



HHS Public Access

Author manuscript

Suicide Life Threat Behav. Author manuscript; available in PMC 2020 April 11.

Published in final edited form as:

Suicide Life Threat Behav. 2020 April ; 50(2): 422–436. doi:10.1111/sltb.12598.

Reaching Those at Highest Risk for Suicide: Development of a Model Using Machine Learning Methods for use With Native American Communities

EMILY E. HAROZ, PHD,

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

COLIN G. WALSH, MD,

Vanderbilt University Medical Center, Nashville, TN, USA

NOVALENE GOKLISH, MS,

Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Center for American Indian Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA and White Mountain Apache Tribe, Whiteriver, AZ, USA.

MARY F. CWIK, PHD,

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

VICTORIA O'KEEFE, PHD,

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

ALLISON BARLOW, PHD

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

Abstract

Objective: Suicide prevention is a major priority in Native American communities. We used machine learning with community-based suicide surveillance data to better identify those most at risk.

Address correspondence to Emily E. Haroz, Center for American Indian Health, 415 N. Washington Street, Baltimore, MD 21231; eharoz1@jhu.edu and Novalene Goklish, Johns Hopkins Center for American Indian Health, P.O. Box 1240, Whiteriver, AZ 85941; ngoklis1@jhu.edu.

Bloomberg American Health Initiative. *Data:* Available by request from the White Mountain Apache Tribe.

SUPPORTINGINFORMATION

Additional Supporting Information may be found in the online version of this article:

Method: This study leverages data from the Celebrating Life program operated by the White Mountain Apache Tribe in Arizona and in partnership with Johns Hopkins University. We examined $N = 2,390$ individuals with a validated suicide-related event between 2006 and 2017. Predictors included 73 variables (e.g., demographics, educational history, past mental health, and substance use). The outcome was suicide attempt 6, 12, and 24 months after an initial event. We tested four algorithmic approaches using cross-validation.

Results: Area under the curves ranged from $AUC = 0.81$ ($95\% \text{ CI} \pm 0.08$) for the decision tree classifiers to $AUC = 0.87$ ($95\% \text{ CI} \pm 0.04$) for the ridge regression, results that were considerably higher than a past suicide attempt ($AUC = 0.57$; $95\% \text{ CI} \pm 0.08$). Selecting a cutoff value based on risk concentration plots yielded 0.88 sensitivity, 0.72 specificity, and a positive predictive value of 0.12 for detecting an attempt 24 months postindex event.

Conclusion: These models substantially improved our ability to determine who was most at risk in this community. Further work is needed including developing clinical guidance and external validation.

Suicide prevention is a high priority for Native American (NA) populations. While rates vary across tribes and communities, NAs overall face a disproportionate burden of suicide and related behavior (CDC, 2005). Suicide rates are increasing for the United States in general, including among NAs, whose rates rose from 10.92 to 13.37 per 100,000 people from 2014 to 2016. Most suicide deaths occur among those 15–24 years old, with rates in this age group estimated to be four times the rate for US All Races (IHS, 2014).

In a national survey among adult NAs, approximately 1.2% reported having attempted suicide in the past year compared with 0.5% of adults in the US All Races population (SAMHSA, 2011). Among NA adolescents, lifetime rates of self-report attempted suicide ranged from 21.8% of girls to 11.8% of boys, with rates on reservations being higher than rates in urban areas (Borowsky, Resnick, Ireland, & Blum, 1999; Freedenthal & Stiffman, 2004). More generally, a past suicide attempt is one of the strongest risk factors for dying by suicide and often goes undetected. Many individuals do not seek medical attention for suicide attempt (Claassen & Larkin, 2005; Kembal, Gasgarth, Johnson, Patil, & Houry, 2008; King, O'mara, Hayward, & Cunningham, 2009), and those who do obtain little or no follow-up mental health treatment (Bridge, Marcus, & Olfson, 2012; Knesper, 2010). This problem is compounded in many reservation settings due to significant barriers to mental health care, including lack of available professional services, long waiting periods, poor access to transportation, long distances to travel, stigma, and poor cultural fit of existing clinical services (Barlow & Walkup, 1998; Freedenthal & Stiffman, 2007; Novins, 2009; O'Keefe, Cwik, Haroz, & Barlow, 2019; Probst et al., 2006; Pullmann, VanHooser, Hoffman, & Heflinger, 2010).

One particularly promising program aimed at addressing suicide is the White Mountain Apache Tribe's (WMAT) Celebrating Life suicide surveillance and prevention program (CL program) (Cwik et al., 2014). In 2001, the WMAT mandated reporting of suicide deaths, attempts, and ideation to a surveillance system after a cluster of 11 youth suicides in six months. In 2007, the Tribal Council passed an amendment to expand the system to include nonsuicidal self-injury. Binge substance use was later included in 2010 (see Cwik et al, 2014

for additional history and methods). Simultaneously, WMAT and the Johns Hopkins Center for American Indian Health partnered through a series of grants to launch a team of specially trained Apache paraprofessional Community Mental Health Specialists (CMHS') to conduct in-person follow-up on every reported surveillance event. CMHS' validate events, assess imminent risk, and connect individuals to available care. In addition, CMHS' provide in-service trainings to all departments and schools and participate in suicide prevention activities in the community (e.g., suicide prevention walks, handing out National Suicide Prevention Lifeline cards). The surveillance system has provided the tribe with a comprehensive data source to monitor suicide trends and patterns in real time, better understand local risk and protective factors (Cwik et al., 2015), and design innovative, culturally derived, and tailored interventions to further reduce rates in this community (Tingey et al., 2016). Comparing rates from 2001–2006 (prior to CL program) to 2007–2012 (after implementation of CL program), the suicide rate decreased by 38.4% among Apache adults and 23% among Apache youth (ages 15–24) (Cwik et al., 2016).

As awareness of the program has grown, so has the number of intakes, resulting in a greater case load for the CMHS'. Given the large geographic area and an increasing number of intakes, prioritization of cases is necessary. Currently, prioritization is based on severity of reported behavior (i.e., those with a reported attempt are visited first by CMHS'). This model attends to those with the most serious reported behavior, but it does not consider future risk. Systematically identifying people at highest risk for suicide after initial contact with the CL program may help in allocating limited resources and attention to those who need it most.

