim JJ, Gutierrez PM, Joiner TE, et al. Economic evaluation of brief cognitive behavioral therapy vs treatment as usual for suicidal US army soldiers. *JAMA Psychiatry* (2019). 77:256–64. doi: 10.1001/jamapsychiatry.2019.3639