

Predicting US Army Soldier Suicide-Related Behaviors at the Time of Periodic Health Assessments

Emily R. Edwards, PhD; Shelby Borowski, MPH, PhD; Matthew K. Nock, PhD; David M. Benedek, MD; Kate H. Bentley, PhD; Joseph Geraci, PhD; Sarah M. Gildea, BS; Marianne Goodman, MD; Amit Gupta, MD; Chris J. Kennedy, PhD; Andrew J. King, MS; Evan M. Kleiman, PhD; Howard Liu, SD; Alex Luedtke, PhD; Sarah Maggio, PhD; James A. Naifeh, PhD; Thomas H. Nassif, PhD; Chris Paine, PhD; Nancy A. Sampson, BA; Nur Hani Zainal, PhD; James Wagner, PhD; Murray B. Stein, MD; Vincent F. Capaldi II, MD, MSc; Robert J. Ursano, MD; Ronald C. Kessler, PhD

[+ Supplemental content](#)

IMPORTANCE Suicide is the leading cause of death among active-duty US Army soldiers. Evidence-based preventive interventions exist but need to be targeted to be cost-effective.

OBJECTIVES To develop machine learning models using administrative data available during periodic health assessments (PHAs) to determine eligibility for a remote group dialectical behavior therapy-based skills training intervention for soldiers with elevated risk of suicide or nonfatal suicide attempt.

DESIGN, SETTING, AND PARTICIPANTS From regular US Army soldier PHAs completed from 2015 to 2019, separate 70% training samples were created for suicides and SAs. Each training sample contained all cases and a stratified equal-probability sample of 10 times as many controls, with inverse-probability of selection weights applied to controls. Model performance was evaluated in the remaining 30% test sample. Data analysis was conducted September 2025 through January 2026.

MAIN OUTCOMES AND MEASURES Suicides were recorded in the National Death Index for soldiers either still in or out of active service. SAs were recorded in US Army records only for soldiers still in active service. A 24-month risk horizon was used because successive PHAs, though designed to occur annually, are sometimes separated by this much time. A loss to follow-up weight was used in the SA model to adjust for leaving service within 24 months.

RESULTS PHA data were available from 668 684 regular US Army soldiers. A total of 1 932 243 PHAs (including 85.5% by male soldiers, 46.7% by soldiers aged 28 years or older, and 58.1% by married soldiers) completed during the 2015 to 2019 period were included in analyses. Over 24 months, suicide prevalence was 66.7 per 100 000 (SE = 3.6) and SA prevalence was 704.1 per 100 000 (SE = 14.7) adjusted for loss to follow-up. Test sample area under the receiver operating characteristic curve (AUROC) for suicide was 0.72 (SE = 0.17), with Integrated Calibration Index (ICI) of 0.0003 and Brier score of 0.0007. Test sample AUROC for SA was 0.81 (SE = 0.02), with ICI of 0.0014 and Brier score of 0.0070. Suicide prevalence was meaningfully elevated (at least twice that expected by chance) only in the 5% of PHAs with highest predicted risk (sensitivity = 18.6%; SE = 2.0%). SA risk was elevated only in the 10% of PHAs with highest predicted risk (sensitivity = 46.5%; SE = 1.0%). Predicted probabilities of suicide and SA were correlated (Pearson $r = 0.29$; $t = 233.8$; $P < .001$). Overall, 11.9% of PHAs revealed US Army soldiers at elevated risk of either suicide or SA, and 3.1% were at elevated risk of both outcomes.

CONCLUSIONS AND RELEVANCE Results of this prognostic study suggest that as suicides and SAs have distinct predictors, an attempt to optimize resource allocation will require thoughtful postintervention considerations of costs and benefits of interventions to optimize net benefit.

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2026.2410
Published online August 12, 2026.

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, Boston, MA 02115 (kessler@hcp.med.harvard.edu).

Suicide has been the leading cause of death among US military service members since 2008, with suicide increasing annually.¹ In 2022, 333 suicides (39% of total deaths) and 1278 nonfatal suicide attempts (SAs) were documented among active-duty military personnel.^{1,2} Although effective suicide prevention interventions exist,³⁻⁶ they must be targeted to high-risk individuals to be cost-effective.⁷ However, military culture discourages disclosure of psychological distress,⁸ causing standard self-report screening to miss most suicide-related behaviors.⁹

Machine learning (ML) risk prediction offers a way to overcome this limitation. In a recent study,⁹ an ML model using a limited set of US Army administrative variables predicted elevated SA risk within 6 months of the periodic health assessment (PHA), an annual US Army health assessment, with good accuracy. Soldier self-reports of suicidality in the PHA, in comparison, were uninformative, as only 0.2% of soldiers admitted any suicidality, and these reports were only weakly related to subsequent SAs.

Based on these results, ML modeling was selected to identify high-risk soldiers for a remote group dialectical behavior therapy skills training intervention to prevent suicide-related behaviors.⁴ This intervention is being evaluated in a randomized clinical trial at Fort Hood, Texas, as part of the Army SAFEGUARD (Suicide Avoidance Focused Enhanced Group Using Algorithm Risk Detection) initiative.¹⁰ The intervention was designed, though, for soldiers with elevated risk of either suicide or SA, whereas the earlier Army Study to Assess Risk and Resilience in Service Members (STARRS) model focused only on the latter. Because predictors differ for suicide and SA,¹¹ we developed separate ML models for each outcome and then cross-classified the 2 predicted risk scores to create a composite risk stratification scheme for the SAFEGUARD initiative. The current study describes the results of that model-building effort.

Predictors were drawn from the consolidated administrative data system developed in the STARRS¹² study, an epidemiological-neurobiological study of suicide risk and protective factors in the US Army. Databases were restricted to those available at the time of a PHA and to variables that emerged as significant predictors in prior STARRS study work.⁹ We developed separate models to predict suicide deaths in the total sample (obtained from the National Death Index and consequently known even if the soldier left military service over the follow-up period) and nonfatal suicide attempts recorded in US Army and Department of Defense records while soldiers were still in service. The follow-up period was set at 24 months because successive PHAs, though intended to be annual, were sometimes found to occur up to 24 months apart.

Methods

Sample and Procedures

The study protocol was approved by the institutional review board of the Henry Jackson Foundation and Harvard Medical School with a waiver of informed consent based on data being deidentified. The data used to train and test the models were

Key Points

Question Can administrative data available during annual periodic health assessments (PHAs) be used in machine learning models to identify US Army soldiers at high risk of subsequent suicide deaths and nonfatal suicide attempts for targeting a preventive intervention?

Findings In this prognostic study including 668 684 US Army soldiers with 1932 243 completed PHAs, model discrimination was fair for suicides and good for nonfatal suicide attempts. Risk for outcomes overlapped only partially, with many PHAs at elevated risk for one outcome but not the other.

