

Predicting Suicidal Behavior From Longitudinal Electronic Health Records

Yuval Barak-Corren, M.S., Victor M. Castro, M.S., Solomon Javitt, M.D., Alison G. Hoffnagle, M.S., Yael Dai, B.A., Roy H. Perlis, M.D., M.Sc., Matthew K. Nock, Ph.D., Jordan W. Smoller, M.D., Sc.D., Ben Y. Reis, Ph.D.

Objective: The purpose of this article was to determine whether longitudinal historical data, commonly available in electronic health record (EHR) systems, can be used to predict patients' future risk of suicidal behavior.

Method: Bayesian models were developed using a retrospective cohort approach. EHR data from a large health care database spanning 15 years (1998–2012) of inpatient and outpatient visits were used to predict future documented suicidal behavior (i.e., suicide attempt or death). Patients with three or more visits (N=1,728,549) were included. ICD-9-based case definition for suicidal behavior was derived by expert clinician consensus review of 2,700 narrative EHR notes (from 520 patients), supplemented by state death certificates. Model performance was evaluated retrospectively using an independent testing set.

Results: Among the study population, 1.2% (N=20,246) met the case definition for suicidal behavior. The model achieved

sensitive (33%–45% sensitivity), specific (90%–95% specificity), and early (3–4 years in advance on average) prediction of patients' future suicidal behavior. The strongest predictors identified by the model included both well-known (e.g., substance abuse and psychiatric disorders) and less conventional (e.g., certain injuries and chronic conditions) risk factors, indicating that a data-driven approach can yield more comprehensive risk profiles.

Conclusions: Longitudinal EHR data, commonly available in clinical settings, can be useful for predicting future risk of suicidal behavior. This modeling approach could serve as an early warning system to help clinicians identify high-risk patients for further screening. By analyzing the full phenotypic breadth of the EHR, computerized risk screening approaches may enhance prediction beyond what is feasible for individual clinicians.

Am J Psychiatry 2017; 174:154–162; doi: 10.1176/appi.ajp.2016.16010077

Suicide is one of the leading causes of death worldwide (1), but several barriers have slowed progress in understanding, predicting, and preventing suicidal behavior. First, attempts to predict suicide have relied almost exclusively on self-reporting of suicidal thoughts and intentions. This is problematic because self-reported data are subject to well-known reporting biases, and especially problematic in the case of suicide given that many people are motivated to deny suicidal thoughts in order to avoid hospitalization (2). Second, there is a limited understanding of what factors actually predict suicidal behavior. To be sure, many risk factors have been identified, including the presence of psychiatric disorders, younger age, and a history of prior suicidal behavior (3). However, these factors predict only a small amount of the variance in suicidal behavior and without a high degree of accuracy. Third, there is currently no accepted model for understanding how risk factors work together to cause suicidal behavior. This is not only a problem for researchers, but for clinicians, as there is currently no algorithm that clinicians

can use to combine information about the multiple risk factors they might assess when trying to determine whether a patient is likely to make a suicide attempt in the near future. Thus, clinicians are left to use their intuition as a guide, which unfortunately is no better than chance at predicting suicidal behaviors (4). Fourth, because suicide is a low base-rate behavior, very large samples are needed to test the complex models that have been proposed to predict suicide. Indeed, virtually all theories of suicide suggest that it is a multidetermined outcome in which many different factors work together to cause suicide attempts. However, the vast majority of prior research on this topic has tested simple bivariate models (5).

What is needed are longitudinal data from large samples that can be used to develop and test new models of suicide risk. Such data are routinely collected in health information systems. However, this resource has been surprisingly underexplored by suicide researchers and represents a promising new direction for future scientific and clinical efforts. The growing adoption of electronic health records

See related features: **Editorial** by Dr. Adkins (p. 93), **Clinical Guidance** (Table of Contents), **CME course** (p. 191), and **AJP Audio** (online)

(EHRs) has created a powerful resource for epidemiologic and risk prediction studies (6–12). We previously showed that data commonly available in EHRs can accurately predict future domestic abuse diagnoses an average of 2 years in advance (13). Using this information to support early detection of individuals at high risk for suicide and self-inflicted injury could help prevent significant morbidity and mortality and ensure that at-risk patients receive the professional care they need. Although some attempts to predict suicidal behavior using electronic health information have been reported, prior studies have had important limitations, including relatively small sample size (14, 15), evaluation of a modest number of potential predictors (14, 16), or limited data on prediction model performance (16).

Here we report the development and validation of a risk prediction model using readily available EHR data to predict suicide attempts or death by suicide in a large health care system. As the data are already collected and readily available in the clinical setting, this study facilitates the goal of constructing a widely adoptable clinical decision support approach.

METHOD

The primary data source for this study was the Partners Healthcare Research Patient Data Registry (17). The Partners Healthcare Research Patient Data Registry is a data warehouse of EHR data covering 4.6 million patients from two large academic medical centers in Boston (Massachusetts General Hospital and Brigham and Women's Hospital), as well as community and specialty hospitals in the Boston area. To assemble the cohort for this study, the Partners Healthcare Research Patient Data Registry was queried for all inpatient and outpatient visits occurring between 1998 and 2012 (inclusive) at Massachusetts General Hospital and Brigham and Women's Hospital. Development of the predictive model was conducted in the stages described below.

