The Journal of Clinical Psychiatry

Supplementary Material

Article Title: Leveraging Natural Language Processing to Improve Electronic Health Record Suicide Risk Prediction for Veterans Health Administration Users

Authors: Maxwell Levis, PhD; Joshua Levy, PhD; Kallisse R. Dent, MPH; Vincent Dufort, PhD; Glenn T. Gobbel, PhD, DVM, MS; Bradley V. Watts, MD, MPH; and Brian Shiner, MD, MPH

DOI Number: 10.4088/JCP.22m14568

LIST OF SUPPLEMENTARY MATERIAL FOR THE ARTICLE

DISCLAIMER

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Supplementary Table 1. REACH-VET algorithm's 61-structured variables[1]

Model Factor	Weight
Demographics	
Male	0.456
>80 yrs. old	-0.081
White race	0.584
Non-White race	-0.118
Marriage status	-0.202
Lives in Western USA	0.206
More than 30% Service Connection	-0.336
More than 70% Service Connection	0.136
Interaction Widowed and Male	-0.109
Interaction between Divorced and Male	0.082
Prior suicide attempts	
Suicide attempt in last mo.	0.093
Suicide attempt in last 6 mos.	0.462
Suicide attempt in last 18 mos.	0.557
Diagnoses	
Arthritis dx. in last 12 mos.	-0.041
Arthritis dx. in last 24 mos.	-0.044
Lupus dx. in last 24 mos.	0.274
Bipolar dx. in last 24 mos.	0.126
Chronic pain dx. in last 24 mos.	0.220
Depression dx. in last 12 mos.	0.145
Depression dx. in last 24 mos.	0.377
Diabetes dx. in last 12 mos.	-0.074
Substance use disorder dx. in last 24 mos.	0.215
Homeless in last 24 mos.	-0.120
Head/neck cancer dx. in last 12 mos.	0.159
Head/neck cancer dx. in last 24 mos.	0.024
Anxiety disorder dx. in last 24 mos.	0.041
Personality disorder dx. in last 24 mos.	0.002
Interaction Other anxiety disorder (prior 24 mos.) and Personality disorder (prior 24 mos.)	0.086
Service utilization	
Emergency Department visit in last 1 mo.	0.125
Service utilization cont.	
Emergency Department visit in last 2 mos.	0.185
Psychiatric discharge in last mo.	0.608
Psychiatric discharge in last 6 mo.	0.221
Psychiatric discharge in last 12 mo.	0.038
Psychiatric discharge in last 24 mo.	0.559
Any Mental Health treatment in last 12 mo.	0.041
Any Mental Health treatment in last 24 mo.	0.002
# days VHA services used in 7th mo. prior	-0.006
# days VHA services used in 13th mo. prior	0.000
# days Emergency Department in last mo.	0.148
# days Emergency Department in last 24 mos.	-0.002
First Use in Prior 5 yrs. was in the Prior yr.	0.054
# days Inpatient Mental Health in last 7 mos. sq.	0.000
# days Outpatient services in 7th mo. prior	-0.006
# days Outpatient services in 8th mo. prior	-0.005
# days Outpatient services in 15th mo. prior	-0.011
# days Outpatient services in 23rd mo. prior	-0.001
Medications	
Alprazolam rx. in last 24 mos.	0.183
Any anti-depressant rx. in last 24 mos.	0.164
Any anti-psychotic rx. in last 12 mos.	0.134
Clonazepam rx. in last 12 mos.	0.114
Clonazepam rx. in last 24 mos.	0.195
Lorazepam rx. in last 12 mos.	0.073
Mirtazapine rx. in last 12 mos.	0.009
Mirtazapine rx. in last 24 mos.	0.050
Mood stabilizer rx. in last 12 mos.	0.018
Opioids rx. in last 12 mos.	0.018
Sedative or anxiolytic rx. in last 12 mos.	0.251
Zolpidem rx. in last 12 mos. Zolpidem rx. in last 12 mos.	0.021
Sedative or anxiolytic rx. in last 24 mos.	0.349
Statin rx. in last 12 mos.	-0.141

Supplementary Table 2. CountVectorizer model

Table presents CountVectorizer[2] output analyzed by Random Forest (RF)[3] and Naïve Bayes (NB)[4] classification models. Each model evaluates notes from different time intervals back from date of death by suicide for cases or matched time points for controls. Overall predictive accuracy is estimated via AUC. Risk concentration for Veterans with the highest predicted risk (10%, 5%, 1%, .1%) is also estimated. Following REACH-VET studies, to evaluate risk concentration, we gauged the proportion of death by suicide to the expected proportion of death by suicide assuming uniform sample distribution, i.e., among Veterans Health Administration patients who scored within the highest 10% of this model, 22% died by suicide. As Countvectorizer models were typically less predictive than TFIDF models, they were not included in additional analyses.

