



Published in final edited form as:

Psychol Med. 2021 June ; 51(8): 1382–1391. doi:10.1017/S0033291720000173.

Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models

Maxwell Levis^{1,2}, Christine Leonard Westgate¹, Jiang Gui², Bradley V. Watts^{2,3}, Brian Shiner^{1,2,3,4}

¹White River Junction VA Medical Center, White River Junction, VT, USA

²Geisel School of Medicine at Dartmouth, Hanover, NH, USA

³VA Office of Systems Redesign and Improvement, White River Junction, VT, USA

⁴National Center for PTSD Executive Division, White River Junction, VT, USA

Abstract

Background.—This study evaluated whether natural language processing (NLP) of psychotherapy note text provides additional accuracy over and above currently used suicide prediction models.

Methods.—We used a cohort of Veterans Health Administration (VHA) users diagnosed with post-traumatic stress disorder (PTSD) between 2004–2013. Using a case-control design, cases (those that died by suicide during the year following diagnosis) were matched to controls (those that remained alive). After selecting conditional matches based on having shared mental health providers, we chose controls using a 5:1 nearest-neighbor propensity match based on the VHA's structured Electronic Medical Records (EMR)-based suicide prediction model. For cases, psychotherapist notes were collected from diagnosis until death. For controls, psychotherapist notes were collected from diagnosis until matched case's date of death. After ensuring similar numbers of notes, the final sample included 246 cases and 986 controls. Notes were analyzed using Sentiment Analysis and Cognition Engine, a Python-based NLP package. The output was evaluated using machine-learning algorithms. The area under the curve (AUC) was calculated to determine models' predictive accuracy.

Results.—NLP derived variables offered small but significant predictive improvement (AUC = 0.58) for patients that had longer treatment duration. A small sample size limited predictive accuracy.

Conclusions.—Study identifies a novel method for measuring suicide risk over time and potentially categorizing patient subgroups with distinct risk sensitivities. Findings suggest leveraging NLP derived variables from psychotherapy notes offers an additional predictive value over and above the VHA's state-of-the-art structured EMR-based suicide prediction model. Replication with a larger non-PTSD specific sample is required.

Author for correspondence: Maxwell Levis, maxwelle.levis@va.gov.

Conflict of interest. Authors have no conflict of interest.

Keywords

Electronic medical records; natural language processing; suicide prediction; veterans mental health

Death by suicide is a major public health concern (World Health Organization, 2018). Given the high rates of death by suicide among military veterans, (Kessler et al., 2017) the U.S. Veterans Health Administration (VHA) has prioritized improving methods to identify individuals at the highest risk for suicide (McCarthy et al., 2015; Torous et al., 2018). As part of this initiative, machine-learning models are increasingly utilized to evaluate complex networks of potential predictor variables, decipher true *v.* false positives, and identify the most predictive variables or constellation of variables (Walsh, Ribeiro, & Franklin, 2017). Although these innovations have led to improvements, even state-of-the-art suicide prediction models have been critiqued for lacking optimal accuracy (Belsher et al., 2019; Kessler et al., 2019). As a means to further increase accuracy, we describe the development and preliminary testing of a novel natural language processing (NLP) approach that includes risk predictor variables extracted from mental health providers' written notes alongside structured variables included in the current VHA state-of-the-art suicide prediction model.

Suicide risk prediction models have been critiqued for a range of reasons, from concerns about specific predictor variables to foundational debates about in-person clinical evaluations *v.* analytically derived risk assessment tools (Kessler, 2019). The accuracy of suicide risk screening tools is further complicated by patients' reluctance to disclose suicidal intent (Ganzini et al., 2013; Husky, Zablith, Fernandez, & Kovess-Masfety, 2016) and concerns about associated stigma (Ganzini et al., 2013; Hom, Stanley, Podlogar, & Joiner, 2017). The direct and indirect costs of provider-administered suicide risk assessments have also been noted (Kessler, 2019). Responding to these concerns, the VHA recently developed Recovery Engagement and Coordination for Health – Veterans Enhanced Treatment (REACH VET; Veterans Affairs Office of Public and Intergovernmental Affairs, 2017), a machine-learning-based suicide prediction model. Drawing from VHA users' Electronic Medical Records (EMR), REACH VET was designed to identify individuals with the highest suicide risk (the top 0.1% tier). REACH VET includes 61 variables abstracted from structured EMR data, ranging from health service usage, to psychotropic medication usage, socio-demographics, and the interaction of demographics and healthcare usage. REACH VET, integrated within an alert system that informs mental health providers about patient risk, has been quickly adopted throughout the VHA network (Lowman, 2019).

REACH VET's predictive ability is defined by its utilization of structured EMR variables, which are easily quantified and analyzable. Although structured EMR variables account for many relevant suicide predictor variables, not all potential predictor variables have been developed into structured formats or have structured formats that are widely used (Barzilay & Apter, 2014; Rudd et al., 2006). To this end, studies have explored the linguistic analysis of unstructured, text-based EMR data for predictive purposes (Leonard Westgate, Shiner, Thompson, & Watts, 2015; Poulin et al., 2014; Rumshisky et al., 2016). Evidence suggests that the linguistic analysis of unstructured EMR data – materials such as clinician's free-text notes and written records – may offer relevant information for suicide risk prediction,

including information about patients' interpersonal patterns and the relationship between the patient and the medical provider. In addition to capturing new and potentially useful clinical information, this indirect approach offers to lessen concerns about patients' self-report bias as well as clinicians' interpretative bias. Finally, these text-based approaches may better account for the dynamic nature of suicide risk in contrast to approaches that largely rely on unchanging demographic variables.

NLP bridges linguistics and machine learning, quantifying written language as vectors that can be statistically evaluated. NLP offers to broaden the reach of computational analysis to better include human experience, emotion, and relationships (Crossley, Kyle, & McNamara, 2017), areas that have been previously linked with suicide risk (Van Orden et al., 2010). Initial research suggests that NLP offers an effective means to mine free-text data for variables that impact suicide (Ben-Ari & Hammond, 2015; Fernandes et al., 2018; Koleck, Dreisbach, Bourne, & Bakken, 2019). This study evaluates whether REACH VET's ability to predict death by suicide can be improved by including NLP-derived variables from unstructured EMR data. To accomplish this objective, we built on established REACH-VET predictor variables to determine whether linguistic analysis of free-text clinical notes could improve prediction of death by suicide within a cohort of veterans that have been diagnosed with post-traumatic stress disorder (PTSD). This study utilized a PTSD cohort because of associations linking PTSD and suicide (McKinney, Hirsch, & Britton, 2017), because excess suicide mortality in the VHA PTSD treatment population has been previously established (Forehand et al., 2019), and because we had a readily available and well-developed cohort (Shiner, Leonard Westgate, Bernardy, Schnurr, & Watts, 2017).

