



External Validation and Updating of a Statistical Civilian-Based Suicide Risk Model in US Naval Primary Care

Michael A. Ripperger, BE; Jhansi Kolli, BS; Drew Wilimitis, BS; Katelyn Robinson, BA; Carrie Reale, MSN, RN-BC; Laurie L. Novak, PhD, MHSA; Craig A. Cunningham, PhD, MSN; Lalon M. Kasuske, PhD, MBA; Shawna G. Grover, PhD, ANP-BC, ACNS-BC, AOCNS; Jessica D. Ribeiro, PhD; Colin G. Walsh, MD, MA

Abstract

IMPORTANCE Suicide remains an ongoing concern in the US military. Statistical models have not been broadly disseminated for US Navy service members.

OBJECTIVE To externally validate and update a statistical suicide risk model initially developed in a civilian setting with an emphasis on primary care.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study used data collected from 2007 through 2017 among active-duty US Navy service members. The external civilian model was applied to every visit at Naval Medical Center Portsmouth (NMCP), its NMCP Naval Branch Health Clinics (NBHCs), and TRICARE Prime Clinics (TPCs) that fall within the NMCP area. The model was retrained and recalibrated using visits to NBHCs and TPCs and updated using Department of Defense (DoD)-specific billing codes and demographic characteristics, including expanded race and ethnicity categories. Domain and temporal analyses were performed with bootstrap validation. Data analysis was performed from September 2020 to December 2022.

EXPOSURE Visit to US NMCP.

MAIN OUTCOMES AND MEASURES Recorded suicidal behavior on the day of or within 30 days of a visit. Performance was assessed using area under the receiver operating curve (AUROC), area under the precision recall curve (AUPRC), Brier score, and Spiegelhalter z-test statistic.

RESULTS Of the 260 583 service members, 6529 (2.5%) had a recorded suicidal behavior, 206 412 (79.2%) were male; 104 835 (40.2%) were aged 20 to 24 years; and 9458 (3.6%) were Asian, 56 715 (21.8%) were Black or African American, and 158 277 (60.7%) were White. Applying the civilian-trained model resulted in an AUROC of 0.77 (95% CI, 0.74-0.79) and an AUPRC of 0.004 (95% CI, 0.003-0.005) at NBHCs with poor calibration (Spiegelhalter $P < .001$). Retraining the algorithm improved AUROC to 0.92 (95% CI, 0.91-0.93) and AUPRC to 0.66 (95% CI, 0.63-0.68). Number needed to screen in the top risk tiers was 366 for the external model and 200 for the retrained model; the lower number indicates better performance. Domain validation showed AUROC of 0.90 (95% CI, 0.90-0.91) and AUPRC of 0.01 (95% CI, 0.01-0.01), and temporal validation showed AUROC of 0.75 (95% CI, 0.72-0.78) and AUPRC of 0.003 (95% CI, 0.003-0.005).

CONCLUSIONS AND RELEVANCE In this cohort study of active-duty Navy service members, a civilian suicide attempt risk model was externally validated. Retraining and updating with DoD-specific variables improved performance. Domain and temporal validation results were similar to external validation, suggesting that implementing an external model in US Navy primary care

(continued)

Key Points

Question How well does a civilian-based suicide risk model generalize in US Navy primary care?

Findings In this cohort study with 260 583 service members, domain and temporal validation showed internal retraining and external validation had similar performance. Updating with US Navy-specific factors added minimal improvement.

Meaning These findings suggest that civilian-based risk models might generalize to military health settings; prior to transferring risk models, external validation might demonstrate adequate performance in new settings and avoid costly internal development.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

clinics may bypass the need for costly internal development and expedite the automation of suicide prevention in these clinics.

JAMA Network Open. 2023;6(11):e2342750. doi:10.1001/jamanetworkopen.2023.42750

Introduction

Suicide remains a costly and complex challenge for the US military.¹ Military veterans in the United States are significantly more at risk of attempting suicide than the general population.² While active-duty service members have historically had comparable suicide rates with the general population, suicide rates for service members have increased at a faster rate than the general population since 2015.³ The largest population of service members are young and enlisted (ie, early in their careers) and have higher rates of deaths from suicide that account for the majority of attempts.³⁻⁵ In addition, these suicide attempts are more likely to result in death.⁶⁻⁸ Even more challenging, disclosure of suicidal ideation by service members may have career-altering consequences, increasing the likelihood of covert distress.⁹⁻¹²

To better understand factors that might increase suicide risk in the military, numerous studies have analyzed protective and risk factors for service members.¹³⁻¹⁶ All service members have been shown to have lower risk when they are focused on a clear mission such as during deployment, yet they experience a higher risk during the postdeployment adjustment period, and this risk increase continues into retirement.¹⁷⁻¹⁹

US service members have contact with primary care practitioners in the month before suicide more than half the time.²⁰⁻²² Primary care practitioners seek tools to aid in identifying at-risk patients whom they otherwise may miss.²³ While those who die from suicide are more likely to perceive treatment barriers, they are also more likely to use health care.^{24,25} Thus, significant attention has been paid to predicting suicide using health care data to guide preventive decisions both within and outside of the military, as meta-analyses suggest traditional screening tools might be inaccurate.²⁶⁻²⁹ Within the military, the Army Study to Assess Risk and Resilience in Servicemembers (STARRS) has published the best known corpus of such models.³⁰⁻³² Broadly, the biomedical literature includes examples of models to predict the emergence, presence, and imminence of suicides.³³⁻³⁵

To implement a suicide risk model, multidisciplinary teams must develop, validate (test), replicate, and then use models in clinical settings to ensure robustness and continued accuracy.³⁶⁻³⁸ In parallel, teams must design effective decision support using risk model data to improve prevention.³⁹⁻⁴¹ The effort to accomplish these steps is significant. Prior to novel model development, testing externally validated models might result in clinically feasible performance with less time and expense.⁴² These models might also be retrained, such that their input predictors are reweighted to reflect the new setting, or updated, such that novel predictors available only in the new setting are added to the initial models.⁴³ Notably, models trained in different settings, such as civilian health systems, might generalize to perform well in military health systems.

In this study, we externally validated a civilian-based statistical risk model currently prompting suicide preventive clinical decision support at Vanderbilt University Medical Center (VUMC) using data from multiple primary care settings serving US Navy sailors.⁴⁴⁻⁴⁶ We then compared model performance after retraining the model on US Navy primary care data and subsequently updated the model with novel factors only relevant in military health systems that were informed by Department of Defense (DoD)-specific health care variables and demographics, including expanded race and ethnicity categories, present only in DoD sources. The transferability of risk models remains an important research question generally and has clinical implications in suicide risk prediction specifically. We sought to understand how transferring a model trained on a heterogeneous civilian cohort to military settings affects its ability to project suicide attempt risk.