However, there are challenges with risk identification. Most risk identification strategies are assessment-based—often requiring intensive training, expertise, and/or time (Linthicum, Schafer & Ribeiro, 2019). Moreover, these assessments are based on rather simple combinations of risk factors known from suicide research (Linthicum, Schafer & Ribeiro, 2019). A recent meta-analysis has shown that despite over 50 years of suicide research, we have seen little improvement in our predictive accuracy for suicide-related behaviors (Franklin et al., 2017). Diagnostic accuracy for suicide attempt and death by suicide prediction is only slightly better than chance (wAUCs = 0.58 and 0.57, respectively (Franklin et al., 2017). Sensitivity analyses show that risk factors only correctly identified suicide attempt 26% of the time and death only 9% of the time (Franklin et al., 2017). Assessments based on these risk factors that attempt to classify people into “high-risk” categories have also been found to have limited clinical value (Carter et al., 2017).

Our own previous descriptive research with ($N=71$) WMAT youth <24 years of age with a previous suicide attempt identified several correlates of risk, including caregivers with substance use problems, youth's substance use history, and recent suicide death or attempt among peers and family members (Cwik et al., 2015). While these single risk factors may be statistically significantly associated with past suicidal behavior, the utility of these factors to determine risk over time and inform response is relatively unknown, given by the complexity of potential interactions between them and the reality that unmeasured factors also contribute to future risk (Franklin et al., 2017).

The application of ML is increasingly being used for many health problems, including suicide (Kessler, Hwang, et al., 2017; Kessler et al., 2016). Machine learning approaches examine the probabilistic relationships between a large set of dependent variables and then test the reliability of the results through cross-validation procedures. Given the complexity of suicide risk and machine learning's ability to model this complexity, ML methods seem to have potential to improve accuracy and scale (Linthicum, Schafer & Ribeiro, 2019). One recent study suggests that ML methods may be better at identifying suicide risk than clinician evaluation (Tran et al., 2014). However, these methods have never been applied in American Indian settings nor with community-based data, limiting our knowledge about how well they perform and how best they can be utilized in different contexts and cultures. Understanding if and how to utilize these methods is even more urgent for health disparity populations given the lack of and barriers to conventional mental health treatment and in the case of WMAT, existing promising community-based suicide prevention efforts. Researching these methods with disparity populations is critical to advancing suicide-related health equity (Rajkomar, Hardt, Howell, Corrado, & Chin, 2018).

THE PRESENT STUDY

The goal of the current research was to develop a risk model that accurately identified NA adults and youth who would attempt suicide within 6, 12, and 24 months of an initial incident report (including ideation, attempt, nonsuicidal self-injury, binge substance use requiring medical care, but not suicide death) to the CL program. We hypothesized that the resulting algorithm would distinguish between those who attempted suicide within 24 months of their initial contact and those who did not attempt with at least acceptable accuracy (e.g., area under the curve = 0.70). We considered applying a previously developed algorithm to this data, but were unable to given the uniqueness of the way in which the data were collected (i.e., community-based vs. clinical), the variables included (e.g., substance use at time of event, resilience factors related to cultural identity), and the underrepresentation of AI/AN populations, who have unique suicide patterns, in previous datasets used to generate existing algorithms (Rajkomar et al., 2018). Clinical implications of our work include potential to embed the resulting algorithm into the surveillance system to help CMHS' identify individuals at higher risk, with a goal of further reducing suicidal behavior in the WMAT community.

METHODS

White Mountain Apache Tribe

The Fort Apache Indian Reservation, located in northeastern Arizona, is home to approximately 17,500 tribal members and is governed by an elected 11-member Tribal Council. While the tribe experiences considerable behavioral and mental health disparities, they have long embraced research-informed public health practices to overcome these challenges. The Celebrating Life program is a prime example of an innovative public health approach to suicide prevention by a sovereign tribal nation. With this system, WMAT health leaders and Tribal Council can monitor real-time trends, identify risk and protective factors,

develop culturally appropriate interventions based on findings, advance a local Native workforce (O’Keefe et al., 2019), and provide care for those at risk of suicide.

Celebrating Life Suicide Surveillance and Prevention Program (CL Program)

The CL program has three main components: (1) tribally mandated active surveillance within a closed reservation system; (2) in-person follow-up and case management (CM) provided by Apache Community Mental Health Specialists (CMHS’); and (3) community outreach and education about the program. All persons, departments, and schools within the WMAT jurisdiction are mandated by tribal law to report any observed or documented suicide ideation, attempts, deaths, nonsuicidal self-injury, and binge substance use to the Celebrating Life program (see below for operational definitions and coding). When an initial report (*intake form*) is received by the CL program, a CMHS seeks out the person at risk to complete an in-person interview and *follow-up* form to validate the event and collect additional risk and protective factor data (i.e., whether the individual is in school or employed, has a history of other suicidal behaviors, recent losses of family or friends to suicide or other violent deaths, reason for act, and substance use).

The follow-up form is not a direct self-report assessment, but rather the CMHS uses it to guide the conversation, elicit self-report, and then make a final decision on coding based on the information gathered from the individual, what is reported from the referral source on the intake form, and applying the standardized definitions used in the CL program (Cwik et al., 2014). For each intake, CMHS’ attempt to complete the follow-up visit as they receive the forms based on the prioritization described above. If they are unable to find the individual within the allotted 90-day window, staff will discontinue looking. This 90-day window is based on research showing the most high-risk period for re-attempt. Follow-up visits take place at the person’s home, school, or another private location. The CL team then refers the individual to appropriate services (*referral*) and makes plans for “wellness check-ins” as needed. All intake and follow-up data are entered into a secure password-protected database (Cwik et al., 2014). The WMAT program relies on Apache paraprofessionals so as to provide culturally responsive care and due to limited mental health services in the area (O’Keefe et al., 2019). Apache paraprofessionals who serve as Community Mental Health Specialists must have a high school diploma, have completed some college or a college degree, previous public health work experience, strong interpersonal skills, and knowledge of Apache and English languages. They receive training in coding and classification, managing and analyzing data, self-care, dissemination methods, engagement, and crisis management strategies and receive ongoing regular supervision by a master’s level clinician or higher. More information on training and supervision can be found in a descriptive article by Cwik et al. (2014).