	RF	Risk concentration at the following risk tiers:				NB	Risk concentration at the following risk tiers:			
Days back	AUC (95% CI)	Top 10%	Top 5%	Top 1%	Top .1%	AUC (95% CI)	Top 10%	Top 5%	Top 1%	Top .1%
30	**.66 (.63 - .68)**	**2.2**	**2.4**	**3.0**	**4.2**	**.61 (.58 - .63)**	**1.4**	**1.4**	**1.7**	**1.2**
90	**.63 (.60 - .65)**	**1.8**	**1.9**	**2.0**	**2.7**	**. 61 (.58 - .63)**	**1.4**	**1.2**	**1.5**	**1.0**
360	**.61 (.59 - .62)**	**1.5**	**1.1**	**1.4**	**1.0**	**.60 (.58 - .62)**	**1.4**	**1.2**	**1.6**	**1.0**

Supplementary Table 3. Parameter tuning

We performed coarse hyperparameter searches to identify ideal Random Forest (RF),[22] XGBoost (XG),[23] and Logistic Regression (LR)[26] model specifications for TFIDF[5] output. Optimal hyperparameters were evaluated based on the loss over each validation set. As follows, we list the hyperparameters scanned for each model through the coarse inspection of validation set statistics. Naïve Bayes models were not subject to cross validations. Final selections were based on sensible recommendations and experimentation.

Hyperparameter tuning for TFIDF (utilized hyperparameters are marked in bold)	
RF	*n_estimators* = 200, **300,** 500, 700, 1000; *max_features* = **auto**, sqrt; *max_depth* = 5, 10, 25, 50, **none**; *min_samples_split* = **2**, 5, 10; *min_samples_leaf* = 1, 2, **10**; *bootstrap* = **true**, false
XG	*n_estimators* = 200, **500**, 700, 1000; *subsample* = .5, **.8**, 1; *num_boost_round* = 2, 10, **50**; *min_child_weight* = 1, 6, **12**; *max_depth* = **5**, 10, 25, 50; *early_stopping_rounds* = 1, 10, **100**; *colsample_bytree* = .6, **.8**, 1
LR	*C* = .001, .01, .1, **1**, 10, 100l; L1, **L2**

Supplementary Table 4. Standardized model

Table presents TFIDF[5] output that was standardized using StandardScaler[6] and then analyzed using Random Forest (RF)[3] and Naïve Bayes (NB)[4] classification models. Each model evaluates notes from different time intervals back from date of death by suicide for cases or matched time points for controls. Overall predictive accuracy is estimated via AUC. Risk concentration for Veterans with the highest predicted risk (10%, 5%, 1%, .1%) is also estimated. Following REACH-VET studies, to evaluate risk concentration, we gauged the proportion of death by suicide to the expected proportion of death by suicide assuming uniform sample distribution, i.e., among Veterans Health Administration patients who scored within the highest 10% of this model, 22% died by suicide. As models that had been standardized were typically less predictive than unstandardized models, they were not included in additional analyses.

	RF	Risk concentration at the following risk tiers:				NB	Risk concentration at the following risk tiers:			
Days back	AUC (95% CI)	Top 10%	Top 5%	Top 1%	Top .1%	AUC (95% CI)	Top 10%	Top 5%	Top 1%	Top .1%
30	**.65 (.63 - .68)**	**2.2**	**2.5**	**2.7**	**4.0**	**.62 (.60 - .64)**	**1.4**	**1.4**	**1.6**	**1.0**
90	**.63 (.61 - .65)**	**1.6**	**1.9**	**2.1**	**2.4**	**. 62 (.60 -.64)**	**1.5**	**1.2**	**1.4**	**1.0**
360	**.60 (.58 - .62)**	**1.5**	**1.1**	**1.3**	**1.0**	**.61 (.58 -.63)**	**1.4**	**1.5**	**1.9**	**2.0**

