

# Evaluation of a Model to Target High-risk Psychiatric Inpatients for an Intensive Postdischarge Suicide Prevention Intervention

Ronald C. Kessler, PhD; Mark S. Bauer, MD; Todd M. Bishop, PhD; Robert M. Bossarte, PhD; Victor M. Castro, MS; Olga V. Demler, PhD; Sarah M. Gildea, BS; Joseph L. Goulet, PhD; Andrew J. King, MS; Chris J. Kennedy, PhD; Sara J. Landes, PhD; Howard Liu, SD; Alex Luedtke, PhD; Patrick Mair, PhD; Brian P. Marx, PhD; Matthew K. Nock, PhD; Maria V. Petukhova, PhD; Wilfred R. Pigeon, PhD; Nancy A. Sampson, BA; Jordan W. Smoller, MD, ScD; Aletha Miller, PhD; Gretchen Haas, PhD; Jeffrey Benware, PhD; John Bradley, MD; Richard R. Owen, MD; Samuel House, MD; Snezana Urošević, PhD; Lauren M. Weinstock, PhD

[+ Supplemental content](#)

**IMPORTANCE** The months after psychiatric hospital discharge are a time of high risk for suicide. Intensive postdischarge case management, although potentially effective in suicide prevention, is likely to be cost-effective only if targeted at high-risk patients. A previously developed machine learning (ML) model showed that postdischarge suicides can be predicted from electronic health records and geospatial data, but it is unknown if prediction could be improved by adding additional information.

**OBJECTIVE** To determine whether model prediction could be improved by adding information extracted from clinical notes and public records.

**DESIGN, SETTING, AND PARTICIPANTS** Models were trained to predict suicides in the 12 months after Veterans Health Administration (VHA) short-term (less than 365 days) psychiatric hospitalizations between the beginning of 2010 and September 1, 2012 (299 050 hospitalizations, with 916 hospitalizations followed within 12 months by suicides) and tested in the hospitalizations from September 2, 2012, to December 31, 2013 (149 738 hospitalizations, with 393 hospitalizations followed within 12 months by suicides). Validation focused on net benefit across a range of plausible decision thresholds. Predictor importance was assessed with Shapley additive explanations (SHAP) values. Data were analyzed from January to August 2022.

**MAIN OUTCOMES AND MEASURES** Suicides were defined by the National Death Index. Base model predictors included VHA electronic health records and patient residential data. The expanded predictors came from natural language processing (NLP) of clinical notes and a social determinants of health (SDOH) public records database.

**RESULTS** The model included 448 788 unique hospitalizations. Net benefit over risk horizons between 3 and 12 months was generally highest for the model that included both NLP and SDOH predictors (area under the receiver operating characteristic curve range, 0.747-0.780; area under the precision recall curve relative to the suicide rate range, 3.87-5.75). NLP and SDOH predictors also had the highest predictor class-level SHAP values (proportional SHAP = 64.0% and 49.3%, respectively), although the single highest positive variable-level SHAP value was for a count of medications classified by the US Food and Drug Administration as increasing suicide risk prescribed the year before hospitalization (proportional SHAP = 15.0%).

**CONCLUSIONS AND RELEVANCE** In this study, clinical notes and public records were found to improve ML model prediction of suicide after psychiatric hospitalization. The model had positive net benefit over 3-month to 12-month risk horizons for plausible decision thresholds. Although caution is needed in inferring causality based on predictor importance, several key predictors have potential intervention implications that should be investigated in future studies.

JAMA Psychiatry. 2023;80(3):230-240. doi:10.1001/jamapsychiatry.2022.4634  
Published online January 18, 2023.

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Corresponding Author:** Ronald C. Kessler, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Ste 215, Boston, MA 02115-5899 (kessler@hcp.med.harvard.edu).

Suicide is the 10th leading cause of death in the US.<sup>1</sup> The highest concentration is among the 1% of the population with a past-year psychiatric hospitalization,<sup>2</sup> who account for 14% of all suicides.<sup>3</sup> Intensive case management might reduce these postdischarge suicides<sup>4-6</sup> but would be cost-effective only if targeted at high-risk patients. A machine learning (ML) model based on electronic health records and geospatial data was developed for this purpose in the Veterans Health Administration (VHA).<sup>7</sup> Based on the strength of that model, an experimental trial was funded for suicide prevention with intensive postdischarge case management of high-risk patients.<sup>8</sup> However, it was decided that further study was needed before implementing the trial to determine if model accuracy could be improved by adding 2 types of predictors found to be important in other suicide prediction models: (1) predictors extracted from clinical notes using natural language processing (NLP) methods<sup>9</sup> and (2) indicators of patient-level social determinants of health (SDOH) extracted from public records.<sup>10</sup> Results of this investigation are presented here along with an analysis of the net benefit (NB) of the expanded model across a range of plausible decision thresholds.

## Methods

### Sample

The models were trained in the 299 050 short-term (less than 365 days) VHA psychiatric hospitalizations that occurred between January 1, 2010, and September 1, 2012 (916 hospitalizations followed within 12 months by a suicide) and tested in a prospective validation sample of the 149 738 short-term VHA psychiatric hospitalizations that occurred between September 2, 2012, and December 31, 2013 (393 hospitalizations followed within 12 months by a suicide). The primary outcome, suicide within 12 months postdischarge, was identified in the National Death Index (NDI).<sup>11</sup> The study protocol was approved by the Research Ethics Committees of the Veterans Administration Center of Excellence for Suicide Prevention and Harvard Medical School with a waiver of informed consent based on data being deidentified. The study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline<sup>12</sup> for reporting analyses designed to develop and validate predictive models.

### Data Sources

The predictors in the original model came from 3 sources<sup>7</sup>: (1) the VHA Corporate Data Warehouse (CDW)<sup>13</sup> (eTable 1 in Supplement 1); (2) the Veterans Administration (VA) Suicide Prevention Applications Network for reporting of suicidal behaviors<sup>14</sup>; and (3) a geospatial SDOH database assembled from diverse government sources (eTable 2 in Supplement 1) about plausible predictors of suicide<sup>15</sup> in the patient's residential neighborhood (ie, block group and census tracts), county, and state. Two new databases were added for the expanded models: (4) a consolidated free text file of VHA clinical notes for all inpatient and outpatient visits in the 12 months before

## Key Points

**Question** Can prediction of posthospitalization suicides be improved by adding information from clinical notes and public records to a standard machine learning model?

**Findings** In this prognostic study including 448 788 hospitalizations, performance of a machine learning model to predict suicides after Veteran Health Administration psychiatric hospital discharges improved significantly when including predictors extracted from clinical notes with natural language processing and from a social determinants of health public records database. The resulting model had positive net benefit for plausible decision thresholds over risk horizons between 3 and 12 months.

**Meaning** In this study, clinical notes and public records improved machine learning model prediction of suicides after psychiatric hospital discharges.

and including hospitalization<sup>16</sup>; and (5) the LexisNexis SDOH database of public records (eTable 3 in Supplement 1) as of the month before hospitalization.<sup>17</sup> Demographic information, including race and ethnicity information, were taken from the VHA VistA Patient file stored in the CDW. Both self-reported and observer-reported race and ethnicity are recorded in this file. Self-reports were taken when available, and observer reports were only taken when self-reports were missing. Data were missing for no more than 3% of records and only in the geospatial SDOH database and LexisNexis SDOH database. In most of these cases, records were nonmissing in earlier records or, in the case of the geospatial SDOH database, in contiguous areas, allowing nearest-neighbor imputations. Remaining missing values and inconsistencies were reconciled using rational imputations and median value imputations.