Case Definition and Validation

There were 1,728,549 patients who met the inclusion criteria of three or more visits, 30 days or more between the first and last visits, and the existence of records after age 10 and before age 90. For each patient, we obtained all the demographic, diagnostic, procedure, laboratory, and medication data recorded at each visit. We excluded 3,658 patients (0.21%) due to lack of historical data (as their first recorded encounter was for suicidal behavior) and 106 individuals for whom gender was not recorded. On average, each patient was followed for a period of 5.27 years, and the analysis included a total of 8,980,954 person-years.

Suicidal behavior was defined according to ICD-9 diagnostic codes and death certificates from the Commonwealth of Massachusetts. ICD-9 codes of E95* (injuries of intentional intent) are the most explicit diagnostic code for suicide attempts. To validate our case definition, we randomly selected 100 patients with an E95* code and reviewed all clinical notes within 1 week of the ICD-9 diagnosis. Three

senior clinicians with expertise in the epidemiology and treatment of suicidal behavior (J.W.S., R.H.P., M.K.N.) manually reviewed narrative notes. Each note was designated as one of six categories using consensus agreement by all three clinicians (see Appendix 1 in the data supplement accompanying the online version of this article), and the positive predictive value of each code was calculated as the proportion of notes classified as either 1) self-harm, suicidal, or 2) self-harm, intentional, nonsuicidal.

Previous reports have indicated that ICD-9 codes matching E950* may have low sensitivity to detect suicidal behavior due to coding practices, reimbursement patterns, and the uncertainty of intent (18). To maximize sensitivity of the case definition, we identified an additional set of 15 ICD-9 injury code categories and E98* (injury of questionable intent) as potential indicators of suicide attempts in the EHR (see Table S1 in the online data supplement). The codes were selected based on prior literature, as well as a review of code descriptions.

For each ICD-9 category, the clinicians reviewed a small sample of patients using the chart review method above. If the prevalence of true cases was >20%, a larger sample of 50 randomly selected patients was subsequently reviewed. Code categories with a positive predictive value >0.70 were selected for our final case definition. These included E95* (positive predictive value: 0.82), 965.* (poisoning by analgesics, antipyretics, and antirheumatics; positive predictive value: 0.80), 967.* (poisoning by sedatives and hypnotics; positive predictive value: 0.84), 969.* (poisoning by psychotropic agents; positive predictive value: 0.80), and 881.* (open wound of elbow, forearm, and wrist; positive predictive value: 0.70). Detailed chart review results, including codes not included in the definition, are available in Table S1 in the online data supplement.

In total, over 2,700 notes for 520 individuals were reviewed to establish codes that best identified suicide attempt cases. These were supplemented by obtaining death certificates from the Commonwealth of Massachusetts, to capture completed suicides not recorded in the EHR. We also included a total of 852 death certificates between 1997 and 2010 with a "manner of death" of suicide (ICD-9: E95* or ICD-10 X60-X84, Y87.0) as cases.

Model Development

To explicitly account for differences by gender, we developed separate and independent models for men and women. We divided our cohort into two subcohorts of 718,793 men and 1,005,992 women. Each subcohort was randomly divided into training and testing (validation) sets of equal sizes. Based on the training sets, we developed naive Bayesian classifier models (19) to estimate a patient's risk for suicidal behavior. Naive Bayesian classifiers are a subclass of Bayesian networks with strong conditional independence of all input features, which greatly reduces model complexity and makes model development highly scalable for handling many independent variables. Naive Bayesian classifier models have been shown to be well-suited for clinical decision support and classification tasks (13) and have the additional benefit of being easy

to interpret. Models were developed using R version 3.1.1 with packages *e1071*, *pROC*, and *ggplot2*. Detailed description of model development is provided in Appendix 2 in the data supplement.

The models included data on demographic characteristics, diagnostic codes, laboratory results (normal/low/high), and prescribed medications (true/false values). Data were collected up to but not including the first suicidal event for the case subjects and for all observed time periods for the control subjects. For each independent input variable in the training data set (e.g., diagnoses, medications, etc.), we assigned a partial risk score based on the ratio of its prevalence among case subjects compared with control subjects. The score was calculated on a logarithmic scale such that negative scores were “protective” (not associated with suicidal behavior), and positive scores were “adverse” (with higher prevalence among cases). In preparation for model validation, thresholds were selected to achieve benchmark specificities of 90% and 95% in the training sets.

Model Validation

We validated the models on the testing set of each gender subcohort, using a simulated prospective approach. For each patient, we calculated an overall risk score at each time point based on the data available for that patient until that time. For each item in the patient’s record, we assigned the appropriate partial risk score based on the model trained above. We then calculated the patient’s overall cumulative risk score by combining these partial risk scores for each subject. The patient’s score was interpreted using the thresholds selected during the training phase to achieve 90% and 95% specificities, respectively, and the sensitivity and timeliness of prediction at these levels of specificity were measured.

The Value of a Comprehensive Data-Driven Approach

To evaluate the usefulness of our comprehensive data-driven approach, we compared our results to the results obtained when looking at three widely accepted risk factors for suicide: 1) depression, 2) substance abuse, and 3) patients having any mental health condition (20). We defined these risk factors using the Clinical Classification Software (21) created and validated by the Healthcare Cost and Utilization Project. The Clinical Classification Software codes used were 657 for depression, 661 for substance abuse, and 650–663, 670 for all mental health conditions. Using the same training and validation sets described above, we tested the predictiveness of depression, substance abuse, depression and substance abuse, and any mental health condition (including depression and substance abuse) for suicidal behavior.

