HHS Public Access

Author manuscript

Am J Psychiatry. Author manuscript; available in PMC 2019 October 01.

Published in final edited form as:

Am J Psychiatry. 2018 October 01; 175(10): 951–960. doi:10.1176/appi.ajp.2018.17101167.

PREDICTING SUICIDE ATTEMPTS AND SUICIDE DEATHS FOLLOWING OUTPATIENT VISITS USING ELECTRONIC HEALTH RECORDS

Gregory E Simon, MD MPH[1], **Eric Johnson, MS**[1], **Jean M Lawrence, ScD**[2], **Rebecca C Rossom, MD MS**[3], **Brian Ahmedani, PhD**[4], **Frances L Lynch, PhD**[5], **Arne Beck, PhD**[6], **Beth Waitzfelder, PhD**[7], **Rebecca Ziebell**[1], **Robert B Penfold, PhD**[1], and **Susan M Shortreed, PhD**[1]

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA

[2]Dept. of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA

[3]HealthPartners Institute, Minneapolis, MN

[4]Center for Health Services Research, Henry Ford Health System, Detroit, MI

[5]Center for Health Research, Kaiser Permanente Northwest, Portland, OR

[6]Institute for Health Research, Kaiser Permanente Colorado, Denver, CO

[7]Center for Health Research, Kaiser Permanente Hawaii, Honolulu, HI

Abstract

Objective—Develop and validate models using electronic health records to predict suicide attempt and suicide death following an outpatient visit.

Methods—Across seven health systems, 2,960,929 patients aged 13 or older (mean age 46, 62% female) made 10,275,853 specialty mental health visits and 9,685,206 primary care visits with mental health diagnoses between 1/1/2009 and 6/30/2015. Health system records and state death certificate data identified suicide attempts (n=24,133) and suicide deaths (n=1240) over 90 days following each visit. Potential predictors included 313 demographic and clinical characteristics extracted from records for up to five years prior to each visit: prior suicide attempts, mental health and substance use diagnoses, medical diagnoses, psychiatric medications dispensed, inpatient or emergency department care, and routinely administered PHQ-9 depression questionnaires. Logistic regression models predicting suicide attempt and death were developed using penalized LASSO variable selection in a random 65% sample of visits and validated in the remaining 35%.

Results—Mental health specialty visits with risk scores in the top 5% accounted for 43% of subsequent suicide attempts and 48% of suicide deaths. Of patients scoring in the top 5%, 5.4% attempted suicide and 0.26% died by suicide within 90 days. C-statistics (equivalent to AUC) for prediction of suicide attempt and suicide death were 0.851 (95% CI 0.848 to 0.853) and 0.861 (95% CI 0.848 to 0.875). Primary care visits with scores in the top 5% accounted for 48% of

Corresponding author: Gregory Simon, Kaiser Permanente Washington Health Research Institute, 1730 Minor Ave. #1600, Seattle, WA 98101. Phone: 206-287-2979, simon.g@ghc.org.

subsequent suicide attempts and 43% of suicide deaths. C-statistics for prediction of suicide attempt and suicide death were 0.853 (95% CI 0.849 to 0.857) and 0.833 (95% CI 0.813 to 0.853).

Conclusions—Prediction models incorporating both health records data and responses to self-report questionnaires substantially outperform existing suicide risk prediction tools.

Suicide accounted for almost 45,000 deaths in the United States in 2016, a 25% increase since 2000[1]. Non-fatal suicide attempts account for almost 500,000 emergency department visits annually[2]. Half of people dying by suicide and two-thirds of people surviving suicide attempts received some mental health diagnosis or treatment during the prior year[3,4]. Mindful of those prevention opportunities, a Joint Commission Sentinel Event Alert now recommends detection of suicide risk across health care[5]. Unfortunately, traditional clinical detection of suicide risk is hardly better than chance[6].

We have previously reported that brief depression questionnaires can accurately predict suicide attempt or death[7]. Outpatients reporting thoughts of death or self-harm "nearly every day" on item 9 of the Patient Health Questionnaire (PHQ-9) are seven times as likely to attempt suicide and six times as likely to die by suicide over the following 90 days compared to patients reporting such thoughts "not at all"[7]. Sensitivity of this tool, however, is only moderate. One-third of suicide attempts and deaths occur among patients reporting suicidal ideation "not at all". Accurate identification of high risk is also only moderate. The 6% of patients reporting suicidal ideation "more than half the days" or "nearly every day" account for only 35% of suicide attempts and deaths. More accurate tools for identifying both low and high risk patients are needed.

Recent research has used various modeling methods to predict suicidal behavior from electronic health records (EHRs). Examples include prediction of suicide death among Veterans Administration service users[8], prediction of suicide death following psychiatric hospitalization among Army soldiers[9], distinguishing patients attempting suicide from those with other injuries or poisonings[10], and prediction of suicide or accidental death following civilian general hospital discharge[11]. Two recent analyses have used health records data to predict suicide attempt or suicide death following outpatient visits. Kessler and colleagues[12] used health records and military service records to predict suicide death among US Army soldiers in the 26 weeks following a mental health visit. Approximately one quarter of suicide deaths occurred after the 5% of visits rated as highest risk. Barak-Corren and colleagues[13] used health records data to predict suicide attempt or death over outpatients making three or more visits in two large academic health systems. One-third of suicide attempts and deaths occurred in the 5% of patients with highest risk scores.

Here we combine data typically available from EHRs with depression questionnaire data in seven large health systems to develop and validate models predicting suicide attempt and suicide death over 90 days following a mental health or primary care visit.

METHODS

The seven health systems participating in this research (HealthPartners; Henry Ford Health System; and the Colorado, Hawaii, Northwest, Southern California and Washington regions of Kaiser Permanente) serve a combined population of approximately eight million members in nine states. Each system provides insurance coverage and comprehensive health care (including general medical and specialty mental health care) to a defined population enrolled through employer-sponsored insurance, individual insurance, capitated Medicaid or Medicare, and subsidized low-income programs. Members are representative of each system's service area in age, race/ethnicity, and socioeconomic status. All systems recommend using the PHQ-9 at mental health visits and primary care visits for depression, but implementation varied across systems during the study period.