Meaning Study results suggest that high risk of suicide death and nonfatal suicide attempt were concentrated in partially overlapping, but distinct, subsets of soldiers, complicating resource allocation to maximize intervention effects.

based on the population of all regular US Army soldiers in active service who completed 1 or more PHAs during the 2015 to 2019 period. A small fraction of PHAs were completed by soldiers who went on to die by some cause other than suicide over the next 24 months. Given that these accounted for only 0.09% of all PHAs, we did not adjust for censoring but instead coded these soldiers as not dying in our person-level suicide model. Censoring was a more serious issue in the SA model. We addressed this problem by training an ensemble ML model to predict not being right censored for SA. All PHAs were given a censoring weight based on this model equal to 1 divided by the predicted probability of not being censored. The right-censored PHAs were deleted from the analysis. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.

Measures

Outcomes

The 2 primary outcomes were administratively documented suicide deaths and SAs within the 24-month period after each PHA. Suicide deaths were identified through the National Death Index¹³ for soldiers in active service and after leaving active service. SAs, in comparison, were known only for soldiers still in active service and were measured via (1) the Department of Defense Suicide Event Report (DoDSER¹⁴) and (2) Department of Defense health care encounters with *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM¹⁵) codes E950-E958 (suicide and self-inflicted injury/poisoning) or *International Statistical Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM¹⁶) codes T14.91XX (SA not otherwise specified), T36.OX2X-T50.992X (intentional self-harm by poisoning), T51.OX2X-T65.92XX (intentional self-harm by toxic effects of substances), T71.112X-T71.232X (intentional self-harm by asphyxiation), or X71.OXXX-X83.8XXX (intentional self-harm) codes.¹⁷

Predictors

Consistent with previous STARRS PHA analyses,⁹ the predictors considered here consisted of the variables that could be

Table 1. Sociodemographic and Army Career Characteristics of Completed PHAs by Regular US Army Soldiers From 2015 to 2019

Characteristic	No. (%) Total sample (N = 1 932 243)	Suicide ^a (n = 1321)	Nonfatal suicide attempt ^b (n = 9454)
Age, y			
18-21	357 845 (18.5)	351 (26.6)	3952 (41.8)
22-24	367 527 (19.0)	338 (25.6)	2259 (23.9)
25-27	303 570 (15.7)	210 (15.9)	1229 (13.0)
28+	903 301 (46.7)	422 (32.0)	2014 (21.3)
Gender			
Female	279 832 (14.5)	77 (5.8)	2517 (26.6)
Male	1 652 411 (85.5)	1244 (94.2)	6937 (73.4)
Race and ethnicity			
Non-Hispanic Black	404 592 (20.9)	211 (16.0)	2323 (24.6)
Hispanic	282 547 (14.6)	166 (12.6)	1591 (16.8)
Non-Hispanic White	1 093 265 (56.6)	849 (64.3)	4823 (51.0)
Other ^c	151 839 (7.9)	95 (7.2)	717 (7.6)
Education			
Less than HS	27 161 (1.4)	42 (3.2)	187 (2.0)
HS	1 295 955 (67.1)	1094 (82.8)	8273 (87.5)
Some college	105 415 (5.5)	50 (3.8)	309 (3.3)
College graduate	503 712 (26.1)	135 (10.2)	685 (7.3)
Marital status			
Married	1 122 645 (58.1)	738 (55.9)	4148 (43.9)
Previously married	78 105 (4.0)	63 (4.8)	317 (3.4)
Never married	731 493 (37.9)	520 (39.4)	4989 (52.8)
Years in service			
1 or less	378 073 (19.6)	329 (24.9)	3955 (41.8)
2	226 564 (11.7)	199 (15.1)	1547 (16.4)
3-4	320 939 (16.6)	228 (17.3)	1504 (15.9)
5 or more	1 006 667 (52.1)	565 (42.8)	2448 (25.9)
Rank			
Junior enlisted (E1-E3)	854 212 (44.2)	747 (56.6)	7085 (74.9)
NCOs (E4-E8)	716 992 (37.1)	491 (37.2)	2053 (21.7)
Officer (warrant or commissioned)	361 039 (18.7)	83 (6.3)	316 (3.3)
Combat deployments, No.			
0	980 617 (50.8)	768 (58.1)	6752 (71.4)
1	392 893 (20.3)	237 (17.9)	1435 (15.2)
2	252 351 (13.1)	155 (11.7)	640 (6.8)
≥3	306 382 (15.9)	161 (12.2)	627 (6.6)
Duty-MOS			
Direct combat arms	608 980 (31.5)	554 (41.9)	2911 (30.8)
Indirect combat arms	114 599 (5.9)	66 (5.0)	325 (3.4)
Combat support	454 772 (23.5)	316 (23.9)	2335 (24.7)
Combat service support	753 892 (39.0)	385 (29.1)	3883 (41.1)
Command			
FORSCOM	1 054 842 (54.6)	860 (65.1)	6292 (66.6)
TRADOC	138 986 (7.2)	77 (5.8)	465 (4.9)
Area Commands	292 738 (15.2)	161 (12.2)	1462 (15.5)
Special Ops	131 967 (6.8)	73 (5.5)	235 (2.5)
MEDCOM	114 257 (5.9)	42 (3.2)	422 (4.5)
AMC, unknown	199 453 (10.3)	108 (8.2)	578 (6.1)

Abbreviations: AMC, US Army Materiel Command; E, enlisted; FORSCOM, US Army Forces Command; HS, high school; MEDCOM, US Army Medical Command; MOS, military occupational specialty; NCO, noncommissioned officer; PHA, periodic health assessment; TRADOC, US Army Training and Doctrine Command.

^a Died by suicide within 24 months of their PHA.

^b Had a nonfatal suicide attempt within 24 months of their PHA.

^c Other race and ethnicity includes American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, or more than 1 race or ethnicity.

accessed in real time as of the time of a PHA abstracted from a series of US Army administrative data systems (eTable 1 in Supplement 1). Included here were demographic, army career, psychopathological, and physical health predictors shown in prior research to predict suicidal outcomes (eTable 2 in Supplement 1).^{9,11,18-22} Demographic data, including race and

ethnicity, were recorded and included in analyses given their potential relevance to predictions of suicidal outcomes. Race and ethnicity were documented in administrative systems as follows: American Indian or Alaska Native, Asian, Black, Hispanic, White, and other (ie, multiple races and/or ethnicities, or missing, unknown, or not reported).

Table 2. Estimates of Model Performance for Suicide and Nonfatal Suicide Attempt in Test Samples^a

Sample	AUROC (SE)	ICI ^b	Brier score ^b
Suicide			
SuperLearner	0.716 (0.174)	0.00019	0.00066
LASSO	0.719 (0.177)	0.00035	0.00068
Best elastic net	0.720 (0.176)	0.00029	0.00067
Best neural net	0.672 (0.194)	0.00177	0.00184
Best random forest	0.711 (0.178)	0.00016	0.00066
Best XGBoost	0.721 (0.171)	0.00028	0.00066
Nonfatal suicide attempt			
SuperLearner	0.806 (0.015)	0.00137	0.00697
LASSO	0.792 (0.015)	0.00154	0.00710
Best elastic net	0.808 (0.015)	0.00157	0.00720
Best neural net	0.777 (0.016)	0.00316	0.00800
Best random forest	0.801 (0.015)	0.00084	0.00702
Best XGBoost	0.772 (0.016)	0.00579	0.00710

Abbreviations: AUROC, area under the receiver operating characteristic curve; ICI, integrated calibration index; LASSO, least absolute shrinkage and selection operator.