### Predictors

Four broad predictor classes, each somewhat expanding the predictors in the earlier model, included: (1) psychopathological risk factors (diagnoses, treatments, suicidality), including information about these variables during the hospitalization; (2) physical disorders and treatments and counts of prescribed medications classified by the US Food and Drug Administration (FDA) as increasing suicide risk (eTable 4 in Supplement 1); (3) facility-level quality indicators (eg, inpatient staff turnover rate) over the 12 months prior to the date of the hospitalization; and (4) indicators of SDOH at both the patient level (from *International Classification of Diseases, Ninth Revision, Clinical Modification* [ICD-9-CM]/ICD-10-CM codes and sociodemographic information) and the geospatial level as of the month before the hospitalization. These variables were selected based on review of earlier research on predictors of suicide after psychiatric hospital discharge<sup>18-21</sup> and in more general patient samples.<sup>22-24</sup> Details about constructs and assessments are presented in the eMethods in Supplement 1. A total of 7902 predictors were extracted from these databases. An additional 1837 predictors were extracted from clinical notes using NLP methods (eMethods in Supplement 1). The LexisNexis SDOH database included an additional 442 predictors. Categorical predictors in all databases were 1-hot encoded as

0-1 dummy variables. Ordinal, interval, and ratio variables were standardized to a mean of 0 and variance of 1, with values more than 3 SDs above or below mean truncated to 3 or -3 to minimize the effects of outliers.

### Statistical Analysis

Analysis was carried out January to August 2022. The Super-Learner (SL) stacked generalization ML method<sup>25</sup> was used to pool results across an ensemble of classifiers (eTable 5 in Supplement 1) to generate optimally weighted composite predictions.<sup>26</sup> Four SL models were trained: (1) the base model, including only predictors from the first 3 databases, and 3 models with additional predictors beyond those in the base model, including (2) the NLP model, (3) the LexisNexis SDOH model, and (4) the combined model, which included the important predictors from the base, NLP, and LexisNexis SDOH models (eMethods in Supplement 1). A lasso model (ie, sparse penalized logistic regression),<sup>27</sup> estimated as a simple benchmark for the more complex stacked generalization models, included the same predictors as the base SL model.

The models were trained to predict suicides over a 12-month risk horizon but were also used in the validation sample to predict suicides over shorter risk horizons based on a prior finding that models predicting postdischarge suicides over a 12-month risk horizon outperform models developed over shorter risk horizons when applied at those shorter horizons.<sup>7</sup> This occurred because the larger number of suicides over 12 months increased statistical power to detect meaningful associations that were comparable over shorter horizons. Meta-learner weights for the learners used to estimate the Super-Learner prediction of the combined model are presented in eTable 9 of Supplement 1. A 2-sided *t* test was used to determine the significance of the 12-month suicide rate between the training and validation samples. One-sided  $\chi^2$  tests were used to determine if AUCs were significantly greater than 0.5. All significance tests were evaluated at  $\alpha = .05$ .

### Model Interpretation

#### Model Performance

Model performance focused on NB, a widely used summary measure of clinical model performance.<sup>28</sup> NB describes the relative benefits (true-positives) vs costs (false-positives) of intervening at a given decision threshold to determine how high predicted suicide risk must be to define a patient as warranting intervention. NB at decision threshold *p* is defined as the ratio of the observed number of true-positives above that threshold to a discounted number of false-positives above the threshold, with discounting defined by the *p/q* break-even point implied when the decision threshold is set at *p*, where  $q = 1 - p$ . Importantly, NB can be highest for a given model at one decision threshold but not others, making the decision threshold of interest explicit when comparing models. Absence of consensus on the minimum risk needed to justify implementing an aggressive suicide prevention intervention makes it impossible to select a single decision threshold for the NB analysis. Presenting results across a plausible range of thresholds is recommended in cases where such uncertainty exists.<sup>29,30</sup> We con-

sidered decision thresholds between roughly half the 12-month suicide rate and 2-fold that rate.

To facilitate comparison across models, NB was divided by the observed suicide rate for the horizon to define standardized NB (SNB), which has an upper bound of 1.0 for each horizon. Although standardized NB usefully combines information about discrimination and calibration, the 2 model performance metrics typically considered in evaluating prediction models, thereby addressing the uncertainty that otherwise arises when one model has better discrimination and another better calibration, we also report comparative model discrimination and calibration. Discrimination was assessed by area under the receiver operating characteristic curve (AUROC) and area under the precision recall curve (AUPRC). Given the low prevalence of suicide, AUPRC was expressed relative to the suicide rate, which is the expected value of AUPRC under a null model. Calibration was assessed with the integrated calibration index (ICI),<sup>31</sup> the weighted mean absolute value difference between observed and predicted probabilities of suicide, expressed relative to the suicide rate.

Although model performance was evaluated in the 30% prospective validation sample, discrimination and calibration were also calculated using 10-fold cross-validation in the total sample to provide a valid comparison with the approach used in most prior ML suicide prediction studies. Estimates based on 10-fold cross-validation are typically a good deal higher than estimates based on prospective validation.<sup>32</sup> Given that some patients had multiple hospitalizations, the cross-validation folds were constructed at the patient level so that all hospitalizations for a specific patient were in the same fold.

#### Predictor Importance

Predictor importance was examined using the model-agnostic kernel Shapley additive explanations (SHAP) method.<sup>33</sup> This method estimates the effect of changing a predictor from its observed score to the sample mean averaged across all logically possible permutations of other predictors. The mean of this SHAP value for a given predictor across all hospitalizations is 0. However, the mean absolute SHAP value provides useful information about the average importance of the predictor. A bee swarm plot of individual-level observed predictor scores by SHAP values shows the dominant direction of association. Proportional mean absolute SHAP values were calculated by dividing mean absolute SHAP values of classes and important predictors within classes by the mean absolute SHAP value of the entire model (eMethods in Supplement 1).

SAS version 9.4 (SAS Institute) was used for data cleaning, imputation, management, estimating suicide prevalence, and calculating AUROC, AUPRC, ICI, and standardized NB. R version 4.0.5 (The R Foundation) was used for NLP feature extraction and to estimate SL models and SHAP values.

## Results

### Sociodemographic and Military Career Variable Distributions

Most training and validation sample patients were male (280 901 of 299 050 [93.9%] and 139 630 of 149 738 [93.2%],



respectively), with a median (IQR) age of 55 (16) and 55 (18) years. Most patients were non-Hispanic White (training sample, 187 745 [62.8%]; validation sample, 92 067 [61.5%]), followed by non-Hispanic Black (training sample, 72 315 [24.2%]; validation sample, 36 205 [24.2%]), Hispanic (training sample, 23 002 [7.7%]; validation sample, 11 687 [7.8%]), and other race (including American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander races, responses of do not know and any other race, and missing data; training sample, 15 987 [5.3%]; validation sample, 9780 [6.5%]) (eTable 6 in Supplement 1). Most served in the Vietnam War era (training sample, 120 365 of 299 050 [40.2%]; validation sample, 53 790 of 149 738 [35.9%]).

### Outcome Distribution

The 448 788 short-term VHA psychiatric hospitalizations from 2010 to 2013 involved 227 801 patients, 835 of whom died by suicide within 12 months of discharge from 1309 psychiatric hospitalizations. The 12-month suicide rate at the level of the hospitalization was significantly higher in the training sample ( $n = 916$ ; mean [SE] rate, 306.3 [10.1] per 100 000 hospitalizations) than the validation sample ( $n = 393$ ; mean [SE] rate, 262.5 [13.2] per 100 000 hospitalizations;  $t = -2.6$ ;  $P = .008$ ). The mean (SE) suicide rate decreased with time since discharge in both the training and validation samples, from highs of 1081.0 (137.3) per 100 000 hospitalization-years and 1149.2 (200.0) per 100 000 hospitalization-years, respectively, in the first week after discharge to lows of 245.9 (13.0) per 100 000 hospitalization-years and 198.3 (16.5) per 100 000 hospitalization-years 7 to 12 months after discharge (training sample,  $\chi^2_{11} = 145.0$ ; validation sample,  $\chi^2_{11} = 61.8$ ;  $P < .001$ ) (Figure 1; eTable 7 in Supplement 1). The reasons for these decreases are unclear but illustrate the importance of using a prospective validation sample rather than a random 30% subsample selected across all study years to avoid overestimating prospective model prediction accuracy.

