Comparing the predictive value of suicide risk screening to the detection of suicide risk using electronic health records in an urban pediatric emergency department

Haroz, E.E., PhD^{1*} Kitchen, C., MS² Nestadt, P.S., MD^{3,4} Wilcox, H.C., PhD^{4,3} DeVlyder, J.E., PhD⁵ & Kharrazi, H., MD, PhD^{2,6}

*Corresponding Author; eharoz1@jhu.edu; 415 N. Washington St. Baltimore, MD 21231, (410) 449-0051

¹ Center for American Indian Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

² Center for Population Health IT, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

³ Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD

⁴ Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

⁵ Graduate School of Social Service, Fordham University, New York, NY

⁶ Division of Health Sciences Informatics, Johns Hopkins School of Medicine, Baltimore, MD

Funding: Author E.E.H. is supported by NIMH grant number: K01MH116335. Authors HK, CK, PSN and HCW are supported by NIMH grants R01MH124724 and R56MH117560. Author P.S.N. is supported by the James Wah Foundation for Mood Disorders Research and the American Foundation for Suicide Prevention.

Conflict of interest: None of the authors have any conflicts of interest to declare.

WORD COUNT: 6370 (All inclusive)

Abstract

Objective: To compare the accuracy, sensitivity and utility of brief screening to predictive modeling for identifying suicide-related outcomes in a pediatric emergency department. Our hypothesis was that predictive modeling would be more accurate and useful compared to brief screening.

Methods: This was a retrospective cohort study at an urban pediatric Emergency Department (PED) in the United States. Patients were aged 8 to 18 years old who presented from January 1, 2017 to June 30, 2019. Predictors included positive suicide risk screen and/or demographic, medical and mental health diagnostic Electronic Health Record (EHR) data at the time of visit. The main outcome was subsequent PED visit with suicide-related presenting problem (i.e. ideation or attempt) within a 3-month follow-up period.

Results: Of the $N=13420$ individuals, $n=141$ had a subsequent suicide-related ED visit. Approximately 63% identified as Black Non-Hispanic. Results showed that a model based only on EHR data performed only slightly worse than the ASQ alone. Combining ASQ screening and EHR data resulted in a 17.4% improvement in sensitivity and 13.4% increase in AUC compared to the ASQ alone. The LASSO models indicated a diagnosis of major depressive disorder in addition to the ASQ improved the AUC by 9%.

Conclusions, Implications, Future Directions: Our findings show that EHR data is helpful either in the absence or as an addition to brief suicide screening. To our knowledge, this is the first study to compare brief screening to EHR based predictive modeling and adds to our understanding of how best to identify youth at risk of suicidal behaviors in clinical care settings.

KEYWORDS: Suicide Screening; Suicide Prevention; Machine Learning

Introduction

Since 2014, suicide has been the 2nd leading cause of death among youth aged 10-14 and 15- 24-years old in the United States.¹ Suicide rates have increased in nearly every state since 1999, with female and Black youth showing especially marked increases over this time period.²⁻⁵ Moreover, 17.2% of high school students nationwide have reported seriously considering attempting suicide in the last year, and 7.4% of students report having attempted suicide in the past 12 months.⁶ Identifying youth most at risk of attempting or dying by suicide is a key part of any preventative approach. The timely identification of those at risk and provision of appropriate and effective care could have a major impact on driving down suicide rates.

Healthcare institutions serve as a critical setting for suicide risk detection. First, healthcare settings are a potential point of contact for someone at risk of attempting or dying by suicide. Several studies focused on examining medical records in the year prior to among suicide decedents have found high rates of utilization among those that died.^{7,8} For example, Ahmedani et al.,⁷ showed that in the month prior to their death, half of all suicide decedents received healthcare services at least once. Second, the acute time period following discharge from a healthcare facility is often characterized by elevated risk for suicide attempt and death.⁹⁻¹¹ Chung et al.¹⁰ showed that across 100 studies, the rate of suicide during the first 3 months after discharge from a psychiatric facility was approximately 100 times higher than the global suicide rate. The increase in risk following healthcare contact is also seen among patients discharged from non-psychiatric settings, including emergency departments (EDs), as well.⁹

The identification of suicide risk at healthcare facilities is a major priority in suicide prevention efforts. However, in recent years two distinct approaches have emerged regarding how best to do this. The first approach emphasizes the use of short, validated screening

measures. The Joint Commission (TJC) requires screening for suicide risk among all patients being evaluated or treated for behavioral health conditions with a validated screening tool, to be in compliance with the National Patient Safety Goal 15.01.01.¹² More broadly, the US Action Alliance’s “Zero Suicide” (ZS) model for suicide prevention in healthcare systems includes universal screening for suicide risk at all points of care in a healthcare setting as one of several essential elements.¹³ Studies have shown that use of brief screeners is feasible,¹⁴ that positive screens are significantly associated with subsequent suicide-related hospital visits,¹⁵ and that screening may double detection of those most at risk.^{15,16} However, implementation of screening involves several important hurdles, including generating buy-in from hospital administrators, programming of screeners into the electronic health record (EHR), training of staff in screening procedures and appropriate triage actions, and ongoing monitoring of fidelity to the instrument and process.^{14,17–19} Many screeners, which generally center on the question of suicidal ideation, can suffer from both high rates of false positives as well as even more dangerous false negatives, which can create a false sense that risk has been appropriately ruled in or out.²⁰