^a Estimates are based on test case-control subsamples. Suicide test case-control subsample n = 4257 (n = 387 suicides and n = 3870 controls). Nonfatal suicide attempt test case-control subsample n = 31 284 (n = 2844 nonfatal suicide attempts and n = 28 440 controls).

^b After prevalence adjustment (recalibrating model intercepts to adjust predicted probabilities to reflect the outcome prevalence of the population).

Statistical Analysis

Data Preparation

Analyses were carried out at the PHA level. Before the analysis, the sample was split into a 70% training sample for model development and a 30% test sample for model evaluation. All PHAs for individual soldiers with multiple PHAs were assigned either to the training sample or the test sample. A 1:10 case-control sampling scheme was used for computational efficiency in the training sample, with controls selected in a stratified equal-probability sample of all PHAs not associated with the outcome of interest. Stratification was based jointly on demographics (sex, age, race and ethnicity, marital status), US Army career characteristics (rank, military occupational specialty, number of deployments), and time of the PHA. Each control PHA was assigned an inverse probability of selection weight so that the sum of weights equaled the population total of control PHAs. The eMethods in Supplement 1 contains more details regarding the case-control subsamples and weighting procedures. This weighting approach corrects for the subsampling of controls and preserves the target population distribution, so that estimation of the regression function is not biased by the case-control sampling.²³ We restricted the ensemble to algorithms that explicitly support observation weights.

Data Analysis

Analysis was conducted September 2025 through January 2026. Unlike most ML studies to predict suicide-related behaviors, which either use a single algorithm or try several different algorithms and then choose the one with the best prediction accuracy,²⁴ we used a stacked generalization approach with the SuperLearner (SL) ensemble ML method that pooled results across multiple algorithms. SL uses a weight generated via cross-validation in a user-specified collection (ensemble) to combine predicted outcome scores across all component algorithms in a way guaranteed in expectation to perform at least as well as the best algorithm in the ensemble according to a prespecified criterion (in our case, minimizing mean-squared error).^{25,26} Consistent with recommendations,²⁷ we used a diverse set of algorithms in the ensemble to capture

nonlinearities and interactions and reduce risk of misspecification (eTable 3 in Supplement 1). We also used a simple least absolute shrinkage and selection operator (LASSO) penalized regression model²⁸ to provide a benchmark comparison of the benefits of the much more complex SL approach.

We evaluated model discrimination (the ability of the model to rank order PHAs by true suicide and SA risk) by applying model predictions based on the training sample to the test sample and examining the area under the receiver operating characteristic curve (AUROC²⁹). We then evaluated model calibration (the ability to estimate probability of suicide and SA accurately) in the test sample by calculating the Integrated Calibration Index (ICI³⁰) based on the locally estimated scatterplot smoothing (LOESS) curve (smoothing span = 0.75) to avoid the need for binning.³¹ Overall model accuracy (ie, combining discrimination and calibration) was evaluated using the Brier score. It is noteworthy that the Brier score, though sensitive to prevalence, can be used to compare combined discrimination and calibration across models when predicting a single outcome in a single sample.³²

We examined concentration of risk by dividing the test samples into 20 risk ventiles and calculating conditional and cumulative sensitivity (the proportion of the observed outcome) and positive predictive value (PPV; observed prevalence of the outcome) within and across ventiles of predicted risk. We then examined overlap in predicted risk scores for the 2 outcomes by generating cross-validated predicted probabilities of each outcome in the total sample and cross-tabulating categorized risk strata. Variation in concentration of risk of each outcome as a function of concentration of risk of the other outcome was examined by computing *lift*, which is the ratio of sensitivity in a subgroup to the proportion of PHAs in the subgroup,³³ based on the common rule of thumb that a lift of 2.0+ is clinically significant.³⁴ Finally, we carried out a variable importance analysis using Shapley Additive Explanations (SHAP³⁵) to identify potentially important predictors associated with overall risk of suicide and SA along with having elevated risk of both outcomes or either outcome (eMethods in Supplement 1).

Table 3. Observed 24-Month Suicides and Nonfatal Suicide Attempts by Ventiles of Predicted Risk in Test Samples^a

Ventile ^b	Suicide			Nonfatal suicide attempt		
	% (SE)	Cumulative SN	Cumulative PPV/100 000, estimate (SE)	% (SE)	Cumulative SN	Cumulative PPV/100 000, estimate (SE)
1	18.6 (2.0)	18.6 (2.0)	248.4 (34.2)	33.4 (1.0)	33.4 (1.0)	4709.7 (242.7)
2	8.5 (1.4)	27.1 (2.3)	181.2 (19.9)	13.1 (0.7)	46.5 (1.0)	3276.6 (127.9)
3	7.7 (1.4)	34.9 (2.4)	155.1 (14.8)	9.4 (0.5)	55.9 (1.0)	2623.8 (88.1)
4	9.8 (1.5)	44.7 (2.5)	149.1 (12.5)	8.3 (0.5)	64.2 (0.9)	2258.4 (68.0)
5	9.8 (1.5)	54.5 (2.5)	145.8 (11.1)	6.4 (0.4)	70.5 (0.8)	1986.0 (55.5)
6	7.2 (1.3)	61.7 (2.5)	137.4 (9.7)	5.1 (0.4)	75.7 (0.8)	1775.7 (46.8)
7	7.0 (1.3)	68.7 (2.4)	131.2 (8.8)	3.9 (0.3)	79.6 (0.7)	1601.4 (40.5)
8	3.1 (0.9)	71.8 (2.3)	119.8 (7.8)	3.2 (0.3)	82.8 (0.7)	1456.9 (35.7)
9	5.9 (1.2)	77.7 (2.1)	115.3 (7.2)	3.2 (0.3)	85.9 (0.6)	1344.5 (32.0)
10	4.4 (1.0)	82.1 (2.0)	109.7 (6.6)	3.0 (0.3)	89.0 (0.6)	1252.8 (29.0)
11	3.6 (1.0)	85.8 (1.8)	104.0 (6.1)	2.3 (0.3)	91.2 (0.5)	1168.0 (26.5)
12	2.6 (0.8)	88.4 (1.6)	98.3 (5.7)	2.0 (0.2)	93.3 (0.4)	1094.7 (24.4)
13	2.6 (0.8)	91.0 (1.5)	93.4 (5.3)	1.6 (0.2)	94.9 (0.4)	1027.5 (22.6)
14	2.8 (0.8)	93.8 (1.2)	89.4 (5.0)	1.0 (0.2)	95.9 (0.3)	964.1 (21.0)
15	0.3 (0.3)	94.0 (1.2)	83.7 (4.7)	0.9 (0.2)	96.8 (0.3)	908.6 (19.6)
16	2.1 (0.7)	96.1 (1.0)	80.2 (4.4)	1.0 (0.2)	97.8 (0.2)	861.0 (18.4)
17	1.8 (0.7)	97.9 (0.7)	76.9 (4.2)	0.7 (0.1)	98.5 (0.2)	816.4 (17.3)
18	1.3 (0.6)	99.2 (0.5)	73.6 (4.0)	0.7 (0.1)	99.2 (0.2)	776.3 (16.4)
19	0.5 (0.4)	99.7 (0.3)	70.1 (3.7)	0.5 (0.1)	99.8 (0.1)	739.5 (15.5)
20	0.3 (0.3)	100.0 (0.0)	66.7 (3.6)	0.2 (0.1)	100.0 (0.0)	704.1 (14.7)

Abbreviations: PPV, positive predicted value (per 100 000 periodic health assessments); SN, sensitivity.