Methods

This retrospective cohort study included encounters for US Navy active-duty service members who received care at Naval Medical Center Portsmouth (NMCP), Naval Branch Health Clinics (NBHCs), and TRICARE Prime Clinics (TPCs) between 2007 through 2017. Demographic characteristics for this population during the study period were obtained from the Defense Manpower Data Center (DMDC). NMCP NBHCs and TPCs provide routine care for service members and their families. The VUMC, NMCP, and Florida State University institutional review boards approved this study. Informed consent was waived given the study's minimal risk and its reliance on deidentified records and data collected in routine care. This study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.⁴⁷

Overview

In this cohort study of active-duty Navy service members, we externally validated, retrained, and updated random forest models of suicide attempt risk. In addition, we performed domain and temporal analyses with random forests. An overview of these methods and relevant acronyms can be found in **Table 1**.

Model and Data Sources

The cohort of Naval health system data was preprocessed based on feature engineering from the published, validated civilian-based model.⁴⁴ Data were obtained from the Navy and Marine Corps Public Health Center, EpiData Center. One data set was created for all NMCP visits and then split in appropriate subsets. In prior work, the civilian-based predictive model was trained at VUMC using the Harrell bias correction method using a case-control data set comprised of electronic health records (EHRs). Case status was determined by 2 suicide experts who manually identified documentation of suicide attempts from 5543 candidate medical records and joining this with a heterogeneous mix of patients seen at VUMC for any reason contemporaneously.^{44,48} Model predictors included counts of diagnostic *International Statistical Classification of Diseases, Tenth Revision, Clinical Modification (ICD-9-CM)* codes filtered through Center for Medicare & Medicaid Services (CMS) Hierarchical Condition Categories; counts of Anatomical Therapeutic Chemical level 4 medications; demographic characteristics; and counts of emergency, inpatient, and outpatient visits for each of the 5 years prior to prediction (eTable 2 in [Supplement 1](#)).⁴⁹⁻⁵¹ This process was repeated in this study with military health system data.

Military health system data were taken from the Standard Inpatient Data Record, the Comprehensive Ambulatory/Professional Encounter Record, the TRICARE Encounter Data Non-Institutional records, and the Theater Medical Data Store. These data sources comprehensively

Table 1. Key to Model Experiments and Commonly Used Terms

Key	Description
NMCP	NMCP provides emergency, outpatient, and inpatient care and is the parent of the NBHCs and TPCs in this study
NBHCs	Outpatient care clinics used for routine care
TPCs	Outpatient care clinics for acute care injuries
External model	Previously published random forest model trained at VUMC and tested on NMCP visits with all available prior Navy health record data for each visit ⁴⁴
Retrained model	Random forest model trained identically as external model using all visits to NBHCs and TPCs (with all prior Navy record data for those visits) and validated with cross validation
Calibrated model	Retrained model with a subsequent layer of calibration using logistic calibration
Updated model	Either the retrained or calibrated model trained again with additional Navy specific features
Domain validation	Random forest trained identically to the external model only at NBHC Little Creek, validated with bootstrapping, and tested at other NBHCs
Temporal validation	Random forest trained identically to the external model with 2 y of NBHC and TPC visits and tested on following 3 y

Abbreviations: NBHC, Naval Branch Health Clinics; NMCP, Naval Medical Center Portsmouth; TPC, TRICARE Prime Clinic; VUMC, Vanderbilt University Medical Center.

recorded Navy-wide inpatient encounters, ambulatory records, out-of-network care, and medical care received while in the military theater of operation (eg, Operation Iraqi Freedom and Operation Enduring Freedom). *ICD-10-CM* codes, used as input features for the model, were transposed into *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* codes using the *ICD-9-CM to ICD-10-CM General Equivalence* mapping from CMS.⁵² Medications were sourced from the Pharmacy Detail Transaction Service, where all pharmacy transactions are logged.

The study data set was prepared recordwise, in which each encounter (record) for each patient (participant) was right-censored at the time of the encounter to replicate clinical data that would have been available to a predictive model running at that time. Demographic variables were retrieved from the most recent record before the prediction date; otherwise the prediction date was discarded. Body mass index (BMI) variables were omitted because of low availability. Race and ethnicity were limited to African American or Black; Asian American; Hispanic; non-Hispanic; White; and unknown, as used in the civilian-based model. All other race and ethnicity variables were encoded as other (eTable 1 in Supplement 2).

Outcomes and Model Evaluation

The outcomes of interest were *ICD-9-CM* and *ICD-10-CM* suicidal behavior codes recorded on the day of the visit or during the next 30 days. Suicidal behavior codes were defined by the National Center for Health Statistics.⁵³ All *ICD-9-CM* suicidal behavior codes used were identical to those used during the development of the statistical model at VUMC.

Statistical Analysis

To evaluate model performance, discrimination was assessed using area under the receiver operating curve (AUROC), area under the precision recall curve (AUPRC), and Brier score. Calibration was assessed using Spiegelhalter z-test statistic, with a significance level of .05. Values less than this significance level indicate miscalibration. Risk concentration was performed and numbers needed to screen (NNS) were calculated for each quantile of risk. The Spiegelhalter z test was performed using R package rms version 6.3-0 (R Project for Statistical Computing).

Retraining and Updating

After external validation of the civilian-based model on Naval health system data, the random forest model was retrained using 5-fold cross-validation using all visits to NBHCs and TPCs. Retraining was nested record-wise, in which patients were split into training, calibration, and testing sets and all records for each patient were sampled together to avoid training and testing on records for the same patients. The initial training set in each fold was cut from 80% to 75% to create a 5% independent calibration set. Recalibration was performed using logistic calibration.

Model updating added US Navy demographics and DoD extender *ICD-9-CM* and unique *ICD-10-CM* codes as features (eTable 2 in Supplement 1).⁵⁴ Armed Forces Qualification Test (AFQT) categories, time in service, rank group, marital status, education level, and expanded race and ethnicity categories were used as recorded by DMDC (eTable 1 in Supplement 2). Sex and age were the only demographic features present in the external model that were not added or expanded in the retrained model. Broadly, race and ethnicity variables were included having found to be important in the external model.

Feature Importance

We compared feature importance weights across external (civilian), retrained, and updated models. Impurity, a common measure of importance used here, measured the amount of variance in the responses at the terminal nodes of the random forest when including or excluding a given feature. Higher variance suggested higher importance. We note impurity assesses importance but not direction, ie, an important feature might be associated with higher or lower risk of the target in question.