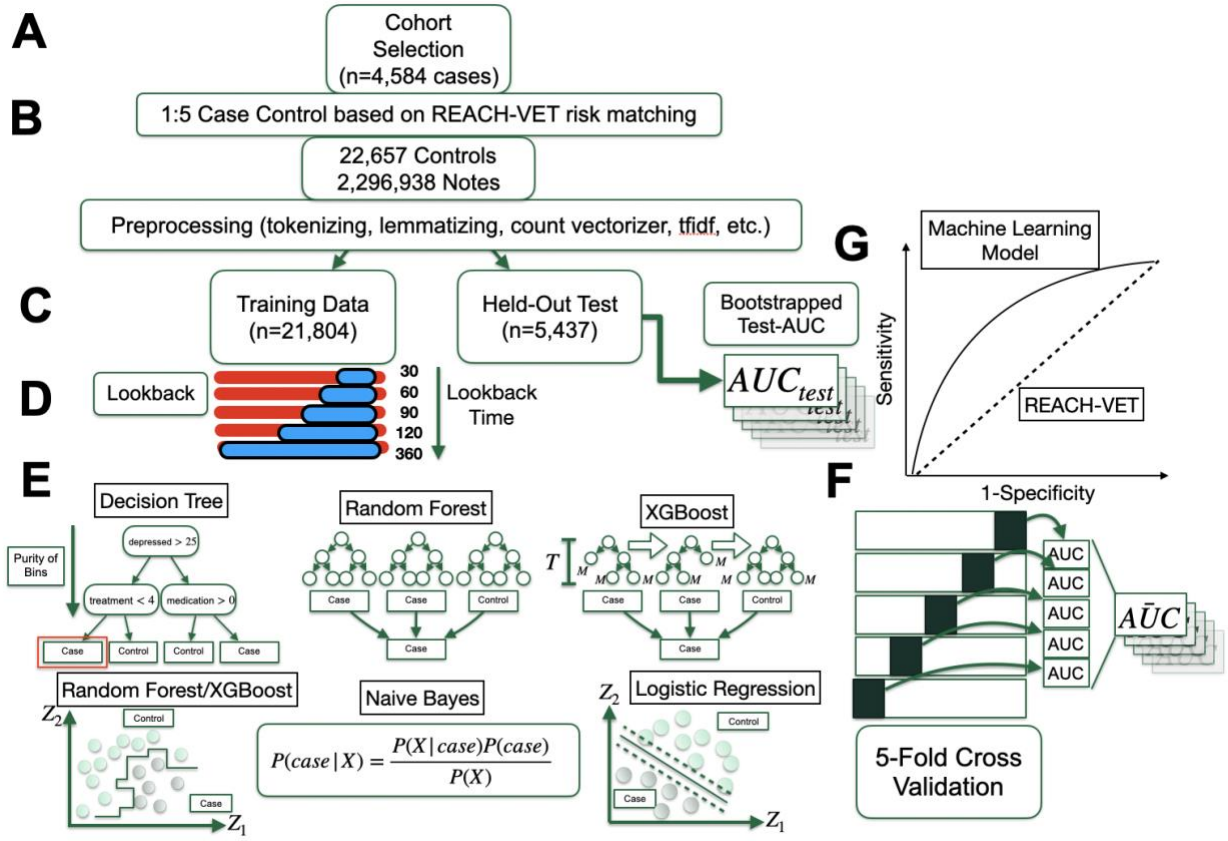
Supplementary Figure 1. Checklist for transparent model reporting

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	3
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	4
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	4
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	4
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	4
	5b	Describe eligibility criteria for participants.	5
	5c	Give details of treatments received, if relevant.	
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	8
	6b	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	9
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	9
Sample size	8	Explain how the study size was arrived at.	5
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	5
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	7
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	7
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	8
Risk groups	11	Provide details on how risk groups were created, if done.	8
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10/ 22
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	10/ 21
Model development	14a	Specify the number of participants and outcome events in each analysis.	22
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	23
	15b	Explain how to the use the prediction model.	10
Model performance	16	Report performance measures (with CIs) for the prediction model.	23
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	13
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	11
Implications	20	Discuss the potential clinical use of the model and implications for future research.	14
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	25/ 26
Funding	22	Give the source of funding and the role of the funders for the present study.	1

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Supplementary Figure 2. Methods overview diagram



1. VA. REACH VET, Predictive Analytics for Suicide Prevention. 2017. Available from: https://www.dspo.mil/Portals/113/Documents/2017%20Conference/Presentations/REACH%20VET%20Predictive%20Modeling.pdf?ver=2017-08-10-132615-843

2. Weinberger K, Dasgupta A, Attenberg J, Langford J, Smola A. Feature Hashing for Large Scale Multitask Learning. 2009; Available from: https://arxiv.org/abs/0902.2206

3. Ho TK. Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. 1995. p. 278–82.

4. Zhang H. Exploring conditions for the optimality of Naïve Bayes. Int J Pattern Recognit Artif Intell. 2005 Mar;19(02):183–98.

5. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inf Process Manag. 1988 Jan;24(5):513–23.

6. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. 2013; Available from: https://arxiv.org/abs/1309.0238