RESULTS

Model Composition

Of the total 1,728,549 patients, we identified 20,246 (1.2%) cases with suicidal behavior. All other patients were labeled as controls. As previously described, we excluded 3,764

subjects from our analysis due to missing data, resulting in a final set of 16,588 case subjects and 1,708,197 control subjects. Of the 852 death certificates with suicide as a cause of death, only 49 did not have one of the ICD-9 codes that comprise the case definition, indicating high sensitivity of our EHR case definition.

The demographic characteristics of all patients recorded within the Partners Healthcare Research Patient Data Registry data warehouse (including the excluded cases) are presented in Table 1. The relative score associated with each demographic factor by gender is summarized in Table 2. Overall, suicidal behavior was more common among men than women (odds ratio=1.75, 95% confidence interval [CI]=1.68–1.82). For both men and women, “separated” marital status was associated with more than a fourfold risk of suicidal behavior compared with married patients ($p<0.001$). Higher risk of suicidal behavior was observed in African American (odds ratio=1.31, 95% CI=1.22–1.41) and Hispanic (odds ratio=1.68, 95% CI=1.58–1.79) patients compared with Caucasian patients. With regard to age, higher prevalence of suicidal behavior was found in women under the age of 25 (odds ratio=1.5, 95% CI=1.37–1.64 compared with other age groups) and in men aged 25–45 (odds ratio=1.83, 95% CI=1.73–1.93 compared with other age groups).

Details regarding the risk scores associated with individual codes are presented in Table S2 in the online data supplement. (It is noteworthy that while Table S2 in the data supplement highlights only the top 100 codes associated with suicidal behavior, the naive Bayesian classifier model actually captures risks associated with all available codes.) Opioid abuse was 16 times more common among case subjects than control subjects (95% CI=14.9–22.8), and personality and bipolar disorders were 7–10 times more common among case subjects ($p<0.001$). Of note, however, a variety of other clinical features beyond mental health diagnoses appeared among the top 100 predictors, including infections such as hepatitis C carrier (odds ratio=6.1, 95% CI=4.5–8.1), alveolitis of the jaw (odds ratio=5.8, 95% CI=3.7–9.4), osteomyelitis (odds ratio=4.85, 95% CI=2.9–8.0), cellulitis (odds ratio=4.6, 95% CI=3.9–5.4), and numerous codes related to wounds and injuries including contusion of the back (odds ratio=4.7, 95% CI=3.1–7.2) (see Table S2 in the data supplement for additional details). A summary view of the effect sizes (odds ratios) of diagnostic codes grouped into 135 categories is presented in Figure 1, as defined by the Clinical Classification Software codes mentioned previously. As expected, suicidal behavior is strongly associated with substance abuse and psychiatric conditions in both men and women (see Table S2A in the data supplement). The lists of top medications and laboratory tests (Table S2B and C in the data supplement) also highlight the elevated suicide risk associated with drug abuse and mental illness. The top laboratory results associated with suicidal behavior were related to standard toxicology screenings, and the top medications associated with suicidal behavior were mostly psychiatric drugs.

Model Performance

Relying solely on coded information commonly available in the EHR, the model successfully predicted suicidal behavior with an overall area under the receiver operating characteristic curve of 0.77 (Figure 2). The model performed similarly in female and male cohorts (area under the curve=0.77 [95% CI=0.77–0.78] compared with 0.76 [95% CI=0.75–0.77], respectively). Detailed results of the naive Bayesian classifier model by gender are summarized in Table 3. With 90% specificity, the model detected 44% and 46% of the suicidal cases among men and women, respectively. Consistent with the low base rate of suicidal behavior in the full cohort, the positive predictive value was 5% and 3% compared with 1.55% and 0.9% baseline prevalence for men and women, respectively. Running the model by gender for specific age groups yielded even better prediction for narrower subpopulations, such as women ages 45–65 where, for 90% specificity, the model achieved 54% sensitivity (see Table S3 in the data supplement).

One of the model's key strengths is its ability to incorporate the full phenotypic breadth for the EHR in making a prediction—beyond what an individual clinician might typically use in a given encounter. To examine the advantages of this approach, we compared the model performance to that of simple models based only on commonly used risk factors. Allowing for a 10% false-positive rate, the full model achieved 45% sensitivity, while models that only used various combinations of widely accepted risk factors performed substantially worse: 29% for depression (area under the curve=0.62 [0.62–0.63]), 25% for substance abuse (area under the curve=0.58 [0.58–0.59]), 34% for depression and substance abuse (area under the curve=0.65 [0.64–0.65]), and 19% for any mental health condition (area under the curve=0.64 [0.63–0.64]). Thus, the relative increase in sensitivity was 32% to 137% compared with these simpler models.

We also examined the average time that models were able to predict suicidal behavior in advance of an individual receiving a case-defining diagnosis. Setting specificity at 90%, the model predicted suicidal behavior an average of 4.0 years before the case-defining code was recorded in the EHR for the 45% of the cases identified by the model at this specificity level. Increasing the model specificity to 95% (i.e., only 5% false positives), our classifier predicted suicidal events an average of 3.5 years prior to the diagnosis for the 33% of the cases identified by the model at this specificity level.

