Comparing the predictive value of screening to the use of electronic health record data for detecting future suicidal thoughts and behavior in an urban pediatric emergency department: A preliminary analysis

Emily E. Haroz, PhD[1], Christopher Kitchen, MS[2], Paul S. Nestadt, MD[3,4], Holly C. Wilcox, PhD[3,4], Jordan E. DeVylder, PhD[5], Hadi Kharrazi, MD, PhD[2,6]

[1]Department of International Health, Center for American Indian Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[2]Department of Health Policy and Management, Center for Population Health IT, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[3]Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

[4]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[5]Graduate School of Social Service, Fordham University, New York, New York, USA

[6]Division of Health Sciences Informatics, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

Abstract

Aim: Brief screening and predictive modeling have garnered attention for utility at identifying individuals at risk of suicide. Although previous research has investigated these methods, little is known about how these methods compare against each other or work in combination in the pediatric population.

Methods: Patients were aged 8–18 years old who presented from January 1, 2017, to June 30, 2019, to a Pediatric Emergency Department (PED). All patients were screened with the Ask

Suicide Questionnaire (ASQ) as part of a universal screening approach. For all models, we used 5-fold cross-validation. We compared four models: Model 1 only included the ASQ; Model 2 included the ASQ and EHR data gathered at the time of ED visit (EHR data); Model 3 only included EHR data; and Model 4 included EHR data and a single item from the ASQ that asked about a lifetime history of suicide attempt. The main outcome was subsequent PED visit with suicide-related presenting problem within a 3-month follow-up period.

Results: Of the $N = 13,420$ individuals, $n = 141$ had a subsequent suicide-related PED visit. Approximately 63% identified as Black. Results showed that a model based only on EHR data (Model 3) had an area under the curve (AUC) of 0.775 compared to the ASQ alone (Model 1), which had an AUC of 0.754. Combining screening and EHR data (Model 4) resulted in a 17.4% (absolute difference = 3.6%) improvement in sensitivity and 13.4% increase in AUC (absolute difference = 6.6%) compared to screening alone (Model 1).

Conclusion: Our findings show that predictive modeling based on EHR data is helpful either in the absence or as an addition to brief suicide screening. This is the first study to compare brief suicide screening to EHR-based predictive modeling and adds to our understanding of how best to identify youth at risk of suicidal thoughts and behaviors in clinical care settings.

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 (National Center for Injury Prevention and Control, n.d.). Suicide rates have increased in nearly every state since 1999, with female and Black youth showing especially marked increases over this time period (CDC, 2017; Congressional Black Caucus, 2019; Curtin et al., 2016; Ruch et al., 2019). 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 (Kann et al., 2018). 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 potential points 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 (Ahmedani et al., 2014; Chock et al., 2015). For example, Ahmedani et al. (2014) 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 (Chung et al., 2017; Olfson et al., 2016; Wang et al., 2019). Chung et al. (2017) 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) (Wang et al., 2019).

The identification of suicide risk at healthcare facilities is a major priority in 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 (The Joint Commission, 2015). 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 (Claassen et al., 2014). Studies have shown that use of brief screeners is feasible (Ballard et al., 2017, Bahraini et al., 2020) that positive screens are significantly associated with subsequent suicide-related hospital visits (DeVylder et al., 2019), and that screening may double detection of those most at risk (Boudreaux et al., 2020; DeVylder et al., 2019). 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 (Ballard et al., 2017; Inman et al., 2019; Labouliere et al., 2018; Snyder et al., 2020). Manycan 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 (Nestadt et al., 2018).

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 healthcare datasets such as the EHR (Walkup et al., 2012), combined with advances in machine learning and an increased understanding of the complexities of co-occurring risk factors for suicide (Franklin et al., 2017), 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 (Belsher et al., 2019; Haroz et al., 2019; Kessler et al., 2017; Walsh et al., 2017, 2018). Accuracy of these models is generally high; however, they are not immune to the challenges of predictive rare events and suffer from similar low levels of predictive validity as screening tools (Belsher et al., 2019). Moreover, these models have rarely been implemented as routine care in health systems (Gordon et al., 2020) as model deployment is often complicated by factors related to preparing data, developing clinician friendly tools and notifications, and ethical and regulatory issues (Bates et al., 2020; Char et al., 2018).

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, fourth, 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; ASQ; Horowitz et al., 2012) 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 data 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. This study can be considered a first step in a process to examine if and how EHR data can be used in lieu of or in combination with brief suicide screening measures. 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 were approved by the Johns Hopkins School of Medicine Institutional Review Board. 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 10 years or above (Horowitz et al., 2012). It includes four screening questions that take 20 seconds to administer (Horowitz et al., 2012). 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 (Taylor et al., 2019).