Simultaneously, the field of suicide prevention has also been pursuing predictive modeling as a method for identifying individuals at risk for attempting or dying by suicide. The availability of large health care datasets such as the EHR,²¹ combined with advances in machine learning and an increased understanding of the complexities of co-occurring risk factors for suicide,²² has led to a growing number of studies aimed at developing, validating and deploying prediction models as routine care to identify individuals at risk for subsequent attempts and deaths.^{23–27} Accuracy of these models is generally high;²³ however, these models have rarely been implemented as routine care in health systems²⁸ as model deployment is often complicated

by factors related to preparing data, developing clinician friendly tools and notifications, and ethical and regulatory issues.^{29,30}

The parallel development and enthusiasm for both universal screening and predictive modeling begs several questions: First, are either of these approaches more accurate at identifying those at risk? Second, how often do these methods agree? Third, how can these methods work in tandem; And, forth, what are the implications for implementation related to these two approaches? The present study compared these two approaches to predicting subsequent suicide related outcomes among a sample of patients aged 8-18 years who presented to the Johns Hopkins Hospital (JHH) pediatric emergency department (PED) from January 1, 2017 to June 30, 2019. The aims were to (1) compare a brief screening measure (the Ask Suicide-Screening Questions; ASQ³¹) to an EHR-based predictive modeling approach on accuracy, sensitivity and utility for identifying those at risk of subsequent return to the PED for suicide-related reasons; and (2) measuring the added value of integrating EHR with ASQ for identify subsequent PED suicide-related visits among our sample. Our hypotheses were that EHR-based predictive modeling would provide similar accuracy, sensitivity and utility as brief suicide screening; and the combination of risk screening and EHR-based predictive modeling would perform the best. To our knowledge, no study has directly compared EHR-based predictive modeling methods to brief suicide risk screeners in terms of their ability to identify those at risk for future suicidal behaviors. Findings from this study will inform our understanding of how best to identify individuals at risk of suicide, as well as implications for implementing such approaches in clinical care settings.

Methods

The study was a retrospective cohort study of all patients seen at Johns Hopkins Children's Center's PED from January 1, 2017 to June 30, 2019. Johns Hopkins Children's Center's PED is part of an urban academic medical center located in Baltimore, MD and provides over 30,000 ED visits a year. The medical record review and analysis included in this study was approved by the Johns Hopkins School of Medicine Institutional Review Board (IRB). Patients did not need to provide informed consent for screening as all data were collected during routine care. All reporting for this study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cohort studies.

Screening Assessment

Universal screening for risk of suicide was accomplished using the Ask Suicide-Screening Questions (ASQ) of all patients ages 10 and older and patients ages 8 and 9 presenting with a behavioral or psychiatric chief complaint. The ASQ is a freely available, brief, validated tool designed for use with patients ages ten years or above.³¹ It includes four screening questions that take 20 seconds to administer.³¹ A positive screen is indicated if a positive (i.e. "yes") response is provided to any of the four questions. If the patient screens positive, a fifth question is asked aimed at understanding acuity of the risk. The ASQ was administered by nurses in the PED during triage. If a patient screened positive, the protocol required that the ED physician was notified and the patient was provided with additional evaluations and referrals as appropriate, although how often this actually happened was variable.³²

Measures

Our outcomes of interest are defined as a positive instance of suicide ideation or attempt at PED visit within three and twelve months of the index visit, independent of responses obtained as part of ASQ screening. We did not rely only on International Classification of Diseases (ICD) codes (e.g. E950-E958) as these were seldom found for the same ED encounters (90 of 25,067 ED encounters between 1/1/2017 and 6/30/2019). Thus, we defined instances of PED suicide ideation or attempt on recorded principal diagnosis description and chief complaint as documented in the EHR, for those missing ICD codes. These fields contained free-text values that reflect reasons for the patient being seen in the ED, and tended to be very brief (on average 13 to 32 characters, or 39-100 max). Our outcome was a binary indicator for pattern matched terms “suicide”, “suicidal”, “suicide ideation”, “suicide gesture”, “self-cutting”, “self-harm”, “self-injury”, “self-injurious”, “self-inflicted”, “self-mutilation”, “self-mutilating”, “suicide attempt”, and “attempted suicide” on either PED chief complaint or presenting diagnosis. As a sensitivity analysis, we compared our pattern matched outcome measure to manually scored suicidal behaviors for overlapping patients in data from DeVlyder et al.¹⁵

