

Confidential. Do not distribute. Pre-embargo material.

JAMA Psychiatry | Original Investigation

Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing

Thomas H. McCoy Jr, MD; Victor M. Castro, MS; Ashlee M. Roberson, BA;
Leslie A. Snapper, BS; Roy H. Perlis, MD, MS

IMPORTANCE Suicide represents the 10th leading cause of death across age groups in the United States (12.6 cases per 100 000) and remains challenging to predict. While many individuals who die by suicide are seen by physicians before their attempt, they may not seek psychiatric care.

OBJECTIVE To determine the extent to which incorporating natural language processing of narrative discharge notes improves stratification of risk for death by suicide after medical or surgical hospital discharge.

DESIGN, SETTING, AND PARTICIPANTS In this retrospective health care use study, clinical data were analyzed from individuals with discharges from 2 large academic medical centers between January 1, 2005, and December 31, 2013.

MAIN OUTCOMES AND MEASURES The primary outcome was suicide as a reported cause of death based on Massachusetts Department of Public Health records. Regression models for prediction of death by suicide or accidental death were compared relying solely on coded clinical data and those using natural language processing of hospital discharge notes.

RESULTS There were 845 417 hospital discharges represented in the cohort, including 458 053 unique individuals. Overall, all-cause mortality was 18% during 9 years, and the median follow-up was 5.2 years. The cohort included 235 (0.1%) who died by suicide during 2.4 million patient-years of follow-up. Positive valence reflected in narrative notes was associated with a 30% reduction in risk for suicide in models adjusted for coded sociodemographic and clinical features (hazard ratio, 0.70; 95% CI, 0.58-0.85; $P < .001$) and improved model fit ($\chi^2_2 = 14.843$, $P < .001$ by log-likelihood test). The C statistic was 0.741 (95% CI, 0.738-0.744) for models of suicide with or without inclusion of accidental death.

CONCLUSIONS AND RELEVANCE Multiple clinical features available at hospital discharge identified a cohort of individuals at substantially increased risk for suicide. Greater positive valence expressed in narrative discharge summaries was associated with substantially diminished risk. Automated tools to aid clinicians in evaluating these risks may assist in identifying high-risk individuals.

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2016.2172
Published online September 14, 2016.

[+ Author Video Interview and JAMA Report Video](#)

[+ Supplemental content](#)

Author Affiliations: Center for Experimental Drugs and Diagnostics, Center for Human Genetic Research and Department of Psychiatry, Massachusetts General Hospital, Boston.

Corresponding Author: Roy H. Perlis, MD, MS, Center for Experimental Drugs and Diagnostics, Center for Human Genetic Research and Department of Psychiatry, Massachusetts General Hospital, 185 Cambridge St, Sixth Floor, Simches Research Building, Boston, MA 02114 (rperlis@partners.org).

Confidential. Do not distribute. Pre-embargo material.

Suicide represents one of the most dreaded outcomes of psychiatric illness. With 41 149 completed suicides reported by the US Centers for Disease Control and Prevention in 2013, suicide is the 10th leading cause of death in the United States (12.6 cases per 100 000) and the second leading cause among individuals aged 15 to 24 years (10.9 cases per 100 000).^{1,2} While epidemiological investigations provide some guidance regarding demographic features, symptoms, or diagnoses associated with greatest risk, such studies tend to focus on suicide attempts among epidemiological cohorts assessed by diagnostic interviews.³⁻⁵ Conversely, small-scale studies^{6,7} have evaluated biomarkers or psychosocial features associated with suicide attempt among high-risk individuals. What is to our knowledge the sole larger study⁸ to examine death by suicide assessed a cohort of US soldiers after psychiatric hospital discharge, reporting on 68 deaths.

The challenge for any of these risk models is automation, namely, efficiently integrating data to facilitate identification of high-risk individuals. While potential interventions exist,⁹⁻¹² it may not be feasible to offer interventions to everyone. To facilitate such efforts, we sought to examine a large, generalizable group of individuals hospitalized over 9 years at 1 of 2 academic medical centers by coupling electronic health records (EHRs) with long-term death certificate data. Because the period after hospital discharge carries elevated risk for suicide, hospital discharge is a moment with increased means of intervention and increased risk of event, making it ideal for risk stratification.¹³ This approach also allows investigation of more representative patient cohorts with greater comorbidity than traditional epidemiological studies while investigating predictors more readily translatable than biomarker or military cohort studies. In particular, focusing on general hospital cohorts allows characterization of individuals typically neglected in the psychiatric literature, namely, those who may not seek psychiatric treatment but who are nonetheless at elevated risk for suicide. If one goal of risk stratification is ultimately translation to develop interventions for preventing suicide, understanding this population may be critical. Almost half of the patients contemplating suicide will see their primary care physician in the month before completion, whereas only one-fifth will see a mental health professional.¹⁴

Recognizing that coded claims data capture only some elements of clinical presentation, we also examined the incremental benefit in prediction from incorporating narrative discharge summaries. In previous work, our group¹⁵ has demonstrated that a natural language processing method that aggregates words conveying positive or negative emotion (ie, valence) improved prediction of all-cause mortality and hospital readmission.¹⁶ Rather than exhaustively mining narrative notes, we sought to examine whether a simple, previously published, and easily scaled method of characterizing notes would improve outcome prediction.