Operational Definition and Coding

Definitions for all suicide-related behaviors are modeled on the Columbia Classification Algorithm for Suicide Assessment (C-CASA) (Posner, Oquendo, Gould, Stanley, & Davies, 2007). Staff are trained on definitions and coding, maintain a codebook for behaviors and methods at the office for reference, and review final coding with the supervision team as needed (Cwik et al., 2014). Suicide attempts include both aborted and interrupted suicide

attempts whereby the person engaged in intentional self-injury with an intent to die. Nonsuicidal self-injury is intentional self-injury without an intent to die. Suicide ideation is thoughts to take one's own life with or without a plan (i.e., both passive and active thoughts). Finally, binge substance use is defined locally as consuming substances with the intention of modifying consciousness and resulting in severe consequence, including being found unresponsive or requiring emergency treatment (Cwik et al., 2014).

Final coding of all events is done based on the initial report and an in-person validation process done by the CMHS using the *follow-up* form. Behaviors and event information are confirmed with information obtained from the person and are supplemented with information from police reports, medical records, first responders, school personnel, and local providers. Even information is based on stated intent, match of method with intent, lethality of method, reported function of the behavior, and substance use at the time of the event. If no other source (e.g., police, medical, and school) contradicts the individual's report, then coding and classification are based on the individual. If there are contradictions, staff review all materials with their supervisors to come to a consensus on final coding (Cwik et al., 2014).

Data

This analysis used all available variables from 10.75 years of surveillance and case management data in the CL system between January 2007 and October 2017. Follow-up data from the first time an individual had in-person contact with CMHS' (index event) were used to examine subsequent events registered in the CL system, thereby representing a community-wide natural cohort of individuals who are at risk of suicide. Those with a subsequent attempt within 24 months were classified as having the outcome, while those with subsequent ideation, binge substance use, nonsuicidal self-injury, or other/unknown behaviors were classified as not having the outcome. Those with subsequent deaths within the 24-month follow-up period were excluded ($n = 6$) (Table 1).

Predictor Types

Predictors included a combination of data collected at the time of event by the CMHS' and data collected over time in the registry. Predictors include 73 different variables across several domains: basic demographics, educational history, household composition, past suicidal behaviors, substance use history, number of events in registry, and time under surveillance. Most predictors were dichotomous or categorical. Categorical variables were transformed into dichotomous dummy variables as part of data preprocessing. Continuous variables included age and frequency of substance use which were treated as continuous variables. While we had age as a continuous predictor, we also created a categorical variable indicating youth (younger than 24) and adult (24 and older) for use in our decision tree analysis.

Statistical Analysis and Modeling Approach

Missing Data and Preprocessing.—Prior to model building and feature selection, we explored variable distributions and missing data mechanisms using STATA 14 (Stata-Corp, 2007). There were no missing data in the outcome variables indicating all behaviors

captured by the surveillance system were classified and coded at time of data entry. Complete data were recorded for $n = 25$ predictors (27.5%). Missing data in the predictors ranged from 16.2% for the variable asking about exposure to domestic violence to 76% for the variables recording who the person lived with at the time of the index event. The high percent of missingness was most associated with the year of incident since it was missing data due to changes in questionnaire items. Imputation of missing data that was found to be missing completely at random (MCAR) or missing at random (MAR) was imputed using Classification and Regression Trees as part of the MICE package in R and using all available data (Buuren & Groothuis-Oudshoorn, 2010). Prior to splitting the data, we removed any zero variance predictors and linear dependencies (correlation $r \geq .95$). Reporting of methods and results followed the TRIPOD guidelines for predictive modeling (Collins, Reitsma, Altman, & Moons, 2015).

Modeling Approach.—We randomly split our dataset into a training and testing dataset using a two-thirds/one-third split, respectively. This resulted in training on $n = 31$ cases for the 6-month outcome, $n = 44$ cases for the 12-month outcome, and $n = 65$ cases for the 24-month outcome. The test dataset had $n = 16$, $n = 23$, and $n = 33$ cases for the 6-, 12- and 24-month data, respectively. We trained and tested five algorithmic approaches: logistic regression, regularized regression using ridge regression, the lasso and elastic net, and a decision tree classifier. We trained all models using repeated cross-validation with ten iterations, up sampling due to the unbalanced nature of our dataset, and selection of tuning parameter that has the lowest root mean squared error (RMSE) between predicted and observed values. Logistic regression benefits from simplicity of calculation and eventual implementation, but has been shown to be inferior in the domain of suicide risk prediction compared with nonparametric algorithms like random forests (Walsh, Ribeiro, & Franklin, 2017, 2018). For penalized regression, we tuned lambda penalties for L1-regularization to generate the most parsimonious model selecting for the fewest predictors. Decision tree classifiers were selected due to their utility in guiding clinical decisions and only included predictor variables that were theoretically grounded and could be used by CMHS' in the field. For decision tree classifiers, we set tune length to 10 and used information gain as our criterion.

Once models were trained, we evaluated their performance on our validation dataset. We compared our model performance based on several statistics: kappa, sensitivity, specificity, receiver operating (ROC) curves and their c-statistic, positive predictive value (PPV), and negative predictive value (NPV). Kappa compares the observe accuracy to accuracy due to random chance (expected accuracy) and is more useful in evaluating model performance than accuracy alone in unbalanced datasets. Sensitivity is the true positive rate or the ability of the model to correctly identify those who attempted. Specificity is the true negative rate or the ability of the model to correctly identify those who did not attempt. The c-statistic from the ROC curve provides an overall estimate about how accurate the test is with 0.5 representing no better than chance and 1 representing nearperfect accuracy. Confidence intervals for ROC Area Under the Curves (AUCs) were computed with 2,000 stratified bootstrap replicates. Positive predictive value is the percentage of those identified as cases that actually are cases, while NPV is the percentage of observations identified as controls

(nonattempts) that are not attempts. Positive predictive value and NPV are particularly important as they tell us how good our model will be at identifying true cases and true controls in the future. To measure calibration, that is, whether model predictions reflected underlying prevalence of suicidality in this cohort, we also examined Brier scores (Rufibach, 2010), calibration slopes, and intercepts (Steyerberg, 2008).

A challenge in suicide research is that PPV and NPV are both impacted by the underlying prevalence of the health issue. Thus with rare events, such as suicide attempt and death, PPV and NPV are often quite low (Belsher et al., 2019). Following recent work by Kessler et al. (2016), Walsh et al. (2018), and Simon et al. (2018), we also looked at risk concentration plots to examine whether cases had higher predicted probabilities than controls. We used the risk concentration plots to generate a cutoff score in our training data based on a sensitivity of 80% and 100%. We then calculated specificity, PPV, and NPV for these cutoff scores in our test data. These methods were selected, because our work is in a community setting, with low resources and high suicide burden, and we needed an operationalized model that would help guide case managers on where to direct extra effort and resources in their follow-up procedures. This would ensure that we would attend to all those who were likely to be cases, while balancing our sensitivity at accurately identifying noncases.