Predictors of interest focused on both acute and chronic psychiatric and medical comorbidities diagnosed at the time of PED visit. Twenty-five medical and psychiatric condition categories were obtained from the Centers for Medicare and Medicaid Services (CMS) Chronic Conditions Warehouse (CCW) and were matched to ICD codes extracted from index PED visits. The CCW markers were chosen on the basis of prior utility in PED suicide risk prediction,¹⁵ plausibility for incidence within the pediatric population (i.e. excluding conditions such as stroke, hip fracture, osteoporosis), and potential relevance for other clinical settings. An additional 17 conditions and 4 comorbidity risk scores (i.e. Charlson comorbidity index) were obtained using the ‘comorbidity’ package³³ for the R programming language (3.6.2). Additional

markers were created for patient age, gender and whether a patient screened positive on the ASQ. We did not include an indicator for race and/or ethnicity for concern about propagating potential “race-based medicine.”³⁴ The ASQ item 4 was also retained for use in a contrast model and consists of asking patients whether they have ever tried to kill themselves. In this manuscript, demographic and diagnostic information, including the Charlson weighted index of comorbidity, are referred to as “EHR” data. We then compared the value of selected EHR data to an ASQ positive screen or affirmative response on ASQ item 4 for predicting suicide-related ED visits.

Data processing

Data preparation consisted of fitting available PED encounter level data to all patients seen in the PED between 1/1/2017 and 6/30/2019. The aggregation of encounter data required identifying case and control groups, defined by whether a patient had an ED visit for suicide ideation or attempts within 3 months of an initialized PED visit. To identify as many cases as possible, we first found all patients that had PED visits in the preceding three months of an ED visit with ideation or attempt. If a patient had multiple spans in which this condition could be met, only the first instances were retained. Relevant encounter data were aggregated within the observation interval such that the presence of any diagnostic condition was flagged positive, leading up to the visit containing ideation or attempt. ‘Controls’ were similarly identified and processed, with indexing on the first observed encounter date and included all single event PED visits. In this way, it was possible for ‘controls’ to have an initial instance of ideation or attempt, but *not* a subsequent one within three months of the indexed date. Finally, ‘control’ patients who

had less than 3 months of observation (i.e. first visit after March, 30th 2019) were removed from the analysis.

Missing Data

Only missing values for the ASQ item 4 were necessary for modeling. These missing values were imputed using the ‘mice’ package in R.³⁵ A total 663 values were imputed for ASQ item 4 using this method. For the other ASQ items missing values existed with progressively more missing values later in the questionnaire (ASQ 1: 102; ASQ 2: 630; ASQ 3: 620; ASQ 4: 663), suggesting that omitted values resulted from questions not being asked at the time of screening.

Statistical Analysis

Exploratory analyses

We compared cases and controls on the basis of age, gender, race/ethnicity, ASQ screening outcome and subsequent ED visit outcome (i.e. ideation or attempt-related or not). All predictors and ASQ questions responses were loaded into point biserial correlation matrices, which aided in removal of collinear or empty variables. Our outcome sensitivity analysis found excellent overall accuracy (97.9%) when compared to previously manually coded outcomes in a similar dataset, but different study population (e.g. DeVlyder et al.¹⁵). The pattern matched definition was also more inclusive, with 309 additional instances of suicidal ideation or attempt over those detected manually and only 6 instances where manual review identified suicide ideation or attempt where our definition did not.

Modeling

The three-month and 12-month analyses consisted of the construction of four models predicting ED visits for suicide ideation or attempt. These included: (1) ASQ positive screen, (2) ASQ positive screens and EHR content, (3) EHR content without the ASQ positive screen, and (4) The EHR content with item 4 of the ASQ. Model 4 was added, given the importance of a history of suicide attempt for future suicide risk²⁰ and the relative ease of asking one question compared to five. Three-month analyses are presented in this paper, while twelve month results are included in the Appendix (see Appendix A1).

For all models, five-fold cross validation was used with binary logistic regression and LASSO regression. Binary logistic regressions were performed using the “cv” method with the train function of R’s caret library.³⁶ LASSO was performed using cv.glmnet with a binomial link function and lambda set at one standard error above average during cross validation. Minimum classification thresholds were set to the probability of a randomly chosen patient having subsequent PED visit for ideation or attempt within three months ($p = .01$). Performance metrics were returned for model accuracy, sensitivity, specificity, positive predictive value, false positive rate and AUC (area under the curve). Model selection demanded consideration of sensitivity, parsimony and fit (AUC) given the high cost of false negative prediction. Variable importance and ROC (receiver operating characteristic) plots were generated to further interpret the performance of the binary logistic regression models.

Secondary Analysis

The above definition of our outcome of interest in PED visits allows for the inclusion of an additional variable: whether a patient had a ED visit for ideation or attempt within the past

three months (twelve months included in Appendix A1) for of their outcome. This was defined as the above pattern matched terms for presenting diagnosis description and chief complaint within the interval leading up to the outcome event. Binary logistic regression and LASSO models were developed with all predictors plus an indicator for ideation or attempt at index visit to inspect the added value of this one marker on model performance (See Appendix A2).