Temporal and Domain Validation

To test temporal and domain effects, a second domain model was retrained using all visits to NHBC Little Creek (NBHC-LC). This domain validation strategy was intended to simulate a common scenario of training on a given site's data with external validation at different sites in a larger health system. The NBHC-LC model was tested at the remaining NBHCs and TPCs and on all NBHC and TPC visits that did not belong to a person who visited NBHC-LC during the study period. Harrell bias correction, modified to test on held out bootstrap samples, similar to Efron 0.632 and 632+ methods, was used to assess performance at NBHC-LC.⁵⁵ A third temporal model was retrained using 2 years of NBHC and TPC visits and tested with the next 3 years of visits on both the same service members, and different service members, than those the model was trained on previously.

Results

Demographic characteristics of service members were determined at the beginning of receiving care at both NMCP and NMCP NBHCs and TPCs (**Table 2**). Of the 260 583 service members, 206 412 (79.2%) were male; 104 835 (40.2%) were aged 20 to 24 years; and 9458 (3.6%) were Asian, 56 715 (21.8%) were Black or African American, and 158 277 (60.7%) were White; 200 568 (77.0%) visited an NMCP NBHC or TPC. In total, the Navy health system data set contained 6 476 555 visits to NMCP after 442 789 visits (6.4%) were removed for lack of demographic characteristics. Of NMCP visits, 255 089 (3.9%) were emergency visits, 233 497 (3.6%) were inpatient visits, and 5 987 969 (92.5%) were outpatient visits. Visits to NMCP NBHCs and TPCs accounted for 2 418 393 (37.3%) of all NMCP visits and were all outpatient. Of NMCP NBHC and TPC visits, 19 694 (0.8%) were to TPCs.

At the beginning of receiving care at NMCP NBHCs and TPCs, service members were slightly older, more frequently married, and had more military experience than those receiving care at NMCP (Table 2). No apparent race, ethnicity, or education differences were found between participants visiting NMCP and NMCP NBHCs and TPCs. Of the 260 583 service members to visit NMCP during this time period, 6529 (2.5%) had a recorded suicidal behavior, of whom 2653 (40.6%) had a recorded suicidal behavior on the day of or within 30 days of a visit. Of the 6 476 555 visits to NMCP, 16 715 (0.26%) preceded a suicidal behavior (**Table 3**). After the data set was developed, and the external model applied to it, performance was assessed.

NMCP and TPCs had a greater service member and visit prevalence than NBHCs. Visit case prevalence was consistent across the NBHCs. Discrimination in terms of AUPRC was consistent across sizable NBHCs and increased at smaller NBHCs and TPCs. NMCP, NBHCs, and TPCs all had uncalibrated performance from the external algorithm. For reference, performance of the external civilian-based model when tested enterprise-wide had an AUROC of 0.836, Brier score of 0.009, and poor initial calibration, as reported previously.⁴⁵

Model Retraining, Calibration, and Updating

For NMCP NBHCs and TPCs, 5-fold cross validation was performed with and without DoD- and Navy-specific variables (**Table 4**). These variables were race (24 categories), ethnicity (23 categories), marital status, education level, AFQT categories, service rank, and time in service. DoD-specific ICD variables added included ICD-9-CM extenders related to traumatic brain injury (TBI) and ICD-10-CM codes tracking the severity, presence, and history of TBI. TBI is known to affect suicide risk.⁵⁶ A full added feature list is available within eTable 2 in [Supplement 1](#).

Mean AUROC and AUPRC within NMCP clinics were 0.92 (95% CI, 0.91-0.93) and 0.62 (95% CI, 0.60-0.65) across the 5 folds before the addition of DoD- and Navy-specific variables. After the addition of DoD- and Navy-specific variables, discrimination improved marginally with a mean AUROC of 0.92 (95% CI, 0.91-0.93) and a mean AUPRC of 0.66 (95% CI, 0.63-0.68).

Logistic calibration was applied to each fold. Mean Spiegelhalter z-test statistic *P* value and Brier score before calibration were <.001 and 0.0004, respectively (Table 4). Calibration was successful,

Table 2. Demographics of US Navy Active-Duty Service Members at NMCP and NMCP NBHCs and TPCs Included in Study Data, 2007 to 2017

Characteristic	Service members, No. (%)	
	NMCP (n = 260 583)	NBHCs and TPCs (n = 200 568)
Age, y		
17-19	26 509 (10.17)	17 867 (8.91)
20-24	104 835 (40.23)	76 535 (38.16)
25-29	57 342 (22.01)	46 947 (23.41)
30-34	31 158 (11.96)	25 052 (12.49)
35-39	23 254 (8.92)	19 069 (9.51)
40-44	11 546 (4.43)	9974 (4.97)
≥45	5939 (2.28)	5124 (2.55)
Sex		
Female	54 171 (20.79)	44 306 (22.09)
Male	206 412 (79.21)	156 262 (77.91)
Race		
American Indian or Alaska Native, Black or African American	2009 (0.77)	1548 (0.77)
American Indian or Alaska Native, White	8440 (3.24)	6464 (3.22)
American Indian or Alaskan Native	9428 (3.62)	7295 (3.64)
Asian	9458 (3.63)	6914 (3.45)
Black or African American	56 715 (21.76)	44 033 (21.95)
Native Hawaiian or Other Pacific Islander	2157 (0.83)	1600 (0.8)
Unknown	7281 (2.79)	5525 (2.75)
White	158 277 (60.74)	122 080 (60.87)
Other ^a	6818 (2.62)	5109 (2.56)
Ethnicity		
Filipino	4263 (1.64)	3070 (1.53)
Latin American with Hispanic descent	9266 (3.56)	6950 (3.47)
Mexican	5025 (1.93)	3724 (1.86)
Unknown	25 237 (9.68)	18 788 (9.37)
No ethnicity	172 558 (66.22)	134 337 (66.98)
Other Hispanic descent	17 896 (6.87)	13 683 (6.82)
Puerto Rican	2915 (1.12)	2333 (1.16)
US or Canadian Indian Tribes	5713 (2.19)	4338 (2.16)
Other ^a	17 710 (6.82)	13 345 (6.65)
Marital status		
Married	118 547 (45.49)	96 708 (48.22)
Never married	141 636 (54.35)	103 497 (51.6)
Other ^a	400 (0.15)	363 (0.18)
Education level		
Adult education diploma	2459 (0.94)	1884 (0.94)
Associate's degree	8940 (3.43)	7335 (3.66)
Baccalaureate degree	24 431 (9.38)	18 935 (9.44)
Completed 1 semester of college, no high school diploma	5048 (1.94)	3928 (1.96)
Doctorate degree	1838 (0.71)	1210 (0.6)
High school diploma	192 126 (73.73)	147 338 (73.46)
Master's degree	6189 (2.38)	5121 (2.55)
Non-high school graduate	1860 (0.71)	1456 (0.73)
Test-based equivalency diploma	4432 (1.7)	3513 (1.75)
Unknown	11 737 (4.5)	8669 (4.32)
Other ^a	1523 (0.58)	1179 (0.59)

(continued)

Table 2. Demographics of US Navy Active-Duty Service Members at NMCP and NMCP NBHCs and TPCs Included in Study Data, 2007 to 2017 (continued)

Characteristic	Service members, No. (%)	
	NMCP (n = 260 583)	NBHCs and TPCs (n = 200 568)
AFQT category		
I	14 495 (5.56)	10 901 (5.44)
II	89 880 (34.49)	70 790 (35.29)
III A	61 575 (23.63)	46 764 (23.32)
III B	55 126 (21.15)	42 058 (20.97)
IV	1386 (0.53)	1152 (0.57)
Unknown or not applicable	38 121 (14.63)	28 903 (14.41)
Rank		
E1-E3	10 720 (4.11)	8386 (4.18)
E4-E6	101 712 (39.03)	70 761 (35.28)
E7-E9	107 980 (41.44)	89 507 (44.63)
O1-O3	15 582 (5.98)	13 682 (6.82)
W1-W3	23 805 (9.14)	17 565 (8.76)
O4, W4, or higher	784 (0.3)	667 (0.33)
Time in service, y		
<1	39 591 (15.19)	26 869 (13.4)
1-2	27 193 (10.44)	21 880 (10.91)
3-4	21 661 (8.31)	17 646 (8.8)
5-9	11 725 (4.5)	10 631 (5.3)
10-14	49 559 (19.02)	38 335 (19.11)
15-19	49 802 (19.11)	43 304 (21.59)
20-24	57 320 (22.0)	38 470 (19.18)
>24	3732 (1.43)	3433 (1.71)

Abbreviations: AFQT, Armed Forces Qualification Test; E, enlisted; NBHC, Naval Branch Health Clinics; NMCP, Naval Medical Center Portsmouth; O, officer; TPC, TRICARE Prime Clinic; W, warrant officer.

^a Other categories are listed in eTable 1 in Supplement 2.

as indicated by an improved Spiegelhalter z-test statistic *P* value and Brier score of .41 and 0.0003, respectively.

NNS in the top decile of risk of the external model at NMCP NBHCs was 366, compared with 200 for the updated model (eFigure in Supplement 2). The civilian-based model had a lower NNS than the updated model in the bottom decile, although it contained fewer cases. NNS in each decile refers to the numbers of individuals who would need to receive a test to identify 1 individual who will have the outcome in question.

Feature Importance

The external, retrained, and updated model features were ranked by importance (impurity) and compared (Table 5). Both the external and retrained models had 1631 features, while the updated model had 1949 (eTable 2 in Supplement 1). Important external model features included age, past suicidal behavior, recent visits, mood disorders, psychotropics, and substance abuse. Simplified race, from the external model, was not important for the retrained or updated model discrimination, while simplified ethnicity, DoD rank, and DoD marital status were important.

Other DoD- and Navy-specific variables, including medical examinations, counseling, time in service, expanded race, expanded ethnicity, and AFQT categories, were important for updated model discrimination. Outpatient visits, pain relievers, antidepressants, hormones, and antibiotics were also important. Posttraumatic stress disorder (PTSD), alcohol dependence, and benzodiazepines were consistently important between the external and updated models. Updating the algorithm after retraining did not affect the most important features compared with retraining alone.

Table 3. Performance of the Civilian-Based External Suicide Risk Model at NMCP and NBHCs and TPCs

Location	Service members, No.	Service member cases, No. (%)	Visits, No.	Visit cases, No. (%)	AUROC (95% CI)	AUPRC (95% CI)	Brier score (95% CI)	Spiegelhalter z-test statistic (95% CI)	Spiegelhalter z-test P value (95% CI)
NMCP	260 583	2653 (1.02)	6 476 555	16 715 (0.26)	0.73 (0.72 to 0.74)	0.013 (0.010 to 0.014)	0.10 (0.10 to 0.10)	-1018.7 (-1028.4 to -1008.8)	<.001 (<.001 to <.001)
NMCP clinics together									
NBHCs	200 299	647 (0.32)	2 398 699	1400 (0.06)	0.77 (0.74 to 0.79)	0.004 (0.003 to 0.005)	0.08 (0.08 to 0.08)	-680.2 (-684.5 to -677.9)	<.001 (<.001 to <.001)
TPCs	4610	30 (0.65)	19 694	56 (0.28)	0.86 (0.78 to 0.90)	0.024 (-0.003 to 0.034)	0.18 (0.17 to 0.18)	-33.7 (-36.3 to -29.4)	<.001 (<.001 to <.001)
Individual NBHCs									
NBHC Naval Station Sewells	128 405	375 (0.29)	1 172 985	737 (0.06)	0.75 (0.72 to 0.77)	0.003 (0.001 to 0.003)	0.08 (0.08 to 0.08)	-475.2 (-477.9 to -474.0)	<.001 (<.001 to <.001)
NBHC Oceana	54 247	123 (0.23)	481 859	257 (0.05)	0.75 (0.71 to 0.80)	0.006 (0.000 to 0.009)	0.07 (0.07 to 0.07)	-305.2 (-306.9 to -302.6)	<.001 (<.001 to <.001)
NBHC Little Creek	61 007	111 (0.18)	444 927	230 (0.05)	0.82 (0.79 to 0.88)	0.005 (0.000 to 0.008)	0.08 (0.08 to 0.08)	-292.4 (-294.6 to -290.0)	<.001 (<.001 to <.001)
NBHC Dam Neck	29 061	50 (0.17)	175 514	102 (0.06)	0.76 (0.73 to 0.83)	0.016 (-0.035 to 0.029)	0.07 (0.06 to 0.07)	-189.6 (-191.8 to -187.6)	<.001 (<.001 to <.001)
NBHC Naval Shipyard Norfolk	19 399	20 (0.10)	50 531	24 (0.05)	0.93 (0.90 to 0.99)	0.014 (-0.057 to 0.024)	0.07 (0.07 to 0.07)	-96.8 (-98.8 to -95.3)	<.001 (<.001 to <.001)
NBHC Yorktown	5610	12 (0.21)	46 662	22 (0.05)	0.73 (0.50 to 0.96)	0.022 (-0.100 to 0.042)	0.07 (0.07 to 0.08)	-95.8 (-99.0 to -93.6)	<.001 (<.001 to <.001)
NBHC Chesapeake	3292	20 (0.61)	26 221	28 (0.11)	0.85 (0.79 to 0.99)	0.007 (0.001 to 0.011)	0.12 (0.11 to 0.12)	-61.5 (-66.0 to -59.4)	<.001 (<.001 to <.001)

Abbreviations: AUPRC, area under the precision recall curve; AUROC, area under the receiver operating characteristic curve; NBHC, Naval Branch Health Clinics; NMCP, Naval Medical Center Portsmouth; TPC, TRICARE Prime Clinic.

Temporal and Domain Validation

Discrimination for domain validation at NBHC-LC, obtained from bootstrapping, was much greater than both the external model and the locations outside of NBHC-LC with an AUROC of 0.96 (95% CI, 0.91-1.00) and AUPRC of 0.58 (95% CI, 0.39-0.77) (eTable 3 in Supplement 2). Discrimination was greatest at NBHC Oceana and at the TPCs. Discrimination dropped to an AUROC of 0.90 (95% CI, 0.90-0.91) and an AUPRC of 0.01 (95% CI, 0.01-0.01) at non-NBHC-LC locations when service

Table 4. Mean Performance of the Retrained, Updated, and Calibrated Suicide Risk Models at Naval Branch Health Clinics and TRICARE Prime Clinics

Metric	Retrained	Updated	Calibrated and retrained	Calibrated and updated
AUROC	0.92 (0.91 to 0.93)	0.92 (0.91 to 0.93)	0.92 (0.91 to 0.93)	0.920 (0.91 to 0.93)
AUPRC	0.62 (0.60 to 0.65)	0.66 (0.63 to 0.68)	0.62 (0.60 to 0.65)	0.66 (0.63 to 0.68)
Brier	3.74×10^{-4} (3.53×10^{-4} to 3.96×10^{-4})	3.52×10^{-4} (3.35×10^{-4} to 3.73×10^{-4})	3.18×10^{-4} (2.96×10^{-4} to 3.38×10^{-4})	2.93×10^{-4} (2.71×10^{-4} to 3.14×10^{-4})
Spiegelhalter z-test statistic	-9.50	-10.41	0.84	0.76
Spiegelhalter z-test P value	<.001	<.001	.41	.45

Abbreviations: AUPRC, area under the precision recall curve; AUROC, area under the receiver operating characteristic curve.

Table 5. Comparison of the Model Features Between the External, Retrained, and Updated Algorithms

Description	Rank ^a		
	External	Retrained	Updated
Important to external			
Age	1	11	11
Suicide and self-inflicted poisoning by tranquilizers and other psychotropic agents	2	99	116
Suicide and self-inflicted poisoning by other specified drugs and medicinal substances	3	50	68
VUMC race	4	1248	1386
VUMC ethnicity	5	29	44
Outpatient visits (previous year)	6	1	1
Inpatient visits (previous year)	7	4	4
Major depressive affective disorder, recurrent episode, severe, without mention of psychotic behavior	8	56	72
Suicide and self-inflicted poisoning by analgesics, antipyretics, and antirheumatics	9	148	163
Emergency visits (previous year)	10	8	7
Diazepines, oxazepines, thiazepines, and oxepines	11	38	52
Other and unspecified alcohol dependence, unspecified	12	25	39
Major depressive affective disorder, single episode, unspecified	13	22	26
Posttraumatic stress disorder	14	44	57
Important to updated			
Outpatient visits (previous year)	6	1	1
Outpatient visits (1-2 y ago)	20	2	2
Outpatient visits (2-3 y ago)	45	3	3
Inpatient visits (previous year)	7	4	4
Propionic acid derivatives	32	5	5
Anilides	27	6	6
Emergency visits (previous year)	10	8	7
Outpatient visits (3-4 y ago)	72	7	8
Opium alkaloids and derivatives	74	9	9
Selective serotonin reuptake inhibitors	25	10	10
Age	1	11	11
DoD time in service	NA ^b	NA ^b	12
Expectorants	140	12	13
Other specified counseling	NA ^b	NA ^b	14
DoD race	NA ^b	NA ^b	15

Abbreviations: DoD, Department of Defense; NA, not applicable; VUMC, Vanderbilt University Medical Center.

^a Lower number for rank indicates higher importance.

^b Feature was added in the updated model.

members who visited NBHC-LC were removed. Calibration was poorest at NBHCs Oceana and Yorktown.

Discrimination for temporal validation resembled external validation when the third model was trained with early NBHC and TPC visits and then tested with later visits. Performance for service members seen by the third model had the same discrimination (AUROC, 0.74 [95% CI, 0.71-0.77]; AUPRC, 0.003 [95% CI, 0.003-0.005]) compared with not seen (AUROC, 0.75 [95% CI, 0.72-0.78]; AUPRC, 0.003 [95% CI, 0.003-0.005]).

Discussion

This cohort study was the first to our knowledge to comprehensively validate, retrain, and update models of suicide risk in US Navy health system data. Primary findings include an assessment of the performance of a civilian-based suicide risk model and a comparison of retrained and updated models specific for a US Navy population. External validation of predictive algorithms has been relatively rare.⁵⁷ We note a generally low rate of suicidal behaviors (2.5%) in this cohort across the study period, a positive finding for military health.

Consistent discrimination with poor calibration was obtained from applying the external model to NMCP, NBHCs, and TPCs. Retraining with US Naval data resulted in improved apparent performance with the key caveat that internal validation (retraining) had higher performance (more optimism) than external validation, but external validation was a more rigorous test. Both internal and external validation were performed in this study. Logistic calibration was successfully able to recalibrate model performance as expected. Model updating was associated with marginal improvement compared with only retraining and indicates a need for further DoD-specific feature collection and engineering. For example, mental health assessments were present in a small percentage of these overall study data, but they might prove important in specific clinical settings or in cohorts enriched for their use.

US Navy data, however, were still important for model performance. Expanded race was helpful, while the externally defined simplified race was not helpful. Valid risk factors were seen in the most important model features. Notable differences between the external and updated model features were conditions, such as PTSD, which when reported during active duty, may result in an earlier transition.⁵⁸ PTSD and other known risk factors, including Vitamin A derivatives, were within the top 1% and 2% of features.^{59,60} Caution should be taken when interpreting importance values, and we underline our reported statistics remain correlative, not causal, in inference.

Greater performance than in external validation was found in the domain validation, with optimistic apparent performance at the training location, NBHC-LC. However, when service members who visited NBHC-LC were removed, performance dropped to a level only slightly better than external validation. This removal ensured performance increases did not result from model bias toward service members available to the model in retraining. In the temporal validation, both service members seen by the model and not seen by the model had the same discrimination. Both were equal to external validation, indicating that adopting an external model in some settings may have similar prospective performance to retraining with less time and cost.

Implications of this study suggest a civilian-based suicide risk model alone would achieve feasible NNS for an automated risk assessment first-pass to be supported or refuted by clinician evaluation and judgment.⁴⁵ In clinical use, a model like this one might have application to prompt further screening or discussion of suicide to prevent covert distress. It also might serve to guide populational health management by quantifying burden of suicidality by clinical setting, geography, or other relevant grouping. With respect to modeling implications, retraining improved precision of this model further (lowering NNS and false positives) while model updating did not, in this case, improve performance notably. Of note, some highly relevant assessments (eg, the mental health assessment) were not present in sufficient quantities in this cohort to justify updating at scale. Future work to study cohorts with these and other assessments is indicated.