Methods

Cohort and Outcome Derivation

The study cohort was defined as all hospital discharges between January 1, 2005, and December 31, 2013, at Massachusetts General Hospital and Brigham and Women's Hospital, Bos-

Key Points

Question To what extent does incorporating natural language processing of narrative discharge notes improve stratification of risk for death by suicide after medical or surgical hospital discharge?

Findings In this health care use study, positive valence reflected in narrative notes was associated with a 30% reduction in risk for suicide in adjusted models and improved model fit.

Meaning Automated tools to aid clinicians in evaluating these risks may assist them in identifying high-risk individuals.

ton. We extracted sociodemographic data, billing codes, and narrative hospital discharge notes for all patients from the hospitals' EHRs. Data were managed using server software (i2b2, version 1.6; Informatics for Integrating Biology and the Bedside).¹⁷⁻¹⁹

The EHRs include patients' vital status based on the US Social Security Death Index (updated monthly) but not cause or circumstances of death. The specific features examined in regression models were standard sociodemographic and clinical cohort descriptors selected by our group a priori, as well as those shown in prior work to be associated with readmission.^{15,20}

The study protocol was reviewed and approved by the Partners Human Research Committee at Massachusetts General Hospital. As a retrospective health care use study, the requirement for informed consent of participants was waived.

Outcome

For all deceased individuals, we queried Massachusetts Department of Public Health records for all individuals in the EHR-defined cohort to determine cause of death. Suicide or accidental death was determined based on reported (coded) cause of death. The primary outcome of interest was suicide as a reported cause of death. However, recognizing that some suicide may be misclassified as accidental death (eg, when circumstances of death cannot be confirmed), we also examined death by either suicide or accidental death as a secondary composite outcome. The analyses of the composite outcome are presented along with the primary results (ie, strictly defined suicide as cause of death) and are noted in the text when they differ substantially from the narrower suicide phenotype.

Development and Application of Natural Language Processing Tools

Our group has previously described the application of a tool for characterizing positive and negative valence expressed in narrative notes.¹⁵ In brief, we used a curated list of almost 3000 subjectively valence-conveying terms to score each narrative note using an open-source opinion mining tool (Pattern, version 2.6; Python).^{21,22} For example, terms with positive valence include *glad*, *pleasant*, and *lovely*. Those with negative valence include *gloomy*, *unfortunate*, and *sad*. The full list is publicly available in the Pattern library. The lexicon has previously been validated as a manually annotated corpus.^{23,24} Each word includes a polarity score (negativity to positivity,

Confidential. Do not distribute. Pre-embargo material.

scored as -1 to 1), as well as subjectivity (not subjective to subjective, scored as 0 to 1). This open-source implementation allows direct inspection of the method and ready replication. As in our prior work, we scored each document based on the mean value for all recognized words after accounting for preceding negation (eg, to distinguish *happy* from *not happy*) and multiplying by the intensity modifier for relevant adjectives. For illustrative purposes, the eTable in the Supplement lists terms associated with positive valence appearing in at least 10% of a random sample of 5000 notes.

Statistical Analysis

The primary analysis used survival methods, with the results censored at the end of follow-up or at death. After confirming that proportional hazards assumptions were met for the primary model, Cox regression was used to examine risk associated with predictors individually and in aggregate. In light of the scale of data, we randomly selected a single hospitalization from individuals with multiple hospitalizations to maximize computability, while minimizing bias. Sensitivity analysis using 2 alternate approaches incorporated all clustered observations: Cox proportional hazards regression and mixed-effects survival with per-individual intercept.^{25,26}

For the primary outcome and the secondary composite outcome, 2 regression models were fit to explain the clinical outcomes. The first model used only coded clinical data,

including age, sex, self-reported race, recent health care use (including outpatient psychiatric visits, overall outpatient visits, and emergency department visits), and overall medical morbidity as estimated by the Charlson Comorbidity Index.^{27,28} Given the pronounced skew in continuous measures (Tables 1, 2, 3, and 4), models used logarithmic (variable plus 1) transformation.

The second model included all features from the first model and then added aggregate measures of positive and negative valence. These measures are considered separately rather than as a continuum because previous analyses indicated that they are not inherently opposing (ie, some notes include terms reflecting both positive and negative valence). As a sensitivity analysis to examine the potential effect of hypothetical missing data due to documentation delay, the 5 suicides occurring within 1 week of discharge were censored and the analysis rerun, with essentially identical results.