Results

Sample Characteristics

A total of 13 420 unique patients in the targeted age range were seen in the PED between January 1, 2017 and June 30, 2019 (**Table 1**). Among our sample, 7040 (52.5%) were female, 6280 (47.5%) were male, and the average age was 14.3 years old. The majority of the sample identified as Black Non-Hispanic (8438; 63%), while a quarter self-identified as White Non-Hispanic (3370; 25%), 914 (6.8%) as Hispanic, and 703 (5.2%) as other or more than one race. All patients seen in the PED were screened with the ASQ. Compliance with each individual ASQ question was high: 99.2% for question 1, 95.3% for question 2; 95.4% for question 3, and 95.1% for question 4. A total of 865 patients (6.4%) presented with an initial complaint of either suicidal ideation and/or suicide attempt during the timeframe for this analysis, from whom 141 patients had a subsequent suicide-related ED visit.

TABLE 1 HERE

Prediction Models

Table 2. shows the comparison of binary logistic regression models based on accuracy, sensitivity, specificity, positive predictive value (PPV), false positive rate and AUC. Model 1, which includes just the binary predictor of ASQ positive vs. negative screen on subsequent suicide-related ED visit had the highest accuracy metric (0.840), but relatively low sensitivity (0.667) and the lowest value for AUC (0.754; **Figure 1.**). A model based only on data from the EHR (Model 3), performed only slightly worse than the ASQ alone as it had the lowest sensitivity (0.601) but slightly higher AUC value (0.775; **Figure 1.**). The best performing model was the combination of screening with the ASQ and use of EHR (Model 2). This model showed an accuracy of 0.825, sensitivity of 0.783 and AUC of 0.855 (**Figure 1.**). When EHR is added to the ASQ screening process, the sensitivity is improved by 17.4%, AUC by 13.4%, while there is a slight loss of specificity (-1.9%). Across all models negative predictive value (NPV) was higher than 0.99. False positive rates ranged from 15.8% for Model 1 to 17.4% for Model 2. False negative rates ranged from 21.7% for Model 2 to 40% for Model 3. Results for our twelve-month analyses are included in Appendix A1. Similar results were observed for the twelve-month follow-up time point: Model 2 (ASQ + EHR) showed the highest sensitivity (0.746; Table A1) and AUC (0.843; Table A1).

TABLE 2 HERE

FIGURE 1 HERE

Penalized Logistic Regression

Table 3. presents the comparison of model accuracy across the same four models but using penalized logistic regression to shrink the coefficients of the less contributive features. The results for Model 1 (i.e. ASQ screening alone) do not change because only one feature exists in the model. However, using the LASSO method, despite the substantially reduced number of features, we continue to see improvement in the sensitivity of the model when EHR is added to information from the ASQ. In fact, just adding one additional feature, a diagnosis of depression at index PED visit, results in this increased sensitivity: Model 2, the ASQ plus EHR has an accuracy of 0.799, sensitivity of 0.804 and AUC of 0.815; while Model 4, using only the 4th question on the ASQ that reflects a self-report history of suicide attempted demonstrated higher accuracy (0.831), but lower sensitivity (0.659) and lower AUC (0.759). False positive rates ranged from 0.082 for Model 3 (i.e. EHR alone) to 0.201 for Model 2 (i.e. ASQ + EHR).

TABLE 3 HERE

Secondary Analysis

For the binary logistic regression, the inclusion of an index PED visit for suicide attempt or ideation coded in the EHR, improved the model accuracy of both Model 2 (ASQ + EHR; accuracy = 0.841) by 2% and Model 3 (EHR only; accuracy = 0.858) by 3% respectively. The AUC improved as well, with Model 2 increasing from 0.855 to 0.874 and Model 3 raising from 0.775 to 0.843 with the addition of this predictor. The largest improvement was observed for sensitivity for Model 3 (EHR only; elevating from 0.601 to 0.717). In summary, the secondary analysis depicted that if an EHR is capable of coding and utilizing this type of data, a model based only on EHR would be more sensitive, specific, and have a higher accuracy than the ASQ alone (see Appendix A2).

Discussion

Main Findings

In this study we sought to compare a brief screening tool, the ASQ, to EHR-based predictive models for their accuracy at identifying individuals at risk of future suicide related ED visits. Our study found that the addition of EHR data to brief suicide screening in a PED setting considerably improves accuracy at identifying individuals at risk of future suicidal behaviors over the next three months. Our logistic regression findings indicate that relying solely on EHR data performs only slightly worse than implementing universal brief suicide screening. And while the best performing models were brief screening and EHR data, using just one item from a brief screener that asks about past history of suicide attempt (i.e., last question of the ASQ) and combining this data with EHR data, leads to substantial improvements in predictive accuracy compared to either screening or EHR models alone.

Given the potential to save lives by identifying those at risk of suicide in hospital settings, a substantial, but parallel effort has been made to develop brief screening measures^{31,37} and predictive models.^{24–26,38} To our knowledge, this is the first study to do a head to head comparison of these approaches. Our findings show that leveraging EHR data can be help improve predictive accuracy and utility when implementing universal screening procedures in a PED. Notably, model accuracy was only around 1% less for the EHR alone model (Model 3) compared to the ASQ (Model 1), though the latter remained 10% more sensitive. However, when added to the ASQ the EHR attributes (Model 2), improved sensitivity by 17.4% and AUC by 13.4% compared to the ASQ only model (Model 1).

Our secondary findings (Appendix 2) indicated that an index visit with diagnosis of suicidal behaviors was important in model accuracy. Inclusion of ideation or attempt at PED

visits as a predictor improved model sensitivity by 9.8% and the AUC by 14% compared to the ASQ alone. Given known problems with documentation of ideation or attempt,³⁹ our main findings indicate that simply asking one question from the ASQ in addition to the data gathered in a PED visit, is comparable to asking a full screening measure in terms of accuracy at identifying future suicide related PED visits. An advantage of screening, even with this one question, is it can give you the historical record and by-passes a potential limitation of EHRs – that if there is no documented history, the EHR by default considers this to mean there is no history, but we know that an absence of evidence is not evidence of absence.

Despite widescale recommendations to routinely screen for suicide risk among children and adolescents,^{12,40} instituting and implementing universal screening requires significant buy in, planning, and continued resources. First, instituting universal screening in a health system requires screening to be built into the work-flow of clinical care including protocols for addressing acute and nonacute screen positives.⁴¹ Given a lack of ready access to community mental health services, some hospital administrators can be reluctant to support institutionalizing this approach. Additional concerns involve the low positive predictive value of such screeners²⁰ or operational modifications needed to implement these approaches. Implementation of screening requires training and ongoing oversight to ensure compliance.^{18,42} While our sample had high ASQ compliance rates, previous studies have shown lower compliance rates indicating a need for ongoing monitoring and quality improvement efforts.⁴³ Notwithstanding, screening for suicide risk has shown promise at identifying patients at risk for suicide.^{14,15}

Predictive modeling is considered potentially less burdensome from a human resource perspective. Instead of ongoing supervision to administer surveys, these modeling techniques leverage data analytics approaches that can be automated with ever advancing computing

capacity through the EHR. However, suicide risk models may have considerable disadvantages. Some patients may feel that this approach is invasive of their privacy, particularly because they may or may not consent to having an algorithm applied to their data. Further, in-person screening may provide therapeutic benefits that would not be present with computational modeling, although the evidence for this is mixed.⁴⁴

Both screening and predictive modeling approaches have limitations. The positive predictive value across all our models were quite low (e.g. 0.04-0.05). This is consistent with a recent systematic review by Belsher et al.²³ focused on predictive models, findings from screening studies, and is reflective of the key challenge in suicide risk screening – suicide is a low base rate problem.⁴⁵ Our false positive rates were considerably lower than those found in previous studies,⁴⁶ but still represent potential inappropriate use of resources and/or a potential source of distrust or source of fatigue in procedures.⁴⁷ While our work and the work of others show suicide risk models can accurately identify risk, depending on how they are operationalized, these models can generate additional workload for clinicians by necessitating response to false positives and strategies for responding to alerts on patients.⁴⁸ Regardless of the approach, careful attention is needed to think about how a risk classification is communicated to the patient as both screening and predictive modeling may feel “out of the blue” to patients, particularly patients who present with medical complaints.

Despite these drawbacks, early and accurate identification of individuals at risk a critical component of a prevention framework. Implementation of either approach should be carefully studied. Developing optimal risk thresholds and care pathways is critical to realizing the value of these procedures in medical settings.²⁸ Focusing on the net benefit of these approaches in a given

setting may be one way to more clearly identify the optimal risk thresholds and support clinical decision making.⁵⁰

Finally, few suicide risk prediction models have been developed in patient populations that represent racial and ethnic diversity. Our sample includes approximately two-thirds Black youth patients. Over the past decade, suicide rates are increasing at a faster rate among Black youth compared to any other racial or ethnic minority.⁴ This crisis needs to be addressed urgently; however, significant concerns exist about current approaches, including predictive models, which may exacerbate health disparities and/or potentially cause individual harm.⁵⁰ Development and implementation of any risk identification strategy should not only consider net benefits, but also focus on incorporation of principles of distributive justice such as equalized outcomes, equal performance and equal allocation.⁵⁰

Limitations

The study has several important limitations worth noting. First, our outcomes did not include suicide deaths. What predicts ideation or attempts could be vastly different from what predicts future deaths by suicide. Moreover, our outcomes were not primarily defined by ICD codes, which may have resulted in some misclassification. Misclassification could have also contributed to too few cases identified given the potential overlap with drug overdoses. However, misclassification with ICD codes is a known issue in suicide research³⁹ and a relative strength of this study is definition of outcomes through review of clinician notes. The generalizability of our findings may also be limited. First, data represents the patient population seen at the Johns Hopkins Hospital (JHH) PED. Generalizing these findings to rural populations, higher-income settings, and/or outpatient populations is not appropriate. Furthermore, our data is

only representative of PED data captured at JHH, which excludes potential visits at other EDs.

Finally, although we used rigorous methods for cross-validation, external validation of our models would further strengthen our findings.

Conclusions

This study represents the first comparison of universal suicide screening and predictive modeling to identify individuals at risk of suicidal behavior in the next three and twelve months. We found that predictive modelling based on EHR may serve as a comparable approach when instituting a brief screening measure is not feasible. The combination of both brief screening, even in the form of a single item (e.g. ASQ4), and predictive modelling with EHR, substantially improves accuracy. Regardless of the method, careful attention should be paid towards optimizing the implementation of either approach, including focusing on net benefits and distributive justice principles.

References

1. Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS) [Online]. Available from: <https://www.cdc.gov/injury/wisqars/>. Accessed November 24, 2020.
2. Ruch DA, Sheftall AH, Schlagbaum P, Rausch J, Campo J V., Bridge JA. Trends in Suicide Among Youth Aged 10 to 19 Years in the United States, 1975 to 2016. *JAMA Netw open*. 2019;2(5):e193886. doi:10.1001/jamanetworkopen.2019.3886
3. Curtin S, Warner M, Hedegaard H. *Increase in Suicide in the United States, 1999-2014. NCHS data brief*. 2016;241:1-8.
4. The Congressional Black Caucus. *Ring the Alarm: The Crisis of Black Youth Suicide in America*. 2019.
5. QuickStats: Suicide Rates*[†] for Teens Aged 15-19 Years, by Sex - United States, 1975-2015. *MMWR Morb Mortal Wkly Rep*. 2017;66(30):816. Published 2017 Aug 4. doi:10.15585/mmwr.mm6630a6
6. Kann L, McManus T, Harris WA, et al. Youth Risk Behavior Surveillance — United States, 2017. *MMWR Surveill Summ*. 2018;67(8):1-114. doi:10.15585/mmwr.ss6708a1
7. Ahmedani BK, Simon GE, Stewart C, et al. Health care contacts in the year before suicide death. *J Gen Intern Med*. 2014;29(6):870-877. doi:10.1007/s11606-014-2767-3
8. Chock MM, Bommersbach TJ, Geske JL, Bostwick JM. Patterns of health care usage in the year before suicide: A Population-Based Case-Control Study. *Mayo Clin Proc*. 2015;90(11):1475-1481. doi:10.1016/j.mayocp.2015.07.023
9. Wang M, Swaraj S, Chung D, Stanton C, Kapur N, Large M. Meta-analysis of suicide rates among people discharged from non-psychiatric settings after presentation with

- suicidal thoughts or behaviours. *Acta Psychiatr Scand.* 2019;139(5):472-483.
doi:10.1111/acps.13023
10. Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, Large MM. Suicide rates after discharge from psychiatric facilities: A systematic review and meta-analysis. *JAMA Psychiatry.* 2017;74(7):694-702. doi:10.1001/jamapsychiatry.2017.1044
11. Olfson M, Wall M, Wang S, et al. Short-Term suicide risk after psychiatric hospital discharge. *JAMA Psychiatry.* 2016;73(11):1119-1126.
doi:10.1001/jamapsychiatry.2016.2035
12. Joint Commission on Accreditation of Healthcare. Hospital national patient safety goal 15.01. 01. 2015.
13. Claassen CA, Pearson JL, Khodyakov D, et al. Reducing the burden of suicide in the U.S.: The aspirational research goals of the national action alliance for suicide prevention research prioritization task force. *Am J Prev Med.* 2014;47(3):309-314.
doi:10.1016/j.amepre.2014.01.004
14. Ballard ED, Cwik M, Van Eck K, et al. Identification of At-Risk Youth by Suicide Screening in a Pediatric Emergency Department. *Prev Sci.* 2017;18(2):174-182.
doi:10.1007/s11121-016-0717-5
15. DeVyllder JE, Ryan TC, Cwik M, et al. Assessment of Selective and Universal Screening for Suicide Risk in a Pediatric Emergency Department. *JAMA Netw open.* 2019;2(10):e1914070. doi:10.1001/jamanetworkopen.2019.14070
16. Boudreaux ED, Haskins BL, Larkin C, et al. Emergency department safety assessment and follow-up evaluation 2: An implementation trial to improve suicide prevention. *Contemp Clin Trials.* 2020;95:106075. doi:10.1016/j.cct.2020.106075