The average cumulative risk scores over time for case subjects compared with control subjects are shown in

TABLE 1. Demographic Features of Case and Control Subjects, With Case Percentage Per Gender

Characteristic	Case Subjects ^a		Control Subjects ^b		% Cases ^c		
	N	Column %	N	Column %	Total	Male	Female
Gender							
Female	9,068	44.8	998,171	58.4	0.90	—	0.90
Male	11,177	55.2	710,027	41.6	1.55	1.55	—
Age group							
<25 years old	2,043	10.1	152,073	8.9	1.33	1.42	1.25
25–45 years old	8,290	40.9	540,394	31.6	1.51	2.31	1.04
45–65 years old	6,768	33.4	588,705	34.5	1.14	1.57	0.82
≥65 years old	3,145	15.5	427,130	25	0.73	0.78	0.69
Race/ethnicity							
Asian	373	1.8	61,079	3.6	0.61	0.82	0.49
African American	1,684	8.3	109,166	6.4	1.52	2.15	1.13
Hispanic	2,490	12.3	126,952	7.4	1.92	2.45	1.53
Other	1,258	6.2	187,482	11.0	0.67	0.95	0.48
White	14,441	71.3	1,223,624	71.6	1.17	1.52	0.90
Veteran status							
Veteran	1,132	5.6	86,552	5.1	1.29	1.31	1.00
Not veteran	13,943	68.9	1,047,040	61.3	1.31	1.79	1.02
Unknown	5,171	25.5	574,711	33.6	0.89	1.22	0.67
Marital status							
Divorced	1,610	8.0	86,329	5.1	1.83	2.75	1.39
Married	5,327	26.3	829,327	48.5	0.64	0.83	0.49
Other/unknown	991	4.9	100,092	5.9	0.98	1.33	0.74
Partner	20	0.1	1,991	0.1	0.99	1.55	0.60
Separated	459	2.3	17,586	1.0	2.54	3.39	2.04
Single	11,013	54.4	591,621	34.6	1.83	2.42	1.37
Widowed	826	4.1	81,357	4.8	1.01	1.27	0.94

^a Data represent the number of case subjects with specific demographic features (absolute numbers, with column percentage per category).

^b Data represent the number of control subjects with specific demographic features (absolute numbers, with column percentage per category).

^c Data represent the percentage of cases within each category out of the entire cohort (total), the male cohort (men), or the female cohort (women).

Figure 2A for women and in Figure 2B for men. As time progresses, there is growing separation between the cumulative score for case subjects compared with control subjects, with a maximal difference after 15 years. A different view of the case subjects is shown in Figure 2C, with risk scores by year leading up to the date of their suicidal event. As shown, there is a noticeable increase in the scores in the 3–4 years in advance of the index event.

DISCUSSION

Using data commonly available in EHRs, our models were able to identify nearly half of all suicides and suicidal behaviors with 90% specificity, an average of 3–4 years in advance. The increasingly widespread adoption of EHRs provides unprecedented opportunities for practical application of precision medicine, including the possibility of risk prediction for major health outcomes. Suicide attempts and suicide deaths are major sources of morbidity and mortality, and prior research has demonstrated that clinicians are generally unable to predict these outcomes (4, 22, 23).

TABLE 2. Relative Risk Scores for Demographic Factors, by Gender

Feature	Control Subjects Per 10,000	Case Subjects Per 10,000	Risk Score ^a	95% CI
Men				
Age group				
<25 years old	976.85	890.24	0.92	0.84–1.00
25–45 years old	2792.08	4206.29	1.52	1.45–1.59
45–65 years old	3461.33	3584.20	1.04	1.00–1.09
≥65 years old	2769.71	1319.27	0.48	0.45–0.52
Race/ethnicity				
Hispanic	758.80	1226.31	1.63	1.51–1.76
African American	585.40	829.46	1.43	1.30–1.57
White	7279.36	7141.58	0.99	0.95–1.02
Other	1062.12	661.42	0.63	0.57–0.70
Asian	314.33	141.22	0.45	0.36–0.56
Veteran status				
Veteran	1161.78	933.14	0.81	0.74–0.88
Marital status				
Separated	90.39	209.15	2.33	1.94–2.80
Divorced	394.02	698.96	1.79	1.61–1.98
Single	3608.09	5675.72	1.58	1.53–1.65
Partner	11.32	12.51	1.11	0.53–2.35
Widowed	232.53	194.85	0.84	0.70–1.02
Other/unknown	578.18	480.87	0.84	0.74–0.95
Married	5085.46	2727.92	0.54	0.51–0.57
Women				
Age group				
<25 years old	831.86	1192.60	1.43	1.31–1.56
25–45 years old	3425.10	3918.86	1.14	1.09–1.20
45–65 years old	3431.29	3178.78	0.93	0.88–0.98
≥65 years old	2311.75	1709.76	0.74	0.69–0.80
Race/ethnicity				
Hispanic	736.28	1243.87	1.69	1.55–1.84
African American	680.62	880.52	1.29	1.17–1.43
White	7071.02	7035.22	0.99	0.96–1.03
Asian	393.04	231.83	0.59	0.49–0.72
Other	1119.04	608.56	0.54	0.48–0.61
Veteran status				
Veteran	45.78	53.50	1.17	0.78–1.75
Marital status				
Separated	114.00	276.42	2.42	2.03–2.90
Single	3361.71	5198.40	1.55	1.48–1.62
Divorced	583.35	896.12	1.54	1.39–1.70
Widowed	648.95	635.31	0.98	0.87–1.10
Other/unknown	593.89	497.10	0.84	0.73–0.96
Partner	12.16	8.92	0.73	0.27–1.96
Married	4685.95	2487.74	0.53	0.50–0.56