Measures

Our outcomes of interest are defined as a positive instance of suicide ideation (SI) or attempt (SA) at PED visit within 3 and 12 months of the index visit, independent of responses obtained as part of ASQ screening. We combined SI and SA, due to small sample sizes. While we used International Classification of Diseases (ICD) codes (e.g., E950-E958) as

much as possible, these were seldom found (90 of 25,067 ED encounters between 1/1/2017 and 6/30/2019). Thus, for those missing ICD codes, we defined instances of PED SI or SA on recorded principal diagnosis description and chief complaint as documented in the EHR. 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). We used the following 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 (see supplemental files for frequencies). To increase our sensitivity of accurately finding cases, we included terms for non-suicidal self-injury (NSSI; e.g., "self-injury" and "self-harm"), but examined these in conjunction with more specific words related to suicidal intent such as "suicidal" and "suicide." As a sensitivity analysis, we compared our pattern-matched outcome measure to manually coded suicidal behaviors (coded using ICD-9 codes due to years covered) for overlapping patients in data from DeVylder et al. (2019).

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 (DeVylder et al., 2019), plausibility for incidence within the pediatric population (i.e., excluding conditions such as stroke, hip fracture, and 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" package (Gasparini, 2018) 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" (Vyas et al., 2020). 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, is referred to as "EHR" data (see Supplemental Files 2 for full list of variables). We then compared the value of EHR-based predictive modeling 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 3 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 con-taining 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 3 months of the indexed date. Finally, "control" patients who had <3 months of observation (i.e., first visit after March 30, 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 (MICE, 2010). The imputation model used all other variables in the dataset coded at the time of visit (e.g., index visit information used to impute ASQ item 4 at index visit time; follow-up values used to imputed ASQ item 4 at follow-up time point). 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; and 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 question 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., DeVylder et al., 2019). Only three outcome events (2%) were coded with only an NSSI related term (e.g., "self-injury" and "self-harm"). The pattern-matched definition was also more inclusive, with 309 additional instances of suicidal ideation or attempt over those detected manually and only six instances where manual review identified suicide ideation or attempt where our pattern-matched definition did not.

Modeling—The 3-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 risk[20] 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, fivefold 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 (Kuhn, 2011). 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 3

months (p = .01). We did not adjust for extreme imbalance in the outcome. Performance metrics were based on evaluation in the entire study sample and 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 an ED visit for ideation or attempt within the past 3 months (12 months included in Appendix A1) as recorded in the EHR. 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 (12 cases identified through ICD codes; 129 identified through pattern matching, see Table S1 for frequency of search terms found among cases).

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 was added to the ASQ screening process, the sensitivity improved

by 17.4%, AUC by 13.4%, while there was 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).

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) did not change because only one feature is included in the model. However, using the LASSO method, despite the substantially reduced number of features, we saw improvement in the sensitivity of the model when EHR data were added to the ASQ. In fact, just adding one additional feature, a diagnosis of depression at index PED visit, resulted in an increased sensitivity: Model 2, the ASQ plus EHR data had 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).

Secondary analysis

For the binary logistic regression, the inclusion of an index PED visit for suicide attempt or ideation as recorded 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 suicide screening tool, the ASQ, to EHR-based predictive models for their accuracy at identifying individuals at risk of future suicide-related ED visits, as a first step to exploring how these approaches may or may not work together. 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 3 months. Our logistic regression findings indicate that relying solely on EHR-based predictive modeling performs only slightly worse than implementing universal brief suicide screening. In addition 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-based predictive 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 suicide screening measures (Horowitz et al., 2012; Posner et al., 2007) and predictive models (Haroz et al., 2019; Kessler et al., 2015; Walsh et al., 2017, 2018). 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. This is consistent with Simon et al.'s findings showing that using a single item from a brief depression screener can improve predictive accuracy of EHR-based models (Simon et al., 2018). 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 a relative 17.4% and AUC by a relative 13.4% compared to the ASQ only model (Model 1).

Our secondary findings (Appendix A2) 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 (Hedegaard et al., 2018), 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.

Unlike many existing models which use historical EHR data (e.gSimon et al., 2018; Walsh et al., 2017; Walsh et al., 2018), our approach focused only on data collected at the time of the visit. This approach was chosen to help facilitate ease of implementation. While EHR platforms are increasing their computational capacity, not all EHRs have the ability to calculate in real-time based on historical data. It may also address some of the limitations with patients who move, visit other health centers, and/or do not have complete data. That being said, the addition of historical variables to this approach may improve accuracy even more, but more research is needed.

Despite widescale recommendations to routinely screen for suicide risk among children and adolescents (The Joint Commission, 2015; Shain, 2016), 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 (Roaten et al., 2018). 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 screeners (Nestadt et al., 2018) or operational modifications needed to implement these approaches. Implementation of screening requires training and ongoing oversight to ensure compliance (Hackfeld, 2020; Inman et al., 2019). 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 (Betz et al., 2016). Notwithstanding, screening for suicide risk has shown promise at identifying patients at risk for suicide (Ballard et al., 2017; DeVylder et al., 2019; King et al., 2021).

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 (Graney et al., 2020).

Both screening and predictive modeling approaches have limitations. The positive predictive value across all our models was quite low (e.g., 0.04–0.05). This is consistent with a recent systematic review by Belsher et al. (2019) 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 (Perlis & Fihn, 2020). Predictive accuracy can vary by the rarity of the behavior selected for the outcome, with lower PPVs observed for deaths, but higher PPVs possible with relatively more common outcomes such as suicide-related ED visits or suicide attempts. Our false-positive rates were considerably lower than those found in previous studies (Nelson et al., 2017), but still represent potential inappropriate use of resources and/or a potential source of distrust or source of fatigue in procedures (Singh et al., 2013). 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 (Kline-Simon et al., 2020). 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.