17. Labouliere CD, Vasan P, Kramer A, et al. "Zero Suicide" – A model for reducing suicide in United States behavioral healthcare. *Suicidologi*. 2018;23(1):22-30.
doi:10.5617/suicidologi.6198
18. Inman DD, Matthews J, Butcher L, Swartz C, Meadows AL. Identifying the risk of suicide among adolescents admitted to a children's hospital using the Ask Suicide-Screening Questions. *J Child Adolesc Psychiatr Nurs*. 2019;32(2):68-72.
doi:10.1111/jcap.12235
19. Snyder DJ, Jordan BA, Aizvera J, et al. From Pilot to Practice: Implementation of a Suicide Risk Screening Program in Hospitalized Medical Patients. *Jt Comm J Qual Patient Saf*. 2020;46(7):417-426. doi:10.1016/j.jcjq.2020.04.011
20. Nestadt PS, Triplett P, Mojtabai R, Berman AL. Universal screening may not prevent suicide. *Gen Hosp Psychiatry*. 2018;S0163-8343(18):30066-5.
doi:10.1016/j.genhosppsy.2018.06.006
21. Walkup JT, Townsend L, Crystal S, Olfson M. A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. *Pharmacoepidemiol Drug Saf*. 2012;21(SUPPL. 1):174-182. doi:10.1002/pds.2335
22. 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
23. 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
24. Haroz EE, Walsh CG, Goklish N, Cwik MF, O'Keefe V, Barlow A. Reaching those at

- highest risk for suicide: Development of a model using machine learning methods for use with Native American communities. *Suicide Life-Threatening Behav.* 2020;50(2):422-436. doi:10.1111/sltb.12598
25. Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *J Child Psychol Psychiatry Allied Discip.* 2018;59(12):1261-1270. doi:10.1111/jcpp.12916
26. 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
27. 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). doi:10.1002/mpr.1575
28. Gordon JA, Avenevoli S, Pearson JL. Suicide prevention research priorities in health care. *JAMA Psychiatry.* 2020;77(9):885-886. doi:10.1001/jamapsychiatry.2020.1042
29. Bates DW, Auerbach A, Schulam P, Wright A, Saria S. Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence. *Ann Intern Med.* 2020;172(11):S137-S144. doi:10.7326/M19-0872
30. Char DS, Shah NH, Magnus D. Implementing machine learning in health care ' addressing ethical challenges. *N Engl J Med.* 2018;378(11):981-983. doi:10.1056/NEJMp1714229
31. Horowitz LM, Bridge JA, Teach SJ, et al. Ask suicide-screening questions (ASQ): A brief instrument for the pediatric emergency department. *Arch Pediatr Adolesc Med.* 2012;166(12):1170-1176. doi:10.1001/archpediatrics.2012.1276

32. Taylor R, Susukida R, Cwik M, et al. Frequent Ed Visits Among Youth Identified at Risk for Suicide. IASR/AFSP Annual Conference. Miami; 2019.
33. Gasparini A. comorbidity: An R package for computing comorbidity scores. *Software Review Repository Archive*. 2018. doi:10.21105/joss.00648
34. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. *N Engl J Med*. 2020;383(9):874-882. doi:10.1056/nejmms2004740
35. MICE: Multivariate Imputation by Chained Equations in R. *Journal of statistical software*. 2010:1-68.
36. Kuhn M. *The Caret Package*. *Journal of statistical software*. 2008:1-26.
37. Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): Classification of Suicidal Events in the FDA’s Pediatric Suicidal Risk Analysis of Antidepressants. *Am J Psychiatry*. 2007;164(7):1035-1043. doi:10.1176/ajp.2007.164.7.1035
38. Kessler RC, Warner CH, Ivany C, et al. 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
39. 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. *National health statistics report*. 2018;108:1-19.
40. Shain B. Suicide and suicide attempts in adolescents. *Pediatrics*. 2016;138(1).

doi:10.1542/peds.2016-1420

41. Roaten K, Johnson C, Genzel R, Khan F, North CS. Development and Implementation of a Universal Suicide Risk Screening Program in a Safety-Net Hospital System. *Jt Comm J Qual Patient Saf.* 2018;44(1):4-11. doi:10.1016/j.jcjq.2017.07.006
42. Hackfeld M. Implementation of a pediatric/adolescent suicide risk screening tool for patients presenting to the Emergency Department with nonbehavioral health complaints. *J Child Adolesc Psychiatr Nurs.* 2020;33(3):131-140. doi:10.1111/jcap.12276
43. Betz ME, Wintersteen M, Boudreaux ED, et al. Reducing Suicide Risk: Challenges and Opportunities in the Emergency Department. *Ann Emerg Med.* 2016;68(6):758-765. doi:10.1016/j.annemergmed.2016.05.030
44. Graney J, Hunt IM, Quinlivan L, et al. Suicide risk assessment in UK mental health services: a national mixed-methods study. *The Lancet Psychiatry.* 2020;7(12):1046-1053. doi:10.1016/S2215-0366(20)30381-3
45. Perlis RH, Fihn SD. Hard Truths About Suicide Prevention. *JAMA Netw open.* 2020;3(10):e2022713. doi:10.1001/jamanetworkopen.2020.22713
46. Nelson HD, Denneson LM, Low AR, et al. Suicide Risk Assessment and Prevention: A Systematic Review Focusing on Veterans. *Psychiatr Serv.* 2017;68:1003-1015. doi:10.1176/appi.ps.201600384
47. Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, Sittig DF. Information overload and missed test results in electronic health record-based settings. *JAMA Intern Med.* 2013;173(8):702-704. doi:10.1001/2013.jamainternmed.61
48. 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

49. 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.

Molecular psychiatry. 2020; 25(1):168-79. doi:10.1038/s41380-019-0531-0

50. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med.* 2018;169(12):866-872.

doi:10.7326/M18-1990

Table 1. Demographic and clinical descriptive data by case status ($N = 13,420$)

	Control	Cases
	13,279 (99%)	141 (1%)
Age (mean, SD)	14.3 (3.1)	14.0 (2.5)
Gender		
Female	6952 (52.4%)	88 (62.4%)
Male	6327 (47.6%)	53 (37.6%)
Race/Ethnicity		
Black Non-Hispanic	8348 (62.9%)	90 (63.8%)
White Non-Hispanic	3330 (25.1%)	40 (28.4%)
Hispanic	903 (6.8%)	11 (7.8%)
Other/more than one race	698 (5.3%)	5 (3.5%)
ASQ Screening ¹		
Positive Screen	2099 (15.8%)	92 (66.7%)
Negative Screen	11,178 (84.2%)	46 (33.3%)
Index ER visit		
Non-suicide related	12,480 (94.0%)	75 (53.2%)
Ideation or attempt	799 (6.0%)	66 (46.8%)

¹Missing 5 ASQ screens

Table 2. Binary Logistic Regression model comparison for accuracy of detecting a suicide related ED event three months after index event ($N = 13,420$)

Model	Accuracy	Sensitivity	Specificity	Positive Predictive Value	False Positive Rate	False Negative Rate	AUC	Number of features
Model 1: ASQ Alone	0.840	0.667	0.842	0.042	0.158	0.333	0.754	1
Model 2: ASQ + EHR	0.825	0.783	0.826	0.045	0.174	0.217	0.855	41
Model 3: EHR alone	0.832	0.601	0.835	0.036	0.165	0.399	0.775	40
Model 4: EHR + ASQ-4	0.827	0.703	0.828	0.041	0.172	0.297	0.820	41

Table 3. LASSO model comparison for accuracy of detecting a suicide related ED event three months after index event ($N = 13,420$)

Model	Accuracy	Sensitivity	Specificity	Positive Predictive Value	False Positive Rate	False Negative Rate	AUC	Number of features
Model 1: ASQ Alone	0.840	0.667	0.842	0.042	0.158	0.333	0.754	1
Model 2: ASQ + EHR	0.799	0.804	0.799	0.040	0.201	0.196	0.815	2
Model 3: EHR alone	0.913	0.428	0.918	0.052	0.082	0.572	0.673	1
Model 4: EHR + ASQ-4	0.831	0.659	0.833	0.039	0.167	0.341	0.759	2

Appendix A1. Twelve-month prediction models

A1. Binary Logistic Regression model comparison for accuracy of detecting a suicide related ED event twelve months after index event ($N = 10\,205$)

Model	Accuracy	Sensitivity	Specificity	Positive Predictive Value	False Positive Rate	False Negative Rate	AUC
Model 1: ASQ Alone	0.836	0.621	0.843	0.100	0.157	0.379	0.732
Model 2: ASQ + EHR	0.815	0.746	0.817	0.103	0.183	0.254	0.843
Model 3: EHR alone	0.818	0.621	0.823	0.090	0.177	0.379	0.789
Model 4: EHR + ASQ-4	0.817	0.686	0.820	0.097	0.180	0.314	0.822

Figure A1. Receiver Operating Curve (ROC) comparing binary logistic regression models ($N = 10\,205$) for prediction accuracy at 12 months

Appendix A2. Model performance including suicide ideation or attempt at index PED visit as coded in the EHR

Table A2. Binary Logistic Regression model performance including the EHR marker for suicide ideation or attempt at index visit for detecting a suicide related ED event three months after index event ($N = 13\,420$).

Model	Accuracy	Sensitivity	Specificity	Positive Predictive Value	False Positive Rate	False Negative Rate	AUC
Model 1: ASQ Alone	0.840	0.667	0.842	0.042	0.158	0.333	0.754
Model 2: ASQ + EHR	0.841	0.768	0.842	0.048	0.159	0.232	0.874
Model 3: EHR alone	0.858	0.717	0.860	0.051	0.140	0.283	0.843
Model 4: EHR + ASQ-4	0.855	0.732	0.856	0.050	0.144	0.268	0.859

Figure A2. Receiver Operating Curve (ROC) comparing binary logistic regression models including the EHR marker for suicide ideation or attempt at index PED visit for prediction accuracy at 3 months ($N = 13,420$)