BMJ Open

Suicide theory-guided natural language processing of clinical progress notes to improve prediction of veteran suicide risk: protocol for a mixed-method study

Esther Lydia Meerwijk ,[1] Suzanne R Tamang,[1,2] Andrea K Finlay,[1,3,4] Mark A Ilgen,[5,6] Ruth M Reeves,[7,8] Alex H S Harris[1,9]

To cite: Meerwijk EL, Tamang SR, Finlay AK, *et al*. Suicide theory-guided natural language processing of clinical progress notes to improve prediction of veteran suicide risk: protocol for a mixed-method study. *BMJ Open* 2022;**12**:e065088. doi:10.1136/bmjopen-2022-065088

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-065088).

Received 24 May 2022
Accepted 02 August 2022

Check for updates

For numbered affiliations see end of article.

Correspondence to
Dr Esther Lydia Meerwijk; esther.meerwijk@va.gov

ABSTRACT

Introduction The state-of-the-art 3-step Theory of Suicide (3ST) describes why people consider suicide and who will act on their suicidal thoughts and attempt suicide. The central concepts of 3ST—psychological pain, hopelessness, connectedness, and capacity for suicide—are among the most important drivers of suicidal behaviour but they are missing from clinical suicide risk prediction models in use at the US Veterans Health Administration (VHA). These four concepts are not systematically recorded in structured fields of VHA's electronic healthcare records. Therefore, this study will develop a domain-specific ontology that will enable automated extraction of these concepts from clinical progress notes using natural language processing (NLP), and test whether NLP-based predictors for these concepts improve accuracy of existing VHA suicide risk prediction models.

Methods and analysis Our mixed-method study has an exploratory sequential design where a qualitative component (aim 1) will inform quantitative analyses (aims 2 and 3). For aim 1, subject matter experts will manually annotate progress notes of clinical encounters with veterans who attempted or died by suicide to develop a domain-specific ontology for the 3ST concepts. During aim 2, we will use NLP to machine-annotate clinical progress notes and derive longitudinal representations for each patient with respect to the presence and intensity of hopelessness, psychological pain, connectedness and capacity for suicide in temporal proximity of suicide attempts and deaths by suicide. These longitudinal representations will be evaluated during aim 3 for their ability to improve existing VHA prediction models of suicide and suicide attempts, STORM (Stratification Tool for Opioid Risk Mitigation) and REACHVET (Recovery Engagement and Coordination for Health - Veterans Enhanced Treatment).

Ethics and dissemination Ethics approval for this study was granted by the Stanford University Institutional Review Board and the Research and Development Committee of the VA Palo Alto Health Care System. Results of the study will be disseminated through several outlets, including peer-reviewed publications and presentations at national conferences.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Our study pairs state-of-the-art natural language processing methods to extract suicide-relevant information from unstructured clinical progress notes with a state-of-the-art theoretical framework—the 3-Step Theory of Suicide—that will guide interpretation of that information regarding who is most likely to act on their suicidal thoughts.
- ⇒ Using expert-derived annotation schema in combination with machine-generated embedding models will help us to capture diverse language forms that correspond with the concepts of interest—psychological pain, hopelessness, connectedness and capacity for suicide.
- ⇒ By implementing multiple approaches to quantify the longitudinal nature of the concepts of interest, we will be able to evaluate which approach is most sensitive to change in proximity of a suicide attempt or death by suicide.
- ⇒ It is unknown if the measures we take to expand annotations for the concepts of interest from selected Veterans Health Administration (VHA) stations will suffice to capture the variation and complexity of clinical language used across all VHA.
- ⇒ While the results of this study are relevant to 6 million veterans who receive healthcare through the VHA, this study does not reach veterans with thoughts of suicide who receive their healthcare outside the VHA network.

INTRODUCTION

Reducing suicide and suicide attempts among veterans is a priority in the USA, as 17 veterans die by suicide every day and 27 attempt suicide.[1–3] Predicting if and especially when a veteran will attempt suicide is among the hardest tasks that clinicians face. Even among patients with many risk factors, most do not attempt suicide or die by suicide.[4] Suicide prediction models rely in large part on known risk and protective factors for suicidal behaviour.[5] Despite our knowledge of these factors, prediction models have done little better than chance in accurately predicting suicide or suicide attempts.[6] The poor performance of suicide prediction models so far suggests that critical drivers of suicidal behaviours described in various theories of suicide,[7–11] like hopelessness, psychological pain, connectedness and capacity for suicide, are not adequately captured by those models. This includes suicide prediction models currently used in clinical practice at the Veterans Health Administration (VHA).[12–14] These models are largely based on structured administrative data from the electronic health records (EHR), such as demographics, psychiatric and medical diagnoses, prescribed medication (eg, opioids, antidepressants), health services provided (eg, hospital discharges, emergency room visits, visits for mental health and substance use services), and prior severe adverse events (eg, suicide attempts, accidental overdose).