No risk identification strategy is useful without linked intervention. Potential interventions that could be linked to these strategies include safety planning (Stanley et al., 2018), brief contact interventions (Milner et al., 2015), and more intensive therapeutic interventions. While these types of interventions hold promise, robust evidence on any single approach or how these work in combination among youth is still emerging (Bahji et al., 2021; Mann et al., 2021). Due to our outcome classification, our results also indicate a need for interventions that address a range of suicidal thoughts and behaviors that place individuals at risk for future suicide-related ED visit. Several prevention strategies linked to EDs may be useful for this range of behaviors (Doupnik et al., 2020), although more research is needed to inform an empirical evidence base (Mann et al., 2021).

Despite these drawbacks, early and accurate identification of individuals at risk is a critical component of any 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 (Gordon et al., 2020, Vickers et al., 2016). 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 (Kessler et al., 2020). Evaluating the cost-effectiveness of screening approaches and predictive modeling is critical. Recent work by Ross et al. (2021) showed that a PPV of 1% or greater was sufficient for showing cost-effectiveness of a predictive model targeting a telephone and brief contact intervention for those flagged as high risk. A higher threshold is needed (2%) for a more expensive intervention such as cognitive behavioral therapy (Ross et al., 2021). This work is promising as it demonstrates that even with low PPV, an issue with both predictive modeling and brief suicide screening, these approaches can facilitate cost-effective preventative care. However, more work is needed to evaluate and compare costs, cost savings, and cost-effectiveness of both screening and predictive modeling approaches across a variety of settings and contexts.

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 (Congressional Black Caucus, 2019). 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 (Rajkomar et al., 2018). 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 (Rajkomar et al., 2018).

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. While we included NSSI terms (e.g., "self-injury" and "self-harm") in an effort to increase our sensitivity, 98% of our outcomes that included NSSI term on primary diagnosis description also included a term directly related to suicide such as "suicidal" or "suicide." Still some misclassification is possible. 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 research (Hedegaard et al., 2018), and a relative strength of this study is definition of outcomes through review of clinician notes. While we compared predictive models to suicide screening on risk for subsequent suicide-related ED visit three-month post initial visit, this outcome is not always the intention of screening programs. Many screeners are developed to identify immediate risk that requires swift action such as hospitalization. We were not able to evaluate the accuracy of predictive models on this type of immediate risk due to the small sample size. Future work could examine this by focusing on risk

within a shorter time period postdischarge. We did not adjust for imbalance in our outcomes, due to small sample sizes as well as, representing an approach that was considered outside of the scope of our research question. The generalizability of our findings may also be limited. Data represent the patient population seen at the JHH PED. Generalizing these findings to rural populations, higher-income settings, and/or outpatient populations is not appropriate. Furthermore, our data are only representative of PED data captured at JHH, which excludes potential visits at other EDs. Finally, although we used rigorous methods for cross-validation, our results most likely suffer from overfitting and 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 3 and 12 months. We found that predictive modeling 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 modeling with EHR, substantially improves accuracy. Further research is needed to explore how models can be used in lieu of or in conjunction with brief suicide risk screening approaches. Regardless of the method, careful attention should be paid toward optimizing the implementation of either approach, including focusing on net benefits and distributive justice principles.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding information

APPENDIX A1

Twelve-month prediction models

TABLE A1

Binary Logistic Regression model comparison for accuracy of detecting a suicide-related ED event 12 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	Accuracy	Sensitivity	Specificity	Positive predictive value	False-positive rate	False-negative rate	AUC
Model 4: HER + ASQ-4	0.817	0.686	0.820	0.097	0.180	0.314	0.822