Methods

Data source

Since 2000, the Department of Veterans Affairs (VA) has employed an electronic medical record for all aspects of patient care. EMR data from all VHA hospitals are stored in the VA Corporate Data Warehouse (CDW). Using the VA CDW, we selected VA users that had been newly diagnosed with PTSD between 2004 and 2013. Individuals in this sample received at least two PTSD diagnoses within 3 months, of which at least one occurred in a mental health clinic, and had not met PTSD diagnostic criterion during the previous 2 years. This original sample ($n = 731\,520$) has been previously described (Shiner et al., 2017). The first of the two qualifying PTSD diagnoses was considered the 'index PTSD diagnosis', and the date of that diagnosis was used as the start of a risk time for analyses (see below). Patients were followed for 1 year following the index diagnosis and if they met criteria for multiple yearly sub-cohorts, they were assigned to the earliest sub-cohort. We obtained information on patient characteristics and service use, as well as clinical note text associated with psychotherapy encounters, from the CDW.

Selection of cases and controls

We used a three-step process to match cases and controls. First, we identified all psychotherapists that had at least one individual psychotherapy session with patients ($n = 436$) who died by suicide within their first treatment year after the index PTSD diagnosis.

Psychotherapists were selected regardless of therapeutic orientation and level of graduate training. As less than 3% of patients in this cohort received evidence-based psychotherapy (EBP) at the recommended treatment level (Shiner et al., 2019), we did not differentiate between EPB and non-EPB providers. We then identified all other patients ($n = 96\,570$) within PTSD cohort that remained alive at the end of their first treatment year who had at least one session with one of these psychotherapists (the same psychotherapist as the case group). Second, to account for individual clinician note-writing differences, familiarity with patients, psychotherapy type, and note-writing style, we performed a conditional match on plurality psychotherapist (the psychotherapist who completed the highest number of administratively-coded psychotherapy sessions with each patient) and a 2-year window. This step resulted in 343 cases and 26 542 potential controls. Third, we selected among conditional matches using propensity scores created with the 61 REACH VET variables. REACH VET is the current standard method for suicide risk identification in the VHA. Minor modifications were made to the REACH VET variables to address variable collinearity, interaction variables, and associated sample cohort differences. Modifications to the REACH VET variables are addressed in Table 1. Because REACH VET's weighted coefficients were not publicly available, we calculated propensity scores from patients' treatment history prior to an end date dependent on the case (date of death occurring during the first year of PTSD treatment) or control (end date of the first year of PTSD treatment). We used greedy matching (Austin, 2011), with the psychotherapist that provided the plurality of psychotherapy visits ('primary psychotherapist') specified as an exact matching constraint to balance the sample such that there were no consequential differences between cases and controls on REACH VET variables (cases and controls were at an equal calculated risk for suicide based on the 61 REACH-VET variables). We utilized a 5:1 nearest-neighbor propensity score match (Austin, 2011) and achieved a caliper of 27.5 and C statistic = 0.806. The initial analytic sample consisted of 252 cases and 1090 controls. Table 1 presents conditional matching and propensity score matching details.

Selection of the text corpus

For cases in the analytic sample, we obtained all clinical notes associated with administratively-coded psychotherapy encounters beginning at the index PTSD diagnosis and ending 5 days before death. Templates and forms were deleted from notes in order to facilitate machine learning on free-text written by clinicians. For each of the five controls, notes were selected from an equivalent range of time as the matched case, so that if a case lived for 6 months, notes from the relevant controls would be evaluated from diagnosis forward for 6 months. We excluded notes from within 5 days before death as the VA EMR often documents calls to or from families following a death by suicide, and dates of death can sometimes be incorrect by several days. We also excluded patients who did not have any notes within the selected date range. Patients who had more than threefold the mean number of notes were removed so as to not to overweight patients that had been seen more frequently. Although we retrieved and processed notes from all psychotherapy encounters, analysis was limited to notes from patients' plurality psychotherapist to account for more developed patient relationships. Due to these exclusions, the number of associated controls for each case varied throughout the sample. The final sample consisted of 246 cases and 986 controls. A total of 10 244 notes were selected for analysis.

Because we were concerned that patients who were in treatment longer may have had more developed relationships with their providers than those that were treated for less time, we grouped patients that lived equivalent lengths of time from diagnosis together. Cases and their associated controls were grouped together as follows: patients that lived up to 3 months after index diagnosis (3-month cohort, cases: $n = 33$, controls: $n = 99$); patients that lived between 3 and 6 months after index diagnosis (6-month cohort, cases: $n = 63$, controls: $n = 222$); patients that lived between 6 and 9 months (9-month cohort, cases: $n = 72$, controls: $n = 297$); and patients that lived between 9 and 12 months (12-month cohort, cases: $n = 78$, controls: 368). Notes from diagnosis until the relevant end date were evaluated.

Calculation of linguistic indices

Notes were processed by Sentiment Analysis and Cognition Engine [SÉANCE; (Crossley et al., 2017)], a Python-based NLP package. SÉANCE utilizes a suite of established linguistic databases including SemanticNet (Cambria, Havasi, & Hussain, 2012; Cambria, Speer, Havasi, & Hussain, 2010), General Inquirer Database (GID; Stone, Dunphy, Smith, & Ogilvie, 1966), EmoLex (Mohammad & Turney, 2010, 2013), Lasswell (Lasswell & Namenwirth, 1969), Valence Aware Dictionary and sEntiment Reasoner (VADER; Hutto & Gilbert, 2014), Hu–Liu (Hu & Liu, 2004), Harvard IV-4 (Stone et al., 1966), and the Geneva Affect Label Coder (GALC; Scherer, 2005). These sources range from expert-derived dictionary lists to rule based systems (Urbanowicz & Moore, 2009), comprise more than 250 unique variables, and can be evaluated in positive and negative iterations. SÉANCE compares positively (Crossley et al., 2017) with Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, and Francis, 2007), a widely used semantic analysis tool. Additionally, in contrast to LIWC, SÉANCE is a downloadable open-source software package that is more easily utilized under VA Office of Information and Technology (VA OI&T) constraints, which restrict the use of cloud-based software packages such as LIWC.

Analysis

Cases and controls were compared over the full range of SÉANCE variables. Data were analyzed using a least absolute shrinkage and selection operator (LASSO) penalized generalized linear mixed model that controlled for varying numbers of psychotherapy sessions. LASSO is a machine-learning algorithm that reduces prediction errors frequently associated with stepwise selection (Tibshirani, 1996). LASSO sets the sum of the absolute values of the regression coefficients to be less than a fixed value, such that less important feature coefficients are reduced to zero and excluded from the model. Bayesian information criterion (BIC; Schwarz, 1978) was utilized to select the tuning parameter.

Each sub-cohort was randomly divided into training (2/3 of sample) and testing (1/3 of sample) sets. LASSO was implemented on the training set to select features, which were in turn utilized in the testing set to estimate prediction scores. Area under the curve (AUC) and confidence interval (95%) statistics were calculated to determine models' predictive accuracy using the c-statistic. Analysis was completed using Python and R's glmnet package (Friedman, Hastie, & Tibshirani, 2010).