Building on work by Durkheim, Shneidman and Joiner, the state-of-the-art 3-step Theory of Suicide (3ST) describes why people consider suicide, and it also distinguishes individuals who will and who will not act on their suicidal thoughts and attempt suicide.[15] Empirical studies corroborate the central tenets of 3ST that (1) hopelessness and psychological pain distinguish people with and without suicide ideation, (2) connectedness is protective against suicidal behaviour and (3) capacity for suicide distinguishes people who attempt suicide from people who do not attempt suicide.[15–20] The main concepts that make up 3ST—hopelessness, psychological pain, connectedness and capacity for suicide—are not systematically recorded in structured fields of the VHA EHR, which may explain why they have not previously been included in VHA's suicide risk prediction models. The VA/DoD clinical practice guidelines for the assessment and management of suicidal patients does recommend that these concepts be evaluated during suicide risk assessments,[21] and our preliminary research has shown that clinicians do record the results of these assessments in clinical progress notes. Unfortunately, clinical progress notes consist of mostly unstructured free texts and their contents are not easily incorporated in suicide prediction models. Information extraction is an area of computer science that involves the development and evaluation of computational techniques to transform unstructured data into structured information that can be used for predictive analytics. Studies using natural language processing (NLP) and other data science methods have shown that suicide-related information can be automatically identified and extracted from large bodies of unstructured text, like clinical progress notes.[22–25] However, none of these studies systematically extracted critical drivers of suicidal behaviour that are suggested by current suicide theory and which are currently missing in suicide prediction models.

Information extraction from clinical progress notes requires NLP as an integral part of the data processing pipeline. Natural language refers to language that is written and spoken, and NLP typically refers to the branch of computer science that aims for computers to understand natural language, as well as to the methods used to achieve that aim. Clinical progress notes are highly complex and include both structured (templated) and unstructured information that must be woven together to create a digital phenotyping algorithm. In recent years, efforts have been devoted to developing NLP methods to extract information from clinical notes, which have resulted in many publicly available NLP systems.[26 27] However, subtle relationships between clinical concepts remain difficult to extract due to the complexity and heterogeneity of the language used and the lack of explicit semantic resources codifying those relationships. To extract this type of specialised information, a combination of domain knowledge and deeper semantic and syntactic analysis of clinical progress notes is required. Use of NLP requires that relevant terms and phrases from the progress notes be mapped to a common language, or ontology. An ontology is a description of concepts and their relationships pertaining to a certain domain, in this case the domain of suicide. Even the well-known and widely used medical ontology Systemized Nomenclature of Medicine[28] does not capture the full complexity of clinical progress notes, and an ontology that captures information related to the concepts of psychological pain, hopelessness, connectedness and capacity for suicide does not exist. Without a domain-specific ontology that describes how language in clinical notes maps to these concepts, NLP cannot reliably classify the potentially potent signals of suicide risk suggested by suicide theory.

Given the aforementioned absence of critical drivers of suicidal behaviours from current VHA suicide risk prediction models and lack of a domain specific ontology, our mixed-method study has the following specific aims: (1) develop a suicide-specific ontology for machine recognition of the 3ST concepts hopelessness, connectedness, psychological pain and capacity for suicide in progress notes of clinical encounters with veterans who attempted or died by suicide, (2) extract information on the presence and intensity of hopelessness, connectedness, psychological pain and capacity for suicide in clinical progress notes and describe change in these concepts in temporal proximity of a suicide or suicide attempt, and (3) determine the predictive validity of hopelessness, connectedness, psychological pain and capacity for suicide regarding veteran suicide attempt and mortality. The NLP and information extraction tools that will be developed will have broad application to the millions of veterans who receive care through the VHA. These tools include ontology-driven representations for the 3ST concepts, which we will derive from expert-driven qualitative analysis and annotation. These representations can be used to translate clinical note content into numerical features that may be predictive of suicide mortality and will be readily available for existing tools to identify veterans at high risk of suicide, such as STORM (Stratification Tool for Opioid Risk Mitigation) and REACHVET (Recovery Engagement and Coordination for Health

- Veterans Enhanced Treatment), which VHA suicide prevention coordinators use daily.[12 29] Enhancing these risk prediction models with potentially potent predictors of suicidal behaviours is likely to improve their accuracy and would ensure that limited intervention resources can be focused on veterans with the highest risk before they attempt or die by suicide.

METHODS AND ANALYSIS

Our mixed-method study has an exploratory sequential design where a qualitative component (aim 1) will inform quantitative analyses (aims 2 and 3). EHR data from VHA's Corporate Data Warehouse (CDW) will be used to address the aims. Our study will not require recruitment of veterans to collect data. Total duration of the study is 3.5 years (June 2021–December 2024).

Aim 1

The goal of aim 1 is to develop a domain-specific ontology for the 3ST concepts hopelessness, psychological pain, connectedness and capacity for suicide, using progress notes of clinical encounters with Veterans who attempted or died by suicide. The deliverables for this aim will be (1) terms-to-concept maps containing instances of the 3ST concepts, and (2) an associated schema that can be used to machine-annotate and label instances of 3ST concepts in progress notes. Machine annotation, also known as tagging, is the process of machine identifying all instances of the concepts (ie, the exact string of text, its location within the text, as well as its classification) in a corpus of text. Labelling is the process of machine interpreting the tagged strings of text as positive or negative for the presence of the concepts they represent.