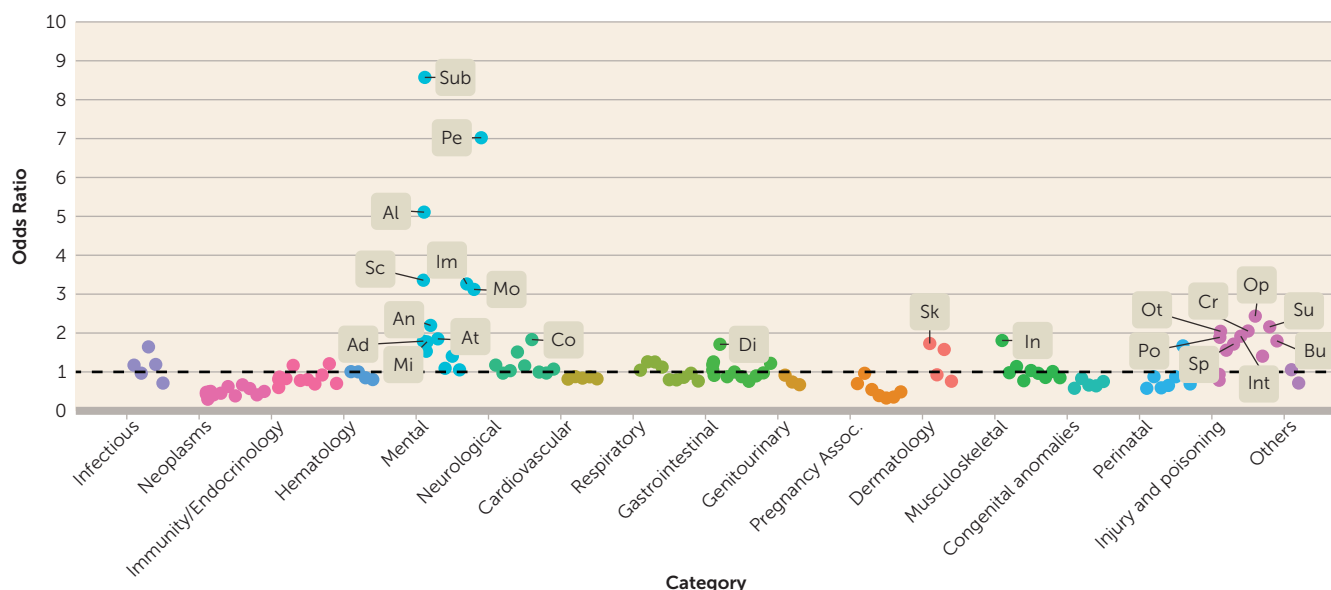
^a The ratio of the likelihood of a case subject having a specific demographic feature compared with a control subject having the same feature (e.g., a man with suicidal behavior is 2.33 times more likely to have a separated marital status than a control male subject).

Our empirical, data-driven modeling approach has a number of key strengths. First, we are able to examine both established and previously unsuspected risk factors by leveraging the full phenotypic breadth offered by the EHR. The most highly weighted variables found by the model were psychological conditions and substance abuse, corresponding to findings from prior epidemiologic studies (1) and supporting the validity of our model. However, rather than limiting the risk profile to known or hypothesized risk variables, the naive Bayesian classifier model assigns risk weights

to all of the coded variables in the health record: diagnoses relating to fractures, wounds, infections, and injuries, as well as certain chronic conditions such as hepatitis, were also associated with elevated suicide risk. Second, this modeling approach assigns separate risk scores not just to general categories of disease (e.g., prior psychiatric disease) but rather to each individual diagnostic, laboratory, and prescription code, allowing for greater insights into which specific codes are associated with higher risk. Third, the longitudinal nature of the EHR allows us to estimate the cumulative effect of risk factors over time and to identify risk profiles well in advance of the index event. Fourth, the model can be tailored to the specific setting and coding environment in which it is implemented: variables' weights can be calibrated using retrospective data from the target site and the selected thresholds modified according to the costs associated with false positive and false negative predictions.

The idea of mining EHR data for suicide risk prediction has been explored in several previous studies. Baca-Garcia and colleagues (14) used several data-mining strategies to reanalyze data from a study of clinician decision-making in the emergency department regarding 509 individuals who attempted suicide. Predictions were based on 139 features that were reduced to five in the best-performing model. The study was based on a smaller sample and number of features compared with the model described in our study. Ilgen and colleagues (16) applied recursive partitioning to a sample of Veteran's Affairs patients treated for depression in order to predict risk of suicide ascertained from the National Death Index. They identified 1,892 deaths by suicide using ICD-9 codes and specified a set of eight candidate predictors derived from treatment records, but model performance metrics were not reported. Using a different approach, Poulin and colleagues (15) derived a machine-learning algorithm

based on clinical notes in the Veteran's Affairs medical record to distinguish three groups (N=70 in each group): those who received mental health treatment and did or did not die by suicide and a control group of those who neither used mental health services nor died by suicide. They achieved an overall classification accuracy of up to 67% (compared with up to 94% in our study). Tran and colleagues (23) applied a penalized regression modeling approach to coded EHR data for 7,399 patients who underwent suicide risk assessment by clinicians and were followed for 180 days.