^a Estimates are from the test case-control subsamples. Suicide test case-control subsample n = 4257 (n = 387 suicides and n = 3870 controls). Nonfatal suicide attempt test case-control subsample n = 31 284 (n = 2844 nonfatal suicide

attempts and n = 28 440 controls).

^b Risk ventiles represent the 20 ordered subgroups of observations based on predicted risk, with risk ventile 1 having the highest predicted risk and risk ventile 20 having the lowest predicted risk.

Data management and calculation of prevalence, sensitivity, and AUROC were carried out in SAS, version 9.4 (SAS Institute).³⁶ The SL and LASSO models, ICI, Brier scores, and variable importance analysis were all carried out in R, version 4.3.2 (R Project for Statistical Computing).³⁷

Results

Sample Characteristics

The demographic and Army career characteristics of the sample are presented in Table 1. For suicide, the 70% training sample before subsampling controls included 1 352 211 PHAs among 468 079 soldiers. The case-control subsample of the training sample included all 934 PHAs completed by the 653 soldiers in the training sample who died by suicide within 24 months of their PHA plus a probability sample of 10 times (ie, n = 9340) as many control PHAs. The 30% suicide test sample included 580 032 PHAs among 200 605 soldiers, including all 387 PHAs completed by the 280 suicide decedents in the test sample.

For SA, the 70% training sample included 1 118 810 PHAs among 388 116 soldiers. The case-control subsample of the training sample included all 6610 PHAs completed by the 4390 soldiers with SAs over the 24-month risk horizon in the training sample plus a probability sample of 10 times (ie, n = 66 610) as many control PHAs. The 30% SA test sample included

479 586 PHAs among 166 337 soldiers, including all 2844 PHAs completed by the 1882 soldiers with SAs over the 24-month risk horizon.

PHA data were available from 668 684 regular US Army soldiers. A total of 1 932 243 PHAs (including 85.5% by male soldiers, 46.7% by soldiers aged 28 years or older, and 58.1% by married soldiers) completed during the 2015 to 2019 period were included in analyses. A total of 903 301 PHAs (46.7%) were completed by soldiers 28 years or older, and 1 122 645 (58.1%) were completed by married soldiers. A total of 1321 PHA visits (for 933 soldiers) were associated with a subsequent suicide, 9454 (for 6272 soldiers) were associated with a subsequent SA, and 126 were associated with both; 1780 PHAs were completed by soldiers who went on to die by some cause other than suicide over the next 24 months. PHA visits followed by either outcome were associated with soldiers being young, having no more than a high school education, being junior enlisted, having had fewer years in military service, never being deployed, and being in the Forces Command. PHA visits followed by a suicide but not a SA were associated with soldiers being male, non-Hispanic White race, and with a Direct Combat Arms military occupational specialty. PHA visits followed by a SA but not a suicide were associated with soldiers being female, never married, non-Hispanic Black race, and Hispanic ethnicity. eFigures 1 and 2 in Supplement 1 contain additional details regarding the sample for the SA.

Table 4. Intersection of Predicted Risk of Suicide and Nonfatal Suicide Attempt of Regular US Army Soldier PHAs (n = 580 032)

Risk level	Sample distribution, %	Distribution of suicides			Distribution of nonfatal suicide attempts		
		SN (SE)	Lift ^a	PPV (SE)	SN (SE)	Lift ^a	PPV (SE)
High risk of suicide ^b							
Highest risk of SA ^c	2.2	11.1 (1.6)	5.2	344.7 (52.5)	11.0 (0.6)	5.1	2517.2 (140.3)
High risk of SA ^d	0.9	1.8 (0.7)	2.0	131.1 (49.5)	1.9 (0.3)	2.1	1011.1 (136.9)
Lower risk of SA ^e	1.9	5.7 (1.2)	2.9	196.6 (41.9)	3.2 (0.3)	1.7	813.3 (84.9)
Total	5.0	18.6 (2.0)	3.7	248.2 (29.2)	16.1 (0.7)	3.2	1582.5 (73.3)
Lower risk of suicide ^f							
Highest risk of SA	2.9	4.4 (1.0)	1.5	102.9 (24.9)	12.1 (0.6)	4.2	2075.3 (110.9)
High risk of SA	4.1	7.5 (1.3)	1.8	122.6 (22.8)	11.7 (0.6)	2.9	1403.2 (76.5)
Lower risk of SA	88.1	69.5 (2.3)	0.8	52.7 (3.2)	60.1 (0.9)	0.7	334.7 (8.1)
Total	95.0	81.4 (2.0)	0.9	57.2 (3.2)	83.9 (0.7)	0.9	432.8 (8.8)
Total							
Highest risk of SA	5.0	15.5 (1.8)	3.1	206.9 (26.7)	23.1 (0.8)	4.6	2265.4 (87.4)
High risk of SA	5.0	9.3 (1.5)	1.9	124.1 (20.7)	13.6 (0.6)	2.7	1330.9 (67.3)
Lower risk of SA	90.0	75.2 (2.2)	0.8	55.7 (3.3)	63.3 (0.9)	0.7	345.0 (8.1)
Total	100.0	100.0 -	-	68.4 (3.4)	100.0 -	-	490.3 (9.2)

Abbreviations: PHA, periodic health assessment; PPV, positive predictive value (number of observed soldiers in the subgroup with the outcome over a 24-month risk period per 100 000 PHAs); SA, nonfatal suicide attempt; SN, sensitivity (the proportion of observed soldiers with the outcome in the subgroup).

^a Lift, the ratio of SN to proportion in the sample distribution.

^b High risk of suicide, the 5% of PHAs with the highest predicted risk of suicide.

^c Highest risk of SA, the 5% of PHAs with highest predicted risk of SA.

^d High risk of SA, the next 5% (6%-10%) with highest predicted risk of SA.

^e Lower risk of SA, lowest predicted risk of SA (11%-100%).

^f Lower risk of suicide, lowest predicted risk of suicide (6%-100%).