Aim 1 annotation overview and progress note selection

A team of trained annotators, including a subject matter expert (first author EM), will manually annotate VHA clinical progress notes of veterans who attempted or died by suicide for instances of the target concepts. A 'progress note' is any text that clinicians enter into the EHR during or after their encounter with a veteran. There can be multiple notes per day, if the veteran sees more than one clinician during the same visit, and multiple notes per clinician encounter, for example, if a clinician enters an addendum after entering the initial note. VHA clinical progress notes are available electronically through the CDW. An instance of a 3ST concept can be a single word, abbreviation, or a phrase that is indicative of the concept in question. We do not need to identify all instances across the entire universe of VHA progress notes, only the set of unique terms and phrases associated with these concepts. We anticipate that it may take progress notes from up to 1000 veterans before new and unique instances of the concepts in question are no longer identified (saturation). If saturation is reached earlier, that is with notes of fewer veterans, we will stop manually annotating clinical notes.

We will employ purposeful sampling to select 40 stations (regionalized facilities or groups of facilities within the VA's network of healthcare localities) among VHA's 130 stations nationwide and then obtain clinical progress notes for 25 randomly selected veterans who attempted suicide or died by suicide per station. Stations will be selected based on an adequate number of suicide attempts or deaths to ensure data availability, and on coverage across the U.S. to maximise heterogeneity, as words, abbreviations and phrases may vary across the VHA system. Veterans who attempted suicide will be identified in the CDW by means of the ICD-9/10 code for suicide attempt and through their being listed in the Suicide Prevention Action Network database[1] and VA Suicide Behavior Overdose Reports. We will use cause-of-death data from the VA/DoD Suicide Data Repository to identify veterans who died by suicide.[30] We will employ a blocking variable for gender and racial minorities, so that women and non-white veterans will be adequately represented. To ensure that justice-involved veterans are adequately represented, we will employ a similar blocking variable. To focus our annotation efforts on information-rich clinical notes, we will annotate progress notes entered for encounters in mental healthcare (including substance use treatment and homelessness), primary care, general (acute) medicine, and urgent and emergency care, as pilot research suggested that clinical progress notes generated in these settings are most likely to contain language related to suicidal thoughts and behaviours. We will focus on inpatient and outpatient progress notes during a 6-month period before and after the suicide attempt or death by suicide, and we will limit the corpus to notes entered after 2013, as the initial VA/DoD clinical practice guidelines for the assessment and management of suicidal patients was released in 2013 and may have affected how suicide assessments are recorded in VHA clinical progress notes.[21]

Aim 1 annotation process

For manual annotation, we will use the extensible Human Oracle Suite of Tools (eHOST).[31] eHOST is an open-source annotation tool that was developed in collaboration between the University of Utah and the VA Consortium for Healthcare Informatics Research.

The annotation team will work in an iterative fashion and continually discuss results and evaluate new annotations against a growing list of annotations that were made before. Annotators will initially annotate and discuss the same notes by way of training. When having demonstrated an acceptable level of understanding of the annotation process, each annotator will annotate different note sets. All annotations will be discussed by the annotation team. Results from our pilot research and our theoretical framework provide a rudimentary ontology that will serve as an initial annotation schema for the annotation of the target concepts. Figure 1 shows the initial schema for capacity for suicide, and similar schema will be developed for hopelessness, psychological pain and connectedness.

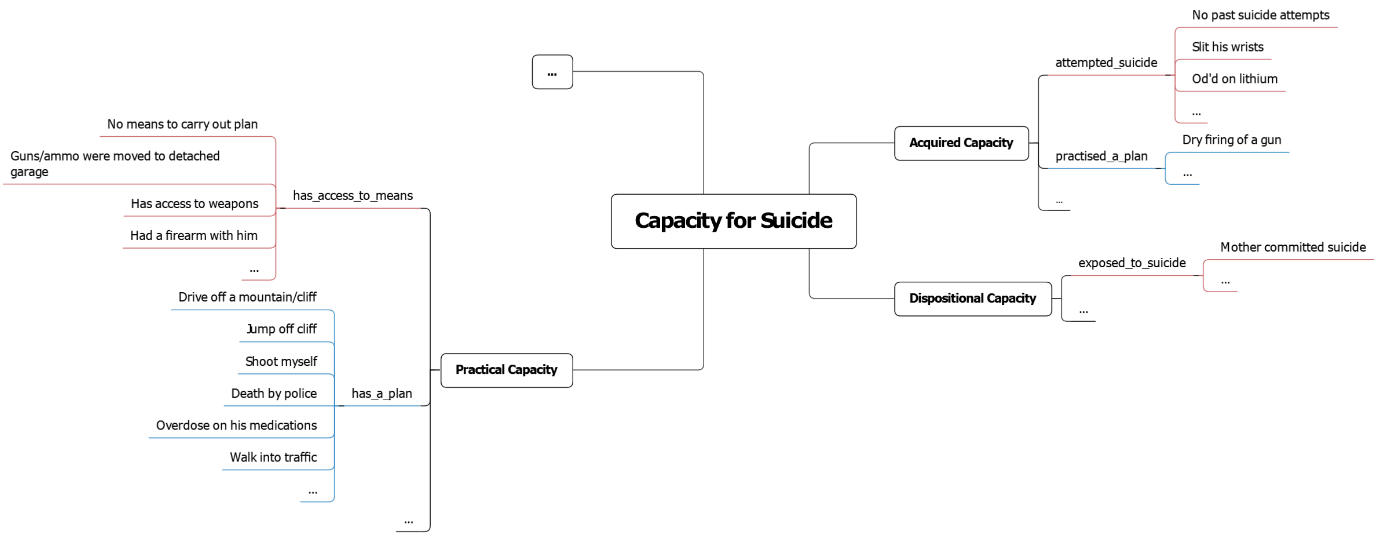

Figure 1 Initial annotation schema for Capacity for Suicide (based on our pilot research), showing annotations, relationships and child concepts.