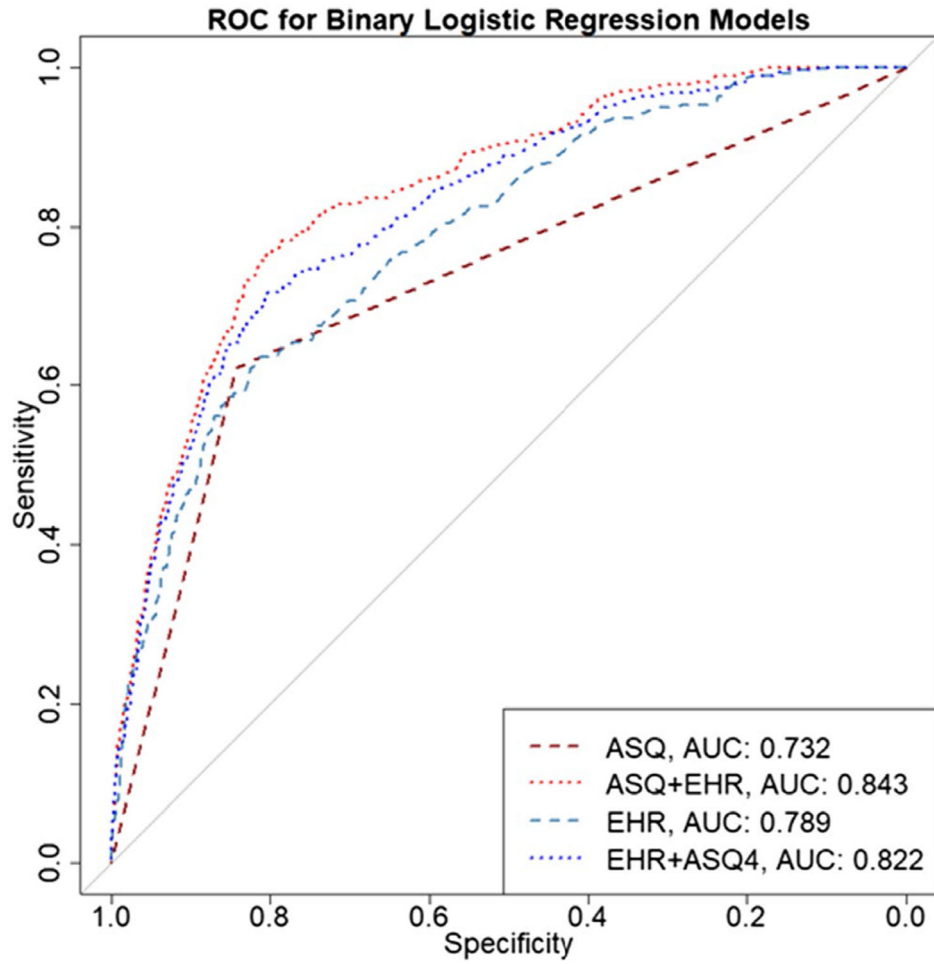**ROC for Binary Logistic Regression Models**

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

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 3 months after index event ($N = 13,420$)

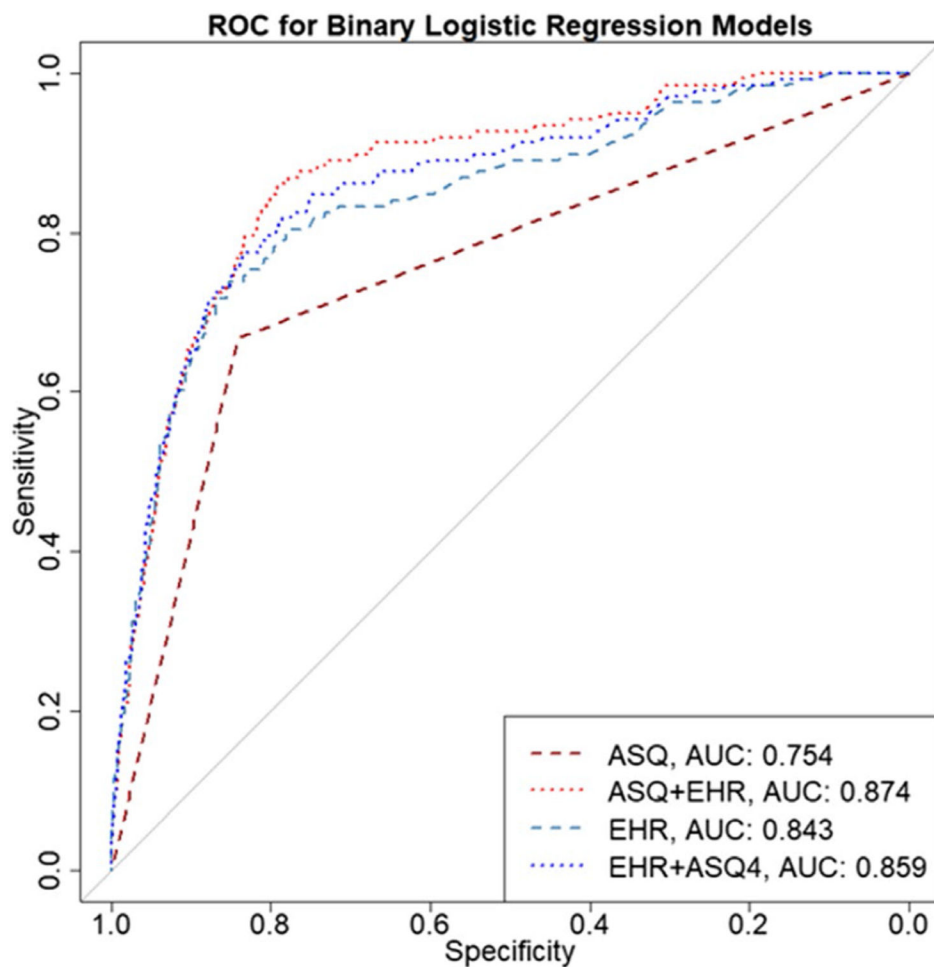| Model	Accuracy	Sensitivity	Specificity	Positive predictive value	False-positive rate	False-negative rate	AUC
Model 1: ASQ Alone	0.840	0.667	0.842	0.042	0.158	0.333	0.754
Model 2: ASQ +EHR	0.841	0.768	0.842	0.048	0.159	0.232	0.874
Model 3: EHR alone	0.858	0.717	0.860	0.051	0.140	0.283	0.843
Model 4: HER + 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).