Models were characterized in terms of standard measures of discrimination, including the C statistic, as well as calibration. The C statistic was calculated with 10-fold cross-validation to minimize optimism. Continuous net reclassification improvement was calculated as a complementary measure of improvement in model fit.²⁹ To aid in understanding the potential application of risk models, we also applied decision curve analysis (<http://decisioncurveanalysis>

Table 1. Description of Hospitalized Cohort, Contrasting Individuals Who Did or Did Not Die by Suicide During Follow-up

| Variable | No. (%) | | |
|--|-----------------------|----------------------------|--------------------------------|
| | Overall (N = 458 053) | Death by Suicide (n = 235) | No Suicide Death (n = 457 818) |
| Male sex | 189 529 (41.4) | 155 (66) | 189 374 (41.3) |
| White race | 346 340 (75.7) | 202 (86) | 346 138 (75.6) |
| Primary psychiatric diagnosis at admission | 5782 (1.3) | 16 (6.8) | 5766 (1.3) |

Table 3. Description of Hospitalized Cohort, Contrasting Individuals Who Did or Did Not Die by Suicide or Accidental Death (Composite Outcome) During Follow-up

| Variable | No. (%) | | |
|--|-----------------------|---|--|
| | Overall (N = 458 053) | Death by Suicide or Accidental Death (n = 2026) | No Suicide or Accidental Death (n = 456 027) |
| Male sex | 189 529 (41.6) | 1254 (61.9) | 188 275 (41.1) |
| White race | 346 340 (75.9) | 1661 (82) | 344 679 (75.2) |
| Primary psychiatric diagnosis at admission | 5782 (1.3) | 92 (4.5) | 5690 (1.2) |

Table 2. Additional Description of Hospitalized Cohort, Contrasting Individuals Who Did or Did Not Die by Suicide During Follow-up

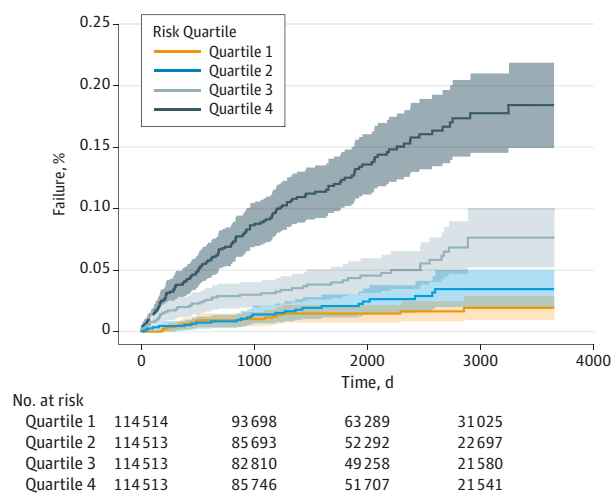
| Variable | Overall (N = 458 053) | | | | No Suicide Death (n = 457 818) | | | | Death by Suicide (n = 235) | | | |
|---|-----------------------|--------|-----------------|---------|--------------------------------|--------|-----------------|---------|----------------------------|--------|-----------------|---------|
| | 25th Percentile | Median | 75th Percentile | Maximum | 25th Percentile | Median | 75th Percentile | Maximum | 25th Percentile | Median | 75th Percentile | Maximum |
| Age at admission, y | 36 | 53 | 67 | 90 | 36 | 53 | 67 | 90 | 36 | 47 | 56.5 | 86 |
| Age-adjusted Charlson Comorbidity Index | 0 | 2 | 4 | 27 | 0 | 2 | 4 | 27 | 0 | 2 | 4 | 27 |
| Psychiatric visits in 12 mo before admission | 0 | 0 | 0 | 135 | 0 | 0 | 0 | 55 | 0 | 0 | 0 | 135 |
| Prior psychiatric visits (ever) | 0 | 0 | 0 | 1068 | 0 | 0 | 1 | 191 | 0 | 0 | 0 | 1068 |
| Outpatient visits (any) in 12 mo before admission | 1 | 3 | 8 | 364 | 1 | 2 | 7 | 63 | 1 | 3 | 8 | 364 |
| Emergency department visits in 12 mo before admission | 0 | 0 | 0 | 67 | 0 | 0 | 1 | 12 | 0 | 0 | 0 | 67 |
| Positive valence | 1.37 | 2.03 | 3.34 | 90.97 | 1.41 | 1.93 | 2.66 | 14.12 | 1.37 | 2.03 | 3.34 | 90.97 |
| Negative valence | 0.74 | 1.13 | 1.78 | 43.24 | 0.86 | 1.24 | 1.83 | 5.60 | 0.74 | 1.13 | 1.78 | 43.24 |

Confidential. Do not distribute. Pre-embargo material.