While the target concepts of 3ST are fixed and well defined, we anticipate that child concepts will emerge from the annotated text that provide more granularity for the target concept in question. For each child concept, we will define its relationship to the annotated text. For example, it can be derived from our theoretical framework that the target concept Capacity for Suicide has three child concepts: (1) Acquired Capacity, (2) Practical Capacity, and (3) Dispositional Capacity (see figure 1). Note text such as '*Slit his wrists*' and '*OD'd on lithium*' would apply to Acquired Capacity with a 'has_attempted_suicide' relationship, whereas text such as '*Had a firearm with him*' and '*Has access to weapons*' would apply to Practical Capacity with an 'access_to_means' relationship. Whenever it is not immediately clear which relationship or child concept should be used for annotation, the team will discuss until consensus is reached. The study team has two suicide experts who can guide this discussion (ELM and MAI). For each annotation, we will indicate whether it evidences the presence or absence of that concept. Each annotator's annotations and relationships will be combined in terms-to-concept maps, one for each 3ST concept, which will be discussed in the project team. The target concepts, child concepts and their relationship to the annotated text constitute a practical ontology that will facilitate machine annotation for 3ST concepts.

Aim 1 embedding models

Using a large corpus of VA clinical progress notes, we will train a word embedding model that can be queried with each of the terms in the initial terms-to-concept maps to identify statistically similar terms that are not in those maps. These new candidate terms may include abbreviations, phrases and lexical variants, including misspellings of the initial terms. The new candidate terms that are closely related, as determined by subject matter experts, will be integrated into the final terms-to-concept maps. For example, our manual annotation process may identify 'divorce' as a term relevant to connectedness. The embedding model may indicate a strong correlation between 'divorce' and 'dvorce' and between 'divorce' and 'divorced'. Recognising 'dvorce' as a potential misspelling of 'divorce' and 'divorced' as past tense of 'divorce', we can then add those terms to our terms-to-concept map for connectedness, so they will be appropriately identified when tagging clinical progress notes in aim 2. Similarly for the concept of psychological pain, the embedding model may indicate a strong correlation between the phrase 'emotional pain' and 'psychological pain'. Knowing that emotional pain is often used as a synonym of psychological pain,[32] we can then add emotional pain as a new term to the terms-to-concept map for psychological pain, so 'emotional pain' will be identified as an instance of psychological pain. Embedding models are essential for solving most modern NLP problems, and we have applied them in prior work to enhance terms-to-concept maps for other clinical areas.[33–35] The process of creating embedding models, that is determining the misspellings and words or phrases that are statistically similar, is based on the proximity of words in clinical progress notes and is relatively straight forward.[36] Because embedding models can contain less relevant terms, our subject matter experts will review the terms that are machine generated for relevancy, before including them in our 3ST terms-to-concept maps. The updated terms-to-concept maps will then be translated into a machine-readable annotation schema that will allow NLP tools to tag and label all instances of the 3ST concepts.

The use of embedding models aims to enhance generalisability of the terms-to-concept maps to the larger corpus of VHA progress notes. To assess the distribution of terms and 3ST concepts in the corpus of VHA notes, we will determine term frequencies of machine-annotated concepts relative to the number of patients and progress notes for all 130 VA stations.

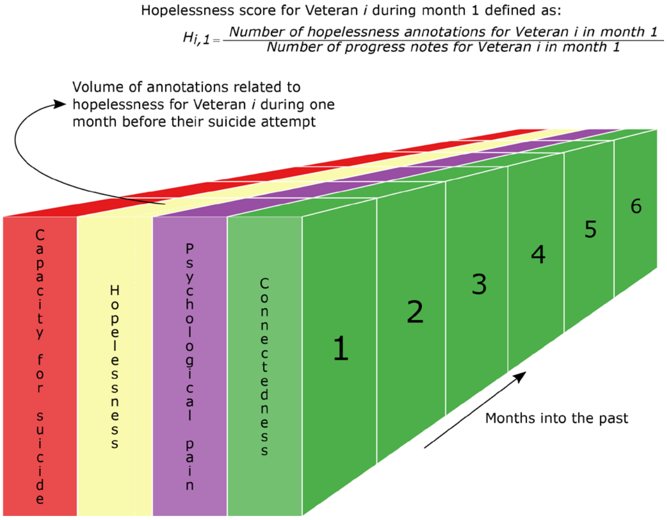
Hopelessness score for Veteran *i* during month 1 defined as:
$H_{i,1} = \frac{\text{Number of hopelessness annotations for Veteran } i \text{ in month 1}}{\text{Number of progress notes for Veteran } i \text{ in month 1}}$
Volume of annotations related to hopelessness for Veteran *i* during one month before their suicide attempt
Capacity for suicide
Hopelessness
Psychological pain
Connectedness
1
2
3
4
5
6
Months into the past

Figure 2 Translation of concept annotations into scores across time, with example of the hopelessness score for one veteran during the month before their suicide or suicide attempt. The numbers indicate months looking back in to the past, with zero (not shown) being the moment of the attempt. Similarly, we will determine scores on a week-by-week basis.