RESULTS

A total of $N = 2,465$ NA youth and adults had a verified index event captured by the WMAT surveillance system between January 2006 and October 2017. There were $n = 69$ individuals who had died by suicide for their index event and were not included in analyses. We also omitted $n = 6$ individuals who had a nonsuicide death index event but later died by suicide as our analysis was focused on nonfatal suicide attempts, leaving a total of $N = 2,390$ individuals in the final analytic sample. Of these $N = 2,390$, 51.8% were male ($n = 1,240$), and average age was 22.1 years old ($SD = 10.9$; Range: 4–75). The final models examined events after an index event. At 6 months postindex event, there were $n = 47$ individuals with an attempt (2.0%); at 12 months, 20 additional people had attempted (2.8%); and at 24 months postindex event, $n = 31$ additional people had attempted (4.1%), for a total of $N = 98$ (4.1%) attempts within the 2-year time follow-up period.

Model Performance

Table 2 shows the comparison of models based on Kappa, sensitivity and specificity, and positive and negative predictive value for each window of follow-up time. Area under the curves (AUCs), calibration intercept and slope, and Brier scores can be found in Table 3. Overall, all models performed substantially better than only a history of past suicide attempt ($AUC > 0.80$ vs. $AUC = 0.57$ for detecting outcomes 24 months post-index event). Ridge regression had the highest AUC at 24 months postindex case ($AUC = 0.87$) indicating good diagnostic accuracy of future risk for suicide attempt (Hajian-Tilaki, 2013). Positive predictive value was quite low across all models with the highest value of $PPV = 0.14$ for the elastic net, decision trees, and unregularized regression for predicating attempt at 6 months postindex event. Based on these results and the feasibility of implementing the algorithm in

the existing system, we selected ridge regression as the most appropriate algorithm for our setting. Beta coefficients for our final model are included in the supplemental material (S1).

Calibration

We assessed the calibration of these approaches during cross-validation by tabulating calibration slope/intercept and Brier scores for 24-month outcomes (Table 3). At 24 months, the lowest Brier score was 0.13 for the ridge regression, lasso regression, and decision tree classifier. However, these results should be interpreted with caution given the relative rarity of suicide attempts (Benedetti, 2010).

Variable Importance

Variable importance is presented in Figure 1. Whether the person had only their index event or other events after their index case was the most important variable across all outcome models. Living status, history of domestic violence, participation in tribal activities, and knowing anyone who died by suicide in their lifetime were next most important variables included in the ridge regression. Variable importance was consistent for all outcome models, indicating stability of the algorithm at identifying those at risk for attempting suicide at different time intervals.

Risk Concentration

Figure 2. shows the percent of cases with predicted probabilities by quantile for the ridge regression in our training and test data. This type of plot allows us to look at whether people with suicide attempts are correctly binned into the highest predicted risk quintiles. All cases in the training dataset had predicted probabilities in the three highest quantiles using the 24-month outcome. Cases in the test dataset all had predicted probabilities in the four highest quantiles. Based on this plot, we selected cutoff points of predicted probabilities in the training data and evaluated their sensitivity, specificity, and PPV using our test dataset. Using a cutoff value of predicted probability = 0.37, sensitivity = 0.88, specificity = 0.72, and PPV = 0.12. Using the cutoff value of predicted probability = 0.56 corresponding to our 9th quantile, sensitivity = 0.58, specificity = 0.82, and PPV = 0.12. Finally using a cutoff of predicted probability = 0.73, specificity = 0.91, sensitivity = 0.33, and PPV = 0.13. Based on the relative severity of missing cases and lack of significant improvement in PPV, we ultimately selected 0.37 as a potential cutoff value to guide clinical decisions based on the ridge regression's output. Prospective validation in future work will reassess this cutoff as models are prone to drift in calibration and performance as factors such as outcome prevalence change over time (Davis, Lasko, Chen, Siew, & Matheny, 2017).

DISCUSSION

Identifying individuals most at risk of suicide in a timely way is critical to reducing suicide. Using machine learning to identify at-risk individuals may be particularly effective among communities with high burden and fewer treatment resources and barriers to existing clinical care. This study aimed to examine historical data gathered from one of the only documented community-based suicide surveillance systems in the world, the White Mountain Apache Tribe's Celebrating Life program, to better understand the constellation of factors that

contribute to NA individuals attempting suicide. We were able to compare different algorithms and their performance at identifying those at risk of suicide attempts within 6, 12, and 24 months of their original validated incident by the CL program. Algorithms showed a high level of accuracy (AUCs > 0.80 for identifying attempts at 24 months). The success of the ML approach appears to represent a meaningful change in our ability to identify risk over time in the CL program as our previous best known risk factor, past suicide attempt, did not distinguish between those who later attempted compared with those who continued to struggle with ideation, nonsuicidal self-injury, or binge substance use (AUC = 0.54).

To our knowledge, this work represents the first real-world machine learning algorithms using community-based data collected by case workers delivering care in a health disparity population. The nature of the data collected through the community-based CL suicide surveillance system made it impossible to apply previous suicide risk algorithms determined with other samples (Kessler, Stein, et al., 2017; Simon et al., 2018; Walsh et al., 2017, 2018). While machine learning methods can be applied to any dataset, the value of results ultimately depends on the underlying integrity and usefulness of the data. There is great concern in the field about bias in existing algorithms potentially exacerbating health disparities and even causing harm (Rajkomar et al., 2018). These biases arise from how models are designed (e.g., using only easily or commonly measured groups) to biases in training data (e.g., a protected group may have insufficient numbers of patients for the model to correctly learn statistical patterns) and to other biases involving interactions with clinicians and patients (Rajkomar et al., 2018). Given the promise of machine learning for medicine (Obermeyer & Emanuel, 2016; Rajkomar, Dean, & Kohane, 2019), ensuring that the most innovative technologies reach populations most in need is critical to reducing suicide-related disparities and promoting health equity. This research is a step toward achieving and advocating for this goal. We emphasize that impact depends on effective translation into care delivery. This latter step requires attention to implementation science, visual analytics, and clinical processes that govern care in this population.