US Navy data in this study are comprehensive with respect to ascertainment of health care data, including purchase claims outside NMCP, pharmacy data, and study outcomes. These data included transitions through clinics and hospitals, into the military theater, and into intensive therapy or treatment regimens. These strengths contrast with most civilian health systems in which care is open and health care information exchange is generally lacking, with key exceptions being payer-based systems and states in which significant investment in information exchange has been made.⁶¹ This study used multiple validation strategies, including replicating the mode in predictive modeling, in which a model is trained in one location and validated elsewhere, either across differing clinical settings (domain validation), time, or geography.⁶²

Limitations

This study has limitations. Different external model performance between VUMC and the US Navy, suggested by AUROC, contrasted with the markedly optimistic performance metrics obtained from retraining, may indicate that a retrained model may be more overfit. Markers of treatment and utilization, or normally undertaken clinical scenarios, may drive model performance instead of imminent risk factors. Self-reported data obtained during routine assessments and known risk factors for suicide in the military, which would likely improve the model, such as predeployment nightmares, were not explicitly represented.⁶³ Further exploration of available DoD-specific features including more specific, localized model development, eg, within specific clinical settings, might demonstrate more value to model updating with US Navy data. Data from within assessments (eg, Pre/Post Deployment Health Assessments) were not used as features, only that information that the assessments were performed.

Further refinement of the retraining approach and application of the model to unseen data are needed. In this study, the retrained algorithm, using a smaller fraction of the data focused on primary care clinics, was able to generalize to other clinics similar to the external model. Work toward assessing the feasibility of obtaining the comprehensive data used here at run time in the US Navy is needed. Additionally, deep learning algorithmic approaches might facilitate transfer learning in which a pretrained model (civilian-based EHR) is improved through learning on the novel data set (US Naval EHR).

Conclusions

In this cohort study of active-duty Navy service members, we externally validated, retrained, and updated civilian-based statistical models of suicide attempt risk. The external civilian-based risk model estimated risk of suicide attempt on the day of or within 30 days of every visit to NMCP, NBHCs, and TPCs. The model had similar discrimination to models trained and updated in temporal and domain validation, suggesting retraining might not be necessary prior to next steps in deployment of algorithms like these in military clinical settings. Evaluating prior published models prior to time-consuming novel model development might facilitate similar applied modeling efforts.

ARTICLE INFORMATION

Accepted for Publication: September 29, 2023.

Published: November 8, 2023. doi:[10.1001/jamanetworkopen.2023.42750](https://doi.org/10.1001/jamanetworkopen.2023.42750)

Open Access: This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2023 Ripperger MA et al. *JAMA Network Open*.

Corresponding Author: Colin G. Walsh, MD, MA, Vanderbilt University Medical Center, 2525 West End Ave, Ste 1475, Nashville, TN 37203 (colin.walsh@vumc.org).

Author Affiliations: Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee (Ripperger, Kolli, Wilimitis, Robinson, Novak, Walsh); Department of Anesthesiology, Vanderbilt

University Medical Center, Nashville, Tennessee (Reale); Retired, US Navy, Portsmouth, Virginia (Cunningham); Daniel K. Inouye Graduate School of Nursing, Uniformed Services University of the Health Sciences, Bethesda, Maryland (Kasuske); Naval Medical Center Portsmouth, US Navy, Portsmouth, Virginia (Grover); Department of Psychology, Florida State University, Tallahassee (Ribeiro); Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (Walsh); Department of Psychiatry, Vanderbilt University Medical Center, Nashville, Tennessee (Walsh).

Author Contributions: Mr Ripperger and Dr Walsh had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Kolli, Wilimitis, Kasuske, Ribeiro, Walsh.

Acquisition, analysis, or interpretation of data: Ripperger, Kolli, Robinson, Reale, Novak, Cunningham, Kasuske, Grover, Ribeiro, Walsh.

Drafting of the manuscript: Ripperger, Kolli, Cunningham, Walsh.

Critical review of the manuscript for important intellectual content: Kolli, Wilimitis, Robinson, Reale, Novak, Cunningham, Kasuske, Grover, Ribeiro, Walsh.

Statistical analysis: Ripperger, Walsh.

Obtained funding: Ribeiro, Walsh.

Administrative, technical, or material support: Kolli, Robinson, Novak, Cunningham, Kasuske, Grover, Ribeiro, Walsh.

Supervision: Robinson, Cunningham, Kasuske, Walsh.

Conflict of Interest Disclosures: Ms Reale reported receiving grants from the Evelyn Selby Stead Fund for Innovation and the National Institute of Mental Health during the conduct of the study. Dr Novak reported receiving grants from the Military Suicide Research Consortium (MSRC) during the conduct of the study. Dr Ribeiro reported receiving grants from the Department of Defense during the conduct of the study. Dr Walsh reported receiving personal fees from Humana and grants from Veranne outside the submitted work. No other disclosures were reported.

Funding/Support: This work was funded by the MSRC, an effort supported by the Office of the Assistant Secretary of Defense for Health Affairs (Award No. W81XWH-10-2-0181). The Vanderbilt University Medical Center study team also received support from National Institute of Mental Health (grant Nos. R01MH121455 and R01MH116269).

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.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of Defense nor the US government.

Data Sharing Statement: See Supplement 3.

Additional Contributions: The study team would like to acknowledge the collaboration of the Navy and Marine Corps Public Health Center, EpiData Center, key personnel Tina Luse, MPH, and Christine Glasheen, MPH, for invaluable, uncompensated aid in understanding and obtaining Navy health system data used in the study.

Additional Information: Capt (ret) Craig A. Cunningham and Cdr Shawna Grover are military service members. This work was prepared as part of their official duties. Title 17 USC 105 provides that "Copyright protection under this title is not available for any work of the United States Government." Title 17 USC 101 defines a US government work as a work prepared by a military service member or employee of the US government as part of that person's official duties. The study protocol (NMCP.2018.0060) was approved by the Navy Medical Center Portsmouth institutional review board in compliance with all applicable federal regulations governing the protection of human subjects.

REFERENCES

1. 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
2. Kang HK, Bullman TA, Smolenski DJ, Skopp NA, Gahm GA, Reger MA. Suicide risk among 1.3 million veterans who were on active duty during the Iraq and Afghanistan wars. *Ann Epidemiol*. 2015;25(2):96-100. doi:10.1016/j.annepidem.2014.11.020
3. Department of Defense. Annual suicide report: calendar year 2020. Accessed April 27, 2022. <https://www.dspo.mil/Portals/113/Documents/CY2020Suicide%20Report/CY%202020%20Annual%20Suicide%20Report.pdf?ver=00wlvDd-PJuA-igow5fBFA%3d%3d>

4. Naifeh JA, Ursano RJ, Kessler RC, et al. Suicide attempts among activated soldiers in the U.S. Army reserve components. *BMC Psychiatry*. 2019;19(1):31. doi:10.1186/s12888-018-1978-2
5. Ursano RJ, Kessler RC, Naifeh JA, et al. Associations of time-related deployment variables with risk of suicide attempt among soldiers: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2018;75(6):596-604. doi:10.1001/jamapsychiatry.2018.0296
6. Naifeh JA, Ursano RJ, Kessler RC, et al. Transition to suicide attempt from recent suicide ideation in U.S. Army soldiers: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Depress Anxiety*. 2019;36(5):412-422. doi:10.1002/da.22870
7. Mash HBH, Ursano RJ, Kessler RC, et al. Predictors of suicide attempt within 30 days after first medically documented suicidal ideation in U.S. Army soldiers. *Am J Psychiatry*. 2021;178(11):1050-1059. doi:10.1176/appi.ajp.2021.20111570
8. Nock MK, Stein MB, Heeringa SG, et al; Army STARRS Collaborators. Prevalence and correlates of suicidal behavior among soldiers: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2014;71(5):514-522. doi:10.1001/jamapsychiatry.2014.30
9. Crouch C, Curley JM, Carreno JT, Wilk JE. Return to duty practices of Army behavioral health providers in Garrison. *Mil Med*. 2018;183(11-12):e617-e623. doi:10.1093/milmed/usy103
10. Curley JM, McDonald JL, Nugent KL, et al. Soldier attitudes toward behavioral health profiles in the US Army. *Mil Med*. 2020;185(1-2):84-91. doi:10.1093/milmed/usz133
11. Heyman RE, Slep AMS, Parsons AM, Ellerbeck EL, McMillan KK. Systematic review of the military career impact of mental health evaluation and treatment. *Mil Med*. 2022;187(5-6):e598-e618. doi:10.1093/milmed/usab283
12. Chu C, Stanley IH, Marx BP, et al. Associations of vulnerability to stressful life events with suicide attempts after active duty among high-risk soldiers: results from the Study to Assess Risk and Resilience in Servicemembers-longitudinal study (STARRS-LS). *Psychol Med*. 2023;53(9):4181-4191. doi:10.1017/S0033291722000915
13. Dempsey CL, Benedek DM, Nock MK, et al. Social closeness and support are associated with lower risk of suicide among U.S. Army soldiers. *Suicide Life Threat Behav*. 2021;51(5):940-954. doi:10.1111/sltb.12778
14. Hyman J, Ireland R, Frost L, Cottrell L. Suicide incidence and risk factors in an active duty US military population. *Am J Public Health*. 2012;102(Suppl 1)(suppl 1):S138-S146. doi:10.2105/AJPH.2011.300484
15. Nock MK, Deming CA, Fullerton CS, et al. Suicide among soldiers: a review of psychosocial risk and protective factors. *Psychiatry*. 2013;76(2):97-125. doi:10.1521/psyc.2013.76.2.97
16. Kessler RC, Warner CH, Ivany C, et al; Army STARRS Collaborators. 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(1):49-57. doi:10.1001/jamapsychiatry.2014.1754
17. Shen YC, Cunha JM, Williams TV. Time-varying associations of suicide with deployments, mental health conditions, and stressful life events among current and former US military personnel. Naval Postgraduate School. Accessed October 11, 2021. <https://calhoun.nps.edu/handle/10945/56669>
18. Ravindran C, Morley SW, Stephens BM, Stanley IH, Reger MA. Association of suicide risk with transition to civilian life among US military service members. *JAMA Netw Open*. 2020;3(9):e2016261. doi:10.1001/jamanetworkopen.2020.16261
19. Kearns JC, Edwards ER, Finley EP, et al. A practical risk calculator for suicidal behavior among transitioning U.S. Army soldiers: results from the Study to Assess Risk and Resilience in Servicemembers-Longitudinal Study (STARRS-LS). *Psychol Med*. Published online March 9, 2023. doi:10.1017/S0033291723000491
20. 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
21. Simon GE, Matarazzo BB, Walsh CG, et al. Reconciling statistical and clinicians' predictions of suicide risk. *Psychiatr Serv*. 2021;72(5):555-562. doi:10.1176/appi.ps.202000214
22. Ribeiro JD, Gutierrez PM, Joiner TE, et al. Health care contact and suicide risk documentation prior to suicide death: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *J Consult Clin Psychol*. 2017;85(4):403-408. doi:10.1037/ccp0000178
23. Bentley KH, Zuromski KL, Fortgang RG, et al. Implementing machine learning models for suicide risk prediction in clinical practice: focus group study with hospital providers. *JMIR Form Res*. 2022;6(3):e30946. doi:10.2196/30946
24. Zuromski KL, Dempsey CL, Ng THH, et al. Utilization of and barriers to treatment among suicide decedents: Results from the Army Study to Assess Risk and Resilience Among Servicemembers (Army STARRS). *J Consult Clin Psychol*. 2019;87(8):671-683. doi:10.1037/ccp0000400

25. Ryan AT, Ghahramanlou-Holloway M, Wilcox HC, Umhau JC, Deuster PA. Mental health care utilization and psychiatric diagnoses in a sample of military suicide decedents and living matched controls. *J Nerv Ment Dis*. 2020;208(9):646-653. doi:10.1097/NMD.0000000000001192
26. Naifeh JA, Mash HBH, Stein MB, Fullerton CS, Kessler RC, Ursano RJ. The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS): progress toward understanding suicide among soldiers. *Mol Psychiatry*. 2019;24(1):34-48. doi:10.1038/s41380-018-0197-z
27. Belsher BE, Smolenski DJ, Pruitt LD, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. *JAMA Psychiatry*. 2019;76(6):642-651. doi:10.1001/jamapsychiatry.2019.0174
28. Kessler RC, Bossarte RM, Luedtke A, Zaslavsky AM, Zubizarreta JR. Suicide prediction models: a critical review of recent research with recommendations for the way forward. *Mol Psychiatry*. 2020;25(1):168-179. doi:10.1038/s41380-019-0531-0
29. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. 2017;143(2):187-232. doi:10.1037/bul0000084
30. Kessler RC, Stein MB, Petukhova MV, et al; Army STARRS Collaborators. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry*. 2017;22(4):544-551. doi:10.1038/mp.2016.110
31. Bernecker SL, Zuromski KL, Gutierrez PM, et al. Predicting suicide attempts among soldiers who deny suicidal ideation in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Behav Res Ther*. 2019;120:103350. doi:10.1016/j.brat.2018.11.018
32. Nock MK, Millner AJ, Joiner TE, et al. Risk factors for the transition from suicide ideation to suicide attempt: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *J Abnorm Psychol*. 2018;127(2):139-149. doi:10.1037/abn0000317
33. O'Connor RC, Kirtley OJ. The integrated motivational-volitional model of suicidal behaviour. *Philos Trans R Soc Lond B Biol Sci*. 2018;373(1754):20170268. doi:10.1098/rstb.2017.0268
34. Pestian J, Santel D, Sorter M, et al. A machine learning approach to identifying changes in suicidal language. *Suicide Life Threat Behav*. 2020;50(5):939-947. doi:10.1111/sltb.12642
35. Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry*. 2018;175(10):951-960. doi:10.1176/appi.ajp.2018.17101167
36. Kessler RC, Hwang I, Hoffmire CA, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration. *Int J Methods Psychiatr Res*. 2017;26(3):e1575. doi:10.1002/mpr.1575
37. Lindsell CJ, Stead WW, Johnson KB. Action-informed artificial intelligence—matching the algorithm to the problem. *JAMA*. 2020;323(21):2141-2142. doi:10.1001/jama.2020.5035
38. 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
39. Kline-Simon AH, Sterling S, Young-Wolff K, et al. Estimates of workload associated with suicide risk alerts after implementation of risk-prediction model. *JAMA Netw Open*. 2020;3(10):e2021189. doi:10.1001/jamanetworkopen.2020.21189
40. Reale C, Novak LL, Robinson K, et al. User-centered design of a machine learning intervention for suicide risk prediction in a military setting. *AMIA Annu Symp Proc*. 2021;2020:1050-1058.
41. Vasey B, Nagendran M, Campbell B, et al; DECIDE-AI expert group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. 2022;377:e070904. doi:10.1136/bmj-2022-070904
42. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005
43. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med*. 2021;385(3):283-286. doi:10.1056/NEJMc2104626
44. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci*. 2017;5(3):457-469. doi:10.1177/2167702617691560
45. Walsh CG, Johnson KB, Ripperger M, et al. Prospective validation of an electronic health record-based, real-time suicide risk model. *JAMA Netw Open*. 2021;4(3):e211428. doi:10.1001/jamanetworkopen.2021.1428

46. Wilimitis D, Turer RW, Ripperger M, et al. Integration of face-to-face screening with real-time machine learning to predict risk of suicide among adults. *JAMA Netw Open*. 2022;5(5):e2212095. doi:10.1001/jamanetworkopen.2022.12095
47. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. 2019;13(suppl 1):S31-S34. doi:10.4103/sja.SJA_543_18
48. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
49. National Center for Health Statistics. *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. Published November 3, 2021. Accessed June 23, 2022. <https://www.cdc.gov/nchs/icd/icd9cm.htm>
50. US Centers for Medicare & Medicaid Services. 2014 Model Software/ICD-9-CM mappings. Accessed June 23, 2022. <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors-Items/Risk2014>
51. WHO Collaborating Centre for Drug Statistics Methodology. ATC alterations from 2005-2021. Accessed June 23, 2022. https://www.whocc.no/atc_ddd_alterations_cumulative/atc_alterations/
52. U.S. Centers for Medicare & Medicaid Services. 2018 ICD-10 CM and GEMs. Accessed July 20, 2022. <https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs>
53. Hedegaard H, Schoenbaum M, Claassen C, Crosby A, Holland K, Proescholdbell S. Issues in developing a surveillance case definition for nonfatal suicide attempt and intentional self-harm using *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)* coded data. *Natl Health Stat Report*. 2018;(108):1-19.
54. American Academy of Professional Coders. Capture active duty diagnoses with DoD unique codes. AAPC Knowledge Center. Published October 1, 2016. Accessed July 21, 2022. <https://www.aapc.com/blog/36452-capture-active-duty-diagnoses-with-dod-unique-codes/>
55. Iba K, Shinozaki T, Maruo K, Noma H. Re-evaluation of the comparative effectiveness of bootstrap-based optimism correction methods in the development of multivariable clinical prediction models. *BMC Med Res Methodol*. 2021;21(1):9. doi:10.1186/s12874-020-01201-w
56. Hostetter TA, Hoffmire CA, Forster JE, Adams RS, Stearns-Yoder KA, Brenner LA. Suicide and traumatic brain injury among individuals seeking Veterans Health Administration services between fiscal years 2006 and 2015. *J Head Trauma Rehabil*. 2019;34(5):E1-E9. doi:10.1097/HTR.0000000000000489
57. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25-34. doi:10.1016/j.jclinepi.2014.09.007
58. Schafer KM, Duffy M, Kennedy G, et al. Suicidal ideation, suicide attempts, and suicide death among Veterans and service members: a comprehensive meta-analysis of risk factors. *Mil Psychol*. 2022;34(2):129-146. doi:10.1080/08995605.2021.1976544
59. Suuberg A. Psychiatric and developmental effects of isotretinoin (retinoid) treatment for acne vulgaris. *Curr Ther Res Clin Exp*. 2019;90:27-31. doi:10.1016/j.curtheres.2019.01.008
60. Cooper SA, Szymanski BR, Bohnert KM, Sripada RK, McCarthy JF. Association between positive results on the primary care-posttraumatic stress disorder screen and suicide mortality among US veterans. *JAMA Netw Open*. 2020;3(9):e2015707. doi:10.1001/jamanetworkopen.2020.15707
61. Han X, Lowry TY, Loo GT, et al. Expanding health information exchange improves identification of frequent emergency department users. *Ann Emerg Med*. 2019;73(2):172-179. doi:10.1016/j.annemergmed.2018.07.024
62. Ursano RJ, Colpe LJ, Heeringa SG, Kessler RC, Schoenbaum M, Stein MB; Army STARRS collaborators. The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Psychiatry*. 2014;77(2):107-119. doi:10.1521/psyc.2014.77.2.107
63. van Liempt S, van Zuiden M, Westenberg H, Super A, Vermetten E. Impact of impaired sleep on the development of PTSD symptoms in combat veterans: a prospective longitudinal cohort study. *Depress Anxiety*. 2013;30(5):469-474. doi:10.1002/da.22054

SUPPLEMENT 1.

eTable 2. External, Retrained, and Updated Model Features

SUPPLEMENT 2.

eTable 1. All Department of Defense (DoD) Categories for Race, Ethnicity, Marital Status, and Education Level

eFigure. Case Prevalence and Numbers Needed to Screen (NNS) for the External and Updated Models

eTable 3. Performance of the NBHC-LC Suicide Risk Model for Domain Validation at Other NBHCs and TPCs and at NBHC-LC via Bootstrapping

SUPPLEMENT 3.

Data Sharing Statement