As members of the Mental Health Research Network, each health system maintains a research data warehouse following the Health Care Systems Research Network Virtual Data Warehouse model[14]. This resource combines data from insurance enrollment records, EHRs, insurance claims, pharmacy dispensings, state mortality records, and census-derived neighborhood characteristics. Responsible institutional review boards for each health system approved use of these de-identified data for this research.

The study sample included any outpatient visit by a member aged 13 or older either to a specialty mental health clinic or to a primary care clinic when a mental health diagnosis was recorded. Sampling was limited to visits to health system clinics (to ensure availability of EHR data) and people insured by the health system's insurance plan (to ensure availability of insurance claims data). All qualifying visits from 1/1/2009 through 6/30/2015 were included, except at Henry Ford where only visits after implementation of a new EHR system (12/1/2012) were included.

Potential predictors extracted from health system records for up to five years prior to each visit included: demographic characteristics (age, sex, race, ethnicity, source of insurance, and neighborhood income and educational attainment), current and past mental health and substance use diagnoses (organized in 12 categories), past suicide attempts, other past injury or poisoning diagnoses, dispensed prescriptions for mental health medication (organized in four categories), past inpatient or emergency department mental health care, general medical diagnoses (by Charlson Comorbidity Index[15] categories), and recorded scores on the PHQ-9 questionnaire[16] (including total scores and item 9 scores).

Potential predictors were represented as dichotomous indicators. Each diagnosis category was represented by three overlapping indicators (recorded at or within 90 days prior to the visit, recorded within one year prior, and recorded within five years prior). Each category of medication or emergency/inpatient utilization was represented by three overlapping indicators (occurred within 90 days prior to the visit, one year prior, or any time prior). To represent temporal patterns of prior PHQ-9 item 9 scores, 24 indicators were calculated for each encounter to represent number of observations, maximum value, and modal value (including value of missing) during three overlapping time periods (previous 90 days, previous 183 days, and previous 365 days). The final set of potential predictors for each encounter included 149 indicators and 164 possible interactions (see Appendix 9a for complete list).

Diagnoses of self-harm or probable suicide attempt were ascertained from all injury or poisoning diagnoses recorded in EHRs and insurance claims accompanied by an ICD-9 cause of injury code indicating intentional self-harm (E950-E958) or undetermined intent (E980-E989). Data from these health systems during the study period indicate that inclusion of injuries and poisonings with undetermined intent increases ascertainment of probable suicide attempts by approximately 25%[7] (see also Appendix 4). While use of E-codes varied across the US during the study period[17], participating health systems were selected for high and consistent rates of E-code use (Appendix 1). Records review[7] also supports the positive predictive value of this definition for identification of true self-harm in these health systems (see also Appendix 2). Furthermore, observation of coding changes across the transition from ICD-9 to the more specific ICD-10 coding scheme indicate that most "undetermined" ICD-9 diagnoses actually reflect self-harm[18] (see also Appendix 3). Ascertainment of suicide attempts was censored at health system disenrollment, after which insurance claims data regarding self-harm diagnoses at external facilities would not be available.

Suicide deaths were ascertained from state mortality records. Following common recommendations[19,20] all deaths with an ICD-10 diagnosis of self-inflicted injury (X60-X84) or injury/poisoning with undetermined intent (Y10-Y34) were considered probable suicide deaths. Inclusion of injury and poisoning deaths with undetermined intent increases ascertainment of probably suicide deaths by 5–10%[7] (see also Appendix 4).

All predictor and outcome variables were completely specified and calculated prior to model training.

Prediction models were developed separately for mental health specialty and primary care visits, with a 65% random sample of each used for model training and 35% set aside for validation. Models included multiple visits per person in order to accurately represent changes in risk within patients over time. For each visit, analyses considered any outcome in the following 90 days, regardless of a subsequent visit in between. This approach uses all data available at the time of the index visit, but avoids informative or biased censoring related to timing of visits following the index date. In the initial variable selection step, separate models predicting risk of suicide attempt and suicide death were estimated using logistic regression with penalized LASSO variable selection[21]. The LASSO penalization factor selects important predictors by shrinking coefficients for weaker predictors toward zero, excluding predictors with estimated zero coefficients from the final sparse prediction model. To avoid over-fitting models to idiosyncratic relationships in the training samples, variable selection used 10-fold cross-validation[22] to select the optimal level of tuning or penalization, measured by the Bayesian Information Criterion[23]. In the second calibration step, generalized estimating equations with a logistic link re-estimated coefficients in the training sample, accounting for both clustering of visits under patients and bias toward the null in LASSO coefficients. In the final validation step, logistic models derived from the above two-step process were applied in the 35% validation sample to calculate predicted probabilities for each visit. Results are reported as receiver operating characteristic (ROC) curves[24] with c-statistics[25,26] along with predicted and observed rates in pre-specified strata of predicted probability. Over-fitting was evaluated by comparing classification performance and in training and validation samples and by comparing predicted risk to observed risk in the validation sample. Variable selection analyses were conducted using the GLMNET[27] and Foreach[28] packages for R statistical software, version 3.4.0. Confidence intervals for c-statistics were calculated via bootstrap with 10,000 replications.

A public repository (www.github.com/MHResearchNetwork) includes: specifications and code for defining predictor and outcome variables, a data dictionary and descriptive statistics for analytic datasets, code for variable selection and calibration steps, coefficients and confidence limits from all final models, and comparison of model performance in training and validation samples.

RESULTS

We identified 19,961,059 eligible visits by 2,960,929 patients during the study period, including 10,275,853 mental health specialty visits and 9,685,206 primary care visits with mental health diagnoses (Table 1). Following the specifications above, health system records identified 24,133 unique probable suicide attempts within 90 days of an eligible visit, and state mortality records identified 1240 unique suicide deaths within 90 days.