Model Performance

Performance of SL models, simple LASSO penalized regression models, and individual algorithms were compared separately for each outcome (eTable 4 in Supplement 1 contains meta-learner weights used in the SL models). For suicide, the best XGBoost model had an AUROC of 0.72 (SE = 0.17), with ICI and Brier score estimates (0.0003 and 0.0007, respectively) that were generally either comparable with or better than those of other algorithms (Table 2). Inspection of calibration plots showed that these models were well calibrated in the range of the intervention threshold, which we identified as 5% to 10% of soldiers with highest predicted risk (eFigure 3 in Supplement 1). For SA, the SL model had an AUROC of 0.81 (SE = 0.02), with ICI and Brier score estimates (0.0014 and 0.0070, respectively) that were generally either comparable with or better than those of other algorithms (Table 2). Inspection of calibration plots again showed that these models were well calibrated in the range of the 5% to 10% intervention threshold (eFigure 4 in Supplement 1). We consequently focused on XGBoost results for suicide and SL results for SA.

Estimates of sensitivity and PPV (per 100 000 PHAs) for suicide and SAs are reported in Table 3. The 24-month suicide prevalence was 66.7 per 100 000 (SE = 3.6/100 000). Inspection of predicted suicide risk ventiles in the XGBoost model showed that only the top ventile had meaningfully elevated sensitivity, which we defined as at least twice that expected by chance. This top ventile accounted for 18.6% (SE = 2.0%) of suicides, with 24-month prevalence of 248.4 per 100 000 (SE = 34.2/100 000). Based on these results, the 5% of PHAs with the highest predicted risk were defined as high risk and the remaining 95% were defined as lower risk.

The 24-month SA prevalence was 704.1 per 100 000 (SE = 14.7/100 000). Inspection of predicted risk ventiles in the SL model showed that only the top 2 ventiles had meaningfully elevated sensitivity (33.4%-13.1%; SE = 1.0%-0.7%) and together accounted for 46.5% (SE = 1.0%) of SAs. The 24-month SA prevalence for these ventiles was 3276.6 per 100 000 (SE = 127.9/100 000). Based on these results, the 5% of PHAs with highest predicted SA risk were defined as highest risk, the next 5% were defined as high risk, and the remaining 90% were defined as lower risk.

Analysis of the Overlap in Risk of Suicide and SA

Although the predicted probabilities of the 2 outcomes were positively correlated, the association was modest in magnitude (Pearson $r = 0.29$; $t = 233.8$; $P < .001$). Examination of the cross-classified 2-category risk scheme for suicide (ie, high and lower risk) and 3-category risk scheme for SA (ie, highest, high, and lower risk) showed that 11.9% of PHAs were at elevated predicted risk of at least one outcome, accounting for (ie, sensitivity) 30.5% (SE = 2.3%) of suicides and 39.9% (SE = 0.9%) of SAs (Table 4). Within this subgroup, 2.2% of PHAs were at high predicted risk of suicide and highest predicted risk of SA, accounting for 11.1% (SE = 1.6%) of suicides and 11.0% (SE = 0.6%) of SAs. Overall, 3.1% of PHAs had high predicted risk of suicide and either highest (5%) or high (10%) predicted risk of SA, accounting for 12.9% (SE = 1.7%) of suicides and 12.9% (SE = 0.6%) of SAs. An additional 1.9% had high predicted risk of suicide only, and 6.9% had either highest or high predicted risk of SA only, together accounting for 17.6% (SE = 1.9%) of suicides and 26.9% (SE = 0.8%) of SAs.

Clinically significant lift (ie, lift ≥ 2.0) for suicide was found for PHAs with high predicted risk of suicide regardless of pre-

dicted SA risk (lift = 2.0-5.2). The 24-month suicide prevalence among these subgroups ranged from 131.1 per 100 000 (SE = 49.5/100 000) to 344.7 per 100 000 (SE = 52.5/100 000). Clinically significant lift for SA was found among those with highest or high predicted risk of SA, regardless of predicted suicide risk (lift = 2.1-5.1). The 24-month SA prevalence among these subgroups with significant lift ranged from 1011.1 per 100 000 (SE = 136.9/100 000) to 2517.2 per 100 000 (SE = 140.3/100 000).

Predictor Importance

We examined proportional SHAP values for predictors of the 2 outcomes (eFigures 5 and 6 in [Supplement 1](#)) as well as of being in the top 5% of predicted risk of both outcomes (eFigure 7 in [Supplement 1](#)) or having high risk of either outcome (defined as the top 5% of suicide risk or the top 10% of SA risk) (eFigure 8 in [Supplement 1](#)). Psychopathological risk factors were the most important class of predictors across all models (proportional SHAP = 70.6%-95.4%), with a wide variety of individual variables in that class emerging as most important in different models and the great majority of the associations being positive. Demographics were strong predictors of both suicide (proportional SHAP = 44.2%) and SA (proportional SHAP = 33.6%), due to strong (proportional SHAP $\geq$ 5%) predictive effects of male sex and non-Hispanic White race and ethnicity (both positively associated with suicide risk), higher education (negatively associated with suicide and SA risk), and older age (negatively associated with SA risk). Army career characteristics were also important for both suicide (proportional SHAP = 40.7%) and SA (proportional SHAP = 31.4%). Physical conditions, finally, were consistently the weakest predictors (proportional SHAP = 3.5%-13.5%).

Discussion

We found that roughly 5% of PHAs were associated with meaningfully elevated predicted risk of suicide, 10% with elevated predicted risk of SA, and 11.9% with either. These PHAs accounted for approximately 31% of subsequent suicides and 40% of SAs. Concentrations of risk as high as these were considered actionable for targeting the SAFEGUARD intervention at Fort Hood, Texas, and have now been used to implement real-time recruitment for this randomized clinical trial. Whether the intervention turns out to be cost-effective at lower risk of the outcomes is being evaluated by including in the experiment a small probability sample of soldiers with lower risk of the outcomes, although the bulk of the sample consists of soldiers with high predicted risk of either outcome to guarantee sufficient statistical power to detect clinically meaningful intervention effects in subsamples defined by the cross-classification of the 2 risk scores as well as to investigate possible heterogeneity of intervention effects.³⁸

Although our results do not guarantee that the intervention will have a positive net benefit, they show that it is possible to identify high-risk soldiers who can be used to identify participants for a randomized clinical trial designed to reduce suicide-related behaviors, thereby addressing a criticism of self-report suicide risk screening.³⁹ The intervention

is designed to be low cost (ie, online and group based)⁴⁰ and has previously been found to be effective in reducing suicide-related behaviors among high-risk veterans.⁴¹⁻⁴³ A formal net benefit analysis will be needed after the intervention is complete to determine whether the intervention has a positive net benefit overall based on the joint consideration of costs, benefits, and potential harms.⁴⁴

To our knowledge, this study was the first to examine patterns of intersecting risk for suicide and SA based on parallel ML models. Consistent with previous evidence of only partially overlapping risk factors,^{11,45,46} we found that overlap of elevated risk was modest, with a notable subset of PHAs with elevated predicted risk for only 1 of the outcomes. Due to low base rates of suicide death, much prior research that ostensibly focused on risk factors,¹⁸ or interventions,⁴⁷ for suicide prevention actually focused on an outcome that includes a broader group of suicide-related behaviors; these behaviors, for the most part, consisted of SAs. These conflation result in an underidentification of persons at high risk of suicide and a reduced ability to identify interventions that are effective in reducing suicides relative to SAs. It is important, given this history, for future studies to clarify the outcome of interest; to distinguish participants at high risk of either suicide, SA, or both at the onset; and to evaluate treatment effects on the separate outcomes as well as in subgroups selected based on high predicted risk of one outcome, the other, or both. This will be done in the SAFEGUARD trial.