REFERENCES

Ahmedani BK, Simon GE, Stewart C, Beck A, Waitzfelder BE, Rossom R, Lynch F, Owen-Smith A, Hunkeler EM, Whiteside U, Operskalski BH, Coffey MJ, & Solberg LI (2014). Health care contacts in the year before suicide death. Journal of General Internal Medicine, 29(6), 870–877. 10.1007/s11606-014-2767-3

Bahji A, Pierce M, Wong J, Roberge JN, Ortega I, & Patten S (2021). Comparative efficacy and acceptability of psychotherapies for self-harm and suicidal behavior among children and adolescents: A systematic review and network meta-analysis. JAMA Network Open, 4(4), e216614. 10.1001/jamanetworkopen.2021.6614 [PubMed: 33861328]

Bahraini N, Brenner LA, Barry C, Hostetter T, Keusch J, Post EP, Kessler C, Smith C, & Matarazzo BB (2020). Assessment of Rates of Suicide Risk Screening and Prevalence of Positive Screening Results Among US Veterans After Implementation of the Veterans Affairs Suicide Risk Identification Strategy. JAMA Netw Open, 3(10), e2022531. 10.1001/jamanetworkopen.2020.22531 [PubMed: 33084900]

Ballard ED, Cwik M, Van Eck K, Goldstein M, Alfes C, Wilson ME, Virden JM, Horowitz LM, & Wilcox HC (2017). Identification of at-risk youth by suicide screening in a pediatric emergency department. Prevention Science, 18(2), 174–182. 10.1007/s11121-016-0717-5 [PubMed: 27678381]

Bates DW, Auerbach A, Schulam P, Wright A, & Saria S (2020). Reporting and implementing interventions involving machine learning and artificial intelligence. Annals of Internal Medicine, 172(11), S137–S144. 10.7326/M19-0872 [PubMed: 32479180]

Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, Morgan RL, Evatt DP, Tucker J, & Skopp NA (2019). Prediction models for suicide attempts and deaths. JAMA Psychiatry, 76(6), 642–651. 10.1001/jamapsychiatry.2019.0174 [PubMed: 30865249]

Betz ME, Wintersteen M, Boudreaux ED, Brown G, Capoccia L, Currier G, Goldstein J, King C, Manton A, Stanley B, Moutier C, & Harkavy-Friedman J (2016). Reducing suicide risk: Challenges and opportunities in the emergency department. Annals of Emergency Medicine, 68(6), 758–765. 10.1016/j.annemergmed.2016.05.030 [PubMed: 27451339]

Boudreaux ED, Haskins BL, Larkin C, Pelletier L, Johnson SA, Stanley B, Brown G, Mattocks K, & Ma Y (2020). Emergency department safety assessment and follow-up evaluation 2: An implementation trial to improve suicide prevention. Contemporary Clinical Trials, 95, 106075. 10.1016/j.cct.2020.106075 [PubMed: 32565041]

Centers for Disease Control (CDC). (2017). Suicide rates*, for teens aged 15–19 years, by sex —United States, 1975–2015. Morbidity and Mortality Weekly Report, 66(30), 816. 10.15585/mmwr.mm6630a6 [PubMed: 28771461]

Char DS, Shah NH, & Magnus D (2018). Implementing machine learning in health care—Addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983. 10.1056/NEJMp1714229

Chock MM, Bommersbach TJ, Geske JL, & Bostwick JM (2015). Patterns of health care usage in the year before suicide: A population-based case-control study. Mayo Clinic Proceedings, 90(11), 1475–1481. 10.1016/j.mayocp.2015.07.023 [PubMed: 26455886]

Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, & Large MM (2017). Suicide rates after discharge from psychiatric facilities: A systematic review and meta-analysis. JAMA Psychiatry, 74(7), 694–702. 10.1001/jamapsychiatry.2017.1044 [PubMed: 28564699]

Claassen CA, Pearson JL, Khodyakov D, Satow PM, Gebbia R, Berman AL, Reidenberg DJ, Feldman S, Molock S, Carras MC, Lento RM, Sherrill J, Pringle B, Dalal S, & Insel TR (2014). 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. American Journal of Preventive Medicine, 47(3), 309–314. 10.1016/j.amepre.2014.01.004 [PubMed: 24750971]

Curtin S, Warner M, & Hedegaard H (2016). Increase in suicide in the United States, 1999-2014. National Center for Health Statistics Data Brief, 241.

DeVylder JE, Ryan TC, Cwik M, Wilson ME, Jay S, Nestadt PS, Goldstein M, & Wilcox HC (2019). Assessment of selective and universal screening for suicide risk in a pediatric emergency department. JAMA Network Open, 2(10), e1914070. 10.1001/jamanetworkopen.2019.14070 [PubMed: 31651971]

Doupnik SK, Rudd B, Schmutte T, Worsley D, Bowden CF, McCarthy E, Eggan E, Bridge JA, & Marcus SC (2020). Association of suicide prevention interventions with subsequent suicide attempts, linkage to follow-up care, and depression symptoms for acute care settings: A systematic review and meta-analysis. JAMA Psychiatry, 77(10), 1021–1030. 10.1001/jamapsychiatry.2020.1586 [PubMed: 32584936]

Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, Musacchio KM, Jaroszewski AC, Chang BP, & Nock MK (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychological Bulletin, 143(2), 187–232. 10.1037/bul0000084 [PubMed: 27841450]

Gasparini A (2018). comorbidity: An R package for computing comorbidity scores Software • Review • Repository • Archive. Journal of Open Source Software, 3(23), 648-10.21105/joss.00648

Gordon JA, Avenevoli S, & Pearson JL (2020). Suicide prevention research priorities in health care. JAMA Psychiatry, 77(9), 885–886. 10.1001/jamapsychiatry.2020.1042 [PubMed: 32432690]

Graney J, Hunt IM, Quinlivan L, Rodway C, Turnbull P, Gianatsi M, Appleby L, & Kapur N (2020). Suicide risk assessment in UK mental health services: A national mixed-methods study. The Lancet Psychiatry, 7(12), 1046–1053. 10.1016/S2215-0366(20)30381-3 [PubMed: 33189221]

Hackfeld M (2020). Implementation of a pediatric/adolescent suicide risk screening tool for patients presenting to the Emergency Department with nonbehavioral health complaints. Journal of Child and Adolescent Psychiatric Nursing, 33(3), 131–140. 10.1111/jcap.12276 [PubMed: 32407581]

Haroz EE, Walsh CG, Goklish N, Cwik MF, O'Keefe V, & Barlow A (2019). Reaching those at highest risk for suicide: Development of a model using machine learning methods for use with Native American communities. Suicide and Life-Threatening Behavior, 50(2), 422–436. 10.1111/sltb.12598 [PubMed: 31692064]

Hedegaard H, Schoenbaum M, Claassen C, Crosby A, Holland K, & Proescholdbell S (2018). 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, 108, 1–19.

Horowitz LM, Bridge JA, Teach SJ, Ballard E, Klima J, Rosenstein DL, Wharff EA, Ginnis K, Cannon E, Joshi P, & Pao M (2012). Ask suicide-screening questions (ASQ): A brief instrument for the pediatric emergency department. Archives of Pediatrics and Adolescent Medicine, 166(12), 1170–1176. 10.1001/archpediatrics.2012.1276 [PubMed: 23027429]

Inman DD, Matthews J, Butcher L, Swartz C, & Meadows AL (2019). Identifying the risk of suicide among adolescents admitted to a children's hospital using the Ask Suicide-Screening Questions. Journal of Child and Adolescent Psychiatric Nursing, 32(2), 68–72. 10.1111/jcap.12235 [PubMed: 31025489]

Joint Commission on Accreditation of Healthcare. (2015). Hospital national patient safety goal 15.01. 01. 2015.

Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Queen B, Lowry R, Chyen D, Whittle L, Thornton J, Lim C, Bradford D, Yamakawa Y, Leon M, Brener N, & Ethier KA (2018). Youth Risk Behavior Surveillance — United States, 2017. MMWR. Surveillance Summaries, 67(8), 1–114. 10.15585/mmwr.ss6708a1

Kessler RC, Bossarte RM, Luedtke A, Zaslavsky AM, & Zubizarreta JR (2020). Suicide prediction models: A critical review of recent research with recommendations for the way forward. Molecular Psychiatry, 25(1), 168–179. 10.1038/s41380-019-0531-0 [PubMed: 31570777]

Kessler RC, Hwang I, Hoffmire CA, McCarthy JF, Petukhova MV, Rosellini AJ, Sampson NA, Schneider AL, Bradley PA, Katz IR, Thompson C, & Bossarte RM (2017). Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration. International Journal of Methods in Psychiatric Research, 26(3), e1575. 10.1002/mpr.1575

Kessler RC, Warner CH, Ivany C, Petukhova MV, Rose S, Bromet EJ, Brown M, Cai T, Colpe LJ, Cox KL, Fullerton CS, Gilman SE, Gruber MJ, Heeringa SG, Lewandowski-Romps L, Li J, Millikan-Bell AM, Naifeh JA, Nock MK, … Ursano RJ (2015). Predicting suicides after psychiatric hospitalization in US army soldiers: The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). JAMA Psychiatry, 72(1), 49–57. 10.1001/jamapsychiatry.2014.1754 [PubMed: 25390793]

King CA, Brent D, Grupp-Phelan J, Casper TC, Dean JM, Chernick LS, Fein JA, Mahabee-Gittens EM, Patel SJ, Mistry RD, Duffy S, Melzer-Lange M, Rogers A, Cohen DM, Keller A, Shenoi R, Hickey RW, Rea M, Cwik M, … Pediatric Emergency Care Applied Research Network (2021). Prospective development and validation of the computerized adaptive screen for suicidal youth. JAMA Psychiatry, 78(5), 540–549. 10.1001/jamapsychiatry.2020.4576 [PubMed: 33533908]

Kline-Simon AH, Sterling S, Young-Wolff K, Simon G, Lu Y, Does M, & Liu V (2020). Estimates of workload associated with suicide risk alerts after implementation of risk-prediction model. JAMA Network Open, 3(10), e2021189. 10.1001/jamanetworkopen.2020.21189 [PubMed: 33084896]

Kuhn M (2011). The caret Package. Journal of Statistical Software, 28(i05), 1–26.