Aim 2

The goal of aim 2 is to derive longitudinal representations for each patient with respect to the presence and intensity of hopelessness, psychological pain, connectedness and capacity for suicide and describe change in these concepts in temporal proximity of suicide attempts and deaths by suicide. The ability of these longitudinal representations to improve accuracy of suicide prediction models will be evaluated in aim 3.

Aim 2 methods

We will determine the longitudinal trajectories of the target concepts in VHA clinical progress notes across 6 months leading up to a suicide attempt or death by suicide. Prior research has shown that the risk of suicide among veterans with VHA utilisation increases during this time frame.[14] Like aim 1, our primary data source will be clinical progress notes from encounters with veterans who attempted or died by suicide after 2013. Based on recent data, we anticipate that this may involve clinical progress notes of about 60 000 veterans.[1 37] We will exclude veterans whose progress notes were used for aim 1, as this will provide a measure of the generalisability of the new ontology, independent from the notes that were used to develop that ontology. For descriptive and comparative analysis, we will also annotate 6 months of clinical notes for veterans without a history of suicide attempt. These veterans will be selected to match the veterans who died by suicide or attempted suicide, in terms of age, gender, race, VA facility and the 6-month time frame.

Using the machine-annotation schema developed for aim 1, we will tag and label the clinical progress notes for the target concepts with a text analysis tool called CLEVER (CL-inical EVE-nt R-ecognizer).[33] Although CLEVER has not been used to detect suicide-related concepts, its applicability to ontology-driven representations and associated expert-rule sets for clinical texts has been demonstrated in several studies.[33 34 38 39] CLEVER was designed to be efficient and flexible for extension by computational scientists or subject matter experts with some coding expertise and a desire to build custom NLP extractors. For each tag, CLEVER will record the applicable semantic relationship, the child and target concept, the surrounding context or 'snippet' entailing the annotation, as well as relevant note metadata such as the clinical note type (eg, mental health note, suicide risk assessment note), entry time, provider type (eg, physician, nurse, social worker), and the patient's ID. CLEVER will also label each tag, indicating whether it is positive or negative for the associated concept, taking into account other common semantic modifiers (eg, history, experiencer).

After all concept information is populated into CLEVER, we will translate each veteran's labelled annotations into a numerical score for each target concept that will be used to test the hypotheses of aim 2. Exploring how to determine concept scores so that they best represent the data over time is part of the proposed study. The scores will be determined for different time windows (eg, week-by-week, month-by-month, going back 6 months prior to the veteran's suicide attempt or death). In determining scores for sequential windows of varying size, we allow for the fact that suicide risk is not necessarily a linear phenomenon. Suicide risk is a dynamic property that may wax and wane depending on internal and external factors.[40 41] The volume of annotations related to each concept depends on the window size; longer windows will capture more clinical encounters and hence more clinical notes and annotations. Therefore, we define concept scores as the number of concept-related annotations over the number of progress notes within a window. Figure 2 visualises this with an example of the score for one veteran's hopelessness during the first month prior to their suicide attempt or death.

Should a window not contain any progress notes, we will carry forward the score determined one window before the current window (ie, going back in time). As 3ST concepts contain child concepts and relationships (see figure 1), we will also explore scores at the child-concept level, defined as the number of child-concept-related annotations over the number of progress notes within a window, and at the relationship level, defined as the number of relationship-related annotations over the number of progress notes within a window. Determining the relationships between annotations and child concepts is done as part of aim 1, and it is unknown beforehand how many will be identified. However, based on our expertise with NLP of clinical notes, we expect the number of relationships and child concepts per 3ST concept to be quite manageable (<< 100). The reason to evaluate multiple approaches to quantify 3ST concepts is that an approach that works for one 3ST concept may not be the best option for another.

Aim 2 hypotheses and analysis

Hierarchical (mixed effects) linear modelling will be used to analyse our repeated measures data across the 6 months leading up to the suicide attempts and deaths. We will address the following hypotheses related to our theoretical framework:

H2a: Hopelessness will increase during the 6 months prior to a suicide attempt or death.

H2b: Psychological pain will increase during the 6 months prior to a suicide attempt or death.

H2c: Connectedness will decrease during the 6 months prior to a suicide attempt or death.

H2d: Capacity for suicide will increase during 6 months prior to a suicide attempt or death.

To test these hypotheses, the fixed-effect part of the models will have the following basic form:

$$Y_{ij} = \text{intercept}_i + \beta_i * \text{time}_{ij}$$

To which random effects will be added for VHA facility and Veterans within those facilities. Y_{ij} represents the outcome (eg, score on hopelessness shown in figure 2) for veteran i, at time j before the suicide or suicide attempt. The time parameter β will be tested for the aim 2 hypotheses. We will test time expressed in weeks and, separately, months. In secondary analyses, we will examine if these relationships differ for veterans without a history of suicide attempt by testing the interaction between time and history of suicide attempt (yes/no).

In descriptive analysis of the target concepts (means, volume of annotations, total number of encounters, etc), we will compare Veteran groups that are of particular interest to VA.[42] Specifically, we will compare data by race, gender, age category, justice-involvement (yes/no), access to means (yes/no) and history of suicide attempt (yes/no).