Limitations

The current report should be considered within the context of several limitations. First, even with a large sample size and multiyear data collection period, numbers of suicides and SAs were relatively low. A larger sample would have increased statistical power and model accuracy. Second, analyses only considered SAs documented through medical records. Previous research suggests that as many as 1 in 3 soldier SAs are undocumented, with risk of undocumented SAs higher among soldiers in their earlier years of service.⁴⁸ Relatedly, reliance on *ICD-10* codes for suicide attempts made it impossible to distinguish between self-directed violence with and without intention of death.¹⁷ Lastly, SAs were assessed only while soldiers were still in active service.

Conclusions

Effective suicide prevention within the US military requires accurate identification of service members at highest risk of suicide-related behaviors. Results of this prognostic study suggest that ML methods applied to US Army administrative data available at the time of PHAs can identify soldiers at comparatively high risk of suicides and SAs over the next 24 months. We found that elevated suicide and suicide attempt risk overlap only partially, with many PHAs at elevated risk for 1 outcome but not the other, attesting to the importance of considering these outcomes separately when developing targeting models and delivering preventive interventions for suicide and suicide-related behaviors.

ARTICLE INFORMATION

Accepted for Publication: June 7, 2026.

Published Online: August 12, 2026.

doi:10.1001/jamapsychiatry.2026.2410

Author Affiliations: Veterans Integrated Service Network 2, Mental Illness Research, Education and Clinical Center, Department of Veterans Affairs, Bronx, New York (Edwards, Geraci, Goodman); Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York (Edwards, Geraci, Goodman); Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Borowski, Gildea, King, Liu, Luedtke, Sampson, Kessler); Department of Psychology, Harvard University, Cambridge, Massachusetts (Nock); Department of Psychiatry, Uniformed Services University of the Health Sciences, Bethesda, Maryland (Benedek, Maggio, Naifeh, Nassif, Paine, Capaldi, Ursano); Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston (Bentley, Kennedy); Center of Excellence for Research on Returning War Veterans, Veterans Integrated Service Network 17, Doris Miller VA Medical Center, Waco, Texas (Geraci); Resilience Center for Veterans & Families, Teachers College, Columbia University, New York, New York (Geraci); Carl R. Darnall Army Medical Center, Fort Hood, Texas. (Gupta); Center for Precision Psychiatry, Massachusetts General Hospital, Boston (Kennedy); Department of Psychology, Rutgers, The State University of New Jersey, New Brunswick (Kleiman); Department of Health Behavior, Society and Policy, Rutgers, The State University of New Jersey, School of Public Health, Piscataway (Kleiman); Department of Statistics, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts (Luedtke); Henry M. Jackson Foundation for the Advancement of Military Medicine Inc, Bethesda, Maryland (Maggio, Naifeh, Paine); Department of Psychology, National University of Singapore, Singapore (Zainal); Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, Ann Arbor (Wagner); VA San Diego Healthcare System, San Diego, California (Stein); Department of Psychiatry, University of California San Diego, La Jolla (Stein); School of Public Health, University of California San Diego, La Jolla (Stein).

Author Contributions: Drs Edwards and Kessler 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: Edwards, Benedek, Bentley, Geraci, Gupta, Kleiman, Luedtke, Maggio, Nassif, Stein, Capaldi, Ursano, Kessler.

Acquisition, analysis, or interpretation of data: Edwards, Borowski, Nock, Gildea, Goodman, Kennedy, King, Liu, Naifeh, Paine, Sampson, Zainal, Wagner, Stein, Capaldi.

Drafting of the manuscript: Edwards, Borowski, King, Liu, Zainal, Capaldi, Kessler.

Critical review of the manuscript for important intellectual content: Edwards, Borowski, Nock, Benedek, Bentley, Geraci, Gildea, Goodman, Gupta, Kennedy, Kleiman, Liu, Luedtke, Maggio, Naifeh, Nassif, Paine, Sampson, Wagner, Stein, Capaldi, Ursano, Kessler.

Statistical analysis: Kennedy, King, Liu, Luedtke.

Obtained funding: Stein, Kessler.

Administrative, technical, or material support: Edwards, Borowski, Nock, Geraci, Gildea, Goodman, Gupta, Maggio, Nassif, Sampson, Wagner, Kessler.

Supervision: Borowski, Gupta, Maggio, Sampson, Capaldi, Ursano, Kessler.

Conflict of Interest Disclosures: Dr Nock reported receiving publication royalties from Macmillan, Pearson, and UpToDate; consultant fees in the past 3 years from Cambridge Health Alliance and Open AI and for legal cases regarding deaths by suicide; having stock options in Cerebral Inc; and serving as an unpaid scientific advisor for Empatica, the JED Foundation, Koko, and TalkLife. Dr Benedek reported receiving royalties from UpToDate outside the submitted work. Dr Bentley reported receiving personal fees (salary) and equity from Spring Health and consultant fees and equity from Visible Health Inc outside the submitted work. Dr Goodman reported receiving personal fees from Boehringer Ingelheim Pharmaceuticals outside the submitted work. Dr Kleiman reported personal fees from Henry M Jackson Foundation for the Advancement of Military Medicine during the conduct of the study. Dr Luedtke reported receiving statistical consulting fees from Harvard University within the last year outside the submitted work. Dr Nassif reported receiving funding for serving as assistant secretary of war for health affairs during the conduct of the study. Dr Wagner reported receiving grants from the Henry M. Jackson Foundation during the conduct of the study. Dr Stein reported receiving consulting income in the past 3 years from atai Life Sciences, Ananda Scientific, BigHealth, Biogen, Bionomics/Neuphoria, Boehringer Ingelheim, EmpowerPharm, Engrail Therapeutics, Jazz Pharmaceuticals, Karuna Therapeutics, Lundbeck, Lykos Therapeutics, Newleos Therapeutics, Orion Pharma, Otsuka US, PureTech Health, Roche/Genentech, Sage Therapeutics, Seaport Therapeutics, Sensorium Therapeutics, and Transcend Therapeutics; having stock options in EpiVario, Newleos Therapeutics, and Oxeia Biopharmaceuticals; receiving pay for his editorial work on *Depression and Anxiety* (editor-in-chief), *Biological Psychiatry* (deputy editor), and UpToDate (co-editor-in-chief for psychiatry); receiving research support from the National Institutes of Health, the Department of Veterans Affairs, and the Department of Defense; and serving on the scientific advisory board of the Brain and Behavior Research Foundation and the Anxiety and Depression Association of America. Dr Kessler reported receiving grants from the Department of War, NIMH, and Department of Veterans Affairs; consultant fees from Cambridge Health Alliance, Child Mind Institute, Massachusetts General Hospital, RallyPoint LLC, Sage Therapeutics, University of Michigan, and University of North Carolina Consultant; stock options from Cerebral Inc, Mirah, PYM (Prepare Your Mind), Verisense Health; and having ownership interest in Menssano LLC during the conduct of the study. No other disclosures were reported.

Funding/Support: Army Study to Assess Risk and Resilience in Servicemembers—Longitudinal Study (STARRS-LS; 2015-present) is funded by the Department of War under Uniformed Services University of the Health Sciences grant HU0001-2520033. Army STARRS (2009-2015) was

sponsored by the Department of the Army and the National Institute of Mental Health (NIMH) and funded under cooperative agreement U01MH087981 with the NIMH. Suicide Avoidance Focused Enhanced Group Using Algorithm Risk Detection (SAFEGUARD) is funded by the Department of War under Federal Award HU00012520025. The analyses reported here were funded by STARRS-LS. The work of Dr Edwards on this project was supported by the United States Department of Veterans Affairs, Suicide Prevention Actively Managed Portfolio, project SPR-002-24F.

Role of the Funder/Sponsor: Both teams collaborated in designing analysis, interpreting results, and preparing the report. Funding for Army Study to Assess Risk and Resilience in Service Members—Longitudinal Study (STARRS-LS) and Suicide Avoidance Focused Enhanced Group Using Algorithm Risk Detection (SAFEGUARD) was provided by the Assistant Secretary of War for Health Affairs. The funder provided oversight to facilitate progress toward established project objectives and reviewed submitted manuscripts to provide organizational feedback. The funder had no role in study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or US government. The opinions and assertions expressed herein are those of the authors and do not reflect the official policy or position of the National Institute of Mental Health, the Department of the Army, the Department of War, the Department of Veteran Affairs, the Henry M. Jackson Foundation for the Advancement of Military Medicine Inc, and the Uniformed Services University of the Health Sciences. The contents of this publication are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions or policies of the Henry M. Jackson Foundation for the Advancement of Military Medicine Inc. Mention of trade names, commercial products, or organizations does not imply endorsement by the US government.

Data Sharing Statement: See [Supplement 2](#).

Additional Information: The Army Study to Assess Risk and Resilience in Service Members (STARRS) team consists of co-Principal Investigators: Robert J. Ursano, MD (Uniformed Services University) and Murray B. Stein, MD, MPH (University of California San Diego and VA San Diego Healthcare System); site Principal Investigators: James Wagner, PhD (University of Michigan) and Ronald C. Kessler, PhD (Harvard Medical School); US Army scientific consultant/liason: Kenneth Cox, MD, MPH (Office of the Assistant Secretary of the Army (Manpower and Reserve Affairs)); other team members: Pablo A. Aliaga, MS (Uniformed Services University); David M. Benedek, MD (Uniformed Services University); Laura Campbell-Sills, PhD (University of California San Diego); Carol S. Fullerton, PhD (Uniformed Services University); Nancy Gebler, MA (University of Michigan); Meredith House, BA (University of Michigan); Paul E. Hurwitz, MPH (Uniformed Services University); Sonia Jain, PhD (University of California San Diego); Tzu-Cheg Kao, PhD (Uniformed Services University); Lisa

Lewandowski-Romps, PhD (University of Michigan); Alex Luedtke, PhD (University of Washington) and Fred Hutchinson Cancer Research Center); Holly Herberman Mash, PhD (Uniformed Services University); James A. Naifeh, PhD (Uniformed Services University); Matthew K. Nock, PhD (Harvard University); Nur Hani Zainal PhD (Harvard Medical School); Nancy A. Sampson, BA (Harvard Medical School); and Alan M. Zaslavsky, PhD (Harvard Medical School). The Suicide Avoidance Focused Enhanced Group Using Algorithm Risk Detection (SAFEGUARD) Life Force Study: team consists of Principal Investigator: Sarah E. Maggio, PhD (Uniformed Services University); Associate Investigators: Vincent F. Capaldi, MD (Uniformed Services University); Ronald C. Kessler, PhD (Harvard Medical School); Joseph C. Geraci, PhD (Bronx Department of Veterans Affairs); Christopher M. Paine, PhD (Uniformed Services University); Joseph C. Geraci, PhD (Bronx Department of Veterans Affairs); Matthew K. Nock, PhD (Harvard University); Joshua C. Gray, PhD (Uniformed Services University); Thomas H. Nassif, PhD (Uniformed Services University); and Stephanie Chardoul, BA (University of Michigan); and collaborators: Marianne S. Goodman, MD (Bronx Department of Veterans Affairs); Emily R. Edwards, PhD (Bronx Department of Veterans Affairs); and Evan K. Kleiman, PhD (Rutgers, The State University of New Jersey).

REFERENCES

- Defense Casualty Analysis System. US active-duty military deaths by year and manner 1980-present. Accessed October 8, 2025. <https://dcas.dmdc.osd.mil/dcas/app/summaryData/deaths/byYearManner>
- Department of Defense. Annual report on suicide in the military: calendar year 2022. Accessed October 8, 2025. https://www.dspo.mil/Portals/113/Documents/ARSM_CY22.pdf?ver=StAk_q6lJgNRUsOlpzVVA%3d%3d
- Decker SE, Kroll-Desrosiers A, Mattocks KM, et al. Study protocol of a hybrid type 1 effectiveness-implementation multisite trial: dialectical behavior therapy skills group for Veterans at high-risk for suicide attempt. *Contemp Clin Trials*. 2025;151:107828. doi:10.1016/j.cct.2025.107828
- Goodman M, Brown GK, Galfalvy HC, et al; Group. Group ("Project Life Force") vs individual suicide safety planning: a randomized clinical trial. *Contemp Clin Trials Commun*. 2020;17:100520. doi:10.1016/j.conctc.2020.100520
- Miller I, Gaudiano B, Weinstock L. *The Coping Long Term with Active Suicide Program (CLASP)—A Multimodal Intervention for Suicide Prevention. Treatments That Work*. Oxford University Press; 2022:144. doi:10.1093/med-psych/9780190095260.001.0001
- Rudd MD, Bryan CJ, Wertenberger EG, et al. Brief cognitive-behavioral therapy effects on posttreatment suicide attempts in a military sample: results of a randomized clinical trial with 2-year follow-up. *Am J Psychiatry*. 2015;172(5):441-449. doi:10.1176/appi.ajp.2014.14070843
- Sawada Y, Ueda M, Matsubayashi T. *Economic Analysis of Suicide Prevention: Toward Evidence-Based Policy Making*. Springer Nature Singapore; 2017. doi:10.1007/978-981-10-1500-7
- Randles R, Finnegan A. Veteran help-seeking behavior for mental health issues: a systematic review. *BMJ Mil Health*. 2022;168(1):99-104. doi:10.1136/bmjilitary-2021-001903
- Naifeh JA, Edwards ER, Bentley KH, et al. Predicting suicide attempts among US Army soldiers using information available at the time of periodic health assessments. *Nat Ment Health*. 2025;3(2):242-252. doi:10.1038/s44220-024-00360-9
- Nassif TH, Adler AB, Atkinson RD, et al. SAFEGUARD: Transforming military suicide prevention through predictive analytics and targeted interventions. *Curr Psychiatry Rep*. 16. In Press. doi:10.1007/s11920-026-01664-6
- 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*. 2021;34(2):129-146. doi:10.1080/0895605.2021.1976544
- Ursano RJ, Colpe LJ, Heeringa SG, Kessler RC, Schoenbaum M, Stein MB; Army STARRS collaborators. The Army Study to Assess Risk and Resilience in Service Members (Army STARRS). *Psychiatry*. 2014;77(2):107-119. doi:10.1521/psyc.2014.77.2.107
- Centers for Disease Control and Prevention National Center for Health Statistics. National Death Index. Accessed October 6, 2025. <https://www.cdc.gov/nchs/ndi/index.html>
- Gahm GA, Reger MA, Kinn JT, Luxton DD, Skopp NA, Bush NE. Addressing the surveillance goal in the national strategy for suicide prevention: the Department of Defense Suicide Event Report. *Am J Public Health*. 2012;102(Suppl 1)(suppl 1):S24-S28. doi:10.2105/AJPH.2011.300574
- Centers for Disease Control and Prevention. *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. Accessed October 6, 2025. <https://archive.cdc.gov/#/details?q=https://www.cdc.gov/nchs/icd/icd9cm.htm&start=0&rows=10&url=https://www.cdc.gov/nchs/icd/icd9cm.htm>
- Centers for Disease Control and Prevention. *ICD-10-CM*. Accessed October 6, 2025. <https://www.cdc.gov/nchs/icd/icd-10-cm/index.html>
- 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.
- 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
- Holliday R, Borges LM, Stearns-Yoder KA, Hoffberg AS, Brenner LA, Monteith LL. Posttraumatic stress disorder, suicidal ideation, and suicidal self-directed violence among US military personnel and veterans: a systematic review of the literature from 2010 to 2018. *Front Psychol*. 2020;11:1998. doi:10.3389/fpsyg.2020.01998
- Klonsky ED, May AM, Saffer BY. Suicide, suicide attempts, and suicidal ideation. *Annu Rev Clin Psychol*. 2016;12(1):307-330. doi:10.1146/annurev-clinpsy-021815-093204
- 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
- Ramchand R, Acosta JD, Burns RM, Jaycox LH, Pernin CG. The war within: preventing suicide in the US military. *Rand Health Q*. 2011;1(1):2.
- Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika*. 1979;66(3):403-411. doi:10.1093/biomet/66.3.403
- Pigoni A, Delvecchio G, Tutulici N, et al. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. *Transl Psychiatry*. 2024;14(1):140. doi:10.1038/s41398-024-02852-9
- Polley E, Rose S, van der Laan M. Super learning. In: Rose S, van der Laan MJ, eds. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer New York; 2011:43-66. doi:10.1007/978-1-4419-9782-1_3
- Polley E, Kennedy C, Moore SE, et al. SuperLearner: SuperLearner prediction (R package). Accessed October 6, 2025. <https://github.com/ecpolley/SuperLearner>
- LeDell E, van der Laan MJ, Petersen M. AUC-maximizing ensembles through meta-learning. *Int J Biostat*. 2016;12(1):203-218. doi:10.1515/ijb-2015-0035
- Tibshirani R. Regression shrinkage and selection via the LASSO. *J R Stat Soc Ser B Methodol*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x
- Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med*. 2013;4(2):627-635.
- 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
- Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med*. 2014;33(3):517-535. doi:10.1002/sim.5941
- Assel M, Sjöberg DD, Vickers AJ. The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagn Progn Res*. 2017;1:19. doi:10.1186/s41512-017-0020-3
- Tufféry S. *Data Mining and Statistics for Decision-Making*. 1st ed. Wiley; 2011:720. doi:10.1002/9780470979174
- 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
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; May 22, 2017; Long Beach, CA.
- SAS Software. Computer program. Version 9.4. SAS Institute; 2013.
- R Core Team. The R project for statistical computing. Accessed October 6, 2025. <https://www.r-project.org/>

38. Angus DC, Chang CH. Heterogeneity of treatment effect: estimating how the effects of interventions vary across individuals. *JAMA*. 2021; 326(22):2312-2313. doi:10.1001/jama.2021.20552
39. Barry MJ, Nicholson WK, Silverstein M, et al; US Preventive Services Task Force. Screening for depression and suicide risk in adults: US Preventive Services Task Force recommendation statement. *JAMA*. 2023;329(23):2057-2067. doi:10.1001/jama.2023.9297
40. Raciborski RA, Hamerling-Potts KK, Mitchell EL, et al. Cost comparison of in-person and telehealth modalities for a suicide safety planning group intervention: interim results from the "Project Life Force" randomized clinical trial. *Front Psychiatry*. 2023;14:1215247. doi:10.3389/fpsy.2023.1215247
41. Goodman M, Strouse M, Boucher C, et al. Pilot implementation of a telehealth safety planning group intervention for suicidal rural veterans enhanced by lived experience veteran peer participation. *Front Psychiatry*. 2025;16:1512523. doi:10.3389/fpsy.2025.1512523
42. Goodman M, Sullivan SR, Spears AP, et al. An open trial of a suicide safety planning group treatment: "Project Life Force". *Arch Suicide Res*. 2021;25(3):690-703. doi:10.1080/13811118.2020.1746940
43. Patel SR, Sullivan SR, Mitchell EL, Jager-Hyman S, Stanley B, Goodman M. Qualitative study of telehealth delivery of suicide-specific group treatment "Project Life Force". *J Technol Behav Sci*. Published online January 4, 2023. doi:10.1007/s41347-022-00297-9
44. Buyse M, Saad ED, Peron J, et al. The net benefit of a treatment should take the correlation between benefits and harms into account. *J Clin Epidemiol*. 2021;137:148-158. doi:10.1016/j.jclinepi.2021.03.018
45. Beautrais AL. Suicides and serious suicide attempts: 2 populations or 1? *Psychol Med*. 2001;31(5):837-845. doi:10.1017/S0033291701003889
46. Ge F, Wang Y, Agerbo E, et al. Contrasting risk profiles for suicide attempt and suicide using Danish registers and genetic data. *JAMA Psychiatry*. 2026;83(1):32-42. doi:10.1001/jamapsychiatry.2025.3444
47. Miller IW, Camargo CA Jr, Arias SA, et al; ED-SAFE Investigators. Suicide prevention in an emergency department population: the ED-SAFE study. *JAMA Psychiatry*. 2017;74(6):563-570. doi:10.1001/jamapsychiatry.2017.0678
48. Naifeh JA, Ursano RJ, Shor R, et al. Undetected suicide attempts among US soldiers: results from the Army Study to Assess Risk and Resilience in Service Members (Army STARRS). *Psychol Med*. 2024;54(11):2947-2955. doi:10.1017/S0033291724001028