FIGURE 1. Summary View of Odds Ratios by Diagnostic Category^a

^a ICD-9 diagnostic codes were grouped together using Clinical Classification Software by the Healthcare Cost and Utilization Project. Odds ratios were calculated for each Clinical Classification Software category. The most prominent categories associated with suicidal behavior were related to mental disorders: "substance-related disorders," "personality disorders," "alcohol-related disorders," "schizophrenia and other psychotic disorders," "open wounds," and "superficial injury." Alongside these more established categories of risk, low to medium levels of risk were also found to be associated with other diagnostic categories (see the Discussion section in the article and Table S2 in the online data supplement). Abbreviations: Ad: adjustment disorders (650.); Al: alcohol-related disorders (660.); An: anxiety disorders (651.); At: attention deficit conduct and disruptive behavior disorders (652.); Bu: burns (240.); Co: coma, stupor, and brain damage (85.); Cr: crushing injury or internal injury (234.); Di: disorders of teeth and jaw (136.); Im: impulse control disorders not elsewhere classified (656.); In: infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease) (201.); Int: intracranial injury (233.); Mi: miscellaneous mental disorders (670.); Mo: mood disorders (657.); Op: open wounds; Ot: other injuries and conditions due to external causes (244.); Pe: personality disorders (658.); Po: poisoning (241–243.); Sc: schizophrenia and other psychotic disorders (659.); Sk: skin and subcutaneous tissue infections (197.); Sp: spinal cord injury (227.); Su: superficial injury, contusion (239.); Sub: substance-related disorders (661).

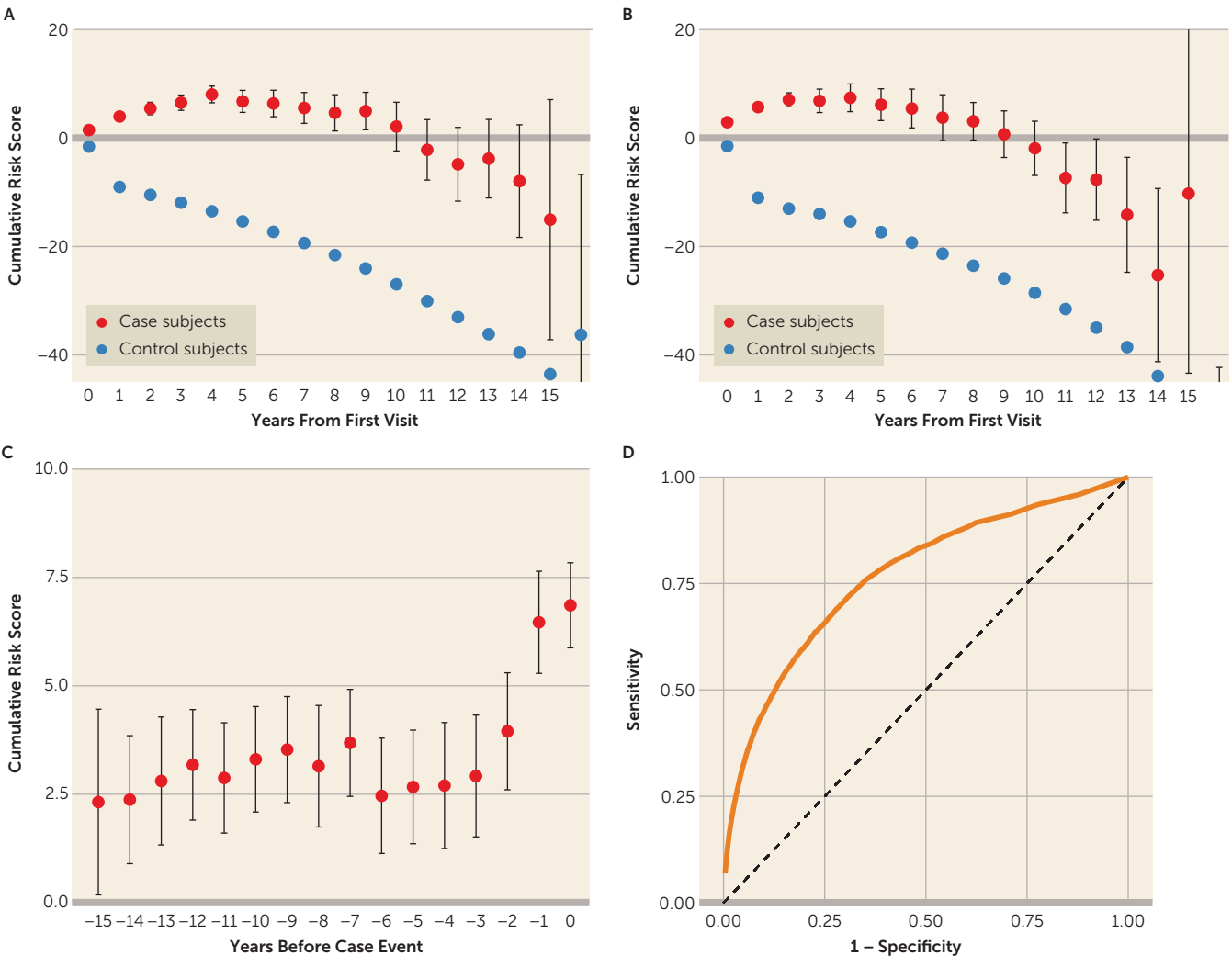
Compared with clinician predictions based on an 18-point suicide assessment checklist, the EHR-based model was more successful in stratifying risk at 30–180 days. Compared with our results, this study focused only on subjects who underwent screening for suicide and predicted risk over a relatively short time frame. Finally, in a study of U.S. Army personnel, Kessler and colleagues (24) applied machine-learning approaches to EHR and administrative data records to predict suicides in the year following a psychiatric hospitalization. In their best-fitting model, 53% of the suicides occurred after the 5% of hospitalizations with the highest predicted suicide risk, but their use of the rich administrative and personal data available in the comprehensive military database limits generalization to standard health care systems. Our model has advantages over these prior studies. First, it uses data readily and widely available in today's EHR systems. Second, it includes a broader range of variables and a larger and more diverse sample than in prior studies. Third, the model incorporates time-varying data, allowing us to determine the timeliness of predictions.