Results

Each sub-cohort's reduced model contained unique linguistic features. AUC statistics evaluating the predictive accuracy of each of these models accounted for little to no improvement above chance except for the 12-month cohort (Table 2). Twelve-month cohort AUC statistics indicated ($C = 58.0$; 95% CI 51.2–64.9) that the associated features offered small (8%) predictive improvements. Within other sub-cohorts, confidence intervals (95%), showed a wide range of values, indicating that the sample was too small for adequate analysis. Figure 1 presents model AUC curves.

The most prominent NLP features in the 3-month cohort dealt with food and life routine issues. The most prominent NLP features in 6-month cohort dealt with vice, anger, and male professional and social roles. The most prominent NLP features in the 9-month cohort dealt with heightened arousal, including issues of affiliation and hostility. The most prominent NLP features in the 12-month cohort dealt with goal directedness and indifference and adjectives evaluating social relationships. Selected NLP features were drawn from GID's (Stone et al., 1966) semantically categorized dictionary list and EmoLex (Mohammad & Turney, 2013) and GALC's (Scherer, 2005) list of emotional specific words. Table 2 includes each sub-cohort's featured model.

Discussion

Whereas REACH VET variables are often static, our results present a novel dynamic method to identify and monitor predictor variables and how they change over time. Following Kleiman et al.'s (2017) findings about the variation and fluctuation of suicide risk factors, this method presents a valuable format to monitor real-time risk factor changes. Results suggest that unique themes were present in the notes of patients that had different life durations after diagnosis. Looking more closely at these selected variables, they parallel Maslow's hierarchy of needs (Maslow, 1943), transitioning over the course of the year after diagnosis from variables associated with basic needs such as food, to those associated with safety and conduct, emotional arousal and social affiliation, and finally to personal fulfillment and achievement. For Maslow, wellbeing rests upon the foundation of satisfied needs, positing that as one level of need is met, other more abstract potentialities become apparent (Heylighen (1992). Supporting this approach, our findings suggest that those that fail to reach certain thresholds of achievement experience increased suicide risk.

Three-month cohort features have strong associations with previous research. Food insecurity has been identified as an important predictor of mental health symptoms and access to health care among veterans (Narain et al., 2018). Indeed, almost 25% of veterans have experienced food insecurity, 40% more than the general population (Widome, Jensen, Bangerter, & Fu, 2015). Food and routine may also be considered proxies for homelessness (Wang et al., 2015), another substantial predictor for suicide (Schinka, Schinka, Casey, Kaspro, & Bossarte, 2012). Food insecure veterans frequently experience heightened levels of mental and physical health disorders, including pronounced rates of substance abuse (Wang et al., 2015; Widome et al., 2015), and limited access to mental health services (McGuire & Rosenheck, 2005). Increased and decreased food intake has been widely linked

with major depression and suicide risk (Brundin, Petersén, Björkqvist, & Träskman-Bendz, 2007; Bulik, Carpenter, Kupfer, & Frank, 1990). We may potentially understand the rapid rates of suicide among this sub-cohort as stemming from the population's acute psychiatric needs, as those with higher rates of psychiatric symptoms tend to complete suicide earlier in treatment (Qin & Nordentoft, 2005).

Six-month cohort features are associated with anger and aggression, substance and physical abuse, and masculine gender norms, all of which are known predictors of suicidality (Bohnert, Ilgen, Louzon, McCarthy, & Katz, 2017; Genuchi, 2019; Wilks et al., 2019). Threats to safety are especially highlighted in the GID Vice Dictionary, which includes words like 'abject', 'abuse', and 'adultery'. Masculine gender roles, especially within the military community, are thought to encourage stoicism as opposed to help-seeking behaviors (Addis, 2008; Genuchi, 2019), and thus add additional suicide risk. Anger, substance abuse, and depression are also cited as correlates of PTSD (Hellmuth, Stappenbeck, Hoerster, & Jakupcak, 2012). As everyone in this sample has received PTSD diagnoses, this may indicate particularly elevated symptomology. Notably the terms described are closely related to core symptoms or PTSD and depression.

Nine-month cohort findings similarly are associated with established suicide predictor variables. Words included within the GID Arousal dictionary (Stone et al., 1966) focus on emotional excitation, including affiliation and hostility related terms. Included words, such as 'abhor', 'affection', and 'antagonism', center on the presence and absence of belongingness and love. Recalling the Interpersonal Theory of Suicide's approach (Van Orden et al., 2010), this relational spectrum could have importance in evaluating suicide risk. Whereas supportive relationships protect against suicidality (Wilks et al., 2019), hostility, anger, and aggression are established suicide risks, especially for a PTSD cohort (McKinney et al., 2017). Similarly, these terms suggest that those patients who died by suicide may have struggled to achieve and maintain relationships.

Twelve-month cohort features include self-fulfillment related constructs. Whereas the GALC Boredom dictionary (Scherer, 2005) highlights words showing lack of interest and indifference, the Harvard Interpersonal Adjective dictionary (Stone et al., 1966), addresses judgments about other people, and constructs connected with competitiveness and externalizing behavior. Based on these word usages, this sub-cohort tends towards depression, hopelessness, and antisocial tendencies, variables recognized as long-term suicide risk factors (Beck, Steer, Kovacs, & Garrison, 1985; Verona, Patrick, & Joiner, 2001). In contrast to individuals that completed suicide earlier in the treatment year whose EMR notes demonstrated acute physical (3-month cohort), psychological (6-month cohort), and relational needs (9-month cohort), those that lived longest had notes that conveyed depression, lack of purpose, and social detachment. While the 3-month, 6-month, and 9-month cohorts had fewer suicide deaths than the 12-month cohort, the fact that these patients died sooner may suggest that failure to address those more basic needs and associated symptoms may result in more rapid progression to lethality.

It is difficult to identify whether these clustered sub-cohorts are connected with personal changes over time, responses to interventions, or different levels of baseline functionality.

Alternatively, differences may be indicative of deepening therapeutic relationships between provider and patient, these shifts being associated with having increased time to explore existential issues. Regardless of the causal mechanism, the described method offers access to theoretically important areas that have remained outside of the purview of computational analysis (Rogers, 2001). An alternate hypothesis is that initial focus in psychotherapy sessions may begin with basic needs, evolve to focus on symptoms, and final begin to address more core interpersonal issues. Given the nature of the data, we are hesitant to derive conclusions regarding whether these trends reflect patients not achieving resolution of these needs or reflect increased focus by the psychotherapist.

Findings reinforce the importance of formally assessing patients' immediate needs as well addressing more abstract patient characteristics, such as goal directedness, hope, and social relationship quality. So as to better identify and mitigate suicide risk, it is recommended that evaluations of these issues be included both as part of initial assessment and at regular intervals during the course of treatment.