Our findings have the potential to translate into several significant public health advances in tribal communities. On the White Mountain Apache Tribal reservation, the number of reports of suicidal behaviors has been increasing over time as the community has come to trust the CL program. As this continues, we need ways to help the CMHS' to identify and reach those at highest risk, and the ML algorithm generated through the present study may aid in this endeavor. It is important to comprehend the long distances CMHS' must travel across rural lands to follow-up, and that current resources only support a team of six CMHS' for a population of 17,500. To increase the CMHS' efficiency, we are working to embed easily interpretable risk flags that appear when individuals' incident reports are entered into the surveillance system that can help CMHS' to prioritize who to see and/or who needs more intensive case management or follow-up care. Moreover, while the CL program has benefitted greatly from dedicated staff who have served their community in this capacity for years, use of easily implementable risk-identifying tool, such as the algorithm generated here, will benefit newly hired CMHS in their immediate work.

Ultimately, implementation of this algorithm will be guided by the local Apache CL team and focused on studying this algorithm within a clinical framework, following recent

recommendations (Belsher et al., 2019). An algorithm that indicates risk 24 months out in time may not be clinically useful. The CL team may want to focus on the 6-month or 12-month risk period as more informative to their work. In the research tested here, we chose cutoff scores to maximize sensitivity, but the CL team may be more inclined to maximize specificity (e.g., how sure they can be that someone not flagged is really not at risk). Given the low PPV of these algorithms, the CL team may still benefit from using this approach, even though resources will be provided to people falsely flagged as high risk. A number of brief interventions have been developed for this population which could be offered to high-risk individuals (Cwik et al., 2019; Tingey et al., 2016). Other brief, low-cost, evidence-based interventions could be developed and implemented (e.g., caring contacts, or others: Zalsman et al., 2016). Ethical implementation of these algorithms would also need to be considered including the implications of falsely flagging someone as high risk within the CL program. Any clinical implementation of this algorithm will involve not just generation of a risk flag but navigating appropriate and ethical clinical care pathways—work that is being developed through our ongoing tribal–academic partnership.

More widely, this algorithm or similar approaches may help to better our understanding of suicide risk in NA communities. With a paucity of mental health professionals and cultural and logistic barriers to care, paraprofessionals are seen as a promising work-force that could be leveraged to provide critical mental and behavioral health care (O’Keefe et al., 2019). Using this algorithm or something similar may help paraprofessionals better identify those at risk and reach them with culturally appropriate care before they engage in serious suicidal behaviors. Several large health care settings have started to use these approaches linked to their electronic medical records data (EHR) (Kessler, Hwang, et al., 2017; Kessler, Stein, et al., 2017; Reger, McClure, Ruskin, Carter, & Reger, 2018; Walsh et al., 2017, 2018), suggesting this is a potentially scalable approach throughout Indian Health Service (IHS) and tribally run hospitals and clinics. More research is needed to validate this model in other tribal settings, continuously improve this model with data from participating tribes, and/or generate new models with EHR data from other tribal settings, involving key stakeholders in every stage of development, implementation, and analysis (Jeffery, 2015).

Limitations

First, it is important to note that use any risk algorithm will not replace professional human resources or thoughtful, culturally informed follow-up care. Apache Community Mental Health Specialists are trusted individuals from the community, and these algorithms are being explored to help them manage their caseloads and increase their ability to reach those at highest risk. Our intent is to develop tools that will help them deliver care and attention to those who need it most. Second, while CMHS’ use the follow-up data collection to guide their case management meetings to elicit self-report, compare this against what is reported on the intake form (e.g., from the ER log, police reports, schools), the final coding may still be subject to misclassification based on the CMHS’ training, mastery of the coding system, and conflicting reports. Third, despite a legal mandate for reporting, it is possible that suicidal behaviors are underreported or misclassified. For example, un-intentional overdose incidents are not included in the current mandate yet may represent additional suicide or self-harm related attempts or deaths in the community. Similarly, if people in the community

are unaware of how or what to report, these events would not be included in this data. The system may also miss transient youth or adults who are more isolated from systems or community members and thus unlikely to get reported.

Fourth, while the algorithm was generated using methods to decrease overfitting, including cross-validation, there is still a need to externally validate this approach with new data. Cross-validation resulted in a small sample of cases in the test data (e.g., $n = 16$ for the 6-month outcome data). This could affect the precision and stability of our algorithm's performance. In particular, models like these are prone to overfitting and we will be diligent to test for this in future work. Moreover, we were unable to split the data based on dates given the decline in suicide attempt rate overtime and resulting sparse data. This may limit the applicability of these flags to ongoing data collection as these time trends may indicate differences in feature importance at different chronologic time. In most suicide risk identification projects due to suicide's low base rate, our cutoff score based on the ridge yielded positive predictive value of only 0.12, which is consistent with other models in the literature (Belsher et al., 2019), and indicates a need for further exploration of the algorithms validity and whether this is an acceptable level of PPV within the CL system. We were not able to compare the performance of the algorithm to other forms of clinical risk assessment, something that could be addressed in a future study. Finally, the results of this study are limited in their generalizability as all of the sample was considered at risk for suicide and limited to one specific reservation-based NA community.

CONCLUSIONS

This is the first study of its kind to apply machine learning methods to better understand suicide risk and response within a NA population. Over the past 30 + years, suicide prevention has been a high priority for NA communities, and due to high rates and lower resources, it is essential to explore how new technologies can help to reduce suicide risk. Our study aimed to apply these methods in a unique setting in order to improve care locally and further suicide prevention research for NA populations to advance health equity by bringing state of the science methods to community-based, culturally informed prevention efforts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

- American Psychiatric Association (2011). Silver and bronze achievement awards. *Psychiatric Services*, 62(11), 1390–1392. [PubMed: 22211225]
- BARLOW A, & WALKUP JT (1998). Developing mental health services for Native American children. *Child and Adolescent Psychiatric Clinics*, 7(3), 555–577. 10.1016/S1056-4993(18)30229-3
- BELSHER BE, SMOLENSKI DJ, PRUITT LD, BUSH NE, BEECH EH, WORKMAN DE, et al. (2019). Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*, 76(6), 642 10.1001/jamapsychiatry.2019.0174 [PubMed: 30865249]