Our results should be interpreted in the context of several limitations. We used 15 years of data from an urban-regional data set including hospital admissions, observation stays, and encounters in emergency and outpatient settings. This data set excluded any patient visits outside this geographical area, time period, or network of hospitals, thus potentially losing

some patients to follow up. As a result, certain codes that may have assisted in identifying high-risk patients may not be recorded in the data set. Furthermore, some of the excluded visits could have been for suicidal behavior that was not recorded in this data set, meaning that these individuals may have been incorrectly classified as control subjects or correctly classified as case subjects but given incorrect onset times. That said, our goal was to determine whether data commonly available in today's real-world EHRs can be used to effectively predict suicidal behavior in a sensitive, specific, and timely manner. Our case definition includes codes that we validated to be highly specific for suicidal behavior. Nevertheless, variability in coding practices could limit the generalizability of our model in some settings. For example, some studies have supported the sensitivity, specificity, and predictive value of E-codes and other suicide-related codes (25, 26), while others have not (27–29).

Alternative approaches (e.g., neural networks, support vector machines, and other machine-learning approaches) might yield comparable (or possibly greater) predictive accuracy but are typically "black box" models that are difficult to interpret. Rather than using complex model selection or data reduction procedures to identify a subset of predictive variables, we demonstrate that the straightforward approach of using all the clinical data available for a patient does very well. In addition, to maximize the generalizability of our tool

FIGURE 2. Cumulative Risk Score Over Time^a



^a Times are shown either relative to the first visit (plots A and B) or to the last index visit (plot C). Each vertical line shows the mean with 95% confidence interval. Plot A shows women's risk scores over time for case subjects (red) compared with control subjects (blue). Plot B shows men's risk scores over time for case subjects (red) compared with control subjects (blue). Plot C shows case subjects' risk scores in the years preceding the index visit. Plot D shows the receiver operating characteristic curve for the naive Bayesian classifier model, showing the overall predictive performance of the model across all test subjects. Overall area under the curve was 0.77. The model was marginally superior in the female cohort compared with the male cohort, with an area under the curve of 0.77 compared with 0.76, respectively.

to other health care systems, we deliberately use codified data that are readily available in EHRs rather than relying on complex text-mining approaches (e.g., natural language processing) that can be more difficult to implement and more sensitive to local documentation practices.

Several aspects of our risk-prediction approach could be enhanced in future research. For example, currently the model yields low to moderate positive predictive value (5%–7% at 95% specificity), although this is to be expected in a condition with a very low baseline probability (1.2%) and represents a 4.5- to 6.5-fold enrichment of suicide risk prediction compared with the base rate. Applying our approach to patient subsamples with high prevalence of suicide risk (e.g., patients in psychiatric care) could enrich the base rate and improve the model's positive predictive value.

TABLE 3. Overall Model Performance by Gender on the Validation Testing Cohorts

Group and Specificity	Accuracy ^a	Sensitivity	Positive Predictive Value	Negative Predictive Value
Men				
90% specificity	0.89	0.44	0.05	0.99
95% specificity	0.94	0.31	0.07	0.99
Women				
90% specificity	0.90	0.46	0.03	1.00
95% specificity	0.95	0.34	0.05	0.99
Overall				
90% specificity	0.90	0.45	0.04	0.99
95% specificity	0.94	0.33	0.06	0.99

^a Data represent the percentage of correct predictions made out of all predictions (true positive plus true negative divided by total subjects).

Additionally, our model currently captures the risk associated with each feature (e.g., diagnoses) separately. More complex models incorporating combinations or interactions of features may improve diagnostic accuracy. With appropriate integration into the clinical workflow, this model can assist already overloaded clinicians to identify high-risk patients who require further in-depth screening. Although a statistical model is never a substitute for clinical evaluation, an early warning system based on our approach may provide a mechanism for identifying patients who are at elevated statistical risk of future suicidal behavior and therefore require screening. This is especially important, since screening rates in clinical settings remain far below desired levels (30).

After further refinement, we envision our models being used as a dashboard element in the EHR at the point of care. In ongoing work, we are designing a user-friendly visualization suitable for incorporation into the EHR interface. For each patient, the system could present the clinician with a high-level summary of short-term, medium-term, and long-term suicide risks, alongside a visualization of the patient's longitudinal history and a list of the most prominent risk factors for that given individual. It will also provide carefully worded messages for clinicians, explaining what the risk alert means and what information it is based on. These messages will be crafted in consultation with clinicians in order to avoid misunderstandings and ensure seamless integration into the clinical workflow. It is important to reiterate that our approach is designed as a screening tool for decision support rather than a specific quantitative prediction of suicide risk. Given the imperfect predictive value of any automated model, it would be inappropriate (and medicolegally imprudent) to base clinical decisions (such as hospitalization) solely on model readouts. Rather, we envision an alert system by which patients exceeding thresholds of predicted risk could be flagged as at relatively higher risk to encourage clinicians to conduct more targeted assessments of suicide risk.

In conclusion, these findings suggest that the vast quantities of longitudinal data accumulating in electronic health information systems present a largely untapped opportunity for improving medical screening and diagnosis. Beyond the direct implications for prediction of suicide risk, this general approach has far-reaching implications for the automated screening of a wide range of clinical conditions for which longitudinal historical information may be beneficial for estimating clinical risk.

AUTHOR AND ARTICLE INFORMATION

From the Predictive Medicine Group, Boston Children's Hospital Informatics Program, Boston; the Technion, Israeli Institute of Technology, Haifa, Israel; the Partners Research Information Systems and Computing, Boston; the Department of Psychiatry, Massachusetts General Hospital, Boston; the Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston; the Department of Psychology, Harvard University, Boston; and Harvard Medical School, Boston.