Models predicting probable suicide attempt over 90 days were developed and validated for both mental health and primary care visits, excluding 0.3% of visits because of disenrollment within 90 days. Clinical variables with the largest positive prediction coefficients are shown in the left portion of Table 2 (see Appendices 9b and 9c for all selected predictors and coefficients). Strongest predictors of suicide attempt were similar in mental health specialty and primary care patients: prior suicide attempt, mental health and substance use diagnoses, responses to PHQ-9 item 9, and prior inpatient or emergency mental health care.

The left portion of Figure 1 shows ROC curves illustrating sensitivity and specificity of suicide attempt predictions in training and validation samples. C-statistics (equivalent to AUC or area under the ROC curve) for prediction of suicide attempt in the validation samples were 0.851 (95% CI 0.848 to 0.853) for mental health specialty visits and 0.853 (95% CI 0.849 to 0.857) for primary care. In each graph, comparison of ROC curves shows no appreciable difference in prediction accuracy between the training and validation samples (i.e. no evidence for model over-fitting). Table 3 compares predicted to observed risk for specific strata selected a priori. Among mental health specialty visits, the lowest two strata included 75% of all visits and 21% of all suicide attempts, while the highest three strata included 5% of visits and 43% of suicide attempts. Among primary care visits, the 75% of visits with lowest risk scores accounted for 21% of suicide attempts, while the 5% of visits with highest scores accounted for 48%. Comparison of predicted risk levels in the training sample and observed risk levels in the validation sample again shows no appreciable decline in model performance or evidence for model over-fitting. Sensitivity analyses limited to diagnoses of definite self-harm slightly improved prediction accuracy (especially among primary care patients) but excluded approximately 25% of probable suicide attempts (Appendix 4). Sensitivity analyses limited to visits preceded by at least 5 years of complete data yielded essentially identical prediction accuracy (Appendix 5). Model fit was consistent across the seven participating health systems and across age and sex subgroups (Appendix 8)

The same process was implemented for prediction of suicide deaths over 90 days, with separate models for mental health specialty and primary care visits. Clinical variables most strongly associated with suicide death in each group are shown in Table 2 (see Appendices 9d and 9e for complete list). Predictors of suicide death were similar in mental health specialty and primary care patients, and were similar to predictors of suicide attempt.

The right portion of Figure 1 shows ROC curves for prediction of suicide death in training and validation samples. C-statistics for prediction of suicide death in the validation samples were 0.861 (95% CI 0.848 to 0.875) for mental health specialty visits and 0.833 (95% CI 0.813 to 0.853) for primary care. Comparison of ROC curves for the training and validation samples shows no evidence of over-fitting in the mental health specialty sample and a minimal separation of training and validation curves in the primary care sample. The right portion of Table 3 compares predicted to observed risk for risk strata selected a priori. Among mental health specialty visits, the lowest two risk strata included 75% of visits and 19% of suicide deaths, while the highest three risk strata included 5% of visits and 48% of suicide deaths. Among primary care visits, the 75% of visits with lowest risk scores accounted for 25% of suicide deaths, while 5% of visits with highest scores accounted for 43%. Comparison of predicted risk levels in the training sample and observed risk levels in the validation sample shows no evidence for over-fitting in the primary care sample and a minimal fall-off between training and validation samples in the primary care sample. Sensitivity analyses limited to deaths coded as due to definite self-inflicted injury or poisoning found no meaningful difference in model fit (Appendix 4).

Table 4 displays sensitivity, specificity, positive predictive value (PPV), and negative predictive value for all four models at cut-points defined by percentiles of the risk score distribution.

DISCUSSION

In a sample of 20 million visits by 3 million patients in seven health systems, data from EHRs accurately stratified mental health specialty and primary care visits according to short-term risk of suicide attempt or suicide death. Observed rates of probable suicide attempt and suicide death were over 200 times as high following visits in the highest 1% compared to visits in the bottom half of predicted risk (Table 3). Strongest predictors included mental health diagnoses, substance use diagnoses, use of mental health emergency and inpatient care, and history of self-harm. Absolute risk was lower in primary care, but predictors selected and accuracy of prediction were similar across care settings. Responses to PHQ-9 questionnaires were selected as important predictors, even though such data were available for only 15% of visits.

Potential Limitations

In interpreting these findings, we should consider both false positive and false negative errors in the ascertainment of probable suicide attempts and deaths. Previous research suggests false positive rates near zero for suicide deaths diagnosed by medical examiners[20] and below 10–20% for diagnoses of definite or possible self-inflicted injury in records from these health systems[7] (also see Appendix 2). Diagnostic data do not distinguish between self-harm with and without intent to die. Consequently, our definition of probable suicide attempt may include a small proportion of self-harm episodes without suicidal intent. False negative errors may be more common. Up to one quarter of suicide deaths may not be identified by medical examiners[19]. Health system records will not capture suicide attempts when people do not seek care or when providers do not recognize and record diagnoses of self-harm. Non-specific error (either false positive or false negative) would lead to under-estimating the accuracy of prediction models (see appendix 4), while selective error in the wrong direction (e.g. under-ascertainment of suicide attempts in patients with low risk scores) could lead to over-estimation of model performance.

Health system records do not reflect important social risk factors for suicidal behavior, such as job loss, bereavement, or relationship disruption. Suicidal behavior likely reflects the intersection of clinical risk factors, negative life events, and access to means of self-harm. Data regarding those social risk factors would certainly improve accuracy of prediction.

Our analyses do not consider the one-third to one half of people attempting suicide or dying by suicide who have no recent mental health treatment or recorded diagnosis[3,4,33]. Prediction using EHR data might also prove useful among patients without recorded mental health diagnoses, but prediction models would necessarily be limited to general medical diagnoses and utilization rather than the mental health diagnoses and treatments selected in this sample.

Methodologic Considerations

We focus on risk over 90 days following an outpatient visit. Risk does vary between visits[29], and near-term risk is most relevant to clinical decisions and quality improvement[30]. The interventions that providers or health systems might provide for high-risk patients would typically be delivered over weeks or months[31,32]. Predictors selected in these models (Table 2) include both recent or short-term factors and long-term factors, consistent with previous research[7,29] indicating that suicidal behavior is influenced by both stable and variable risk factors. Sensitivity analyses using a 30-day outcome window (Appendix 7) yielded similar results regarding both predictors selected and accuracy of prediction. Analyses regarding longer-term risk might identify different predictors of suicidal behavior.

Of predictive modeling methods, parametric methods like LASSO lie closest to traditional regression. Non-parametric methods[34] such as random forest could theoretically improve accuracy of prediction. Direct comparisons to date[12,35], however, have found equal or superior prediction using parametric methods similar to those used here. Non-parametric methods may have little advantage when predictors are dichotomous, such as the diagnosis and utilization indicators included in our models. Parametric models are usually more transparent to clinicians[36] and simpler to implement in EHRs, as is now underway in these health systems and the Veterans Health Administration[35].

Variable selection models are subject to over-fitting or selection of predictive relationships idiosyncratic to a specific sample. The large sample used for training of these models offers some protection against over-fitting. In addition, we present explicit comparisons of performance in the training and randomly selected validation samples for all four models

(Table 3 and Figure 1), finding no indication of over-fitting in prediction of suicide attempts or prediction of suicide deaths following mental health specialty visits. We do find a slight indication of over-fitting in prediction of suicide deaths following primary care visits, likely reflecting the smaller number of events included in these models. Nevertheless, overall accuracy of prediction (c-statistic) in the independent validation sample still exceeds 80%.

In addition to evaluating over-fitting within this sample, we should consider generalizability to other care settings or patient populations. This sample included almost 20 million visits in seven health systems serving patients in nine states – including states with high and low rates of suicide mortality. Patients were broadly representative of those service areas in race/ethnicity, socioeconomic status, and source of insurance coverage – including substantial numbers insured by Medicare and Medicaid. Methods could be easily transported to health systems with standard electronic health records and insurance claims databases. Predicted risk levels, however, could be over- or under-estimated in settings with higher or lower average risk of suicidal behavior. Predictors selected and accuracy of prediction could differ in settings with different patterns of mental health care, especially if patterns of diagnosis or utilization are less closely linked to risk of suicidal behavior. Intervention of effective suicide prevention programs could also weaken the relationship between these identified risk predictors and subsequent suicidal behavior. Consequently, we recommend replication in other health systems prior to broad application. All information necessary for replication is available via our online repository.

Context

These empirically derived risk scores outperformed risk stratification based solely on PHQ-9 item 9. Regarding sensitivity: selecting mental health visits with any positive response to item 9 would identify only two thirds of subsequent suicide attempts and deaths[7], while selecting visits with risk scores above the 75th percentile would identify 80%. Regarding efficient identification of high risk: selecting the 6% of visits with a response of “more than half the days” or “nearly every day” would identify one-third of subsequent suicide attempts and deaths[7], while selecting the 5% of visits with highest risk scores would identify almost half.

Predictors identified in these models included a range of demographic characteristics, mental health diagnoses, and historical indicators of mental health treatment; generally similar to those identified in previous research[9,12,13]. Based on results in validation samples, performance of these prediction models equaled or exceeded that of other published models using health records to predict suicidal behavior, where c-statistics ranged from 0.67 to 0.84[8–13]. These models significantly outperformed other published models predicting suicidal behavior after an outpatient visit, a question of high interest to a wide range of mental health and primary care providers. In this sample, mental health specialty visits with risk scores in the top 5% accounted for 43% of suicide attempts and 48% of suicide deaths in the following 90 days, while primary care visits in the top 5% accounted for 48% of subsequent suicide attempts and 43% of subsequent suicide deaths. For comparison, in two previous models predicting suicidal behavior following outpatient visits, the top 5% of patients accounted for between one quarter and one third of subsequent suicide attempts and deaths[12,13]. This improved prediction likely reflects differences in data and methods. First, longitudinal records in integrated health systems may allow more complete ascertainment of risk factors. Second, our analyses consider a larger number of potential predictors and more detailed temporal encoding. Third, responses to PHQ-9 item 9 contributed to prediction, even though such data were available for only 10–20% of visits. Prediction accuracy would likely improve with greater use of the PHQ-9 or similar measures, as is expected with new initiatives promoting routine outcome assessment[37] and identification of suicidal ideation[5].

C-statistics for these suicide prediction models also exceed those for models using health records data to predict re-hospitalization for heart failure[38], in-hospital mortality from sepsis[39], and high emergency department utilization[40]. Suicidal behavior may be more predictable than many adverse medical outcomes.

Among mental health specialty visits, a cut-point at the 95th percentile of risk had a positive predictive value of 5.4% for suicide attempt within 90 days. While that predictive value would be inadequate for a diagnostic test, it is similar or superior to widely accepted tools for prediction of major medical outcomes such as stroke in atrial fibrillation[41], or cardiovascular events[42]. Furthermore, predictive values or expected event rates for widely accepted medical prediction tools often include adverse outcomes accumulated over many years[41,42], rather than the 90-day risk period considered in these analyses.

Clinical Implications

Some recent discussions of predictive modeling in healthcare warn that reliance on algorithms could lead to inappropriate causal inference[43–45] or atrophy of clinician judgement[43]. Regarding the first point, associations identified by our model should certainly not be interpreted as evidence for independent or causal relationships. For example, a recent benzodiazepine prescription is more likely a marker of increased risk than a cause of suicidal behavior. We report predictors selected (Table 2) to demonstrate that all are expected correlates of suicidal behavior, albeit in specific combinations within specific time periods. Regarding the second point, our model and other models predicting suicidal behavior from records data rely largely on the diagnostic and treatment decisions of treating clinicians. The predictors identified by our analyses would be well-known to most mental health providers. Predictive models simply allow us to consistently combine millions of providers' individual judgements to accurately predict an important but rare event[45].

Prediction models cannot replace clinical judgement, but risk scores can certainly inform both individual clinical decisions and quality improvement programs. Participating health systems now recommend completion of a structured suicide risk assessment[46] following any response of "more than half the days" or "nearly every day" to PHQ9 item 9 – implying a 90-day risk of suicide attempt of 2–3%[7]. A predicted 90-day risk exceeding 5% (i.e. above the 95th percentile for mental health specialty visits) would seem to warrant a similar level of additional assessment. A predicted 90-day suicide attempt risk exceeding 10% (i.e. above the 99th percentile for mental health specialty visits) should warrant creation of a personal safety plan and counseling regarding reducing access to means of self-harm[47,48]. Accurate risk stratification can also inform providers' and health systems' decisions regarding frequency of follow-up, referral for intensive treatment, or outreach following missed or cancelled appointments[30,49]. Implementing these risk-based care pathways and outreach programs is a central goal of the Zero Suicide prevention model recommended by the U.S. National Action Alliance for Suicide Prevention[48]. Empirically derived risk predictions can be an important component of that national suicide prevention strategy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by cooperative agreement U19 MH092201 with the National Institute of Mental Health

References

1. Kochanek KD, Murphy SL, Xu JQ, Arias E. NCHS Data Brief: Mortality in the United States, 2016. Hyattsville, MD: National Center for Health Statistics; 2017.
2. WISQARS Nonfatal Injury Reports, 200–2014. 2017. (Accessed April 4, 2017, at https://webappa.cdc.gov/sasweb/ncipc/nfirates.html.)
3. Ahmedani BK, Simon GE, Stewart C, et al. Health care contacts in the year before suicide death. J Gen Intern Med. 2014; 29:870–7.
4. Ahmedani BK, Stewart C, Simon GE, et al. Racial/Ethnic differences in health care visits made before suicide attempt across the United States. Med Care. 2015; 53:430–5. [PubMed: 25872151]
5. Patient Safety Advisory Group. Detecting and treating suicidal ideation in all settings. The Joint Commission Sentinel Event Alerts. 2016:56.
6. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychol Bull. 2017; 143:187–232. [PubMed: 27841450]
7. Simon GE, Coleman KJ, Rossom RC, et al. Risk of suicide attempt and suicide death following completion of the Patient Health Questionnaire depression module in community practice. J Clin Psychiatry. 2016; 77:221–7. [PubMed: 26930521]
8. McCarthy JF, Bossarte RM, Katz IR, et al. Predictive Modeling and Concentration of the Risk of Suicide: Implications for Preventive Interventions in the US Department of Veterans Affairs. Am J Public Health. 2015; 105:1935–42. [PubMed: 26066914]
9. Kessler RC, Warner CH, Ivany C, et al. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study To Assess Risk and rEsilience in Servicemembers (Army STARRS). JAMA Psychiatry. 2015; 72:49–57. [PubMed: 25390793]
10. Walsh CG, Ribeiro JD, Franklin JC. Predicting Risk of Suicide Attempts Over Time Through Machine Learning. Clinical Psychological Science. 2017; 5:457–69.
11. McCoy TH Jr, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. JAMA Psychiatry. 2016; 73:1064–71. [PubMed: 27626235]
12. Kessler RC, Stein MB, Petukhova MV, et al. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). Mol Psychiatry. 2016
13. Barak-Corren Y, Castro VM, Javitt S, et al. Predicting Suicidal Behavior From Longitudinal Electronic Health Records. Am J Psychiatry. 2017; 174:154–62. [PubMed: 27609239]
14. Ross TR, Ng D, Brown JS, et al. The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. eGEMs. 2014:2.
15. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. J Clin Epidemiol. 1994; 47:1245–51. [PubMed: 7722560]

16. Kroenke K, Spitzer RL, Williams JB, Lowe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry. 2010; 32:345–59. [PubMed: 20633738]
17. Lu CY, Stewart C, Ahmed AT, et al. How complete are E-codes in commercal plan claims databases? Pharmacoepidemiol Drug Saf. 2014; 23:218–20. [PubMed: 24453020]
18. Stewart C, Crawford PM, Simon GE. Changes in Coding of Suicide Attempts or Self-Harm With Transition From ICD-9 to ICD-10. Psychiatr Serv. 2017; 68:215. [PubMed: 27903145]
19. Bakst SS, Braun T, Zucker I, Amitai Z, Shohat T. The accuracy of suicide statistics: are true suicide deaths misclassified? Soc Psychiatry Psychiatr Epidemiol. 2016; 51:115–23. [PubMed: 26364837]
20. Cox KL, Nock MK, Biggs QM, et al. An Examination of Potential Misclassification of Army Suicides: Results from the Army Study to Assess Risk and Resilience in Servicemembers. Suicide Life Threat Behav. 2016
21. Tibshirani R. Regression shrinkage and selection via the lasso. J Royal Stat Soc (B). 1996; 58:267–88.
22. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2nd. New York: Springer; 2009.
23. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc. 1995; 90:773–95.
24. Egan JP. Signal Detection Theory and ROC Analysis. New York: Springer Academic Press; 1975.
25. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]
26. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. 1997; 30:1145–59.
27. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33:1–22. [PubMed: 20808728]
28. Weston S. Foreach looping construct for R. R Package Version. 2015; 143
29. Simon GE, Shortreed SM, Johnson E, et al. Between-visit changes in suicidal ideation and risk of subsequent suicide attempt. Depress Anxiety. 2017
30. Olfson M, Marcus SC, Bridge JA. Focusing suicide prevention on periods of high risk. JAMA. 2014; 311:1107–8. [PubMed: 24515285]
31. Brown GK, Ten Have T, Henriques GR, Xie SX, Hollander JE, Beck AT. Cognitive therapy for the prevention of suicide attempts: a randomized controlled trial. JAMA. 2005; 294:563–70. [PubMed: 16077050]
32. Comtois KA, Linehan MM. Psychosocial treatments of suicidal behaviors: a practice-friendly review. J Clin Psychol. 2006; 62:161–70. [PubMed: 16342292]
33. Han B, Compton WM, Gfroerer J, McKeon R. Mental health treatment patterns among adults with recent suicide attempts in the United States. Am J Public Health. 2014; 104:2359–68. [PubMed: 25322299]
34. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform. 2002; 35:352–9. [PubMed: 12968784]
35. Kessler RC, Hwang I, Hoffmire CA, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. Int J Methods Psychiatr Res. 2017:26.
36. Adkins DE. Machine Learning and Electronic Health Records: A Paradigm Shift. Am J Psychiatry. 2017; 174:93–4. [PubMed: 28142275]
37. HEDIS Depression Measures Specified for Electronic Clinical Data Systems. 2016. (Accessed February 16, 2016, at http://www.ncqa.org/HEDISQualityMeasurement/HEDISLearningCollaborative/HEDISDepressionMeasures.aspx.)
38. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure: Comparison of Machine Learning and Other Statistical Approaches. JAMA Cardiol. 2017; 2:204–9. [PubMed: 27784047]
39. Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Acad Emerg Med. 2016; 23:269–78. [PubMed: 26679719]

40. Frost DW, Vembu S, Wang J, Tu K, Morris Q, Abrams HB. Using the Electronic Medical Record to Identify Patients at High Risk for Frequent Emergency Department Visits and High System Costs. Am J Med. 2017

41. Lip GY. Can we predict stroke in atrial fibrillation? Clin Cardiol. 2012; 35(Suppl 1):21–7.

42. Rana JS, Tabada GH, Solomon MD, et al. Accuracy of the Atherosclerotic Cardiovascular Risk Equation in a Large Contemporary, Multiethnic Population. J Am Coll Cardiol. 2016; 67:2118–30. [PubMed: 27151343]

43. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017; 318:517–8. [PubMed: 28727867]

44. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. N Engl J Med. 2017; 376:2507–9. [PubMed: 28657867]

45. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016; 375:1216–9. [PubMed: 27682033]

46. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011; 168:1266–77. [PubMed: 22193671]

47. Rossom RC, Simon GE, Beck A, et al. Facilitating Action for Suicide Prevention by Learning Health Care Systems. Psychiatr Serv. 2016; 67:830–2. [PubMed: 27032667]

48. Hogan MF, Grumet JG. Suicide Prevention: An Emerging Priority For Health Care. Health Aff (Millwood). 2016; 35:1084–90. [PubMed: 27269026]

49. Miller IW, Camargo CA Jr, Arias SA, et al. Suicide Prevention in an Emergency Department Population: The ED-SAFE Study. JAMA Psychiatry. 2017

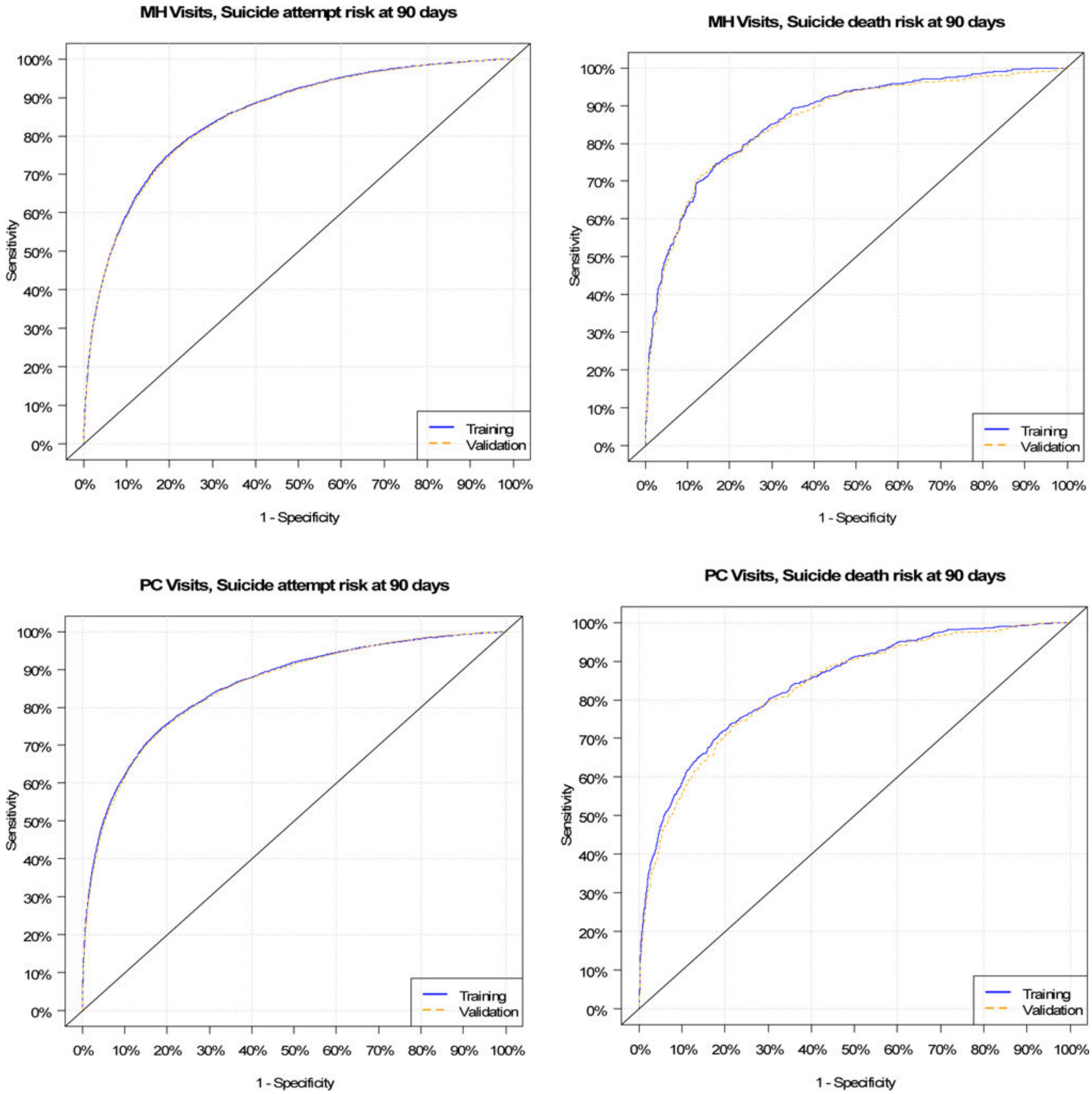

Figure 1.

Receiver operating characteristic curves illustrating model performance in validation dataset for prediction of suicide attempts and suicide deaths within 90 days of visit in seven health systems, 2009–2015. The area below the training curve and above the validation curve indicates potential over-fitting in the training sample.

Table 1

Characteristics of sampled visits to specialty mental health and primary care providers, randomly divided into model training (65%) and validation (35%) samples.

	Mental Health Specialty				Primary Care			
	Training		Validation		Training		Validation	
VISITS	6,679,128		3,596,725		6,297,465		3,387,741	
Female	4,157,997	62%	2,239,213	62%	3,872,830	61%	2,083,4 24	61 %
Age								
13–17	671,313	10%	360,619	10%	250,878	4%	135,070	4%
18–29	1,118,492	17%	603,044	17%	822,668	13%	442,774	13%
30–44	1,744,704	26%	939,431	26%	1,337,686	21%	720,878	21%
45–64	2,453,509	37%	1,321,986	37%	2,466,992	39%	1,326,237	39%
65 or older	691,110	10%	371,645	10%	1,419,241	23%	762,782	23%
Race								
White	4,562,203	68%	2,455,211	68%	4,162,033	66%	2,237,952	66%
Asian	302,231	5%	162,400	5%	379,910	6%	204,272	6%
Black	600,219	9%	324,233	9%	514,021	8%	276,260	8%
Hawaiian/Pacific Islander	74,473	1%	40,118	1%	103,420	2%	55,833	2%
Native American	65,309	1%	35,332	1%	69,425	1%	37,717	1%
More than one or Other	38,223	1%	20,485	1%	43,445	1%	23,391	1%
Not Recorded	1,036,470	16%	558,946	16%	1,025,211	16%	552,316	16%
Ethnicity								
Hispanic	1,486,400	22%	800,547	22%	1,430,611	23%	769,498	23%
Insurance Type								
Commercial Group	5,057,328	76%	2,724,286	76%	4,198,138	67%	2,258,974	67%
Individual	827,218	12%	445,749	12%	1,079,401	17%	580,225	17%
Medicare	363,598	5%	194,773	5%	576,184	9%	310,001	9%
Medicaid	213,573	3%	114,767	3%	297,710	5%	160,063	5%
Other	217,411	3%	117,150	3%	146,032	2%	78,478	2%
PHQ9 Item 9 score recorded								

	Mental Health Specialty				Primary Care			
	Training		Validation		Training		Validation	
At index visit	657,998	10%	354,918	10%	312,065	5%	168,569	5%
At any visit in past year	1,328,571	20%	714,693	20%	671,643	11%	362,438	11%
Length of enrollment prior to visit								
1 year or more	5,810,841	87%	3,129,151	87%	5,352,845	85%	2,879,580	85%
5 years or more	3,772,409	56%	2,031,916	56%	3,542,358	56%	1,907,063	56%
Visits followed by								
Suicide Attempt within 90 days	41,470	0.62%	22,329	0.62%	16,302	0.26%	8688	0.26%
Suicide Death within 90 days	1529	0.02%	854	0.02%	856	0.01%	445	0.01%

Table 2

Clinical characteristics selected for prediction of suicide attempt and suicide death within 90 days of visit, listed in order of coefficients in logistic regression models. Interaction terms are indicated by “with”. See Appendices 8b–8e for complete list.

SUICIDE ATTEMPT FOLLOWING:		SUICIDE DEATH FOLLOWING:	
MENTAL HEALTH SPECIALTY VISIT (of 94 predictors selected)	**PRIMARY CARE VISIT (of 102 predictors selected)**	**MENTAL HEALTH SPECIALTY VISIT (of 43 predictors selected)**	**PRIMARY CARE VISIT (of 29 predictors selected)**
Depression diagnosis in last 5 yrs.	Depression diagnosis in last 5 yrs.	Suicide attempt diagnosis in last year	Mental health ER visit in last 3 mos.
Drug abuse diagnosis in last 5 yrs.	Suicide attempt diagnosis in last 5 yrs.	Benzodiazepine Rx. in last 3 mos	Alcohol abuse diagnosis in last 5 yrs.
PHQ-9 Item 9 score =3 in last year	Drug abuse diagnosis in last 5 yrs.	Mental health ER visit in last 3 mos	Benzodiazepine Rx. in last 3 mos.
Alcohol use disorder Diag. in last 5 yrs	Alcohol abuse diagnosis in last 5 yrs.	2nd Gen. Antipsychotic Rx in last 5 years	Depression diagnosis in last 5 yrs.
Mental health inpatient stay in last yr.	PHQ-9 Item 9 score=3 in last year	Mental health inpatient stay in last 5 years	Mental health inpatient stay in last year
Benzodiazepine Rx. in last 3 mos.	Suicide attempt diagnosis in last 3 mos.	Mental health inpatient stay in last 3 mos	Injury/Poisoning diagnosis in last year
Suicide attempt in last 3 mos.	Suicide attempt diagnosis in last year	Mental health inpatient stay in last year	Anxiety disorder diagnosis in last 5 yrs.
Personality disorder diag. in last 5 yrs.	Personality disorder diag. in last 5 yrs.	Alcohol use disorder Diag. in last 5 years	PHQ-9 Item 9 score=1 with PHQ8 score
Eating disorder diagnosis in last 5 yrs.	Anxiety Disorder diagnosis in last 5 yrs.	Antidepressant Rx in last 3 mos	PHQ-9 item 9 score=3 with Age
Suicide Attempt in last year	Suicide attempt diagnosis in last 5 yrs with Schizophrenia diag. in last 5 yrs.	PHQ-9 Item 9 score = 3 with PHQ8 score	Suicide attempt diag. in past 5 yrs with Age
Mental health ER visit in last 3 mos.	Benzodiazepine Rx. in last 3 mos.	PHQ-9 item 9 score = 1 with Age	Mental health ER visit in past year
Self-inflicted cutting/piercing in last year	Eating Disorder diagnosis in last 5 yrs.	Depression diag. in last 5 yrs. with Age	PHQ-9 Item 9 score=2 with Age
Suicide attempt in last 5 yrs.	Mental health ER visit in last 3 mos.	Suicide attempt diag. in last 5 yrs. with Charlson Score	PHQ-9 Item 9 score=3 with PHQ8 score
Injury/poisoning diagnosis in last 3 mos.	Injury/Poisoning diagnosis in last year	PHQ-9 Item 9 score = 2 with Age	Bipolar disorder diagnosis in last 5 yrs with Age
Antidepressant Rx. in last 3 mos.	Mental health ER visit in last year	Anxiety disorder diag. in last 5 yrs. with Age	Depression diagnosis in last 5 yrs with Age

Table 3

Classification accuracy in pre-defined strata for prediction of suicide attempts and suicide deaths within 90 days of a mental health or primary care visit in seven health systems, 2009–2015. Potential over-fitting in training sample is indicated by differences between predicted and actual risks.