### Model Performance

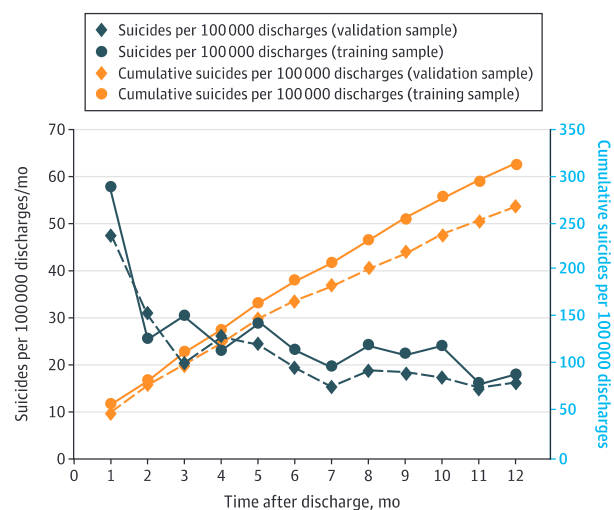
#### Discrimination and Calibration

Validation sample AUROC was in the range traditionally considered acceptable in all models other than the lasso model (AUROC range, 0.721-0.784) and poor in the lasso model (AUROC range, 0.642-0.668).<sup>34</sup> AUROC was generally highest in the combined model (range, 0.747-0.780) (Table 1). AUPRC relative to the suicide rate was 3-fold to 6-fold higher for a random classifier in all models other than the lasso model and 1.5-fold to 2-fold higher in the lasso model. Model discrimination was substantially higher when based on 10-fold cross-validation in the total sample (AUROC range, 0.824-0.883; AUPRC relative to the suicide rate range, 49.5-81.6) (eTable 8 in Supplement 1). Model calibration was best for the lasso model (ICI relative to the suicide rate range, 0.215-0.269) (Table 1).

#### Standardized NB

The proportion of hospitalizations screened in at the optimal standardized NB decreased as the decision threshold increased and as the risk horizon decreased, from highs of

**Figure 1. Monthly Suicide Hazard Rates and Cumulative Incidence Rates Over the 12 Months After Psychiatric Hospital Discharge in the Training Sample (January 1, 2010, to September 1, 2012) and Prospective Validation Sample (September 2, 2012, to December 31, 2013)**



29.9% to 35.2% (capturing 69.2% to 71.4% of true suicides) at the lowest threshold to lows of 0.7% to 4.4% (capturing 6.8% to 23.7% of true suicides) at the highest threshold. Although the base model consistently had the highest standardized NB at the lowest decision threshold (Table 2), the combined model had the highest standardized NB at all higher decision thresholds for 6-month and 12-month risk horizons. Less consistency occurred for the 3-month horizon. All models had negative standardized NB for horizons less than 3 months (results not reported), implying that the number of true-positives was so small relative to the number of false-positives that the benefit of detecting and intervening to prevent suicides was outweighed by the cost of intervening with more than  $p/q$  false-positives per suicide.

### Predictor Importance

Significant cross-validated univariable associations (area under the curve greater than 0.50) with 12-month suicide in the training sample were found for 1666 base model predictors (21.1%), 442 NLP predictors (24.1%), and 29 LexisNexis SDOH predictors (6.6%) (Table 3). These predictors overlapped substantially, resulting in 80% of total model proportional SHAP being captured by the 100 most important predictors in the base model and 110 most important predictors in the NLP model. Only those most important predictors and the significant LexisNexis SDOH predictors (191 after excluding duplicates across models) were used to estimate the combined model.

Class-level proportional SHAP for these important predictors in the combined model were highest for the NLP (64.0%) and SDOH (49.3%) predictors and lowest for facility-level quality indicators (11.4%) (Figure 2). These proportional SHAP values sum to more than 100% because most patients had some characteristics associated with increased suicide risk and others associated with decreased suicide risk. The key NLP pre-

**Table 1. Estimated Area Under the Receiver Operating Characteristic Curve (AUROC) and Area Under the Precision Recall Curve (AUPRC) Relative to the Suicide Rate of the 5 Models Over a Range of Risk Horizons in the Prospective Validation Sample**

Risk horizon	Estimate (SE)				
	Lasso model	Base model	NLP model	SDOH model	Combined model
<b>AUROC</b>					
1 wk	0.642 (0.271)	0.770 (0.274)	0.721 (0.297)	0.775 (0.266)	0.777 (0.259)
1 mo	0.655 (0.146)	0.784 (0.132) <sup>a</sup>	0.756 (0.144) <sup>a</sup>	0.781 (0.132) <sup>a</sup>	0.780 (0.142) <sup>a</sup>
3 mo	0.657 (0.097)	0.763 (0.087) <sup>a</sup>	0.730 (0.097) <sup>a</sup>	0.756 (0.092) <sup>a</sup>	0.774 (0.089) <sup>a</sup>
6 mo	0.668 (0.073) <sup>a</sup>	0.744 (0.070) <sup>a</sup>	0.726 (0.072) <sup>a</sup>	0.743 (0.071) <sup>a</sup>	0.757 (0.071) <sup>a</sup>
12 mo	0.665 (0.059) <sup>a</sup>	0.728 (0.054) <sup>a</sup>	0.723 (0.057) <sup>a</sup>	0.726 (0.057) <sup>a</sup>	0.747 (0.057) <sup>a</sup>
<b>AUPRC relative to the suicide rate</b>					
1 wk	1.556 (0.082) <sup>a</sup>	3.926 (0.393) <sup>a</sup>	3.489 (0.380) <sup>a</sup>	4.038 (0.415) <sup>a</sup>	4.955 (0.807) <sup>a</sup>
1 mo	1.675 (0.077) <sup>a</sup>	5.026 (0.426) <sup>a</sup>	3.757 (0.224) <sup>a</sup>	4.996 (0.385) <sup>a</sup>	4.342 (0.306) <sup>a</sup>
3 mo	1.756 (0.050) <sup>a</sup>	3.671 (0.144) <sup>a</sup>	3.285 (0.123) <sup>a</sup>	3.676 (0.143) <sup>a</sup>	3.837 (0.130) <sup>a</sup>
6 mo	1.796 (0.032) <sup>a</sup>	3.750 (0.146) <sup>a</sup>	3.688 (0.435) <sup>a</sup>	6.398 (0.680) <sup>a</sup>	4.170 (0.242) <sup>a</sup>
12 mo	1.859 (0.028) <sup>a</sup>	3.281 (0.121) <sup>a</sup>	3.189 (0.201) <sup>a</sup>	5.342 (0.339) <sup>a</sup>	4.897 (0.293) <sup>a</sup>
<b>ICI relative to the suicide rate<sup>b</sup></b>					
Logistic	0.269	0.712	0.587	0.902	0.600
Isotonic	0.215	0.866	0.431	1.401	0.629

Abbreviations: ICI, integrated calibration index; NLP, natural language processing; SDOH, social determinants of health.

<sup>a</sup> Significantly higher than expected by chance at  $P < .05$ ; 1-sided test when compared with a null model.