Table 4. Additional Description of Hospitalized Cohort, Contrasting Individuals Who Did or Did Not Die by Suicide or Accidental Death (Composite Outcome) During Follow-up

| Variable | Overall (N = 458 053) | | | | No Suicide or Accidental Death (n = 456 027) | | | | Death by Suicide or Accidental Death (n = 2026) | | | |
|---|-----------------------|--------|-----------------|---------|--|--------|-----------------|---------|---|--------|-----------------|---------|
| | 25th Percentile | Median | 75th Percentile | Maximum | 25th Percentile | Median | 75th Percentile | Maximum | 25th Percentile | Median | 75th Percentile | Maximum |
| Age at admission, y | 36 | 53 | 67 | 90 | 44 | 56 | 76 | 90 | 36 | 53 | 67 | 90 |
| Age-adjusted Charlson Comorbidity Index | 0 | 2 | 4 | 27 | 1 | 3 | 6 | 21 | 0 | 2 | 4 | 27 |
| Psychiatric visits in 12 mo before admission | 0 | 0 | 0 | 135 | 0 | 0 | 0 | 55 | 0 | 0 | 0 | 135 |
| Prior psychiatric visits (ever) | 0 | 0 | 0 | 1068 | 0 | 0 | 0 | 213 | 0 | 0 | 0 | 1068 |
| Outpatient visits (any) in 12 mo before admission | 1 | 3 | 8 | 364 | 0 | 2 | 6 | 75 | 1 | 3 | 8 | 364 |
| Emergency department visits in 12 mo before admission | 0 | 0 | 0 | 67 | 0 | 0 | 1 | 23 | 0 | 0 | 0 | 67 |
| Positive valence | 1.37 | 2.03 | 3.34 | 90.97 | 1.29 | 1.79 | 2.56 | 26.00 | 1.37 | 2.03 | 3.34 | 90.97 |
| Negative valence | 0.74 | 1.13 | 1.78 | 43.24 | 0.79 | 1.12 | 1.64 | 13.19 | 0.74 | 1.13 | 1.78 | 43.24 |

Figure 1. Kaplan-Meier Curves for Time to Death by Suicide Among 458 053 Individuals With at Least 1 Hospital Discharge by Predicted Risk Quartile



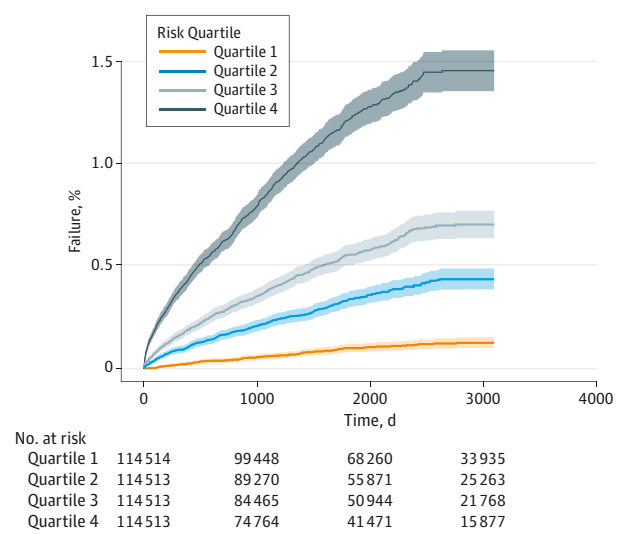
The axes are rescaled inside the figure to improve interpretability.

.org) as implemented in a software package (Stata, stdca; StataCorp LP) using the default settings that assume no direct harm from testing itself.^{30,31}

Results

There were 845 417 hospital discharges represented in the cohort, including 458 053 unique individuals. Overall, all-cause mortality was 18% for 9 years, and the median follow-up was 5.2 years. For the cohort as a whole, 235 (0.1%) died by suicide during follow-up (Figure 1). These deaths included 77 in the first year of follow-up (11 who died within 30 days and 50 who died within 180 days), 46 in the second

Figure 2. Kaplan-Meier Curves for Time to Death by Suicide or Accidental Death Among 458 053 Individuals With at Least 1 Hospital Discharge by Predicted Risk Quartile



The axes are rescaled inside the figure to improve interpretability.

year of follow-up, and 192 within the first 5 years. The features of the cohort, contrasting individuals who did vs did not die by suicide, are summarized in Tables 1 and 2. A total of 2026 individuals (0.4%) died by either suicide or accidental death during follow-up (Figure 2), including 712 in the first year of follow-up, of which 178 occurred within the first 30 days after discharge. The features of the cohort, contrasted by the composite outcome, are summarized in Tables 3 and 4.

Table 5 lists adjusted hazard ratios for suicide from Cox proportional hazards regression using the coded data only (model 1). In the fully adjusted models, significant features

Confidential. Do not distribute. Pre-embargo material.

Table 5. Cox Proportional Hazards Regression Model of Suicide Risk Without (Model 1) and With (Model 2) Inclusion of Natural Language Processing of Narrative Discharge Summary

| Variable | Adjusted Hazard Ratio (95% CI) ^a | |
|---|---|---------------------|
| | Model 1 | Model 2 |
| Male sex | 2.80 (2.11-3.71) | 2.57 (1.94-3.41) |
| White race | 2.27 (1.56-3.30) | 2.23 (1.53-3.24) |
| Age at admission, y | 0.80 (0.65-0.97) | 0.79 (0.65-0.96) |
| Age-adjusted Charlson Comorbidity Index | 0.95 (0.74-1.21) | 0.89 (0.69-1.15) |
| Primary psychiatric diagnosis at admission | 1.63 (0.93-2.85) | 1.39 (0.79-2.45) |
| Psychiatric visits in 12 mo before admission | 1.18 (1.07-1.31) | 1.17 (1.06-1.30) |
| Prior psychiatric visits (ever) | 1.15 (1.01-1.31) | 1.14 (1.01-1.30) |
| Outpatient visits (any) in 12 mo before admission | 0.79 (0.67-0.94) | 0.84 (0.71-1.00) |
| Emergency department visits in 12 mo before admission | 2.37 (1.83-3.06) | 2.19 (1.68-2.84) |
| Positive valence | NA | 0.70 (0.58-0.85) |
| Negative valence | NA | 1.15 (0.99-1.32) |

Abbreviation: NA, not applicable.

^a Adjusted for all other terms in the model.

associated with greater suicide risk included white race, male sex, and more emergency department visits and psychiatric outpatient visits in the 12 months before admission. Alternative models that included frailty-clustered Cox proportional hazards regression and mixed-effects regression with per-individual intercepts yielded similar results. Overall, the C statistic for the model using the coded data only was 0.737 (95% CI, 0.734-0.741). The results from the suicide and accidental death composite outcome are summarized in **Table 6** as model 1. The C statistic was 0.728 (95% CI, 0.727-0.728). In general, coefficients are similar across the 2 outcome definitions, with the exception of age and Charlson Comorbidity Index, which may reflect truly accidental death disproportionately affecting older and sicker patients rather than misclassified suicide.

In a second model (model 2) created by adding positive and negative valence extracted from discharge summaries by natural language processing to model 1, positive valence was also associated with the primary suicide risk outcome (hazard ratio, 0.70; 95% CI, 0.58-0.85) (Table 5). In other words, a 1-SD increase was associated with a 30% reduction in risk for suicide. Adding this feature to the coded data model improved model fit ($\chi^2_2 = 14.843$, $P < .001$ by log-likelihood test), and the resulting C statistic was 0.741 (95% CI, 0.738-0.744). Continuous net reclassification improvement was 0.10. From the lowest-risk quartile to the highest-risk quartile (Figure 1), the total numbers of suicides observed were 17 (7.2%), 25 (10.6%), 52 (22.1%), and 141 (60%), respectively (ie, the top 50% of risk identified 82.1% of suicides). To facilitate qualitative comparison, eFigure 1 in the Supplement shows model 1 adjacent to model 2.

Table 6. Cox Proportional Hazards Regression Model of Suicide or Accidental Death (Composite Outcome) Risk Without (Model 1) and With (Model 2) Inclusion of Natural Language Processing of Narrative Discharge Summary

| Variable | Adjusted Hazard Ratio (95% CI) ^a | |
|---|---|---------------------|
| | Model 1 | Model 2 |
| Male sex | 1.88 (1.71-2.06) | 1.82 (1.66-1.99) |
| White race | 1.37 (1.22-1.54) | 1.37 (1.22-1.53) |
| Age at admission, y | 1.21 (1.13-1.30) | 1.19 (1.11-1.28) |
| Age-adjusted Charlson Comorbidity Index | 1.39 (1.27-1.53) | 1.35 (1.23-1.48) |
| Primary psychiatric diagnosis at admission | 1.84 (1.47-2.31) | 1.69 (1.35-2.11) |
| Psychiatric visits in 12 mo before admission | 1.05 (1.01-1.10) | 1.05 (1.01-1.10) |
| Prior psychiatric visits (ever) | 1.16 (1.11-1.21) | 1.16 (1.10-1.21) |
| Outpatient visits (any) in 12 mo before admission | 0.63 (0.60-0.67) | 0.65 (0.62-0.69) |
| Emergency department visits in 12 mo before admission | 2.98 (2.71-3.26) | 2.84 (2.58-3.11) |
| Positive valence | NA | 0.80 (0.75-0.85) |
| Negative valence | NA | 1.06 (1.01-1.12) |

Abbreviation: NA, not applicable.

^a Adjusted for all other terms in the model.

Likewise, addition of the valence feature to the model of the composite outcome (suicide or accidental death) improved model fit ($\chi^2_2 = 45.269$, $P < .001$ by log-likelihood test), with continuous net reclassification improvement of 0.02. A 1-SD increase in positive valence was associated with approximately a 20% reduction in risk (hazard ratio, 0.80; 95% CI, 0.75-0.85) (model 2 in Table 6 and Figure 2). For this model, the C statistic was 0.732 (95% CI, 0.731-0.732). From the lowest-risk quartile to the highest-risk quartile, the total numbers of events were 107 (5.3%), 339 (16.7%), 534 (26.4%), and 1046 (51.6%), respectively (ie, an intervention targeting the top 50% of risk could prevent at most 78% of suicides or accidental deaths, while one targeting the highest-risk quartile could prevent at most 52%). As with the suicide-only models, eFigure 2 in the Supplement shows model 1 adjacent to model 2. Decision curve analysis (eFigure 3 in the Supplement) suggests that the greatest benefit (vs intervening in all individuals or not at all) is observed with interventions in individuals with predicted probability of an event between approximately 0.25% and 1%.

Discussion

In this cohort, which spans approximately 2.4 million patient-years, we developed a model based on coded clinical data that predicts suicide and accidental death among patients discharged from academic medical centers at a rate substantially exceeding chance, with an area under the curve of approximately 0.73. To our knowledge, postdischarge risk for

Confidential. Do not distribute. Pre-embargo material.

suicide death in large nonpsychiatric cohorts has not previously been modeled. We further found that addition of uncoded clinical data reflecting positive and negative valence available in general hospital discharge notes modestly improved prediction of these outcomes, suggesting more generally the potential usefulness of augmenting models using coded data only with concepts extracted from narrative clinical notes.

Among the coded data, we confirmed multiple clinical features previously associated with risk, such as male sex and white race, illustrating assay sensitivity and consistency with prior epidemiological investigations of suicide.^{3,10} Likewise, as anticipated, any psychiatric visit and prior psychiatric treatment were individually associated with substantial increase in risk. Because we sought to predict suicide death (and not unsuccessful attempt, as in most past efforts), our results are difficult to compare directly with prior studies.^{4,5} In one study⁸ examining suicide death among a cohort of active-duty soldiers discharged after psychiatric hospitalization, there were 68 deaths within 12 months. Machine learning models incorporating administrative and clinical details were effective in identifying high-risk hospitalizations. A systematic literature review did not identify any similar efforts among patients without psychiatric hospitalization.⁸

Notably, 115 of 235 (48.9%) suicide deaths in the present study occurred among individuals with no coded data reflecting psychiatric *International Classification of Diseases, Ninth Revision* diagnostic codes in this health system. This finding is consistent with prior reports that, while individuals who die by suicide often have contact with a health professional, the clinician is likely to not be a psychiatrist or therapist,¹⁴ which underscores the importance of psychiatric expertise in the general hospital setting. We cannot exclude the possibility that some of these individuals had sought treatment in the community (eg, in a private practice or another health system). Even if so, this treatment was not documented in hospital records and so presumably was not known by hospital staff. These omissions may reflect failure to inquire about psychiatric history, a potential target for intervention meriting further study.

In examining the hypothesis that narrative notes could improve prediction, we find that incorporating a simple natural language processing strategy improved the ability to estimate risk for suicide and accidental death. The particular strategy used herein identifies valence reflected in narrative notes. Such words may reflect individual symptoms, as well as clinician perception not reflected elsewhere.¹⁵ The modest improvement in discrimination (eg, continuous net reclassification improvement of 0.10 and 0.02) afforded by the natural language processing suggests that, while statistically significant, these additional features may not yet be clinically significant. As the eTable in the Supplement summarizes, there is likely substantial opportunity to better capture words reflecting emotion using more curated data sets and thereby improve discrimination further.³² It is also possible that some proxies among coded data could be identified. We describe this concept as “valence” for consistency with long-standing psychological literature,¹⁶ in the sense of words conveying positive or negative emotion rather than in reference to the di-

mensional feature of psychopathology described in the National Institute of Mental Health’s Research Domain Criteria. Still, despite the granularity of the method we applied, we capture a set of patient-level features not otherwise reflected in coded data and associated with a 30% change in hazard. We present a generalizable approach spanning approximately 2.4 million patient-years and multiple hospitals in anticipation that others will apply and advance such methods. Our results suggest that moving beyond coded data to capture details of clinical presentation may be valuable in efforts to develop more stratified or patient-specific interventions. This value may be in extracting codifiable facts from free text that were merely not coded at the time of the encounter²⁷ or in extracting metrics from notes that were never candidates for coding.³³ In either case, developing methods that do not rely on additional data gathering and entry by treatment teams but rather do more with the results of routine care should lower the barriers to translation of risk stratification models into clinical practice.

Developing a risk stratification model represents only part of a continuum from defining a clinical problem to validation and model presentation.^{34,35} For example, a previous effort⁷ to predict suicide among high-risk psychiatric patients based on a constellation of manually scored psychosocial features subsequently did not replicate,³⁶ illustrating the challenges in disseminating such models. In the absence of prior risk prediction models among medically hospitalized patients, our model using coded data only may be considered a starting point with which efforts at improvement may be compared. Assessing the clinical usefulness of such models requires additional considerations that take into account the risks and benefits of the specific intervention contemplated in the high-risk groups, which may be presented using decision curves.³⁰ The rarity of suicide and accidental death precludes reliable estimates of risk without multiple, larger data sets. We anticipate that these results will encourage other investigators to apply these methods to their own cohorts. Still, it is promising that a strategy targeting the highest-risk quartile identifies more than 50% of individuals with suicide or accidental death, while targeting the top 2 quartiles identifies approximately 80% of subsequent deaths.

Several limitations in the present data must be considered. First, cause of death was not available for all individuals, so some degree of misclassification must be assumed. In general, this limitation should bias the results toward failing to detect effects, particularly if those at greatest risk are also those most likely to be missed (eg, because they moved out of state or disappeared as a result of psychosis, mania, or substance abuse). To directly examine misclassification, we considered a broader outcome that included accidental death. Second, we were unable to examine the specific features of psychopathology (eg, as represented in the National Institute of Mental Health’s Research Domain Criteria) systematically because of the absence of structured research assessments or experimental testing paradigms. On the other hand, these results reflect our group’s earlier observation that clinicians document clinically meaningful symptoms in a dimensional fashion even in the absence of formal measures.³³ Third, we could

Confidential. Do not distribute. Pre-embargo material.

not directly examine clinician-level features in these notes, which are the product of multiple admitting and discharging resident physicians. Fourth, while the results span a large cohort observed for up to a decade, they still reflect the population of 2 academic medical centers in a single region, albeit a large one. As a result, further investigation to understand the extent to which these results generalize to other populations will be required.

Notwithstanding these important limitations, our results suggest the outlines of an approach to characterizing risk feasibly and cost-effectively. While discharge summaries are not necessarily available at discharge, with modern EHRs they are typically available between 1 day and 1 week after discharge. Alternatively, it may be possible to find proxies for these features among coded data, although such data may also be unavailable at discharge. As such, under a population management strategy, scores might be generated for each individual, with those individuals in the highest-risk quartile (or some other threshold determined based on the intensity of the intervention and available resources) targeted for a follow-up telephone call, letter to the primary care practice, or office visit to assess risk and assist with psychiatric referral if needed. The

intensity of the intervention might further be tailored to the risk quartile (eg, providing a more intensive intervention to the top quartile and a less intensive one to the next quartile).

Conclusions

In aggregate, the present study demonstrates the feasibility of characterizing suicide risk based on data available as part of routine clinical care as a possible step toward clinical risk stratification. Even limited to coded data, our prediction substantially improves on chance or on the current standard of no systematic assessment. Furthermore, it illustrates the application of simple machine learning techniques to extract additional data from clinician notes as a means of capturing more detail than is available in coded data sets and crucially shows that even a coarse measure may substantially improve risk stratification. While the value of large data sets in health care has undoubtedly been the subject of substantial hyperbole, our results add to a growing body of work indicating the feasibility of leveraging such data sets with standard computational tools to make predictions that may be applied to stratify risk.

ARTICLE INFORMATION

Accepted for Publication: July 23, 2016.

Published Online: September 14, 2016.
doi:10.1001/jamapsychiatry.2016.2172

Author Contributions: Dr Perlis had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: McCoy, Snapper, Perlis.

Acquisition, analysis, or interpretation of data: McCoy, Castro, Roberson, Perlis.

Drafting of the manuscript: McCoy, Roberson, Perlis.

Critical revision of the manuscript for important intellectual content: McCoy, Castro, Snapper, Perlis.

Statistical analysis: McCoy, Castro, Perlis.

Administrative, technical, or material support: Castro, Roberson, Snapper.

Study supervision: Perlis.

Conflict of Interest Disclosures: Dr Perlis reported serving on scientific advisory boards or consulting for Genomind LLC, Healthrageous, Pfizer, Perfect Health, Proteus Biomedical, PsyBrain Inc, and RID Ventures LLC and reported receiving royalties through Massachusetts General Hospital from Concordant Rater Systems (now Bracket/Medco). No other disclosures were reported.

Funding/Support: Dr McCoy is supported in part by a Stanley Center fellowship and grant R25MH094612 from the National Institute of Mental Health. Dr Perlis is supported in part by grants P50 MH106933 and R01MH106577 from the National Institute of Mental Health and by grant P50 MH106933 from the National Human Genome Research Institute.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, choice to submit, or approval of the manuscript.

REFERENCES

- National Vital Statistics System, National Center for Health Statistics, Centers for Disease Control and Prevention. 10 Leading causes of death by age group, United States: 2013. Produced by National Center for Injury Prevention and Control using WISQARS. www.cdc.gov/injury/wisqars/pdf/leading_causes_of_death_by_age_group_2013-a.pdf. Accessed August 3, 2016.
- Xu J, Murphy SL, Kochanek KD, Bastian BA. Deaths: final data for 2013. *Natl Vital Stat Rep*. 2016;64(2):1-119.
- Borges G, Nock MK, Haro Abad JM, et al. Twelve-month prevalence of and risk factors for suicide attempts in the World Health Organization World Mental Health Surveys. *J Clin Psychiatry*. 2010;71(12):1617-1628.
- Nock MK, Hwang I, Sampson NA, Kessler RC. Mental disorders, comorbidity and suicidal behavior: results from the National Comorbidity Survey Replication. *Mol Psychiatry*. 2010;15(8):868-876.
- Nock MK, Green JG, Hwang I, et al. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry*. 2013;70(3):300-310.
- Niculescu AB, Levey D, Le-Niculescu H, Niculescu E, Kurian SM, Salomon D. Psychiatric blood biomarkers: avoiding jumping to premature negative or positive conclusions. *Mol Psychiatry*. 2015;20(3):286-288.
- Motto JA, Heilbron DC, Juster RP. Development of a clinical instrument to estimate suicide risk. *Am J Psychiatry*. 1985;142(6):680-686.
- Kessler RC, Warner CH, Ivany C, et al. Army STARRS Collaborators. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2015;72(1):49-57.
- Dobscha SK, Denneson LM, Kovas AE, et al. Correlates of suicide among veterans treated in primary care: case-control study of a nationally representative sample. *J Gen Intern Med*. 2014;29(suppl 4):853-860.
- Turecki G, Brent DA. Suicide and suicidal behavior. *Lancet*. 2016;387(10024):1227-1239.
- Hirschfeld RM, Russell JM. Assessment and treatment of suicidal patients. *N Engl J Med*. 1997;337(13):910-915.
- Pirkis J, Too LS, Spittal MJ, Krysinska K, Robinson J, Cheung YT. Interventions to reduce suicides at suicide hotspots: a systematic review and meta-analysis. *Lancet Psychiatry*. 2015;2(11):994-1001.
- Tidemalm D, Långström N, Lichtenstein P, Runeson B. Risk of suicide after suicide attempt according to coexisting psychiatric disorder: Swedish cohort study with long term follow-up. *BMJ*. 2008;337(7682):1328-1333.
- Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. *Am J Psychiatry*. 2002;159(6):909-916.
- McCoy TH, Castro VM, Cagan A, Roberson AM, Kohane IS, Perlis RH. Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: an electronic health record study. *PLoS One*. 2015;10(8):e0136341. doi:10.1371/journal.pone.0136341.
- Bourne LE. Stimulus selection and memory strategies in concept problems. *Mem Cognit*. 1974;2(2):270-274.
- Murphy SN, Mendis M, Hackett K, et al. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. *AMIA Annu Symp Proc*. 2007:548-552.

Confidential. Do not distribute. Pre-embargo material.

18. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with Informatics for Integrating Biology and the Bedside (i2b2). *J Am Med Inform Assoc*. 2010;17(2):124-130.
19. Murphy S, Churchill S, Bry L, et al. Instrumenting the health care enterprise for discovery research in the genomic era. *Genome Res*. 2009;19(9):1675-1681.
20. Castro VM, McCoy TH, Cagan A, et al. Stratification of risk for hospital admissions for injury related to fall: cohort study. *BMJ*. 2014;349:g5863.
21. De Smedt T, Daelemans W. Pattern for Python. *J Mach Learn Res*. 2012;13:2063-2067.
22. De Smedt T, Daelemans W. "Vreselijk mooi!" (terribly beautiful): a subjectivity lexicon for Dutch adjectives. www.clips.ua.ac.be/sites/default/files/desmedt-subjectivity.pdf. Accessed August 3, 2016.
23. Pang B, Lee L, Vaithyanathan S. Thumbs up? sentiment classification using machine learning techniques. In: Proceedings from the Association for Computational Linguistics' Empirical Methods in Natural Language Processing; July 7-12, 2002; Philadelphia, PA.
24. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS One*. 2013;8(5):e64417. doi:10.1371/journal.pone.0064417.
25. Therneau T, Grambsch P, Pankratz V. Penalized survival models and frailty. *J Comput Graph Stat*. 2003;12(1):156-175.
26. Therneau T. Mixed effects Cox models. <https://cran.r-project.org/web/packages/coxme/vignettes/coxme.pdf>. Published June 15, 2015. Accessed August 3, 2016.
27. Perlis RH, Iosifescu DV, Castro VM, et al. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychol Med*. 2012;42(1):41-50.
28. 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(6):727-734.
29. Pencina MJ, D'Agostino RB, Pencina KM, Janssens AC, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol*. 2012;176(6):473-481.
30. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.
31. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8(1):53.
32. Rumshisky A, Ghassemi M, Naumann TJ, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry*. In press.
33. McCoy TH, Castro VM, Rosenfield HR, Cagan A, Kohane IS, Perlis RH. A clinical perspective on the relevance of research domain criteria in electronic health records. *Am J Psychiatry*. 2015;172(4):316-320.
34. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.
35. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.
36. Clark IA, Cowden WB, Butcher GA, Hunt NH. Possible roles of tumor necrosis factor in the pathology of malaria. *Am J Pathol*. 1987;129(1):192-199.