Labouliere CD, Vasan P, Kramer A, Brown G, Green K, Kammer J, Finnerty M, & Stanley B (2018). «Zero Suicide» – A model for reducing suicide in United States behavioral healthcare. Suicidologi, 23(1), 22–30. 10.5617/suicidologi.6198 [PubMed: 29970972]

Mann JJ, Michel CA, & Auerbach RP (2021). Improving suicide prevention through evidence-based strategies: A systematic review. American Journal of Psychiatry. 178(7), 611–624. 10.1176/appi.ajp.2020.20060864

Milner AJ, Carter G, Pirkis J, Robinson J, & Spittal MJ (2015). Letters, green cards, telephone calls and postcards: Systematic and meta-analytic review of brief contact interventions for reducing self-harm, suicide attempts and suicide. The British Journal of Psychiatry, 206(3), 184–190. 10.1192/bjp.bp.114.147819 [PubMed: 25733570]

National Center for Injury Prevention and Control. WISQARS (Web-based Injury Statistics Query and Reporting System). Retrieved from https://www.cdc.gov/injury/wisqars/

Nelson HD, Denneson LM, Low AR, Bauer BW, O'Neil M, Kansagara D, & Teo AR (2017). Suicide risk assessment and prevention: A systematic review focusing on veterans. Psychiatric Services, 68, 1003–1015. 10.1176/appi.ps.201600384 [PubMed: 28617209]

Nestadt PS, Triplett P, Mojtabai R, & Berman AL (2018). Universal screening may not prevent suicide. General Hospital Psychiatry, 63, 14–15. 10.1016/j.genhosppsych.2018.06.006 [PubMed: 30072036]

Olfson M, Wall M, Wang S, Crystal S, Liu SM, Gerhard T, & Blanco C (2016). Short-term suicide risk after psychiatric hospital discharge. JAMA Psychiatry, 73(11), 1119–1126. 10.1001/jamapsychiatry.2016.2035 [PubMed: 27654151]

Perlis RH, & Fihn SD (2020). Hard truths about suicide prevention. JAMA Network Open, 3(10), e2022713. 10.1001/jamanetworkopen.2020.22713 [PubMed: 33084895]

Posner K, Oquendo MA, Gould M, Stanley B, & Davies M (2007). Columbia Classification Algorithm of Suicide Assessment (C-CASA): Classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. American Journal of Psychiatry, 164(7), 1035–1043. 10.1176/ajp.2007.164.7.1035

Rajkomar A, Hardt M, Howell MD, Corrado G, & Chin MH (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866–872. 10.7326/M18-1990 [PubMed: 30508424]

Roaten K, Johnson C, Genzel R, Khan F, & North CS (2018). Development and implementation of a universal suicide risk screening program in a safety-net hospital system. Joint Commission Journal on Quality and Patient Safety, 44(1), 4–11. 10.1016/j.jcjq.2017.07.006 [PubMed: 29290245]

Ross EL, Zuromski KL, Reis BY, Nock MK, Kessler RC, & Smoller JW (2021). Accuracy requirements for cost-effective suicide risk prediction among primary care patients in the US. JAMA Psychiatry. 78(6):642–650. 10.1001/jamapsychiatry.2021.0089 [PubMed: 33729432]

Ruch DA, Sheftall AH, Schlagbaum P, Rausch J, Campo JV, & Bridge JA (2019). Trends in suicide among youth aged 10 to 19 years in the United States, 1975 to 2016. JAMA Network Open, 2(5), e193886. 10.1001/jamanetworkopen.2019.3886 [PubMed: 31099867]

Shain B (2016). Suicide and suicide attempts in adolescents. Pediatrics, 138(1), 669–676. 10.1542/peds.2016-1420

Simon GE, Johnson E, Lawrence JM, Rossom RC, Ahmedani B, Lynch FL, Beck A, Waitzfelder B, Ziebell R, Penfold RB, & Shortreed SM (2018). Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. American Journal of Psychiatry, 175(10), 951–960. 10.1176/appi.ajp.2018.17101167

Singh H, Spitzmueller C, Petersen NJ, Sawhney MK, & Sittig DF (2013). Information overload and missed test results in electronic health record-based settings. JAMA Internal Medicine, 173(8), 702–704. 10.1001/2013.jamainternmed.61 [PubMed: 23460235]

Snyder DJ, Jordan BA, Aizvera J, Innis M, Mayberry H, Raju M, Lawrence D, Dufek A, Pao M, & Horowitz LM (2020). From pilot to practice: Implementation of a suicide risk screening program in hospitalized medical patients. Joint Commission Journal on Quality and Patient Safety, 46(7), 417–426. 10.1016/j.jcjq.2020.04.011 [PubMed: 32473966]

Stanley B, Brown GK, Brenner LA, Galfalvy HC, Currier GW, Knox KL, Chaudhury SR, Bush AL, & Green KL (2018). Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. JAMA Psychiatry, 75(9), 894–900. 10.1001/jamapsychiatry.2018.1776 [PubMed: 29998307]

Taylor R, Susukida R, Cwik M, Goldstein M, Wilson M, Nestadt P, & Wilcox H (2019). Frequent Ed Visits Among Youth Identified at Risk for Suicide (IASR/AFSP Conference).