SUICIDE ATTEMPTS FOLLOWING:					SUICIDE DEATHS FOLLOWING:				
A Ment al Health Specialty Visit					**A Mental Health Specialty Visit**				
Risk Score Percentile Strata	**Predicted Risk[1]**	**Actual Risk[2]**	**% of All Attempts[3]**	**Standardized Event Ratio[4]**	**Risk Score Percentile Strata**	**Predicted Risk[1]**	**Actual Risk[2]**	**% of All Deaths[3]**	**Standardized Event Ratio[4]**
>99.5th	13.0%	12.7%	10%	20.7	>99.5th	0.654%	0.694%	12%	24.6
99^{th} to 99.5^{th}	8.5%	8.1%	6%	12.9	99^{th} to 99.5^{th}	0.638%	0.595%	11%	21.5
95^{th} to 99^{th}	4.1%	4.2%	27%	6.7	95^{th} to 99^{th}	0.162%	0.167%	25%	6.3
90^{th} to 95^{th}	1.9%	1.8%	15%	3.0	90^{th} to 95^{th}	0.068%	0.088%	16%	2.3
75^{th} to 90^{th}	0.9%	0.9%	21%	1.4	75^{th} to 90^{th}	0.031%	0.029%	16%	1.1
50^{th} to 75^{th}	0.3%	0.3%	13%	0.51	50^{th} to 75^{th}	0.014%	0.015%	13%	0.54
$<50^{th}$	0.1%	0.1%	8%	0.16	$<50^{th}$	0.003%	0.003%	6%	0.12
A Primary Care Visit with Mental Health Diagnosis					A Primary Care Visit with Mental Health Diagnosis				
Risk Score Percentile Strata	Predicted Risk[1]	Actual Risk[2]	% of All Attempts[3]	Standardized Event Ratio[4]	Risk Score Percentile Strata	Predicted Risk[1]	Actual Risk[2]	% of All Deaths[3]	Standardized Event Ratio[4]
$>99.5^{th}$	8.6%	8.0%	15%	30.5	$>99.5^{th}$	0.536%	0.435%	14%	28.8
99^{th} to 99.5^{th}	4.1%	4.2%	8%	16.3	99^{th} to 99.5^{th}	0.181%	0.197%	7%	13.0
95^{th} to 99^{th}	1.6%	1.6%	25%	6.2	95^{th} to 99^{th}	0.092%	0.083%	22%	5.6
90^{th} to 95^{th}	0.7%	0.7%	13%	2.6	90^{th} to 95^{th}	0.035%	0.038%	13%	2.5
75^{th} to 90^{th}	0.3%	0.3%	18%	1.2	75^{th} to 90^{th}	0.018%	0.019%	19%	1.3
50^{th} to 75^{th}	0.1%	0.1%	12%	0.49	50^{th} to 75^{th}	0.009%	0.009%	15%	0.62
<50th	0.04%	0.04%	9%	0.17	<50th	0.003%	0.003%	10%	0.19

Notes:

[1] Predicted risk in this stratum using final model predictors and coefficients in the training sample

[2] Observed risk in this stratum using final model predictors and coefficients in the validation sample

[3] Percentage of all suicide attempts or deaths occurring in this stratum in validation sample

[4]Ratio of observed risk in this stratum of the validation sample to average risk in the full validation sample

Table 4

Performance characteristics at various cut-points for prediction of suicide attempts and suicide deaths within 90 days of visit in seven health systems, 2009–2015.

SUICIDE ATTEMPTS FOLLOWING:					**SUICIDE DEATHS FOLLOWING:**				
Mental Health Specialty Visits					**Mental Health Specialty Visits**				
Risk Score Percentile Cut-Points	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Risk Score Percentile Cut-Points**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
>99th	16.8%	99.1%	10.4%	99.4%	>99th	23.1%	99.0%	0.62%	99.9%
>95th	43.7%	95.2%	5.4%	99.6%	>95th	48.1%	95.0%	0.26%	99.9%
>90th	58.3%	90.3%	3.6%	99.7%	>90th	64.3%	90.0%	0.17%	99.9%
>75th	79.2%	75.2%	2.0%	99.8%	>75th	80.4%	75.1%	0.08%	99.9%
>50th	92.1%	50.0%	1.1%	99.9%	>50th	94.0%	50.0%	0.05%	99.9%
Primary Care Visits with Mental Health Diagnosis					Primary Care Visits with Mental Health Diagnosis				
Risk Score Percentile Cut-Points	Sensitivity	Specificity	PPV	NPV	Risk Score Percentile Cut-Points	Sensitivity	Specificity	PPV	NPV
>99th	23.5%	99.1%	6.1%	99.8%	>99th	20.9%	99.0%	0.31%	99.9%
>95th	48.2%	95.1%	2.5%	99.9%	>95th	43.1%	95.0%	0.13%	99.9%
>90th	61.0%	90.1%	1.6%	99.9%	>90th	55.7%	90.0%	0.08%	99.9%
>75th	79.1%	75.1%	0.8%	99.9%	>75th	74.8%	75.1%	0.05%	99.9%
>50th	91.4%	50.1%	0.5%	99.9%	>50th	90.3%	50.0%	0.03%	99.9%