Aim 3

Traditional approaches to developing suicide prediction models rely in large part on known risk and protective factors.[6 43 44] Suicide prediction models currently used in VHA clinical practice, STORM and REACHVET described elsewhere,[12–14] extract these factors from structured data in the EHR and do not capture the key drivers of suicidal behaviour known from 3ST. We expect that enhancing these suicide prediction models with our longitudinal representations of the 3ST concepts will improve the models' accuracy and will better distinguish who will and who will not act on their suicidal thoughts and potentially die by suicide. The goal of aim 3 is to determine the predictive validity of the 3ST concepts regarding veteran suicide and suicide attempts.

Aim 3 methods

The primary outcomes of the models used for this aim are suicide attempt and death by suicide, as determined by the same data sources described for aim 1. For the analysis, we will predict suicide attempts and deaths during the most recent year for which cause-of-death data are available (currently, 2019). Suicide attempts will be assessed for medical seriousness of the attempt, as promising evidence exists that medically serious suicide attempts may be more predictive of future death by suicide than medically non-serious attempts.[45] Briefly, medically serious suicide attempts are defined as suicide attempts that require hospitalisation for more than 24 hours in combination with one of the following criteria: (1) a highly lethal method was used, like hanging or gunshot, or (2) treatment in specialised unit, like an intensive care unit or burns unit, was required, or (3) surgery under general anaesthesia was required.[46] In contrast to aims 1 and 2, aim 3 will draw from the entire population of VHA patients alive at the beginning of 2019 (> 5 million veterans), including patients without a history of suicide ideation and patients with a history of suicide ideation but who never attempted suicide.

Using the information extraction system developed through aim 1, we will machine-annotate clinical progress notes for the 3ST concepts of all veterans who received care through VHA in 2018. These annotations will then be translated for each veteran and multiple time windows into a numeric score per concept, as described for aim 2. From these scores, we will derive model features and assess their value in the prediction of suicide attempt and mortality. At the most basic level, the score for each 3ST concept in each time window will serve as a feature, but we will also test more complex features based on interactions between concepts (eg, the product of two or more features) and between features from different time windows; a process known as feature engineering.[47] These features will be added to STORM and REACHVET's existing set of predictors, to evaluate whether their accuracy can be improved.[12] This will enable us to evaluate our methods in a general veteran population (REACHVET) and in a high-risk Veteran population (STORM: veterans with active opioid prescriptions). It also enables us to evaluate different outcomes and prediction horizons, as REACHVET predicts the risk of death by suicide within 30 days and STORM predicts the risk of suicide attempt or death by suicide within 1 year.[12 13]

For each risk prediction model, we will create two feature sets that can be explored in predictive analytics. Feature set 1 will contain the predictors that are currently used in STORM, or REACHVET, and will function as our baseline. Feature set 2 will contain the features derived from the 3ST concepts as well as the features from feature set 1. For each set of features, we will construct machine learning models for the respective outcome of STORM and REACHVET, using penalised regression with L1/L2 (elastic net) regularisation, support vector machines and random forests. A recent systematic review supports the use of various machine learning models to further suicide risk prediction.[44]

Aim 3 hypotheses and analysis

We will use 60% of the sample for model training, 20% of the sample for model selection/tuning, and the final 20%

of the data for model evaluation. To increase the number of training events for REACHVET, we will train the models on death by suicide or medically serious suicide attempt, instead of death by suicide alone. Standard classification performance measures will be used to compare the alternative models (eg, precision, recall (sensitivity), specificity, area under the curve). We will address the following hypotheses:

H3a: Models that contain STORM feature set 2 will have better classification accuracy of suicide attempt or death by suicide than models that contain STORM feature set 1.

H3b: Models that contain REACHVET feature set 2 will have better classification accuracy of death by suicide than models that contain REACHVET feature set 1.

In a secondary regression analysis, we will address the following hypothesis underlying 3ST by testing the three-way interaction between connectedness, psychological pain and capacity for suicide:

H3c: Capacity for suicide will moderate the likelihood of death by suicide, such that among Veterans with high psychological pain and a low level of connectedness, those with low capacity for suicide are less likely to die by suicide.

As a secondary analysis, we will assess algorithmic bias of the suicide prediction models with feature set 2 for age, race or gender. Note that variables for age, race, and gender are already included in feature set 2, and that support vector machines, and random forests will automatically identify and include interactions for these subgroups if they are significant.[47] For models based on penalised regression, we will manually add interaction terms with age, race and gender to assess dependency on these variables. If evidence of interaction is found, we will develop separate models for these groups and assess their performance compared with the combined models, as we seek to develop software tools that generalise to the entire population of veterans who receive care through VHA.

Patient and public involvement

To gain important insights from veterans who have received healthcare through VHA, we presented the study proposal at the Palo Alto Veteran and Family Research Council. Their feedback has been incorporated in this protocol, in particular the selection process to ensure that gender and racial minorities, as well as high-risk groups such as justice-involved veterans, will be adequately represented in our sample of clinical progress notes for aim 1. Also, a veteran representative has agreed to participate in the project advisory group to provide guidance on the patient experience and perspective when analysing and interpreting the data.

ETHICS AND DISSEMINATION

Ethics approval for this study was granted by the Stanford University Institutional Review Board (IRB protocol #58991) and the Research and Development Committee of the VA Palo Alto Health Care System (RDIS No. HAX0046), who provided a waiver of consent.