Leveraging NLP variables extracted from EMR data only substantially improved REACH VET's predictive model for patients that received the most care. It is worth noting, however, that while AUC statistics for all sub-cohorts remained close to 0.50, ranging from 0.46 to 0.58, there was considerable confidence interval variance. The confidence interval range became smaller as the sample size increased, inferring that the sample was not uniform and that with a larger sample size, estimates would likely be different. Only the 12-month cohort, the sub-cohort containing the largest sample size and quantity of free-text data, showed significant improvements above REACH VET. Findings suggest that with an adequate sample size and quantity of free-text data, NLP-derived variables added to REACH-VET's ability to predict death by suicide.

Recent research has led to significant advances in understanding the centrality of therapeutic alliance in predicting psychotherapy outcome in general (Lambert & Barley, 2001; Norcross & Lambert, 2018), and PTSD (Keller, Zoellner, & Feeny, 2010) and suicide treatment outcomes (Dunster-Page, Haddock, Wainwright, & Berry, 2017) in particular. While progress has been made in the development of psychometrically robust therapeutic alliance measures (Mallinckrodt & Tekie, 2016), alliance can be difficult to adequately monitor (Elvins & Green, 2008). In line with current trends (Colli & Lingardi, 2009; Martinez et al., 2019), our research suggests the value of assessing therapeutic alliance through indirect mechanisms. Leveraging NLP derived variables may offer additional ability to indirectly monitor therapeutic alliance over the course of treatment. That being noted, at least within our sample, the presence or absence of therapeutic alliance related linguistic themes could be associated with psychotherapist or theory-related factors as opposed to patient factors.

Several considerations should be acknowledged regarding the unique nature of this study and its design. Firstly, as we utilized a PTSD sample, as opposed to the broader VA population that REACH VET was developed on, results are not necessarily comparable. Secondly, as REACH VET variables' coefficients were not publicly available, we ran a regression model using selected cases and potential control pool to implement propensity score matching. While this approach sufficiently accounted for all REACH VET variables, it

is likely that our method yielded different results than the weighted REACH VET model and may have been overly conservative or liberal in restricting the sample. Thirdly, and perhaps most importantly, our sample size was insufficient to optimally develop testing and training sets, which in turn may have impacted the accuracy of the feature selection algorithm. As the sample size was dependent on suicide, we could not correct for this limitation. Even with these constraints, as the REACH VET variables were precisely developed and evaluated (Kessler et al., 2017; McCarthy et al., 2015), any improvements beyond its predictive model should be taken seriously. Given the importance of effective and timely suicide intervention, even small improvements in prediction may make lifesaving differences.

Several additional concerns are worth noting and should be addressed in future studies. Our sample may not be comparable with the population used to develop SÉANCE and associated toolkits' variables. To avoid this confound, deep-learning NLP approaches (Geraci et al., 2017) could be used to develop population specific linguistic references. Although our note extraction method took steps to eliminate template and information that was copied and pasted from other sources, it is difficult to fully exclude this content. Not knowing to what extent information has been duplicated, limits ability to extrapolate personalized information. As duplication is relatively common (Cohen, Elhadad, & Elhadad, 2013), preventative steps could be taken to pre-evaluate if EMR notes include copy and pasted materials.

We acknowledge that the discussed method necessitates patients' having EMR psychotherapy notes. As such, the method does not offer increased predictive accuracy for patients without said notes. It is unclear if an adapted method could be usefully applied to non-psychotherapy mental health notes or to notes from general medical visits. Lastly, we plan to re-run the study and evaluate comparative impact if and when REACH VET coefficients become publically available.

Conclusion

The study identified a novel method for measuring suicide risk over time and potentially categorizing patient subgroups with distinct risk sensitivities. Results suggest that modest improvements above and beyond REACH VET's predicative capability were achieved during select times over the treatment year. In particular, for the 12-month cohort, the group with the largest sample size and the greatest number of psychotherapy sessions, NLP-derived variables provided an 8% predictive gain above state-of-the-art standards. Future research is necessary to parse whether the strengths of the 12-month cohort are associated with sample size, length of treatment, or number of notes. Despite its shortcoming, this study broadens domains of inquiry, contributing new methods to assess patients' feelings and experiences. These domains present opportunities to inform theory and practice, and potentially save lives.

Acknowledgments

Financial support. This work was funded by the VA National Center for Patient Safety Center of Inquiry Program (PSCI-WRJ-Shiner). Dr Levis' time was supported by the VA Office of Academic Affiliations Advanced

Fellowship in Health Systems Engineering. Dr Shiner's time was supported by the VA Health Services Research and Development Career Development Award Program (CDA11–263).

References

- Addis ME. (2008). Gender and depression in men. *Clinical Psychology: Science and Practice*, 15(3), 153–168.
- Austin PC (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. [PubMed: 21818162]
- Barzilay S, & Apter A. (2014). Psychological models of suicide. *Archives of Suicide Research*, 18(4), 295–312. [PubMed: 24568371]
- Beck A, Steer R, Kovacks M, & Garrison B. (1985). Hopelessness and eventual suicide: A 10-year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry*, 142, 559–563.
- Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, ... Skopp NA (2019). Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*, 76(6), 642–651. [PubMed: 30865249]
- Ben-Ari A, & Hammond K. (2015). Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 Persian Gulf War. 2015 48th Hawaii international conference on system sciences (pp. 3168–3175), January 2015.
- Bohnert KM, Ilgen MA, Louzon S, McCarthy JF, & Katz IR (2017). Substance use disorders and the risk of suicide mortality among men and women in the US Veterans Health Administration. *Addiction*, 112(7), 1193–1201. [PubMed: 28301070]
- Brundin L, Petersén Å, Björkqvist M, & Träskman-Bendz L. (2007). Orexin and psychiatric symptoms in suicide attempters. *Journal of Affective Disorders*, 100(1–3), 259–263. [PubMed: 17141878]
- Bulik CM, Carpenter LL, Kupfer DJ, & Frank E. (1990). Features associated with suicide attempts in recurrent major depression. *Journal of Affective Disorders*, 18(1), 29–37. [PubMed: 2136867]
- Cambria E, Havasi C, & Hussain A. (2012). SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis. Twenty-Fifth international FLAIRS conference, May 2012.
- Cambria E, Speer R, Havasi C, & Hussain A. (2010). Senticnet: A publicly available semantic resource for opinion mining. 2010 AAAI fall symposium series, November 2010.
- Cohen R, Elhadad M, & Elhadad N. (2013). Redundancy in electronic health record corpora: Analysis, impact on text mining performance and mitigation strategies. *BMC Bioinformatics*, 14(1), 10–25. [PubMed: 23323800]
- Colli A, & Lingiardi V. (2009). The Collaborative Interactions Scale: A new transcript-based method for the assessment of therapeutic alliance ruptures and resolutions in psychotherapy. *Psychotherapy Research*, 19(6), 718–734. [PubMed: 19742381]
- Crossley SA, Kyle K, & McNamara DS (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, 49(3), 803–821. [PubMed: 27193159]
- Dunster-Page C, Haddock G, Wainwright L, & Berry K. (2017). The relationship between therapeutic alliance and patient's suicidal thoughts, self-harming behaviours and suicide attempts: A systematic review. *Journal of Affective Disorders*, 223, 165–174. [PubMed: 28755624]
- Elvins R, & Green J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. *Clinical Psychology Review*, 28 (7), 1167–1187. [PubMed: 18538907]
- Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, & Chandran D. (2018). Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. *Scientific Reports*, 8(1), 7426. [PubMed: 29743531]
- Forehand JA, Peltzman T, Westgate CL, Riblet NB, Watts BV, & Shiner B. (2019). Causes of excess mortality in veterans treated for post-traumatic stress disorder. *American Journal of Preventive Medicine*. 57(2), 145–152. [PubMed: 31248740]
- Friedman J, Hastie T, & Tibshirani R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. [PubMed: 20808728]

- Ganzini L, Denneson LM, Press N, Bair MJ, Helmer DA, Poat J, & Dobscha SK (2013). Trust is the basis for effective suicide risk screening and assessment in veterans. *Journal of General Internal Medicine*, 28(9), 1215–1221. [PubMed: 23580131]
- Genuchi MC (2019). The role of masculinity and depressive symptoms in predicting suicidal ideation in homeless men. *Archives of Suicide Research*, 23(2), 289–311. [PubMed: 29461153]
- Geraci J, Wilansky P, de Luca V, Roy A, Kennedy JL, & Strauss J. (2017). Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression. *Evidence-Based Mental Health*, 20(3), 83–87. [PubMed: 28739578]
- Hellmuth JC, Stappenbeck CA, Hoerster KD, & Jakupcak M. (2012). Modeling PTSD symptom clusters, alcohol misuse, anger, and depression as they relate to aggression and suicidality in returning US veterans. *Journal of Traumatic Stress*, 25(5), 527–534. [PubMed: 23073972]
- Heylighen F. (1992). A cognitive-systemic reconstruction of Maslow's theory of self-actualization. *Behavioral Science*, 37(1), 39–58.
- Hom MA, Stanley IH, Podlogar MC, & Joiner TE Jr (2017). 'Are you having thoughts of suicide?' Examining experiences with disclosing and denying suicidal ideation. *Journal of Clinical Psychology*, 73(10), 1382–1392. [PubMed: 28085200]
- Hu M, & Liu B. (2004). Mining and summarizing customer reviews. In Kim W & Kohavi R (Eds.), *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177). Washington, DC: ACM Press.
- Husky MM, Zablith I, Fernandez VA, & Kovess-Masfety V. (2016). Factors associated with suicidal ideation disclosure: Results from a large population-based study. *Journal of Affective Disorders*, 205, 36–43. [PubMed: 27400193]
- Hutto CJ, & Gilbert E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Eighth international AAAI conference on weblogs and social media, May 2014.
- Keller SM, Zoellner LA, & Feeny NC (2010). Understanding factors associated with early therapeutic alliance in PTSD treatment: Adherence, childhood sexual abuse history, and social support. *Journal of Consulting and Clinical Psychology*, 78(6), 974–979. [PubMed: 20873895]
- Kessler RC (2019). Clinical epidemiological research on suicide-related behaviors – where we are and where we need to go. *JAMA Psychiatry*, 76(8), 777–778. [PubMed: 31188420]
- Kessler RC, Bernecker SL, Bossarte RM, Luedtke AR, McCarthy JF, Nock MK, ... Zuromski KL (2019). The role of big data analytics in predicting suicide. In Passos I, Mwangi B, Kapczynski F. (eds), *Personalized psychiatry* (pp. 77–98). Cham, CH: Springer Nature.
- Kessler RC, Hwang I, Hoffmire CA, McCarthy JF, Petukhova MV, Rosellini AJ, ... Thompson C. (2017). Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration. *International Journal of Methods in Psychiatric Research*, 26(3), e1575. [PubMed: 28675617]
- Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, & Nock MK (2017). Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. *Journal of Abnormal Psychology*, 126(6), 726–738. [PubMed: 28481571]
- Koleck TA, Dreisbach C, Bourne PE, & Bakken S. (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. *Journal of the American Medical Informatics Association*, 26(4), 364–379. [PubMed: 30726935]
- Lambert MJ, & Barley DE (2001). Research summary on the therapeutic relationship and psychotherapy outcome. *Psychotherapy: Theory, Research, Practice, Training* (new York, N Y), 38(4), 357–361.
- Lasswell HD, & Namenwirth JZ (1969). *The Lasswell value dictionary*. New Haven, CT: Yale Press.
- Leonard Westgate C, Shiner B, Thompson P, & Watts BV (2015). Evaluation of veterans' suicide risk with the use of linguistic detection methods. *Psychiatric Services*, 66(10), 1051–1056. [PubMed: 26073409]
- Lowman CA (2019). Optimizing clinical outcomes in VA mental health care. In Ritchie E, Llorente M. (eds), *Veteran psychiatry in the US* (pp. 29–48). New York, NY: Springer.
- Mallinckrodt B, & Tekie YT (2016). Item response theory analysis of Working Alliance Inventory, revised response format, and new Brief Alliance Inventory. *Psychotherapy Research*, 26(6), 694–718. [PubMed: 26549302]

- Martinez VR, Flemotomos N, Ardulov V, Somandepalli K, Goldberg SB, Imel ZE, ... Narayanan S. (2019). Identifying therapist and client personae for therapeutic alliance estimation. *Proc. Interspeech* (pp. 1901–1905).
- Maslow AH (1943). A theory of human motivation. *Psychological Review*, 50 (4), 370–396.
- McCarthy JF, Bossarte RM, Katz IR, Thompson C, Kemp J, Hannemann CM, ... Schoenbaum M. (2015). Predictive modeling and concentration of the risk of suicide: Implications for preventive interventions in the US Department of Veterans Affairs. *American Journal of Public Health*, 105(9), 1935–1942. [PubMed: 26066914]
- McGuire J, & Rosenheck R. (2005). The quality of preventive medical care for homeless veterans with mental illness. *Journal for Healthcare Quality*, 27, 26–32.
- McKinney JM, Hirsch JK, & Britton PC (2017). PTSD Symptoms and suicide risk in veterans: Serial indirect effects via depression and anger. *Journal of Affective Disorders*, 214, 100–107. [PubMed: 28288403]
- Mohammad SM, & Turney PD (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26–34). Association for Computational Linguistics, June 2010.
- Mohammad SM, & Turney PD (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465.
- Narain K, Bean-Mayberry B, Washington DL, Canelo IA, Darling JE, & Yano EM (2018). Access to care and health outcomes among women veterans using veterans administration health care: Association with food insufficiency. *Women's Health Issues*, 28(3), 267–272. [PubMed: 29475630]
- Norcross JC, & Lambert MJ (2018). Psychotherapy relationships that work III. *Psychotherapy*, 55(4), 303–315. [PubMed: 30335448]
- Pennebaker JW, Booth RJ, & Francis ME (2007). *Linguistic inquiry and word count: LIWC* [Computer software]. Austin, TX: liwc.net, 135.
- Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B, ... McAllister T. (2014). Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS One*, 9(1), e85733.
- Qin P, & Nordentoft M. (2005). Suicide risk in relation to psychiatric hospitalization: Evidence based on longitudinal registers. *Archives of General Psychiatry*, 62(4), 427–432. [PubMed: 15809410]
- Rogers JR (2001). Theoretical grounding: The 'missing link' in suicide research. *Journal of Counseling & Development*, 79(1), 16–25.
- Rudd MD, Berman AL, Joiner TE Jr, Nock MK, Silverman MM, Mandrusiak M, ... Witte T. (2006). Warning signs for suicide: Theory, research, and clinical applications. *Suicide and Life-Threatening Behavior*, 36(3), 255–262. [PubMed: 16805653]
- Rumshisky A, Ghassemi M, Naumann T, Szolovits P, Castro VM, McCoy TH, & Perlis RH (2016). Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational Psychiatry*, 6(10), e921. [PubMed: 27754482]
- Scherer KR (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729.
- Schinka JA, Schinka KC, Casey RJ, KasproW W, & Bossarte RM (2012). Suicidal behavior in a national sample of older homeless veterans. *American Journal of Public Health*, 102(Suppl 1), S147–S153. [PubMed: 22390590]
- Schwarz G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Shiner B, Leonard Westgate C, Bernardy NC, Schnurr PP, & Watts BV (2017). Trends in opioid use disorder diagnoses and medication treatment among veterans with posttraumatic stress disorder. *Journal of Dual Diagnosis*, 13(3), 201–212. [PubMed: 28481727]
- Shiner B, Westgate CL, Gui J, Cornelius S, Maguen SE, Watts BV, & Schnurr PP (2019). Measurement strategies for evidence-based psychotherapy for posttraumatic stress disorder delivery: Trends and associations with patient-reported outcomes. *Administration and Policy in Mental Health and Mental Health Services Research*, 1–17. 10.1007/s10488-019-01004-2. [PubMed: 29948427]

- Stone PJ, Dunphy DC, Smith MS, & Ogilvie DM (1966). *The general inquirer: A computer approach to content analysis: Studies in psychology, sociology, anthropology, and political science*. Cambridge, MA: MIT Press.
- Tibshirani R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
- Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, & Firth J. (2018). Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. *Current Psychiatry Reports*, 20(7), 51–67. [PubMed: 29956120]
- Urbanowicz RJ, & Moore JH (2009). Learning classifier systems: A complete introduction, review, and roadmap. *Journal of Artificial Evolution and Applications*, 2009(1), 1–25.
- Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, & Joiner TE Jr (2010). The interpersonal theory of suicide. *Psychological Review*, 117(2), 575–600. [PubMed: 20438238]
- VA Office of Public and Intergovernmental Affairs. (2017). VA Office of Public and Intergovernmental Affairs. VA REACH VET Initiative helps save veterans lives: Program signals when more help is needed for at-risk veterans. Retrieved from https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/3527-notes.pdf.
- Verona E, Patrick CJ, & Joiner TE (2001). Psychopathy, antisocial personality, and suicide risk. *Journal of Abnormal Psychology*, 110(3), 462–470. [PubMed: 11502089]
- Walsh CG, Ribeiro JD, & Franklin JC (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3), 457–469.
- Wang EA, McGinnis KA, Goulet J, Bryant K, Gibert C, Leaf DA, ... Fiellin DA (2015). Food insecurity and health: Data from the Veterans Aging Cohort Study. *Public Health Reports*, 130(3), 261–268. [PubMed: 25931630]
- Widome R, Jensen A, Bangerter A, & Fu SS (2015). Food insecurity among veterans of the US wars in Iraq and Afghanistan. *Public Health Nutrition*, 18(5), 844–849. [PubMed: 24806818]
- Wilks CR, Morland LA, Dillon KH, Mackintosh MA, Blakey SM, Wagner HR, ... Elbogen EB (2019). Anger, social support, and suicide risk in US military veterans. *Journal of Psychiatric Research*, 109, 139–144. [PubMed: 30537566]
- World Health Organization. (2018). Retrieved from http://www.who.int/mental_health/prevention/suicide/sui-cideprevent/en/.

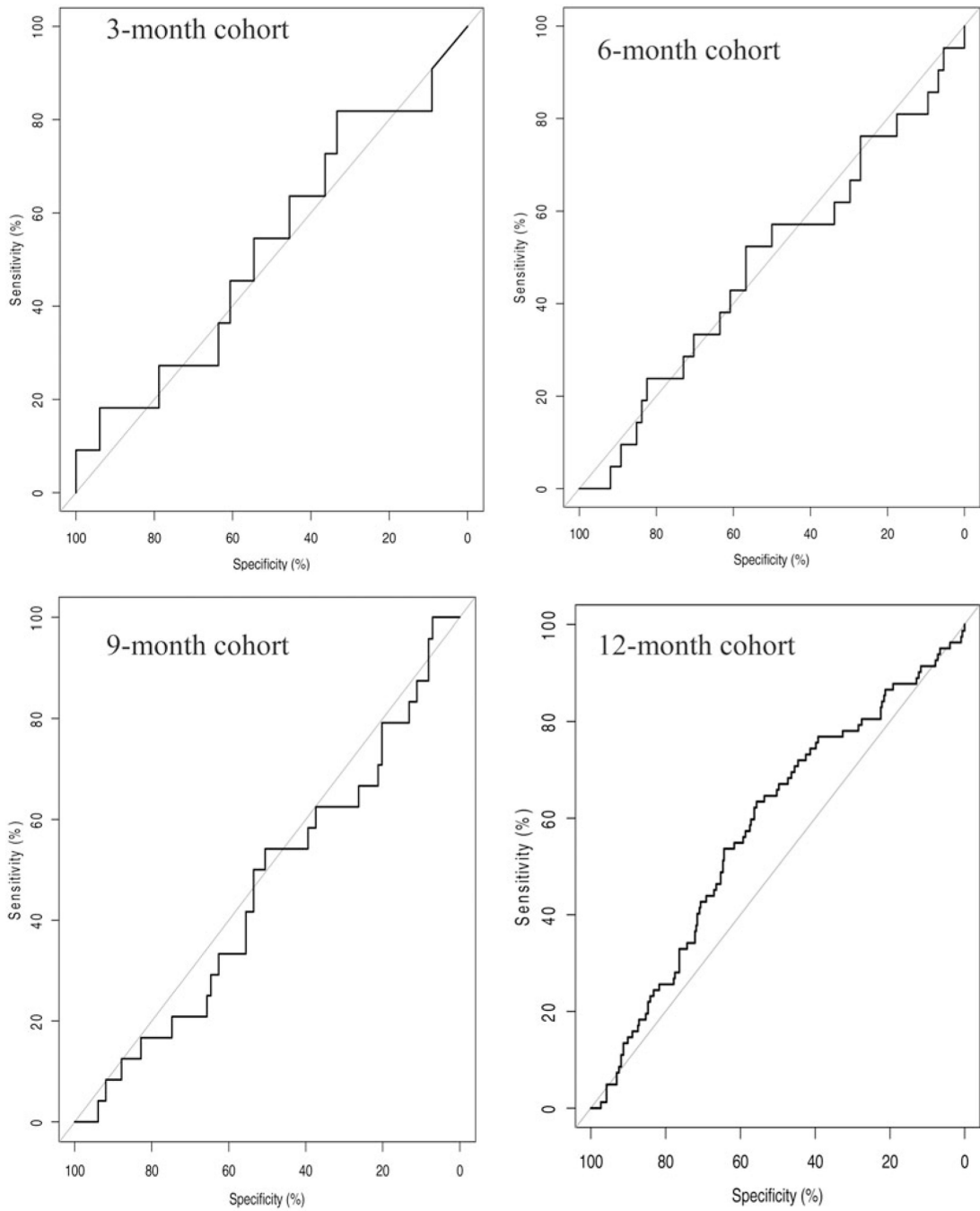


Fig. 1. Figure presents each sub-cohort’s AUC curve.

Table 1.

Case and control matching based on established suicide risk variables

	Conditional match			Propensity match		
	Case (n = 343)	Control (n = 27 356)	Stand. diff.	Case (n = 252)	Control (n = 1133)	Stand. diff.
Mean (SD)						
Emergency Department visit in last 1 month	0.17 (0.76)	0.05 (0.31)	0.20	0.09 (0.39)	0.06 (0.3473)	0.05
Emergency Department visit in last 2 months	0.29 (0.94)	0.10 (0.47)	0.18	0.15 (0.54)	0.14 (0.5826)	-0.04
Number of days Inpatient Mental Health used in last 7 months	6.29 (-22.46)	1.87 (14.59)	0.23	2.17 (11.33)	2.01 (14.8802)	0.01
Number of days Inpatient Mental Health used in last 7 months squared	542.50 (3310.50)	216.50 (2503.10)	0.11	132.60 (1403.60)	225.30 (2630.2)	-0.03
Number of days Outpatient Mental Health used in last month	1.84 (2.90)	0.91 (1.77)	0.39	1.36 (2.37)	1.23 (2.1372)	0.05
Number of days Outpatient Mental Health used in last 1 month squared	11.24 (42.28)	3.94 (21.22)	0.22	7.45 (39.28)	6.08 (25.9583)	0.04
Number of days Outpatient services used in last 7 months	26.62 (22.14)	17.25 (16.77)	0.46	21.80 (18.26)	20.69 (18.6728)	0.06
Number of days Outpatient services used in last 8 months	29.67 (24.64)	20.02 (19.01)	0.42	24.71 (20.90)	23.82 (21.2843)	0.04
Number of days Outpatient services used in last 15 months	45.71 (38.45)	40.29 (32.54)	0.15	42.87 (38.69)	43.49 (35.0809)	-0.02
Number of days Outpatient services used in last 23 months	56.36 (49.54)	49.92 (40.79)	0.14	53.63 (49.64)	54.79 (44.8523)	-0.03
Number of days VHA services used in last 7 months	33.67 (41.10)	20.17 (28.48)	0.39	24.57 (26.72)	23.13 (29.3099)	0.04
Number of days VHA services used in last 13 months	54.06 (71.38)	42.78 (53.25)	0.19	45.76 (51.88)	47.27 (57.8569)	-0.03
n (%)						
Male gender	321 (93.59)	25 514 (93.27)	-0.01	238 (94.44)	1052 (92.85)	-0.06
Less than 80 years old	339 (98.83)	26 718 (97.67)	-0.09	248 (98.41)	1118 (98.68)	0.02
More than 80 years old	4 (1.17)	638 (2.33)		4 (1.59)	15 (1.32)	
White Race	260 (75.8)	20 148 (73.65)	-0.05	190 (75.4)	852 (7)	0.00

	Conditional match		Propensity match			
	Case (n = 343)	Control (n = 27 356)	Stand. diff.	Case (n = 252)	Control (n = 1133)	Stand. diff.
Non-White Race ^a	37 (10.79)	5137 (18.78)		27 (10.71)	209 (18.45)	
Married	136 (39.65)	15 504 (56.67)	0.35	108 (42.86)	495 (43.69)	0.02
Divorced ^a	107 (31.2)	6033 (22.05)	-0.21	74 (29.37)	311 (27.45)	-0.04
Widowed ^a	8 (2.33)	497 (1.82)	-0.03	5 (1.98)	22 (1.94)	0.00
Lives in Western USA ^a	121 (35.28)	7998 (29.24)	-0.14	87 (34.52)	363 (26.21)	-0.05
More than 30% Service Connection	158 (46.06)	17 053 (62.34)	0.35	133 (52.78)	618 (54.55)	0.04
More than 70% Service Connection	96 (27.99)	9124 (33.35)	0.13	82 (32.54)	382 (33.72)	0.03
Suicide attempt in last month	9 (2.62)	31 (0.11)	-0.22	2 (0.79)	2 (0.18)	-0.05
Suicide attempt in last 6 months	20 (5.83)	134 (0.49)	-0.21	8 (3.17)	14 (1.24)	-0.10
Suicide attempt in last 18 months	30 (8.75)	413 (1.51)	-0.25	14 (5.56)	31 (2.74)	-0.12
Arthritis diagnosis in last 12 months	158 (46.06)	14 029 (51.28)	0.12	118 (46.83)	560 (49.43)	0.05
Arthritis diagnosis in last 24 months	186 (54.23)	16 706 (61.07)	0.15	142 (56.35)	659 (58.16)	0.04
Lupus diagnosis in last 24 months	4 (1.17)	31 (0.11)	-0.13	1 (0.4)	3 (0.26)	-0.02
Bipolar diagnosis in last 24 months	63 (18.37)	1969 (7.2)	-0.34	35 (13.89)	134 (11.83)	-0.06
Chronic pain diagnosis in last 24 months	33 (9.62)	1549 (5.66)	-0.14	19 (7.54)	94 (8.3)	0.03
Depression diagnosis in last 12 months	274 (79.88)	17 220 (62.95)	-0.37	189 (75/00)	848 (74.85)	0.00
Depression diagnosis in last 24 months	283 (82.51)	18 876 (69)	-0.31	197 (78.17)	881 (77.76)	-0.01
Diabetes diagnosis in last 12 months	31 (9.04)	5348 (19.55)	0.31	29 (11.51)	139 (12.27)	0.02
Substance use disorder diagnosis in last 24 months	172 (50.15)	8170 (29.87)	-0.42	110 (43.65)	474 (41.84)	-0.04