<sup>b</sup> ICI relative to the suicide rate was estimated only for the 12-month models, as the predicted values from those models were used to estimate prevalence over all shorter risk horizons.

dictors (ie, the 5 highest proportional SHAP values) included counts of the terms *suicide attempt* and *suicide*, both positively associated with suicide risk, and counts of several topic clusters (2 of them including the word *suicide*), all negatively associated with suicide risk. Component proportional SHAP for SDOH was higher for LexisNexis (proportional SHAP = 48.9%), geospatial (proportional SHAP = 48.4%), and sociodemographic (proportional SHAP = 31.7%) variables than patient-level ICD-9-CM/ICD-10-CM codes (proportional SHAP = 8.4%). The key LexisNexis predictors included 4 variables negatively associated with suicide risk (high identity fraud risk, derogatory public records, number of household residents, and number of relative/associate households on record) and 1 positively associated (patient income). Recent income changes were considered but were not key predictors. The key geospatial predictors, although involving disadvantage either at the county or census tract levels, all had negative associations with suicide risk. The key sociodemographics included being Non-Hispanic White and most recently in service after the September 11, 2001, attack (both positively associated with suicide risk) and age, being Non-Hispanic Black, and being Baptist (all negatively associated with suicide risk).

Interestingly, physical disorders had larger proportional SHAP values (40.3%) than psychopathological risk factors (34.2%). However, by far the single most important predictor in the physical disorder class was a count of medications classified by the FDA as increasing suicide risk prescribed in the year before hospitalization. Four key psychopathological predictors were positively associated with suicide risk: suicidality diagnosed in the past year, number of suicide attempts in the past 2 years, recent diagnoses of recurrent major depression, and recent antidepressant prescriptions. The fifth key psychopathological predictor was negatively associated with suicide risk: number of visits for tobacco-related disorders in the past 2 years. Facility-level quality indicators had the lowest effect (proportional SHAP = 11.4%), with the key predic-

tors including past year staff turnover rate and ratio and driving time to the nearest VHA primary and tertiary care facilities.

## Discussion

The objective of this study was to evaluate the effects of adding predictors from clinical notes extracted using NLP and from public records to a previously developed ML model to improve model performance before using the model in an experimental intensive postdischarge suicide prevention intervention. Results showed clearly that both types of predictors improved model prediction accuracy. We also carried out an analysis that documented a positive NB of the preferred model for plausible decision thresholds over risk horizons between 3 and 12 months. Although numerous prior studies used ML methods to predict suicides from electronic health records,<sup>35-37</sup> suicides after hospital discharges were only seldom the focus,<sup>38</sup> making it difficult to draw direct comparisons between our results and those of prior studies.

Our finding of positive NB speaks directly to the criticism that suicide prediction models are not actionable given the rarity of suicides among high-risk patients.<sup>39-43</sup> As we<sup>7,44</sup> and others<sup>45,46</sup> have noted elsewhere, this criticism is misplaced because it focuses on the rarity of the outcome, whereas a more appropriate focus is NB. The positive NB we found is broadly consistent with findings in other areas of medicine that interventions for relatively rare but severe outcomes are actionable when the benefits of preventing the outcome far exceed the costs of intervention. For example, statins are recommended for adults aged 40 to 75 years with mildly elevated total cholesterol levels even though prevalence of extreme cardiovascular events over a 12-month risk horizon at the intervention decision threshold specified for statin use is no higher than prevalence of suicide at the deci-



Table 2. Estimated Net Benefit (NB) of the 5 Models Over a Range of Decision Thresholds and Risk Horizons in the Prospective Validation Sample

Decision threshold <sup>a</sup>	%								
	12 mo			6 mo			3 mo		
	Total sample	True-positives	Standardized NB	Total sample	True-positives	Standardized NB	Total sample	True-positives	Standardized NB
<b>150/100 000</b>									
Treat all patients	100	100	42.9	100	100	8.3	100	100	-52.9
Lasso model	60.5 <sup>b</sup>	83.0 <sup>b</sup>	48.4 <sup>b</sup>	60.5 <sup>b</sup>	84.9 <sup>b</sup>	29.5 <sup>b</sup>	60.5 <sup>b</sup>	82.3 <sup>b</sup>	-10.1 <sup>b</sup>
Base model	35.2	69.2	49.2	35.2	72.2	40.1	29.9 <sup>b</sup>	71.4 <sup>b</sup>	25.9 <sup>b</sup>
NLP model	30.4	61.8	44.5	30.4	61.2	33.4	30.4	62.6	16.2
SDOH model	100 <sup>b</sup>	100 <sup>b</sup>	42.9 <sup>b</sup>	23.8	56.7	35.0	23.8	60.5	24.2
Combined model	100 <sup>b</sup>	100 <sup>b</sup>	42.9 <sup>b</sup>	17.6	51.8	35.8	17.6	52.4	25.5
<b>200/100 000</b>									
Treat all patients	100	100	23.8	100	100	-22.3	100	100	-103.9
Lasso model	60.6	83.0	36.8	36.2 <sup>b</sup>	56.3 <sup>b</sup>	12.2 <sup>b</sup>	36.2 <sup>b</sup>	54.4 <sup>b</sup>	-19.3 <sup>b</sup>
Base model	29.4	61.8	39.5	29.4	65.3	29.4	17.9 <sup>b</sup>	49.0 <sup>b</sup>	12.5 <sup>b</sup>
NLP model	30.4	61.8	38.8	30.4	61.2	24.2	30.4	62.6	0.7
SDOH model	23.0	53.4	36.0	23.0	56.3	28.3	23.0	59.9	13.0
Combined model	40.7 <sup>b</sup>	75.6 <sup>b</sup>	44.7 <sup>b</sup>	17.6	51.8	30.4	17.6	52.4	16.6
<b>300/100 000</b>									
Treat all patients	100	100	-14.3	100	100	-83.6	100	100	-206.2
Lasso model	15.8 <sup>b</sup>	35.1 <sup>b</sup>	17.1 <sup>b</sup>	15.4	33.9	5.6	15.4	32.0	-15.3
Base model	10.1	30.0	18.5	10.1	34.3	15.8	10.1	35.4	4.5
NLP model	17.3	42.5	22.8	6.7 <sup>b</sup>	28.2 <sup>b</sup>	15.9 <sup>b</sup>	6.7 <sup>b</sup>	29.9 <sup>b</sup>	9.4 <sup>b</sup>
SDOH model	18.9	46.3	24.8	18.9	49.4	14.9	1.1 <sup>b</sup>	10.2 <sup>b</sup>	6.8 <sup>b</sup>
Combined model	12.1	40.2	26.5	12.1	41.6	19.6	4.8 <sup>b</sup>	25.2 <sup>b</sup>	10.5 <sup>b</sup>
<b>400/100 000</b>									
Treat all patients	100	100	-52.6	100	100	-145.1	100	100	-308.7
Lasso model	7.4 <sup>b</sup>	18.6 <sup>b</sup>	7.3 <sup>b</sup>	7.4 <sup>b</sup>	17.1 <sup>b</sup>	-1.0 <sup>b</sup>	7.4 <sup>b</sup>	17.7 <sup>b</sup>	-12.6 <sup>b</sup>
Base model	10.1	30.0	14.7	10.1	34.3	9.6	6.1 <sup>b</sup>	23.1 <sup>b</sup>	-1.6 <sup>b</sup>
NLP model	17.3	42.5	16.2	1.0 <sup>b</sup>	7.3 <sup>b</sup>	4.9 <sup>b</sup>	1.0 <sup>b</sup>	7.5 <sup>b</sup>	3.4 <sup>b</sup>
SDOH model	11.1	32.1	15.2	11.1	36.7	9.6	0.8 <sup>b</sup>	8.2 <sup>b</sup>	4.8 <sup>b</sup>
Combined model	12.1	40.2	21.9	12.1	41.6	12.2	1.9 <sup>b</sup>	12.2 <sup>b</sup>	4.4 <sup>b</sup>
<b>500/100 000</b>									
Treat all patients	100	100	-91.0	100	100	-206.6	100	100	-411.4
Lasso model	4.7 <sup>b</sup>	12.2 <sup>b</sup>	3.3 <sup>b</sup>	3.7	8.6	-2.8	3.7	8.8	-10.1
Base model	9.9	29.5	10.8	4.3 <sup>b</sup>	18.4 <sup>b</sup>	5.2 <sup>b</sup>	4.3 <sup>b</sup>	19.0 <sup>b</sup>	-2.9 <sup>b</sup>
NLP model	11.6	33.1	11.0	0.5 <sup>b</sup>	3.3 <sup>b</sup>	1.8 <sup>b</sup>	0.5 <sup>b</sup>	3.4 <sup>b</sup>	1.0 <sup>b</sup>
SDOH model	11.1	32.1	10.9	0.7 <sup>b</sup>	6.9 <sup>b</sup>	4.7 <sup>b</sup>	0.7 <sup>b</sup>	6.8 <sup>b</sup>	3.1 <sup>b</sup>
Combined model	4.4	22.9	14.6	4.4	23.7	10.3	1.3 <sup>b</sup>	8.2 <sup>b</sup>	1.6 <sup>b</sup>

Abbreviations: NLP, natural language processing; SDOH, social determinants of health.

<sup>a</sup> For each decision threshold, the assumption is that the decision to implement a preventive intervention would be made only with patients who had a

predicted risk above the threshold. See the eMethods in [Supplement 1](#) for a discussion.

<sup>b</sup> Logistic calibration. All others were isotonic calibration.

sion thresholds over the same risk horizons that we considered here.<sup>47</sup> Furthermore, postdischarge interventions, such as case management for patients at high suicide risk, have other benefits besides suicide prevention.<sup>48</sup>

The fact that NB was negative over short risk horizons, in comparison, is broadly consistent with previous research on

imminent risk of suicidal behaviors.<sup>49</sup> Importantly, though, this negative NB was not because our model did not predict imminent risk. Quite the contrary: AUROC was 0.777 for 1-week suicides. The negative NB instead reflected the high number of false-positives over short risk horizons relative to the number of true-positives. For example, a 1-week risk model with

Table 3. Predictors Used in the Analysis

Predictor	Distribution of potential predictors		Proportion and distribution of statistically significant predictors			Proportion and distribution of combined model predictors		
	No.	Distribution, % <sup>a</sup>	No.	Proportion, % <sup>b</sup>	Distribution, % <sup>a</sup>	No.	Proportion, % <sup>b</sup>	Distribution, % <sup>a</sup>
Total	10 181	100	2137	21.0	100	191	8.9	100
<b>Psychopathological risk factors</b>								
Diagnoses	3364	33.0	649	19.3	30.4	12	1.8	6.3
Treatments	224	2.2	80	35.7	3.7	11	13.8	5.8
Suicidality	46	0.5	38	82.6	1.8	4	10.5	2.1
Total	3634	35.7	767	21.1	35.9	27	3.5	14.1
<b>Physical disorders</b>								
Diagnoses	3716	36.5	666	17.9	31.2	22	3.3	11.5
Treatments	230	2.3	96	41.7	4.5	7	7.3	3.7
FDA medications increasing suicide risk	28	0.3	21	75.0	1.0	3	14.3	1.6
Total	3974	39.0	783	19.7	36.6	32	4.1	16.8
<b>Facility-level quality indicators</b>								
Total	6	0.1	6	100	0.3	4	66.7	2.1
<b>SDOH<sup>c</sup></b>								
Geospatial indicators	90	0.9	53	58.9	2.5	33	62.3	17.3
LexisNexis public records	442	4.3	29	6.6	1.4	29	100	15.2
Patient-level factors (ICD-9-CM/ICD-10-CM codes)	174	1.7	40	23.0	1.9	6	15.0	3.1
Sociodemographic characteristics	24	0.2	17	70.8	0.8	9	52.9	4.7
Total	733	7.2	142	19.4	6.6	80	56.3	41.9
<b>NLP term/topic frequency</b>								
Terms	1687	16.6	344	20.4	16.1	24	7.0	12.6
Topics	150	1.5	98	100	4.6	27	100	14.1
Total	1837	18.0	442	24.1	20.7	51	11.5	26.7

Abbreviations: FDA, US Food and Drug Administration; ICD-9-CM, *International Classification of Diseases, Ninth Revision, Clinical Modification*; ICD-10-CM, *International Classification of Diseases, Tenth Revision, Clinical Modification*; NLP, natural language processing; SDOH, social determinants of health.

<sup>a</sup> Entries in the distribution columns represent the contribution of predictors in the row heading to the total in the column. The percentage estimates in each distribution column sum to 100%.

<sup>b</sup> Entries in the proportion column represent the proportion of variation in the prior column for the same row that continue to exist in the current column. For example, the 649 significant psychiatric diagnoses in the first row and second column represent 19.3% of all the 3364 psychiatric diagnoses included in the initial potential predictor set, and the 12 psychiatric diagnoses in the final predictor set represent 1.8% of those 649 significant predictors.

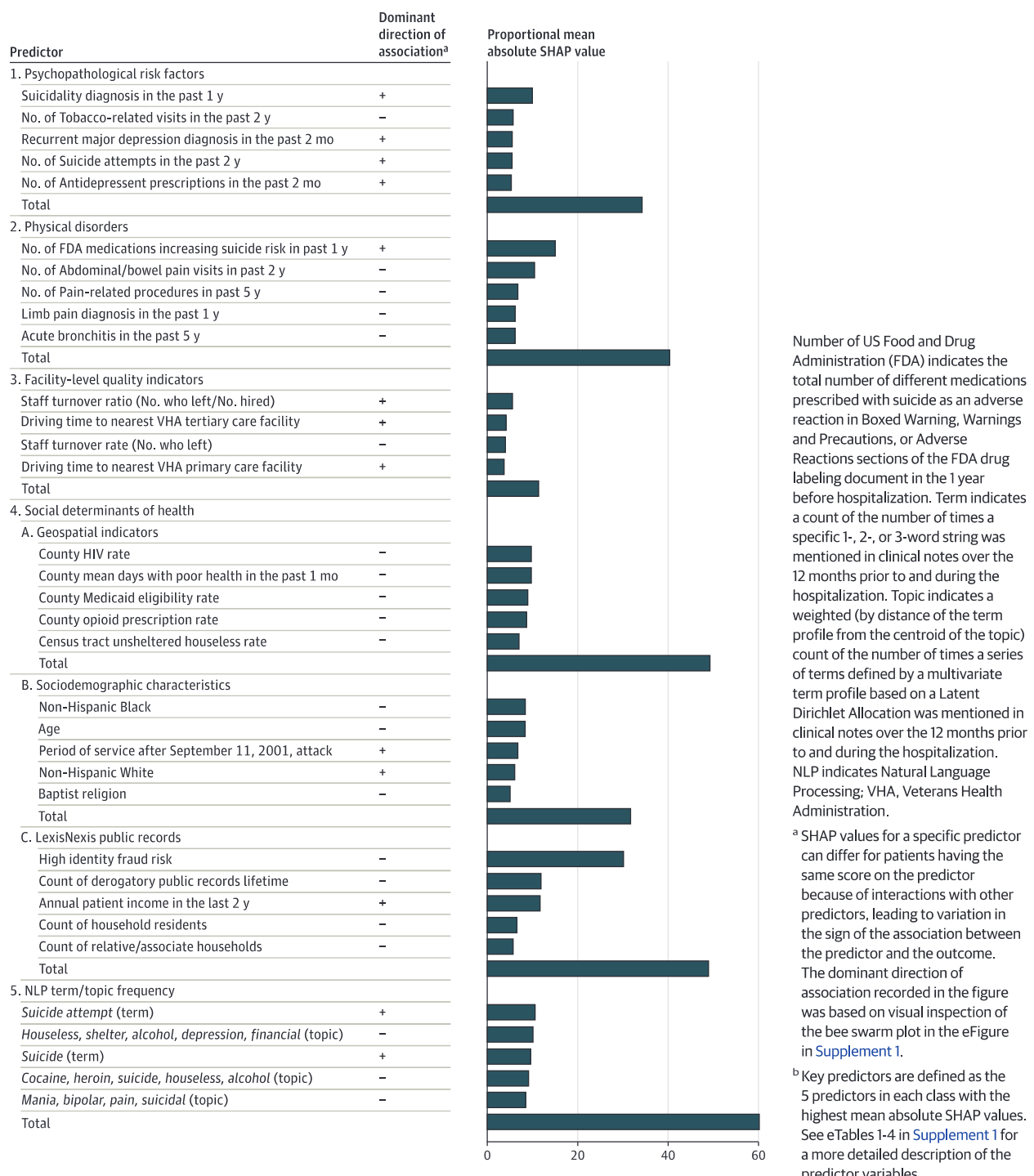
<sup>c</sup> Three of the NLP variables in the predictor set for the Combined model are included in this total as well as in the NLP total. One is the term *trauma*. The other 2 are topics in which the prominent terms are either *homeless/shelter/homelessness/lack of housing* or *divorce/stressors*.

75% sensitivity at 95% specificity, which would be considered excellent based on conventional criteria,<sup>34</sup> would result in roughly 6000 false-positives for every true-positive, given the low 1-week prevalence of postdischarge suicides. No plausible cost-effectiveness scenario would justify implementing an intensive case management program if the only goal was to prevent suicides over such a short horizon.<sup>29</sup> Fortunately, though, the model to predict 12-month suicides was also optimal for predicting imminent suicide risk, making it unnecessary to focus on separate segments of the patient population for short-term and longer-term prevention. This discovery is a valuable contribution to the field in that it both documents that prediction of imminent risk can be improved by using a longer risk horizon and shows that the cost-effectiveness of intervening with patients at high imminent risk could be increased to the extent that the intervention might reduce suicides over a much longer risk period.

Our finding that NLP and LexisNexis SDOH variables improve model prediction accuracy is broadly consistent

with prior studies of NLP and SDOH variables predicting suicide,<sup>9,50-53</sup> although to our knowledge, no prior studies investigated these associations among postdischarge psychiatric inpatients. Four results about these predictors are especially noteworthy in terms of potential clinical implications.

First, although the importance of SDOH predictors supports the recent emphasis on targeting SDOH for suicide prevention,<sup>54-56</sup> caution is needed in interpreting these associations unequivocally as being causal given that modifiable SDOH predictors could also be noncausal risk markers.<sup>57</sup> Second, the low proportional SHAP of psychopathological variables is most plausible interpreted as due to constrained variance in that all inpatients have serious psychiatric problems. This might also help explain the third noteworthy result: that many predictors typically considered risk factors for suicide had significant negative associations with suicide among recently discharged psychiatric inpatients (eg, pain-related physical disorders; NLP topics involving homelessness, substance use disorder, and bipolar disorder;

**Figure 2. Dominant Direction of Association With Suicide and Proportional Mean Absolute Shapley Additive Explanations (SHAP) Values of Key Predictors of the Combined Model in the Prospective Validation Sample<sup>a,b</sup>**

public records involving fraud risk and derogatory public records), whereas income was positively associated with suicide risk. It is noteworthy that these predictors have the same signs in their zero-order associations with suicide among recent inpatients, whereas these zero-order associations are generally the opposite sign in the larger VHA population. At least 2 possibilities exist for this pattern: (1) the

clinical severity threshold might be lower for hospitalizing patients with these disadvantaged characteristics than other patients due to these disadvantaged characteristics themselves increasing probability of hospitalization, resulting in lower suicide risk among hospitalized patients with these characteristics and/or (2) postdischarge interventions are more intensive for patients with these disadvantaged char-



acteristics than other patients. Both possibilities warrant further study given their implications for changing clinical practice guidelines. Fourth, the highest positive variable-level proportional SHAP was for a count of medications prescribed to the patient in the prior year classified by the FDA as increasing suicide risk. It was impossible to disaggregate this association to specific medications given that each medication was prescribed only rarely, but it is noteworthy that other potential predictors that did not have strong proportional SHAP values were counts of prescribed antidepressants and other types of psychotropic medications. Further disaggregation of these medications might be possible in the total VHA patient population, though, and should be carried out in future studies.

### Limitations

This study has limitations. First, absence of consensus on the minimum level of risk needed to justify implementing an aggressive suicide prevention intervention made it impossible to select a single decision threshold for the NB analysis. Presenting results across a plausible range of thresholds, as we did here, is recommended in cases where such uncertainty exists.<sup>30</sup> Second, generalization beyond the VHA is likely to be low

because of the unique characteristics of veterans, the comparatively high access to high-quality care enjoyed by VHA patients, and the fact that some administrative predictors in our model are unavailable in health systems other than the VHA. Third, prospective generalization within the VHA over time is also uncertain, as the model was trained and validated during years well before the COVID-19 pandemic and the changes in health care associated with the pandemic. We will be able to investigate this issue by applying the model to the training sample of our trial.

### Conclusions

Within the context of these limitations, we found that improvements could be made to a standard ML model for postdischarge suicides by adding predictors extracted from clinical notes and public records. The expanded model had positive NB at plausible decision thresholds for risk horizons of 3 to 12 months. Several results regarding important predictors might have clinical implications, but further research is needed on these possibilities before such implications would be actionable.

### ARTICLE INFORMATION

**Accepted for Publication:** November 9, 2022.

**Published Online:** January 18, 2023.

doi:10.1001/jamapsychiatry.2022.4634

**Author Affiliations:** Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Kessler, Gildea, King, Liu, Petukhova, Sampson); Department of Psychiatry, Harvard Medical School, Boston, Massachusetts (Bauer, Kennedy, Smoller, Bradley); VA Boston Healthcare System, Boston, Massachusetts (Bauer, Bradley); Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center, Canandaigua, New York (Bishop, Bossarte, Liu, Pigeon); Department of Psychiatry, University of Rochester Medical Center, Rochester, New York (Bishop, Pigeon); Department of Psychiatry and Behavioral Neurosciences, University of South Florida, Tampa (Bossarte); Research Information Science and Computing, Mass General Brigham, Somerville, Massachusetts (Castro); Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (Demler); Department of Computer Science, ETH Zurich, Zurich, Switzerland (Demler); Pain, Research, Informatics, Multi-morbidities and Education Center, VA Connecticut Healthcare System, West Haven (Goulet); Department of Emergency Medicine, Yale School of Medicine, New Haven, Connecticut (Goulet); Department of Psychiatry, Massachusetts General Hospital, Boston (Kennedy, Smoller); Behavioral Health Quality Enhancement Research Initiative (QUERI), Central Arkansas Veterans Healthcare System, North Little Rock (Landes); Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock (Landes); Department of Statistics, University of Washington, Seattle (Luedtke); Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington (Luedtke); Department of Psychology, Harvard University, Cambridge,

Massachusetts (Mair, Nock); National Center for PTSD, VA Boston Healthcare System, Boston, Massachusetts (Marx); Department of Psychiatry, Boston University School of Medicine, Boston, Massachusetts (Marx); Center for Precision Psychiatry, Massachusetts General Hospital, Boston (Smoller); Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts (Smoller); Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Smoller); VA North Texas Healthcare System, Dallas (Miller); VA Pittsburgh Healthcare System, Pittsburgh, Pennsylvania (Haas); Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania (Haas); VA St Louis Healthcare System, St Louis, Missouri (Benware); Central Arkansas Veterans Healthcare System, Little Rock (Owen, House); Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock (Owen, House); Minneapolis VA Healthcare System, Minneapolis, Minnesota (Urosevic); Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis (Urosevic); Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University, Providence, Rhode Island (Weinstock).

**Author Contributions:** Dr Kessler had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Kessler, Bauer, Bishop, Bossarte, Landes, Pigeon, Bradley, Weinstock.

**Acquisition, analysis, or interpretation of data:** Kessler, Bauer, Bishop, Bossarte, Castro, Demler, Gildea, Goulet, King, Kennedy, Liu, Luedtke, Mair, Marx, Nock, Petukhova, Sampson, Smoller, Miller, Haas, Benware, Owen, House, Urosevic, Weinstock.

**Drafting of the manuscript:** Kessler, Bishop, Bossarte, Gildea, King, Liu, Sampson, Haas, Bradley. **Critical revision of the manuscript for important intellectual content:** Kessler, Bauer, Bishop,

Bossarte, Castro, Demler, Gildea, Goulet, Kennedy, Landes, Liu, Luedtke, Mair, Marx, Nock, Petukhova, Pigeon, Sampson, Smoller, Miller, Haas, Benware, Bradley, Owen, House, Urosevic, Weinstock. **Statistical analysis:** Kessler, Bossarte, Castro, Demler, Goulet, King, Kennedy, Liu, Luedtke, Mair, Petukhova, Sampson. **Obtained funding:** Kessler, Bishop, Bradley, Weinstock. **Administrative, technical, or material support:** Kessler, Bauer, Bishop, Bossarte, Marx, Nock, Pigeon, Miller, Benware, Bradley, Owen, Weinstock. **Study supervision:** Kessler, Bossarte, Luedtke, Pigeon, Benware, Bradley, Owen, House, Weinstock.

**Conflict of Interest Disclosures:** Dr Kessler has received grants from Warren Alpert Foundation; personal fees from Cambridge Health Alliance, Canandaigua VA Medical Center, Holmusk, Partners Healthcare, RallyPoint Networks, Sage Therapeutics during the conduct of the study; and holds stock options in Cerebral, Mirah, PYM, and Roga Sciences. Dr Goulet has received grants from Department of Veterans Affairs during the conduct of the study. Dr Landes has received grants from QUERI as well as personal fees from RAND and UHealth Houston outside the submitted work. Dr Liu has received grants from Warren Alpert Foundation during the conduct of the study. Dr Luedtke has received personal fees from Harvard University during the conduct of the study. Dr Mair has received personal fees from Canandaigua VA Medical Center, RallyPoint Networks, Rehab Zentrum Liesing, and Simon Fraser University. Dr Nock has received personal fees from Microsoft, Veterans Health Administration, and COMPASS Pathways; royalties from Macmillan, Pearson, and UpToDate; is an unpaid scientific advisor for Empatica, Koko, and TalkLife; and owns stock options in Cerebral outside the submitted work. Dr Sampson has received grants from Warren Alpert Foundation during the conduct of the study. Dr Urosevic has received

grants from Warren Alpert Foundation during the conduct of the study. No other disclosures were reported.

**Funding/Support:** This work was supported by a grant from the Warren Alpert Foundation. Drs Weinstock and Kessler are joint principal investigators. Dr Demler was supported in part by grant 5K01HL135342 from the National Heart, Lung, and Blood Institute and by grant 7IGMV33860009 from the American Heart Association. Dr Landes was supported in part by grant UL1 TR003107 from the National Center for Advancing Translational Sciences of the National Institutes of Health and by grants QUE 20-026, PII 18-195, and EBP 22-104 from the Department of Veterans Affairs Quality Enhancement Research Initiative. Dr Smoller and Mr Castro were supported in part by grant R01 MH117599 from the National Institute of Mental Health. Dr Goulet was supported by grant IIR 16-262 from the VA Health Services Research and Development Service.

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Data Sharing Statement:** See Supplement 2.

## REFERENCES

- Heron M. Deaths: leading causes for 2016. Accessed July 26, 2022. [https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67\\_06.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_06.pdf)
- Ahmedani BK, Simon GE, Stewart C, et al. Health care contacts in the year before suicide death. *J Gen Intern Med*. 2014;29(6):870-877. doi:10.1007/s11606-014-2767-3
- 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(7):694-702. doi:10.1001/jamapsychiatry.2017.1044
- 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(6):709-717. doi:10.1111/sltb.12334
- Miller IW, Gaudiano BA, Weinstock LM. The coping long term with active suicide program: description and pilot data. *Suicide Life Threat Behav*. 2016;46(6):752-761. doi:10.1111/sltb.12247
- Stanley B, Brown GK, Brenner LA, 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(9):894-900. doi:10.1001/jamapsychiatry.2018.1776
- Kessler RC, Bauer MS, Bishop TM, et al. Using administrative data to predict suicide after psychiatric hospitalization in the Veterans Health Administration system. *Front Psychiatry*. 2020;11:390. doi:10.3389/fpsy.2020.00390
- Veterans Coordinated Community Care (3C) Study (3C). ClinicalTrials.gov identifier: NCT05272176. Updated October 31, 2022. Accessed August 2, 2022. <https://www.clinicaltrials.gov/ct2/show/NCT05272176?term=3C&draw=2&rank=2>
- Cohen J, Wright-Berryman J, Rohlfes L, Trocinski D, Daniel L, Klatt TW. Integration and validation of a natural language processing machine learning suicide risk prediction model based on open-ended interview language in the emergency department. *Front Digit Health*. 2022;4:818705. doi:10.3389/fdgh.2022.818705
- Blosnich JR, Montgomery AE, Dichter ME, et al. Social determinants and military veterans' suicide ideation and attempt: a cross-sectional analysis of electronic health record data. *J Gen Intern Med*. 2020;35(6):1759-1767. doi:10.1007/s11606-019-05447-z
- US Centers for Disease Control and Prevention. National Death Index. Accessed March 2, 2022. <https://www.cdc.gov/nchs/ndi/index.htm>
- Collins GS, Reitsma JB, Altman DG, Moons KG; TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Circulation*. 2015;131(2):211-219. doi:10.1161/CIRCULATIONAHA.114.014508
- US Department of Veterans Affairs. Corporate Data Warehouse (CDW). Accessed March 1, 2022. [https://www.hsrd.research.va.gov/for\\_researchers/vinci/cdw.cfm](https://www.hsrd.research.va.gov/for_researchers/vinci/cdw.cfm)
- 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(6):816-821. doi:10.1177/0033354916670133
- Wong MS, Steers WN, Hoggatt KJ, Ziaieian B, Washington DL. Relationship of neighborhood social determinants of health on racial/ethnic mortality disparities in US veterans—mediation and moderating effects. *Health Serv Res*. 2020;55(suppl 2):851-862. doi:10.1111/1475-6773.13547
- VA Health Services Research and Development Service; Internet Archive. Consortium for Healthcare Informatics Research (CHIR). Accessed February 22, 2022. [https://web.archive.org/web/20221028115247/https://www.hsrd.research.va.gov/for\\_researchers/chir.cfm](https://web.archive.org/web/20221028115247/https://www.hsrd.research.va.gov/for_researchers/chir.cfm)
- LexisNexis Risk Solutions Group. Social determinants of health. Accessed March 2, 2022. <https://risk.lexisnexis.com/healthcare/social-determinants-of-health>
- 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(7):653-659. doi:10.1176/appi.ps.2012.00026
- 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(8):619-628. doi:10.3109/00048674.2011.590465
- 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(2):145-150. doi:10.1016/j.psychres.2012.09.039
- 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(1):60-65. doi:10.1007/s11920-008-0011-8
- Bachmann S. Epidemiology of suicide and the psychiatric perspective. *Int J Environ Res Public Health*. 2018;15(7):1425. doi:10.3390/ijerph15071425
- Klonsky ED, May AM, Saffer BY. Suicide, suicide attempts, and suicidal ideation. *Annu Rev Clin Psychol*. 2016;12:307-330. doi:10.1146/annurev-clinpsy-021815-093204
- O'Connor RC, Nock MK. The psychology of suicidal behaviour. *Lancet Psychiatry*. 2014;1(1):73-85. doi:10.1016/S2215-0366(14)70222-6
- van der Laan MJ, Rose S. *Targeted Learning: Casual Inference for Observational and Experimental Data*. Springer; 2011.
- Polley E, LeDell E, Kennedy C, Lendle S, van der Laan M. SuperLearner: super learner prediction. Version 2.0-24. Accessed February 18, 2022. <https://cran.r-project.org/web/packages/SuperLearner/index.html>
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22. doi:10.18637/jss.v033.i01
- Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18. doi:10.1186/s41512-019-0064-7
- Ross EL, Zuromski KL, Reis BY, Nock MK, Kessler RC, Smoller JW. Accuracy requirements for cost-effective suicide risk prediction among primary care patients in the US. *JAMA Psychiatry*. 2021;78(6):642-650. doi:10.1001/jamapsychiatry.2021.0089
- Wynants L, van Smeden M, McLernon DJ, Timmerman D, Steyerberg EW, Van Calster B; Topic Group 'Evaluating Diagnostic Tests and Prediction Models' of the STRATOS Initiative. Three myths about risk thresholds for prediction models. *BMC Med*. 2019;17(1):192. doi:10.1186/s12916-019-1425-3
- Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat Med*. 2019;38(21):4051-4065. doi:10.1002/sim.8281
- Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. 2006;7:91. doi:10.1186/1471-2105-7-91
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Accessed August 4, 2022. <https://dl.acm.org/doi/10.5555/3295222.3295230>
- Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. Wiley; 2013. doi:10.1002/9781118548387
- Grendas LN, Chiappella L, Rodante DE, Daray FM. Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour. *J Psychiatr Res*. 2021;145:85-91. doi:10.1016/j.jpsychires.2021.11.029
- Kirtley OJ, van Mens K, Hoogendoorn M, Kapur N, de Beurs D. Translating promise into practice: a review of machine learning in suicide research and prevention. *Lancet Psychiatry*. 2022;9(3):243-252. doi:10.1016/S2215-0366(21)00254-6
- Lejeune A, Le Glaz A, Perron PA, et al. Artificial intelligence and suicide prevention: a systematic review. *Eur Psychiatry*. 2022;65(1):1-22. doi:10.1192/j.eurpsy.2022.8

38. Jiang T, Rosellini AJ, Horváth-Puhó E, et al. Using machine learning to predict suicide in the 30 days after discharge from psychiatric hospital in Denmark. *Br J Psychiatry*. 2021;219(5):440-447. doi:10.1192/bjp.2021.19
39. Bolton JM. Suicide risk assessment in the emergency department: out of the darkness. *Depress Anxiety*. 2015;32(2):73-75. doi:10.1002/da.22320
40. Hoge CW. Suicide reduction and research efforts in service members and veterans-sobering realities. *JAMA Psychiatry*. 2019;76(5):464-466. doi:10.1001/jamapsychiatry.2018.4564
41. Mulder R, Newton-Howes G, Coid JW. The futility of risk prediction in psychiatry. *Br J Psychiatry*. 2016;209(4):271-272. doi:10.1192/bjp.bp.116.184960
42. Owens D, Kelley R. Predictive properties of risk assessment instruments following self-harm. *Br J Psychiatry*. 2017;210(6):384-386. doi:10.1192/bjp.bp.116.196253
43. Large M, Myles N, Myles H, et al. Suicide risk assessment among psychiatric inpatients: a systematic review and meta-analysis of high-risk categories. *Psychol Med*. 2018;48(7):1119-1127. doi:10.1017/S0033291717002537
44. Kessler RC. Clinical epidemiological research on suicide-related behaviors—where we are and where we need to go. *JAMA Psychiatry*. 2019;76(8):777-778. doi:10.1001/jamapsychiatry.2019.1238
45. Simon GE, Shortreed SM, Coley RY. Positive predictive values and potential success of suicide prediction models. *JAMA Psychiatry*. 2019;76(8):868-869. doi:10.1001/jamapsychiatry.2019.1516
46. Kennedy CJ, Mark DG, Huang J, van der Laan MJ, Hubbard AE, Reed ME. Development of an ensemble machine learning prognostic model to predict 60-day risk of major adverse cardiac events in adults with chest pain. *medRxiv*. Preprint posted online March 13, 2021. doi:10.1101/2021.03.08.21252615
47. Stone NJ, Robinson JG, Lichtenstein AH, et al; American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 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(25)(suppl 2):S1-S45. doi:10.1161/01.cir.0000437738.63853.7a
48. McCarthy JF, Cooper SA, Dent KR, et al. Evaluation of the recovery engagement and coordination for health-veterans enhanced treatment suicide risk modeling clinical program in the Veterans Health Administration. *JAMA Netw Open*. 2021;4(10):e2129900. doi:10.1001/jamanetworkopen.2021.29900
49. Ribeiro JD, Huang X, Fox KR, Walsh CG, Linthicum KP. Predicting imminent suicidal thoughts and nonfatal attempts: the role of complexity. *Clin Psychol Sci*. 2019;7(5):941-957. doi:10.1177/2167702619838464
50. Levis M, Leonard Westgate C, Gui J, Watts BV, Shiner B. Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models. *Psychol Med*. 2021;51(8):1382-1391. doi:10.1017/S0033291720000173
51. Levis M, Levy J, Dufort V, Gobbet GT, Watts BV, Shiner B. Leveraging unstructured electronic medical record notes to derive population-specific suicide risk models. *Psychiatry Res*. 2022;315:114703. doi:10.1016/j.psychres.2022.114703
52. Alemi F, Avramovic S, Renshaw KD, Kanchi R, Schwartz M. Relative accuracy of social and medical determinants of suicide in electronic health records. *Health Serv Res*. 2020;55(suppl 2):833-840. doi:10.1111/1475-6773.13540
53. Blois JR, Montgomery AE, Taylor LD, Dichter ME. Adverse social factors and all-cause mortality among male and female patients receiving care in the Veterans Health Administration. *Prev Med*. 2020;141:106272. doi:10.1016/j.ypmed.2020.106272
54. Blanco C, Wall MM, Olfson M. A population-level approach to suicide prevention. *JAMA*. 2021;325(23):2339-2340. doi:10.1001/jama.2021.6678
55. Iskander JK, Crosby AE. Implementing the national suicide prevention strategy: time for action to flatten the curve. *Prev Med*. 2021;152(pt 1):106734. doi:10.1016/j.ypmed.2021.106734
56. National Academies of Sciences, Engineering, and Medicine; Division of Behavioral and Social Sciences and Education; Board on Behavioral, Cognitive, and Sensory Sciences; Yoder L, eds. *Community Interventions to Prevent Veteran Suicide: The Role of Social Determinants: Proceedings of a Virtual Symposium*. National Academies Press; 2022.
57. Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, Kupfer DJ. Coming to terms with the terms of risk. *Arch Gen Psychiatry*. 1997;54(4):337-343. doi:10.1001/archpsyc.1997.01830160065009