Address correspondence to Dr. Smoller (jsmoller@hms.harvard.edu).

Drs. Smoller and Reis contributed equally to this article.

Supported by a gift from the Tommy Fuss Fund.

Dr. Perlis has served on scientific advisory boards for or as a consultant to Genomind, Perfect Health, Proteus Biomedical, PsyBrain, and RID Ventures; and he has also received support from Healthracious, Massachusetts General Hospital, and Pfizer. All other authors report no financial relationships with commercial interests.

Received Jan. 20, 2016; revision received May 2, 2016; accepted May 20, 2016; published online Sept. 9, 2016.

REFERENCES

1. Nock MK, Borges G, Bromet EJ, et al: Suicide and suicidal behavior. *Epidemiol Rev* 2008; 30:133–154
2. Busch KA, Fawcett J, Jacobs DG: Clinical correlates of inpatient suicide. *J Clin Psychiatry* 2003; 64:14–19
3. Nock MK, Borges G, Bromet EJ, et al: Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry* 2008; 192:98–105
4. Nock MK, Park JM, Finn CT, et al: Measuring the suicidal mind: implicit cognition predicts suicidal behavior. *Psychol Sci* 2010; 21: 511–517
5. Glenn CR, Nock MK: Improving the short-term prediction of suicidal behavior. *Am J Prev Med* 2014; 47(suppl 2):S176–S180
6. Reis BY, Kohane IS, Mandl KD: An epidemiological network model for disease outbreak detection. *PLoS Med* 2007; 4:e210
7. Wang JF, Reis BY, Hu MG, et al: Area disease estimation based on sentinel hospital records. *PLoS One* 2011; 6:e23428
8. Cami A, Reis BY: Concordance and predictive value of two adverse drug event data sets. *BMC Med Inform Decis Mak* 2014; 14:74
9. Reis BY, Brownstein JS: Measuring the impact of health policies using Internet search patterns: the case of abortion. *BMC Public Health* 2010; 10:514
10. Reis BY, Mandl KD: Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. *Ann Emerg Med* 2004; 44:235–241
11. Castro VM, Gallagher PJ, Clements CC, et al: Incident user cohort study of risk for gastrointestinal bleed and stroke in individuals with major depressive disorder treated with antidepressants. *BMJ Open* 2012; 2:e000544
12. Clements CC, Castro VM, Blumenthal SR, et al: Prenatal antidepressant exposure is associated with risk for attention-deficit hyperactivity disorder but not autism spectrum disorder in a large health system. *Mol Psychiatry* 2015; 20:727–734.
13. Reis BY, Kohane IS, Mandl KD: Longitudinal histories as predictors of future diagnoses of domestic abuse: modelling study. *BMJ* 2009; 339:b3677
14. Baca-García E, Perez-Rodriguez MM, Basurte-Villamor I, et al: Using data mining to explore complex clinical decisions: a study of hospitalization after a suicide attempt. *J Clin Psychiatry* 2006; 67: 1124–1132
15. Poulin C, Shiner B, Thompson P, et al: Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS One* 2014; 9:e85733
16. Ilgen MA, Downing K, Zivin K, et al: Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. *J Clin Psychiatry* 2009; 70:1495–1500
17. Nalichowski R, Keogh D, Chueh HC, et al: Calculating the benefits of a Research Patient Data Repository. *AMIA Annu Symp Proc* 2006; 1044
18. Ting SA, Sullivan AF, Boudreaux ED, et al: Trends in US emergency department visits for attempted suicide and self-inflicted injury, 1993–2008. *Gen Hosp Psychiatry* 2012; 34:557–565
19. Kononenko I: Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* 2001; 23:89–109
20. US Public Health Service: Suicide: Risk and Protective Factors. Atlanta, Centers for Disease Control and Prevention, 1999. <http://www.cdc.gov/violenceprevention/suicide/riskprotectivefactors.html>
21. Elixhauser A, Steiner C, Palmer L: Clinical Classifications Software (CCS)

22. Grove WM, Zald DH, Lebow BS, et al: Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess* 2000; 12:19–30
23. Tran T, Luo W, Phung D, et al: Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry* 2014; 14:76
24. 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
25. Patrick AR, Miller M, Barber CW, et al: Identification of hospitalizations for intentional self-harm when E-codes are incompletely recorded. *Pharmacoepidemiol Drug Saf* 2010; 19:1263–1275
26. Callahan ST, Fuchs DC, Shelton RC, et al: Identifying suicidal behavior among adolescents using administrative claims data. *Pharmacoepidemiol Drug Saf* 2013; 22:769–775
27. Walkup JT, Townsend L, Crystal S, et al: A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. *Pharmacoepidemiol Drug Saf* 2012; 21(suppl 1):174–182
28. Haerian K, Salmasian H, Friedman C: Methods for identifying suicide or suicidal ideation in EHRs. *AMIA Annu Symp Proc* 2012; 2012:1244–1253
29. Lu CY, Stewart C, Ahmed AT, et al: How complete are E-codes in commercial plan claims databases? *Pharmacoepidemiol Drug Saf* 2014; 23:218–220
30. O'Connor E, Gaynes B, Burda B, et al: Screening for suicide risk in primary care: a systematic evidence review for the US Preventive Services Task Force. Rockville (Md): Agency for Healthcare Research and Quality (US); 2013 Apr Report No: 13-05188-EF-1. US Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews.