IMPLEMENTING REAL-TIME RISK PREDICTION FOR CLINICAL DECISION SUPPORT TO PREVENT SUICIDE

VANDERBILT UNIVERSITY MEDICAL CENTER

MICHAEL RIPPERGER[1], DREW WILIMITIS[1], COLIN G. WALSH[1]

[1]VANDERBILT UNIVERSITY MEDICAL CENTER, NASHVILLE, TENNESSEE

Objective: To validate a suicide risk model within a vendor-supplied electronic health record (EHR)

INTRODUCTION

- Many suicide prediction models have been published internationally yet few have been implemented into clinical systems.
- Adoption challenges include generalizability, data availability, reliability, complexity, and transparency.
- The notable exception of an operational system is REACH VET, a prevention program within the Veteran's Health Adminstration.[1]

METHODS

- A web service was developed internally, linked to a random forest algorithm, and predictions were logged at the beginning of each encounter ('check-in').
- Demographics, diagnoses, medications, and inpatient, outpatient, and emergency visits were pulled from production databases to calculate risk.
- Performance was evaluated using Area Under the Precision Recall Curve (AURPC), sensitivity, specificity, precision, risk concentration, calibration, Spiegelhalter's Z-statistic, and Number Needed to Screen (NNS).

- Following the five-stage Action-Informed Artificial Intelligence Framework (Figure 1), our team deployed a suicide risk model in a vendor-supplied electronic health record.[2]

1 Anticipation of clinical outcomes the AI tool will address

- Engage clinicians, patients, and operational leaders
- Define characteristics of affected patients and clinical settings
- Define how and to whom the algorithm's results will be provided

Stage 1: Clinical psychologists advised us that the prediction of suicide has stayed at near chance for decades.

2 Research and development of the AI tool

- Obtain data for algorithm development
- Develop algorithms using collected data
- Confirm early validation of algorithm

Stage 2: Past suicide attempts, identified by codes, were hand validated. Multiple algorithms were developed using varying time points and evaluated.

3 Replication

- Identify similar data sources
- Identify similar patients
- Replication by computer simulation

Stage 3: One suicide risk algorithm was replicated retrospectively using non-fatal adolescent suicide attempts at Vanderbilt University Medical Center (VUMC).

4 Design, testing, and deployment of the AI tool

- Design the platform for use
- Test usability and feasibility for operational deployment
- Create the operational platform

Stage 4: An observational, prospective cohort of inpatient, emergency department, and ambulatory surgery encounters was silently produced from June 2019 to April 2020 using production data systems.

5 Improvement of determined outcomes

- Implement the operational platform
- Test effectiveness in a pragmatic trial
- Implement the AI tool and algorithm-guided practice systemwide

Stage 5: Clinical decision support to prompt screening will be assessed in a clinical trial (ClinicalTrials.gov NCT05312437). Safety plan utilization will also be measured.

REFERENCES

1. 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 Network Open.* 2021;4(10):e2129900. doi:10.1001/jamanetworkopen.2021.299002.
2. Lindsell CJ, Stead WW, Johnson KB. Action-Informed Artificial Intelligence-Matching the Algorithm to the Problem. JAMA. Published online May 1, 2020. doi:10.1001/jama.2020.5035

RESULTS

- Our algorithm generated 115,905 predictions over 296 days with 129 subsequent suicide attempts (Table 1).

Care site	Total	Race					Sex			Age, median, y	Utilization			
		White	Black	Asian	Unknown	Other	Male	Female	Unknown		Median encounters per month	Length of stay, d Median	Length of stay, d Mean	Medical record length, median, y
Medical center wide	77 973	60 586	12 620	1454	1233	2080	42 490	35 404	79	52.0	10 923	0.3	1.9	9.0
Behavioral health	2905	2278	497	55	53	23	1532	1373	0	37.0	317	0.3	1.6	5.2
Emergency department	33 235	23 650	7409	646	387	1143	15 862	17 305	68	46.7	3551	0.3	2	5.3
Adult hospital	46 389	38 047	5848	850	678	966	20 540	25 841	8	57.1	6390	2.3	4.5	5.2

Table 1. Baseline characteristics of cohort

- Performance was good across VUMC, in the Adult Hospital, and Emergency Department (ED). Performance was poor in psychiatric settings (Table 2).

Care Site	AUROC (95% CI)	Brier (95% CI)	Spiegelhalter *z* statistic (95% CI)	Spiegelhalter *z*, *P* value (95% CI)	Calibration slope (95% CI)	Calibration intercept (95% CI)
Suicide attempt						
Medical center wide	0.797 (0.796 to 0.798)	0.001 (0.001 to 0.001)	−24.683 (−24.933 to −24.433)	<.001 (<.001 to <.001)	0.189 (0.186 to 0.191)	−5.492 (−5.499 to −5.485)
Emergency department	0.7 (0.699 to 0.7)	0.002 (0.002 to 0.002)	−18.235 (−18.373 to −18.097)	<.001 (<.001 to <.001)	0.113 (0.112 to 0.113)	−5.788 (−5.793 to −5.784)
Adult hospital	0.842 (0.841 to 0.842)	0.001 (0.001 to 0.001)	−14.828 (−15.05 to −14.605)	<.001 (<.001 to <.001)	0.142 (0.141 to 0.142)	−6.462 (−6.48 to −6.444)
Behavioral health	0.544 (0.539 to 0.548)	0.011 (0.011 to 0.011)	−5.539 (−5.567 to −5.511)	<.001 (<.001 to <.001)	0.175 (0.167 to 0.183)	−3.882 (−3.914 to −3.85)

Table 2. Prospective validation by hospital setting

- In settings with universal screening (ED) and without universal screening (non-ED settings), NNS in the top risk deciles were 225 and 450 (all settings: 271). Lower is better.
- Predictions were miscalibrated (S. Z statistic -3.1, p-value 0.001) during the first five months and improved after recalibration with logistic calibration (S. Z statistic 1.1, p-value 0.26).

- The Columbia Suicide Severity Rating Scale (C-SSRS) was then added to the suicide risk model out of production and tested for predictive improvement.

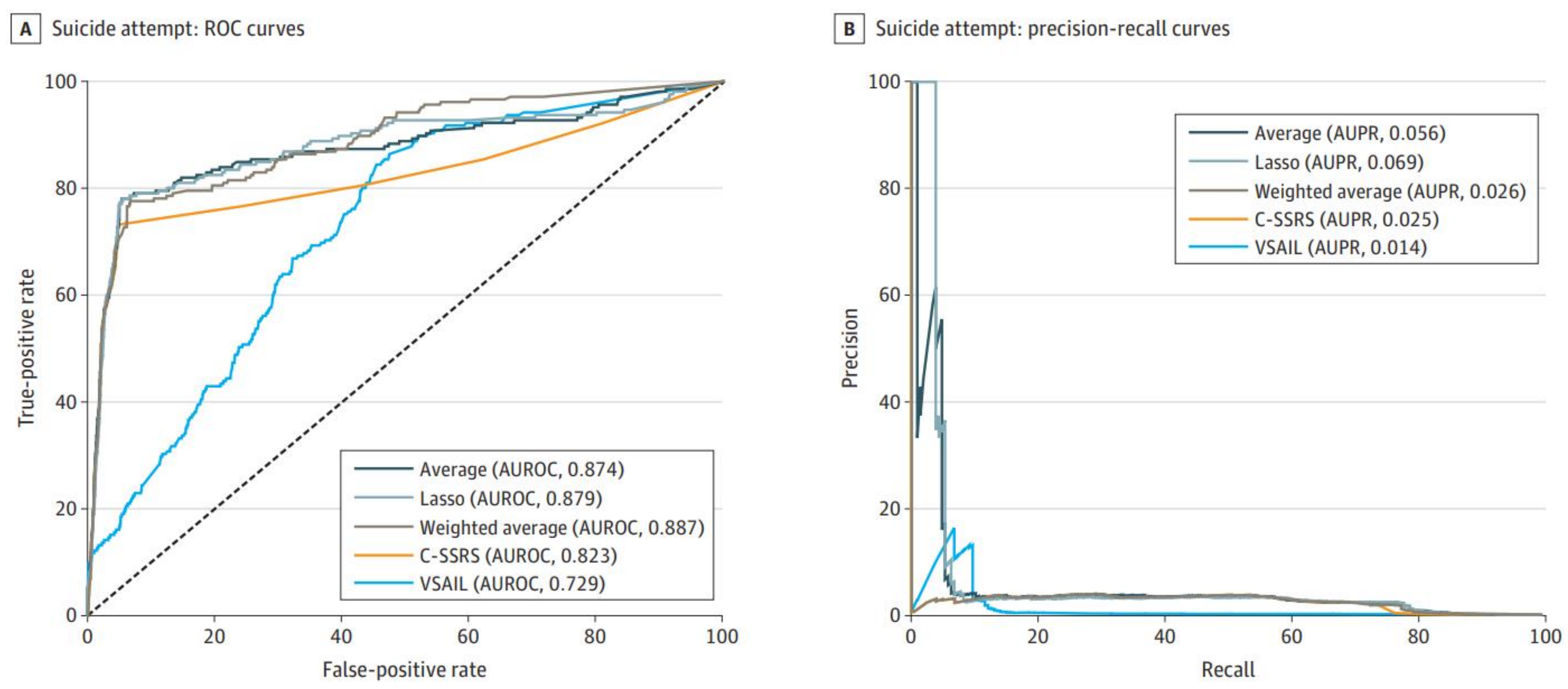

Figure 2. Synergistic effect of incorporating C-SSRS (AUROC: Area Under the Receiver Operating Curve)

- Adding C-SSRS had higher AUROC and AUPR scores than either the algorithm or C-SSRS alone.
- Lasso ensembling performed comparable to other methods and had the highest AUPR (Figure 2).

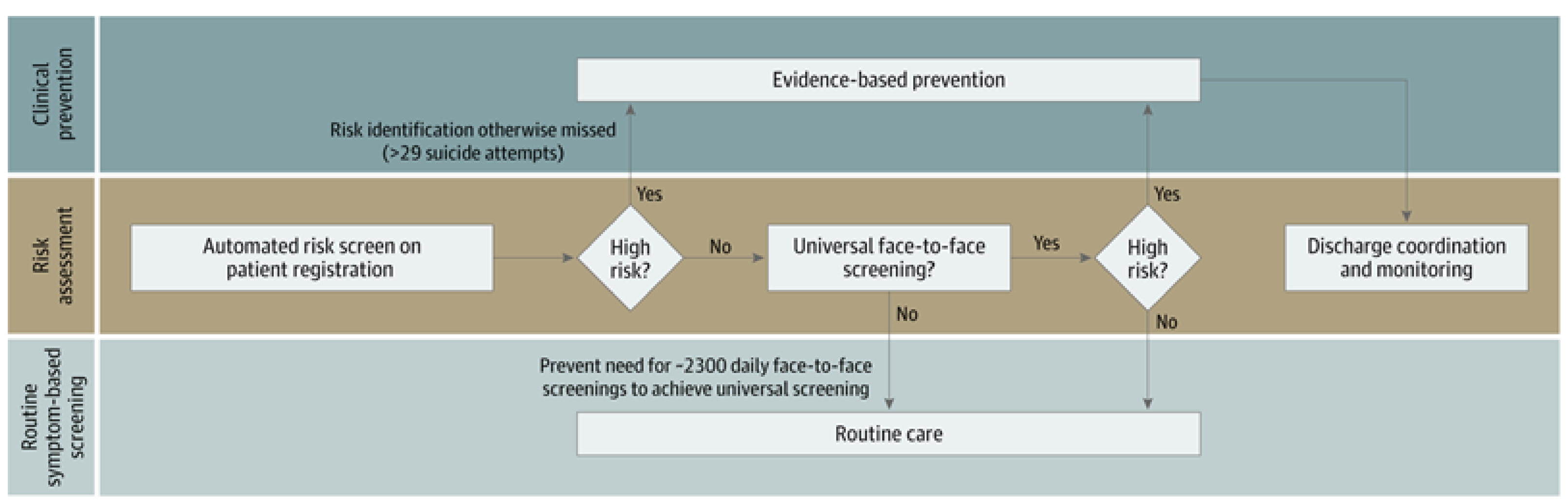

Figure 3. Predictive modeling-enabled suicide screening protocol

- Limitation: Only one medical center and structured data were used.
- Limitation: Death from suicide was not collected in this prospective study (open health system EHR).
- Natural Language Processing (NLP) will be used to ascertain additional suicide events and outcomes.

CONCLUSIONS

- Large amounts of true negatives in the lowest risk tier support automated risk screening.
- Large amounts of false positives were seen in the highest risk tier, but feasible NNSs exist for screening.
- Site-specific modeling may improve prediction in psychiatric settings.

- Recalibration was necessary to achieve calibration.
- Adding C-SSRS to the suicide risk model improved risk prediction.

VANDERBILT UNIVERSITY MEDICAL CENTER