Results of the study will be disseminated through several outlets, including peer-reviewed publications, presentation at national conferences (eg, American Association of Suicidology Annual Conference, VA/DoD Suicide Prevention Conference), and as a cyber seminar in the suicide prevention series of VA Health Services Research & Development. We will be in regular contact with our colleagues at the VA Office of Mental Health and Suicide Prevention to discuss progress and facilitate implementation of our results in the operational versions of STORM and REACHVET.

Data scientists and machine learning experts have applied computing power to large unstructured text corpora, often without regard for suicide theory, in hopes of advancing the ability to predict suicide. Thus far, these approaches have not led to accurate prediction models of suicidal behaviours. The key methodological innovation of our study is to pair a state-of-the-art theoretical framework (3ST) that predicts who is most likely to act on their suicidal thoughts with state-of-the-art methods to extract information from unstructured clinical progress notes. As 3ST concepts are not currently represented in structured VHA patient data, we will develop novel NLP and information extraction tools and apply them to clinical progress notes, the potential of which has not been explored to improve suicide prediction models. While our study uses VHA data, the ontology for 3ST concepts that will be developed may generalise to populations other than veterans (eg, active-duty military) and will be made available to other healthcare systems.

Author affiliations

[1]VA Health Services Research & Development, Center for Innovation to Implementation, VA Palo Alto Health Care System, Palo Alto, California, USA
[2]Department of Biomedical Data Science, Stanford University, Stanford, California, USA
[3]Schar School of Policy and Government, George Mason University, Arlington, Virginia, USA
[4]VA National Center on Homelessness Among Veterans, Durham, North Carolina, USA
[5]Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA
[6]VA Health Services Research & Development, Center for Clinical Management Research, VA Ann Arbor Health Care System, Ann Arbor, Michigan, USA
[7]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[8]VA Health Sevices Research & Development, VA Tennessee Valley Health Care System, Nashville, Tennessee, USA
[9]Stanford–Surgical Policy Improvement Research and Education Center, Stanford University School of Medicine, Stanford, California, USA

Contributors ELM, SRT and AHSH made substantial contributions to the conception and design of the study. All authors significantly contributed to the development of the protocol. All authors read, revised and approved the final manuscript.

Funding This study is funded by U.S. Veterans Health Administration Research & Development grant SDR 19-345.

Disclaimer The funding agency had no role in the design of the study nor in collection, analysis, and interpretation of data or in writing of the manuscript.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

ORCID iD

Esther Lydia Meerwijk http://orcid.org/0000-0002-5367-0978

REFERENCES

1 Hoffmire C, Stephens B, Morley S, *et al*. Va suicide prevention applications network: a national health care system-based suicide event tracking system. *Public Health Rep* 2016;131:816–21.
2 The White House. *Reducing military and veteran suicide: advancing a comprehensive cross-sector, evidence-informed public health strategy*. Washington, DC, 2021.
3 U.S. Department of Veterans Affairs. Office of mental health and suicide prevention. 2021 national veteran suicide prevention annual report; 2021.
4 Klonsky ED, May AM. Differentiating suicide attempters from suicide ideators: a critical frontier for suicidology research. *Suicide Life Threat Behav* 2014;44:1–5.
5 Turecki G, Brent DA. Suicide and suicidal behaviour. *Lancet* 2016;387:1227–39.
6 Franklin JC, Ribeiro JD, Fox KR, *et al*. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull* 2017;143:187–232.
7 Van Orden KA, Witte TK, Cukrowicz KC, *et al*. The interpersonal theory of suicide. *Psychol Rev* 2010;117:575–600.
8 Maltsberger JT. The descent into suicide. *Int J Psychoanal* 2004;85:653–68.
9 Leenaars AA. Suicide: a multidimensional malaise. *Suicide Life Threat Behav* 1996;26:221–36.
10 Baumeister RF. Suicide as escape from self. *Psychol Rev* 1990;97:90–113.
11 Shneidman ES, suicidology Pon. Perspectives on suicidology. further reflections on suicide and psychache. *Suicide Life Threat Behav* 1998;28:245–50.
12 Oliva EM, Bowe T, Tavakoli S, *et al*. Development and applications of the veterans health administration's stratification tool for opioid risk mitigation (storm) to improve opioid safety and prevent overdose and suicide. *Psychol Serv* 2017;14:34–49.
13 Kessler RC, Hwang I, Hoffmire CA, *et al*. Developing a practical suicide risk prediction model for targeting high-risk patients in the veterans health administration. *Int J Methods Psychiatr Res* 2017;26. doi:10.1002/mpr.1575. [Epub ahead of print: 04 07 2017].
14 McCarthy JF, Bossarte RM, Katz IR, *et al*. Predictive modeling and concentration of the risk of suicide: implications for preventive interventions in the US department of veterans affairs. *Am J Public Health* 2015;105:1935–42.
15 Klonsky ED, May AM. The three-step theory (3ST): a new theory of suicide rooted in the "ideation-to-action" framework. *Int J Cogn Ther* 2015;8:114–29.
16 Yang L, Liu X, Chen W, *et al*. A test of the three-step theory of suicide among Chinese people: a study based on the ideation-to-action framework. *Arch Suicide Res* 2019;23:648–61.
17 Dhingra K, Klonsky ED, Tapola V. An empirical test of the three-step theory of suicide in U.K. university students. *Suicide Life Threat Behav* 2019;49:478–87.
18 Hagan CR, Muehlenkamp JJ. Retracted: the three-step theory of suicide: an independent replication and conceptual extension. *Suicide Life Threat Behav* 2020;50:751.
19 Wolford-Clevenger C, Flores LY, Stuart GL. Proximal correlates of suicidal ideation among transgender and gender diverse people: a preliminary test of the three-step theory. *Suicide Life Threat Behav* 2021;51:1077–85.
20 Pachkowski MC, Hewitt PL, Klonsky ED. Examining suicidal desire through the lens of the three-step theory: a cross-sectional and longitudinal investigation in a community sample. *J Consult Clin Psychol* 2021;89:1–10.
21 U.S. Departments of Veterans Affairs & Defense. *VA/DoD clinical practice guideline for the assessment and management of patients at risk for suicide. version 2.0 ED*. The Assessment and Management of Suicide Risk Work Group, 2019.
22 Leonard Westgate C, Shiner B, Thompson P, *et al*. Evaluation of veterans' suicide risk with the use of linguistic detection methods. *Psychiatr Serv* 2015;66:1051–6.
23 Hammond KW, Ben-Ari AY, Laundry RJ, *et al*. The feasibility of using large-scale text mining to detect adverse childhood experiences in a VA-Treated population. *J Trauma Stress* 2015;28:505–14.
24 Poulin C, Shiner B, Thompson P, *et al*. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS One* 2014;9:e85733.
25 Levis M, Leonard Westgate C, Gui J, *et al*. Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models. *Psychol Med* 2020:1–10.
26 Reátegui R, Ratté S. Comparison of MetaMap and cTAKES for entity extraction in clinical notes. *BMC Med Inform Decis Mak* 2018;18:74.
27 Chen ES, Hripcsak G, Friedman C. Disseminating natural language processed clinical narratives. *AMIA Annu Symp Proc* 2006:126–30.
28 National Library of Medicine. Overview of the Systemized Nomenclature of medicine (SNOMED) CT, 2022. Available: https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html [Accessed 12 Apr 2022].
29 McCarthy JF, Cooper SA, Dent KR, *et al*. Evaluation of the recovery engagement and coordination for Health-Veterans enhanced treatment suicide risk modeling clinical program in the veterans health administration. *JAMA Netw Open* 2021;4:e2129900.
30 U.S. Defense Suicide Prevention Office. Suicide data Repository. Available: https://www.dspo.mil/About-Suicide/Suicide-Data-Repository/ [Accessed 03 Apr 2019].
31 Leng C. Annotation tool: the extensible human Oracle suite of tools (eHOST). Available: github.com/chrisleng/ehost.
32 Meerwijk EL, Weiss SJ. Toward a unifying definition of psychological pain. *J Loss Trauma* 2011;16:402–12.
33 Ling AY, Kurian AW, Caswell-Jin JL. A Semi-supervised machine learning approach to detecting recurrent metastatic breast cancer cases using linked cancer registry and electronic medical record data. *ArXiv* 2019.
34 Tamang SR, Hernandez-Boussard T, Ross EG, *et al*. Enhanced quality measurement event detection: an application to physician reporting. *EGEMS* 2017;5:5.
35 Mikolov T, Chen K, Corrado G. Efficient estimation of word representations in vector space. *arXiv* 2013.
36 Chen P-H. Essential elements of natural language processing: what the radiologist should know. *Acad Radiol* 2020;27:6-12.
37 U.S. Department of Veterans Affairs. Office of mental health and suicide prevention. 2019 national veteran suicide prevention annual report; 2019.
38 Lupus Erythematosis. *187 application of text mining methods to identify lupus nephritis from electronic health records. 13th International Conference on systemic*. San Francisco, CA, 2019.
39 Tamang S, Patel MI, Blayney DW, *et al*. Detecting unplanned care from clinician notes in electronic health records. *J Oncol Pract* 2015;11:e313–9.
40 Klonsky ED, Saffer BY, Bryan CJ. Ideation-to-action theories of suicide: a conceptual and empirical update. *Curr Opin Psychol* 2018;22:38–43.
41 Bryan CJ, Rudd MD. The importance of temporal dynamics in the transition from suicidal thought to behavior. *Clin Psychol Sci Pract* 2016;23:21–5.
42 VA Health Services Research and Development Service. Health services research and development updated research priorities. Available: https://www.hsrd.research.va.gov/funding/PriorityDomains2019.pdf [Accessed 18 Mar 2019].
43 Belsher BE, Smolenski DJ, Pruitt LD, *et al*. Prediction models for suicide attempts and deaths: a systematic review and simulation. *JAMA Psychiatry* 2019;76:642-651.
44 Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: a systematic review. *J Affect Disord* 2019;245:869–84.
45 Gvion Y, Levi-Belz Y. Serious suicide attempts: systematic review of psychological risk factors. *Front Psychiatry* 2018;9:56.
46 Beautrais AL. Suicides and serious suicide attempts: two populations or one? *Psychol Med* 2001;31:837–45.
47 Kuhn M, Johnson K. *Feature engineering and selection: a practical approach for predictive models*. CRC, 2019.