	Conditional match			Propensity match		
	Case (n = 343)	Control (n = 27 356)	Stand. diff.	Case (n = 252)	Control (n = 1133)	Stand. diff.
Homeless in last 24 months	44 (12.83)	1746 (6.38)	-0.22	24 (9.52)	85 (7.5)	-0.07
Head/neck cancer diagnosis in last 24 months ^b	3 (0.87)	139 (0.51)	-0.04	3 (1.19)	16 (1.41)	0.03
Anxiety disorder diagnosis in last 24 months ^a	155 (45.19)	9270 (33.89)	-0.23	99 (39.29)	464 (40.95)	0.03
Personality disorder diagnosis in last 24 months ^a	47 (13.7)	1246 (4.55)	-0.32	28 (11.11)	91 (8.03)	-0.11
Prescription for Alprazolam in last 24 months	42 (12.24)	1761 (6.44)	-0.20	27 (10.71)	73 (6.44)	-0.15
Prescription for any anti-depressant in last 24 months	319 (93)	23 888 (87.32)	-0.18	233 (92.46)	1036 (91.44)	-0.03
Prescription for any anti-psychotic in last 24 months	146 (42.57)	6923 (25.31)	-0.36	101 (40)	380 (33.54)	-0.14
Prescription for Clonazepam in last 12 months	87 (25.36)	3307 (12.09)	-0.34	53 (21.03)	200 (17.65)	-0.09
Prescription for Clonazepam in last 24 months	91 (26.53)	3566 (13.04)	-0.34	56 (22.22)	217 (19.15)	-0.08
Prescription for Lorazepam in last 12 months	38 (11.08)	2757 (10.08)	-0.03	27 (10.71)	97 (8.56)	-0.07
Prescription for Mirtazapine in last 12 months	58 (16.91)	3349 (12.24)	-0.13	38 (15.08)	191 (16.86)	0.05
Prescription for Mirtazapine in last 24 months	70 (20.41)	3791 (13.86)	-0.17	43 (17.06)	220 (19.42)	0.06
Prescription for mood stabilizer in last 12 months	164 (47.81)	9245 (33.8)	-0.28	111 (44.05)	471 (41.57)	-0.05
Prescription for opioids in last 12 months	157 (45.77)	10 094 (36.9)	-0.17	112 (44.44)	516 (45.54)	0.02
Sedative or anxiolytic use disorder diagnosis in last 12 months	222 (64.72)	12 173 (44.5)	-0.40	156 (61.9)	633 (55.87)	-0.12
Sedative or anxiolytic use disorder diagnosis in last 24 months	229 (66.76)	12 956 (47.36)	-0.38	160 (63.49)	667 (58.87)	-0.10
Prescription for statin in last 12 months	87 (25.36)	9787 (35.78)	0.24	72 (28.57)	332 (29.3)	0.02
Prescription for Zolpidem in last 12 months	74 (21.67)	4266 (15.59)	-0.15	50 (19.84)	226 (19.95)	0.00
Any Mental Health treatment in last 12 month	343 (100)	27 348 (99.97)	0.02	252 (100)	1133 (100)	0.00

	Conditional match			Propensity match		
	Case (n = 343)	Control (n = 27 356)	Stand. diff.	Case (n = 252)	Control (n = 1133)	Stand. diff.
Number of days Emergency Department used in last month	28 (8.16)	1006 (3.68)	0.19	17 (6.75)	48 (4.24)	0.11
Number of days Emergency Department used in last 24 months	48 (13.99)	1825 (6.67)	0.18	25 (9.92)	95 (8.38)	-0.04
Any Mental Health treatment in last 1 month	15 (4.37)	173 (0.63)	-0.05	4 (1.59)	10 (0.88)	0.01
Any Mental Health treatment in last 6 months	68 (19.83)	907 (3.32)	-0.45	23 (9.13)	53 (4.68)	-0.13
Any Mental Health treatment in last 24 months	91 (26.53)	2189 (8)	-0.42	38 (15.08)	37 (12.09)	-0.07
Any Mental Health treatment in last 12 months	101 (29.45)	2690 (9.83)	-0.43	45 (17.86)	165 (14.56)	-0.07
Any Mental Health treatment in last 24 months	343 (100)	2732 (99.99)	0.02	252 (100)	1133 (100)	0.00

Table presents conditional match and propensity match for cases and control based on 61-variable electronic medical record (EMR) suicide prediction metric.^c

^aThe original 61-variable suicide risk model includes three interaction variables. As the weighted regression model associated with these variables was not publically available, we could not effectively evaluate these interaction variables. To best approximate, we included both parts of the interaction variable as unique variables. As the original model included Interaction between anxiety disorder and personality disorders diagnoses in the last 24 months, we instead included both anxiety disorder diagnosis and personality disorders diagnoses in the last 24 months. Similarly, as the original model included Interaction between Divorced and Male and the Interaction between Widowed and Male, we instead included Divorced, Male, and Widowed.

^bAK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY.

^cThe original 61-variable suicide risk model also included 'Head/neck cancer diagnosis in last 12 months' and 'First VHA visit in last 5 years was in the prior year'. 'Head/neck cancer diagnosis in last 12 months' was excluded from our model due to collinearity with the 'Head/neck cancer diagnosis in last 24 months'. 'First VHA visit in last 5 years was in the prior year' was excluded from our model as we did not have access to data going back 5 years.

Table 2.

Selected features for each sub-cohort

	3-month cohort	6-month cohort	9-month cohort	12-month cohort
<i>N</i> (training set) CASE/CONTROL	22/66	42/148	48/198	52/245
<i>N</i> (test set) CASE/CONTROL	11/33	21/74	24/99	26/123
Selected variables	Food ^a Ritual ^a Food ^{a,b}	Male ^a Anger ^c Vice ^a Male ^{a,b}	Arousal ^a Arousal ^{a,b}	Boredom ^d Ipadj ^a Boredom ^c Interpersonal Adjectives ^{a,b}
AUC	52.75	48.58	46.04	58.04
95% CI (DeLong)	31.89–73.62%	33.98–63.19%	33.17–58.91%	51.15–64.93%

Table presents SÉANCE (Crossley et al., 2017) selected features. Features were selected by a Lasso-penalized generalized linear mixed model that used BIC to select tuning parameters. Data were randomly divided into training (2/3) and test (1/3) sets. The following table presents evaluation of features using the test set.

^a Features were selected from the GID (Stone et al., 1966) semantic dictionary.

^b Repeated items correspond to positive and negative articulations of selected features.

^c Features were selected from EmoLex (Mohammad & Turney, 2010, 2013) semantic dictionary.

^d Features were selected from Geneva Affect Label Coder (GALC; Scherer, 2005) semantic dictionary.