The Congressional Black Caucus (2019). Ring the alarm: The crisis of black youth suicide in America. The Congressional Black Caucus.

van Buuren S, Groothuis-Oudshoorn K (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1–67. https://www.jstatsoft.org/v45/i03/

Vickers AJ, Van Calster B, & Steyerberg EW (2016). Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ, 352, i6. 10.1136/bmj.i6 [PubMed: 26810254]

Vyas DA, Eisenstein LG, & Jones DS (2020). Hidden in plain sight — Reconsidering the use of race correction in clinical algorithms. New England Journal of Medicine, 383(9), 874–882. 10.1056/nejmms2004740

Walkup JT, Townsend L, Crystal S, & Olfson M (2012). A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. Pharmacoepidemiology and Drug Safety, 21(1), 174–182. 10.1002/pds.2335 [PubMed: 22262604]

Walsh CG, Ribeiro JD, & Franklin JC (2017). Predicting risk of suicide attempts over time through machine learning. Clinical Psychological Science, 5(3), 457–469. 10.1177/2167702617691560

Walsh CG, Ribeiro JD, & Franklin JC (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. Journal of Child Psychology and Psychiatry and Allied Disciplines, 59(12), 1261–1270. 10.1111/jcpp.12916

Wang M, Swaraj S, Chung D, Stanton C, Kapur N, & Large M (2019). Meta-analysis of suicide rates among people discharged from non-psychiatric settings after presentation with suicidal thoughts or behaviours. Acta Psychiatrica Scandinavica, 139(5), 472–483. 10.1111/acps.13023 [PubMed: 30864183]
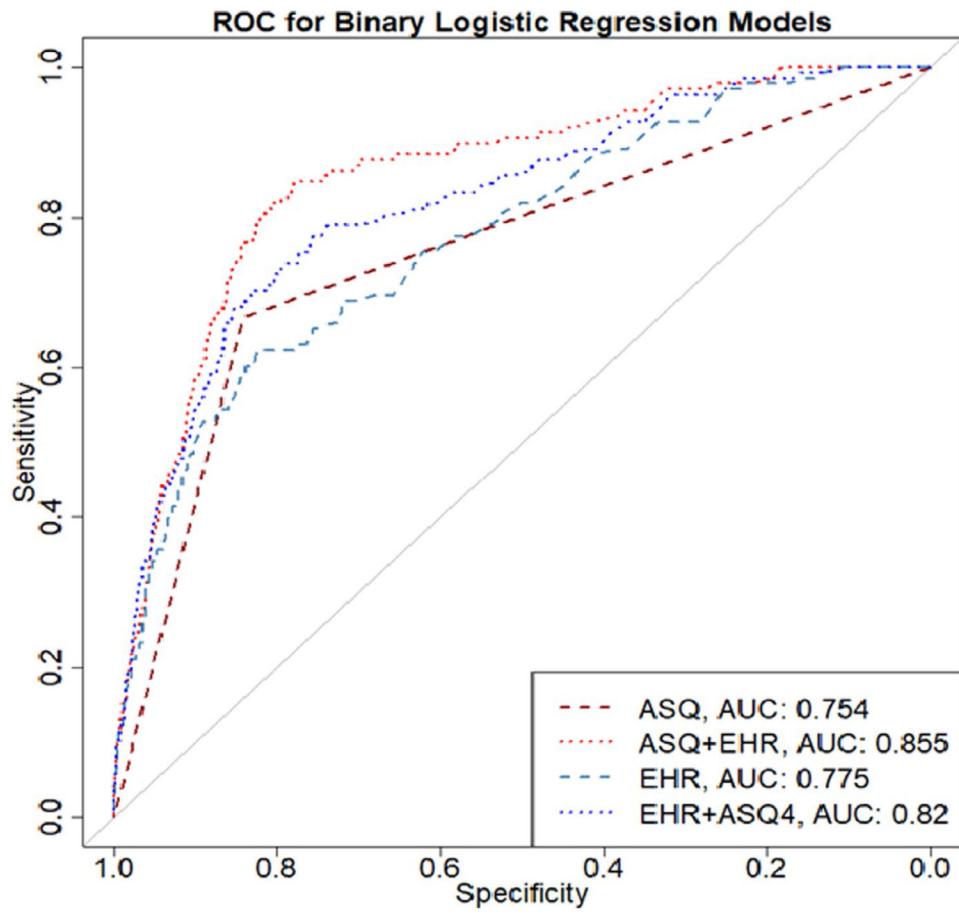
FIGURE 1.
Receiver operating curve (ROC) comparing binary logistic regression models (N = 13,420)

TABLE 1

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

	Control 13,279 (99%)	Cases 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[a]		
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%)

[a] Missing 5 ASQ screens.

TABLE 2

Binary Logistic Regression model comparison for accuracy of detecting a suicide-related ED event 3 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 3 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