- BENEDETTI R (2010). Scoring rules for forecast verification. *Monthly Weather Review*, 138 (1), 203–211. 10.1175/2009MWR2945.1
- BOROWSKY IW, RESNICK MD, IRELAND M, & BLUM RW (1999). Suicide attempts among American Indian and Alaska Native youth: Risk and protective factors. *Archives of Pediatrics and Adolescent Medicine*, 153(6), 573–580. 10.1001/archpedi.153.6.573 [PubMed: 10357296]
- BRIDGE JA, MARCUS SC, & OLFSON M (2012). Outpatient care of young people after emergency treatment of deliberate self-harm. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(2), 213–222.e211. 10.1016/j.jaac.2011.11.002 [PubMed: 22265367]
- BUUREN S, & GROOTHUIS-OUDSHOORN K (2010). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–68.
- CARTER G, MILNER A, MCGILL K, PIRKIS J, KAPUR N, & SPITTAL MJ (2017). Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *The British Journal of Psychiatry*, 210(6), 387–395. 10.1192/bjp.bp.116.182717 [PubMed: 28302700]
- CLAASSEN CA, & LARKIN GL (2005). Occult suicidality in an emergency department population. *The British Journal of Psychiatry*, 186(4), 352–353. 10.1192/bjp.186.4.352 [PubMed: 15802695]
- COLLINS GS, REITSMA JB, ALTMAN DG, & MOONS KG (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMC medicine*, 13(1), 1 10.1186/s12916-014-0241-z [PubMed: 25563062]
- CWIK MF, BARLOW A, GOKLISH N, LARZELERE-HINTON F, TINGEY L, CRAIG M et al. (2014). Community-based surveillance and case management for suicide prevention: An American Indian tribally initiated system. *American Journal of Public Health*, 104(S3), e18–e23. 10.2105/AJPH.2014.301872
- CWIK M, BARLOW A, TINGEY L, GOKLISH N, LARZELERE-HINTON F, CRAIG M, et al. (2015). Exploring risk and protective factors with a community sample of American Indian adolescents who attempted suicide. *Archives of Suicide Research*, 19(2), 172–189. 10.1080/13811118.2015.1004472 [PubMed: 25909321]
- CWIK M, GOKLISH N, MASTEN K, LEE A, SUTTLE R, ALCHEPAY M, et al. (2019). “Let our apache heritage and culture live on forever and teach the young ones”: Development of the elders’ resilience curriculum, an upstream suicide prevention approach for American Indian Youth. *American Journal of Community Psychology*, 64, 137–145. 10.1002/ajcp.12351. [PubMed: 31313327]
- CWIK MF, TINGEY L, MASCHINO A, GOKLISH N, LARZELERE-HINTON F, WALKUP J, et al. (2016). Decreases in suicide deaths and attempts linked to the White Mountain Apache suicide surveillance and prevention system, 2001–2012. *American Journal of Public Health*, 106(12), 2183–2189. 10.2105/AJPH.2016.303453 [PubMed: 27736202]
- DAVIS SE, LASKO TA, CHEN G, SIEW ED, & MATHENY ME (2017). Calibration drift in regression and machine learning models for acute kidney injury. *Journal of the American Medical Informatics Association*, 24(6), 1052–1061. 10.1093/jamia/ocx030 [PubMed: 28379439]
- FRANKLIN JC, RIBEIRO JD, FOX KR, BENTLEY KH, KLEIMAN EM, HUANG X, et al. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, 143(2), 187–233. 10.1037/bul0000084 [PubMed: 27841450]
- FREEDENTHAL S, & STIFFMAN AR (2004). Suicidal behavior in urban American Indian adolescents: A comparison with reservation youth in a southwestern state. *Suicide and Life-Threatening Behavior*, 34(2), 160–171. 10.1521/suli.34.2.160.32789 [PubMed: 15191272]
- FREEDENTHAL S, & STIFFMAN AR (2007). “They might think I was crazy”: Young American Indians’ reasons for not seeking help when suicidal. *Journal of Adolescent Research*, 22(1), 58–77. 10.1177/0743558406295969
- HAJIAN-TILAKI K (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 4(2), 627–635. [PubMed: 24009950]
- Indian Health Service (2014). Trends in Indian Health. Retrieved May 1, 2019, from https://www.ihs.gov/dps/includes/themes/newihstheme/display_objects/documents/Trends2014Book508.pdf.

- JEFFERY AD (2015). Methodological challenges in examining the impact of healthcare predictive analytics on nursing-sensitive patient outcomes. *Computers, Informatics, Nursing*, 33(6), 258–264. 10.1097/CIN.000000000000154
- KEMBALL RS, GASGARTH R, JOHNSON B, PATIL M, & HOURY D (2008). Unrecognized suicidal ideation in ED patients: are we missing an opportunity? *The American Journal of Emergency Medicine*, 26(6), 701–705. 10.1016/j.ajem.2007.09.006 [PubMed: 18606326]
- KESSLER RC, HWANG I, HOFFMIRE CA, MCCARTHY JF, PETUKHOVA MV, ROSELLINI AJ, et al. (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, STEIN MB, PETUKHOVA MV, BLIESE P, BOSSARTE RM, BROMET EJ, et al. (2017). Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Molecular Psychiatry*, 22(4), 544–551. 10.1038/mp.2016.110 [PubMed: 27431294]
- KESSLER RC, VAN LOO HM, WARDENAAR KJ, BOSSARTE RM, BRENNER LA, CAI T, et al. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*, 21(10), 1366–1371. 10.1038/mp.2015.198 [PubMed: 26728563]
- KING CA, O'MARA RM, HAYWARD CN, & CUNNINGHAM RM (2009). Adolescent suicide risk screening in the emergency department. *Academic Emergency Medicine*, 16(11), 1234–1241. 10.1111/j.1553-2712.2009.00500.x [PubMed: 19845554]
- KNESPER D (2010). American Association of Suicidology, & Suicide Prevention Resource Center Continuity of care for suicide prevention and research: Suicide attempts and suicide deaths subsequent to discharge from the emergency department or psychiatry inpatient unit. Newton, MA: Education Development Center. Education Development Center Inc.
- LINTHICUM KP, SCHAFFER KM, & RIBEIRO JD (2019). Machine learning in suicide science: Applications and ethics. *Behavioral Sciences and the Law*, 37(3), 214–222. 10.1002/bsl.2392 [PubMed: 30609102]
- NOVINS D (2009). Participatory research brings knowledge and hope to American Indian communities. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(6), 585–586. 10.1097/CHI.0b013e3181a1f575 [PubMed: 19454913]
- OBERMEYER Z, & EMANUEL EJ (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375(13), 1216–1219. 10.1056/NEJMp1606181 [PubMed: 27682033]
- O'KEEFE VM, CWIK M, HAROZ E, & BARLOW A (2019). Increasing culturally responsive care and mental health equity with indigenous community mental health workers. *Psychological Services*, 10.1037/ser0000358 [Epub ahead of print]
- 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 [PubMed: 17606655]
- PROBST JC, LADITKA SB, MOORE CG, HARUN N, POWELL MP, & BAXLEY EG (2006). Rural-urban differences in depression prevalence: implications for family medicine. *Family Medicine-Kansas City*, 38(9), 653.
- PULLMANN MD, VANHOOSER S, HOFFMAN C, & HEFLINGER CA (2010). Barriers to and supports of family participation in a rural system of care for children with serious emotional problems. *Community Mental Health Journal*, 46(3), 211–220. 10.1007/s10597-009-9208-5 [PubMed: 19551506]
- RAJKOMAR A, DEAN J, & KOHANE I (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358. 10.1056/NEJMr1814259 [PubMed: 30943338]
- 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]

- REGER GM, MCCLURE ML, RUSKIN D, CARTER SP, & REGER MA (2018). Integrating predictive modeling into mental health care: an example in suicide prevention. *Psychiatric Services*, 70, 71–74. [PubMed: 30301448]
- RUFIBACH K (2010). Use of Brier score to assess binary predictions. *Journal of Clinical Epidemiology*, 63(8), 938–939. 10.1016/j.jclinepi.2009.11.009 [PubMed: 20189763]
- SIMON GE, JOHNSON E, LAWRENCE JM, ROSSOM RC, AHMEDANI B, LYNCH FL, et al. (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 [PubMed: 29792051]
- StataCorp L (2007). *Stata data analysis and statistical Software*. Special Edition Release, 10, 733.
- STEYERBERG EW (2008). *Clinical prediction models: A practical approach to development, validation, and updating*. Berlin, Germany: Springer Science & Business Media.
- Substance Abuse and Mental Health Services Administration (2011). *Results from the 2010 National Survey on Drug Use and Health: Mental health findings*. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- TINGEY L, LEE A, SUTTLE R, LAKE K, WALKUP JT, & ALLISON BARLOW M (2016). Development and piloting of a brief intervention for suicidal American Indian adolescents. *American Indian and Alaska Native Mental Health Research*, 23 (1), 105–124. 10.5820/aian.2301.2016.105 [PubMed: 28562844]
- TRAN T, LUO W, PHUNG D, HARVEY R, BERK M, KENNEDY RL, et al. (2014). Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry*, 14(1), 76–85. 10.1186/1471-244X-14-76 [PubMed: 24628849]
- WALSH CG, RIBEIRO JD, & FRANKLIN JC (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3), 457–469. 10.1186/1471-244X-14-76
- 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*, 59(12), 1261–1270. 10.1111/jcpp.12916 [PubMed: 29709069]
- Web-based Injury Statistics Query and Reporting System (WISQARS). (2005). National center for injury prevention and control. Atlanta, GA: Web-based Injury Statistics Query and Reporting System (WISQARS).
- ZALSMAN G, HAWTON K, WASSERMAN D, VAN HEERINGEN K, ARENSMAN E, SARCHIAPONE M, et al. (2016). Suicide prevention strategies revisited: 10-year systematic review. *The Lancet Psychiatry*, 3(7), 646–659. 10.1016/S2215-0366(16)30030-X [PubMed: 27289303]

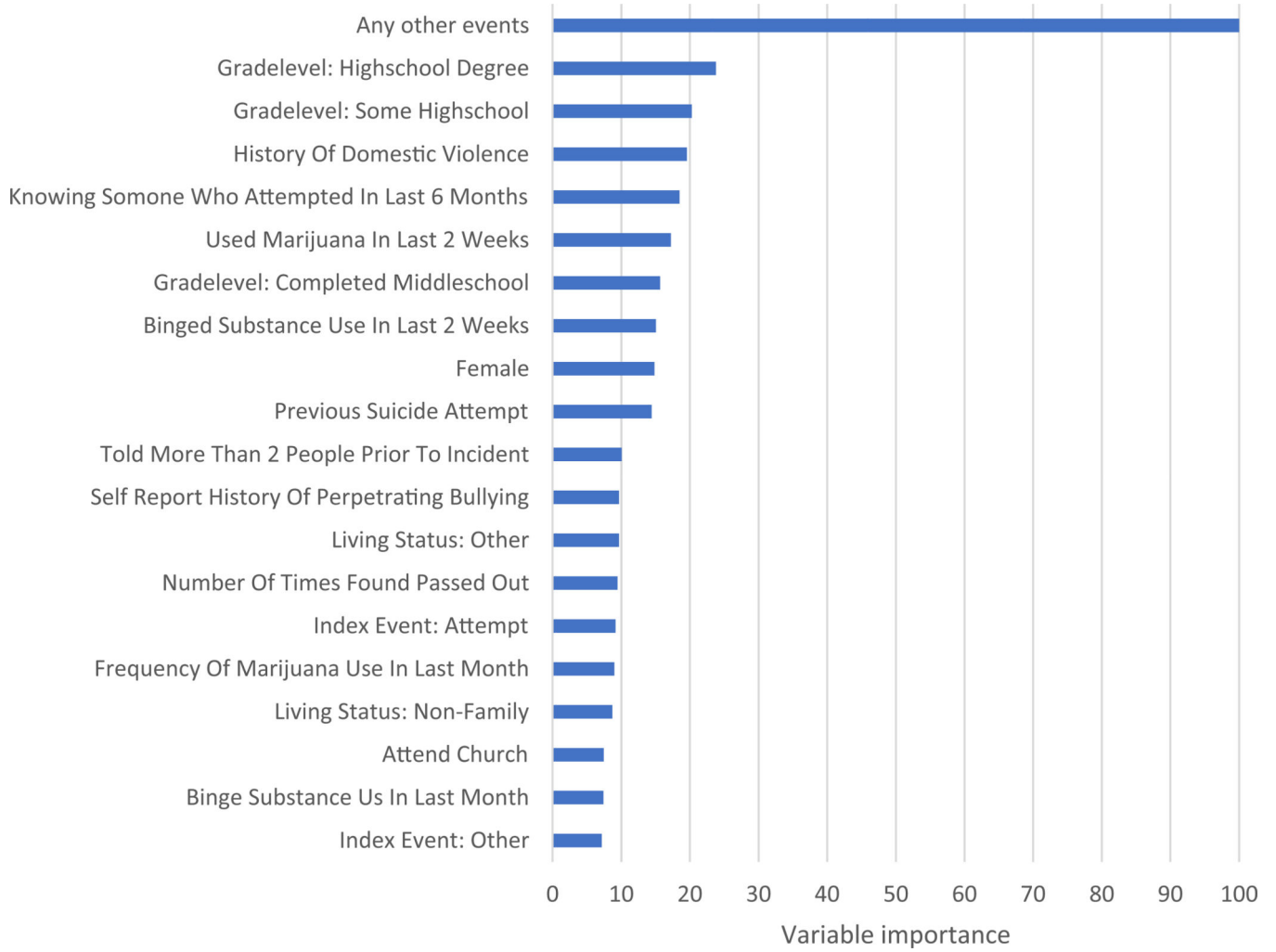


Figure 1. Feature importance for top 20 most important features from ridge regression at 24 months.

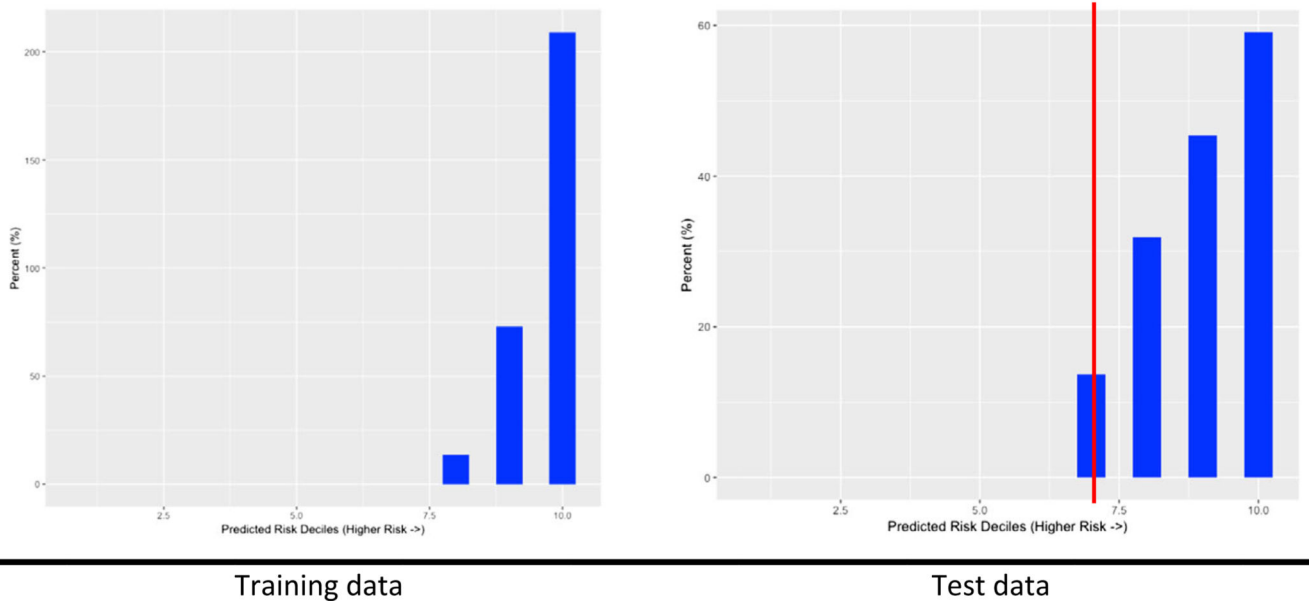


Figure 2. Rates of validated suicide attempts by decile of predicted risk at 24 months using ridge regression in training data and in test data showing cutoff value.

TABLE 1*Index Events in Celebrating Life System 2006–2017 (N = 2,390)*

Gender	
Male	1240 (51.9)
Female	1150 (48.1)
Age (Mean, SD); Range	22.1 (10.9); 4–74
Types of suicidal behavior at index event	
Ideation	610 (35.2)
Nonsuicidal self-injury	613 (25.7)
Suicide Attempt	333 (13.9)
Binge substance use	287 (12.0)
Unspecified	458 (19.2)
Suicide attempts over time <i>N</i> (%) of individuals	
6 months postindex	47 (2.0)
12 months postindex	67 (2.8)
24 months postindex	98 (4.1)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2
 Comparison of Models for Identifying Attempts Based on Event Data in Test Dataset

Algorithm	Kappa	Sens	Spec	PPV	NPV
Attempt within six months					
Ridge regression	0.13	0.64	0.78	0.11	0.98
Lasso regression	0.14	0.58	0.82	0.12	0.98
Elastic net	0.18	0.76	0.81	0.14	0.99
Unregularized logistic regression	0.17	0.67	0.82	0.14	0.98
Decision trees	0.14	0.30	0.92	0.14	0.97
Previous attempt	0.04	0.39	0.75	0.06	0.97
Attempt within 12 months					
Ridge regression	0.15	0.70	0.78	0.12	0.98
Lasso regression	0.15	0.67	0.79	0.12	0.98
Elastic net	0.16	0.64	0.82	0.13	0.98
Unregularized logistic regression	0.13	0.52	0.83	0.12	0.98
Decision trees	0.14	0.28	0.93	0.14	0.97
Previous attempt	0.02	0.30	0.75	0.05	0.96
Attempt within 24 months					
Ridge regression	0.16	0.76	0.76	0.13	0.99
Lasso regression	0.14	0.58	0.82	0.12	0.98
Elastic net	0.13	0.52	0.83	0.11	0.98
Unregularized logistic regression	0.13	0.52	0.83	0.12	0.98
Decision trees	0.17	0.70	0.81	0.13	0.98
Previous attempt	0.04	0.36	0.78	0.07	0.97

TABLE 3

Comparison of Model Accuracy at 24 months

Algorithm	AUC	95% CI for AUC	Calibration slope	Intercept	Brier score
Ridge regression	0.87	0.83, 0.90	0.95	-2.72	0.13
Lasso regression	0.86	0.82, 0.90	0.31	-2.42	0.13
Elastic net	0.85	0.81, 0.89	0.28	-2.27	0.13
Unregularized logistic regression	0.83	0.78, 0.88	0.14	-2.13	0.14
Decision tree classifier	0.81	0.73, 0.88	0.13	-4.60	0.13
Previous attempt	0.57	0.49, 0.65	1.00	-2.13	0.24

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript