

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-065088
Article Type:	Protocol
Date Submitted by the Author:	24-May-2022
Complete List of Authors:	Meerwijk, Esther; VA Palo Alto Health Care System, Center for Innovation to Implementation (Ci2i) Tamang, Suzanne R.; VA Palo Alto Health Care System; Stanford University, Department of Biomedical Data Science Finlay, Andrea K.; VA Palo Alto Health Care System; George Mason University, Schar School of Policy and Government Ilgen, Mark A.; VA Ann Arbor Healthcare System, Center for Clinical Management Research; University of Michigan, Department of Psychiatry Reeves, Ruth M.; VA Tennessee Valley Healthcare System; Vanderbilt University Medical Center, Department of Biomedical Informatics Harris, Alex H. S.; VA Palo Alto Health Care System, Center for Innovation to Implementation; Stanford University School of Medicine, Stanford-Surgical Policy Improvement Research and Education Center
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MENTAL HEALTH, PREVENTIVE MEDICINE, Suicide & self-harm < PSYCHIATRY

SCHOLARONE™
Manuscripts

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

Esther L. 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}

¹ VA Health Services Research & Development, Center for Innovation to Implementation (Ci2i), VA Palo Alto Health Care System, Menlo Park, CA, USA.

² Department of Biomedical Data Science, Stanford University, Stanford, CA, USA.

³ VA National Center on Homelessness Among Veterans, USA.

⁴ Schar School of Policy and Government, George Mason University, Arlington, VA, USA.

⁵ VA Health Services Research & Development, Center for Clinical Management Research, Ann Arbor, MI, USA.

⁶ Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA.

⁷ VA Tennessee Valley Healthcare System, Nashville, TN, USA.

⁸ Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA.

⁹ Department of Surgery, Stanford University, Stanford, CA, USA.

Corresponding Author: Esther L. Meerwijk, VA Health Services Research & Development, Center for Innovation to Implementation (Ci2i), VA Palo Alto Health Care System, 795 Willow Road, Menlo Park, CA 94025, USA. E-mail: esther.meerwijk@va.gov, phone: (650) 614-9997 ext. 27275, fax: (650) 617-2736.

1
2
3
4 Word count: 4991
5
6
7
8
9

10 **ABSTRACT**

11
12 **Introduction** The state-of-the-art 3-step Theory of Suicide (3ST) describes why people
13 consider suicide and who will act on their suicidal thoughts and attempt suicide. The central
14 concepts of 3ST - psychological pain, hopelessness, connectedness, and capacity for suicide - are
15 among the most important drivers of suicidal behavior but they are missing from clinical suicide
16 risk prediction models in use at the US Veterans Health Administration (VHA). These four
17 concepts are not systematically recorded in structured fields of VHA's electronic healthcare
18 records. Therefore, this study will develop a domain-specific ontology that will enable automated
19 extraction of these concepts from clinical progress notes using natural language processing
20 (NLP), and test whether NLP-based predictors for these concepts improve accuracy of existing
21 VHA suicide risk prediction models.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 **Methods and analysis** Our mixed-methods study has an exploratory sequential design
39 where a qualitative component (Aim 1) will inform quantitative analyses (Aims 2 and 3). For
40 Aim 1, subject matter experts will manually annotate progress notes of clinical encounters with
41 Veterans who attempted or died by suicide to develop a domain-specific ontology for the 3ST
42 concepts. During Aim 2 we will use NLP to machine-annotate clinical progress notes and derive
43 longitudinal representations for each patient with respect to the presence and intensity of
44 hopelessness, psychological pain, connectedness, and capacity for suicide in temporal proximity
45 of suicide attempts and deaths by suicide. These longitudinal representations will be evaluated
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 during Aim 3 for their ability to improve existing VHA prediction models of suicide and suicide
5 attempts, STORM and REACHVET.
6
7
8
9
10

11 **Ethics and dissemination** Ethics approval for this study was granted by the Stanford
12 University Institutional Review Board and the Research and Development Committee of the VA
13 Palo Alto Health Care System. Results of the study will be disseminated through several outlets,
14 including peer-reviewed publications and presentations at national conferences.
15
16
17
18
19
20
21

22 **STRENGTHS AND LIMITATIONS**

- 23
24 • Our study pairs state-of-the-art Natural Language Processing methods to extract suicide-
25 relevant information from unstructured clinical progress notes with a state-of-the-art
26 theoretical framework – the 3-Step Theory of Suicide - that will guide interpretation of that
27 information regarding who is most likely to act on their suicidal thoughts.
28
29
30
31
32
- 33 • Using expert-derived annotation schema in combination with machine-generated embedding
34 models will help us to capture diverse language forms that correspond with the concepts of
35 interest - psychological pain, hopelessness, connectedness, and capacity for suicide.
36
37
38
39
- 40 • By implementing multiple approaches to quantify the longitudinal nature of the concepts of
41 interest, we will be able to evaluate which approach is most sensitive to change in proximity
42 of a suicide attempt or death by suicide.
43
44
45
46
- 47 • The ontology that will be developed for the concepts of interest and that will enable
48 automated extraction of these concepts from clinical progress notes will apply to contexts
49 other than VHA, where unstructured text is available in electronic form.
50
51
52
53
54
55
56
57
58
59
60

- 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.

Keywords

Natural language processing, clinical progress notes, veterans, 3-step theory of suicide, information extraction, veterans health services.

INTRODUCTION

Reducing suicide and suicide attempts among veterans is a priority in the United States, 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 behavior.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 behaviors 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 (e.g., opioids, antidepressants), health services provided (e.g., hospital discharges, emergency room visits,

1
2
3
4 visits for mental health and substance use services), and prior severe adverse events (e.g., suicide
5 attempts, accidental overdose).
6
7

8
9 Building on work by Durkheim, Shneidman and Joiner, the state-of-the-art 3-step
10 Theory of Suicide (3ST) describes why people consider suicide, and it also distinguishes
11 individuals who will and who will not act on their suicidal thoughts and attempt suicide.¹⁵
12
13 Empirical studies corroborate the central tenets of 3ST that (a) hopelessness and psychological
14 pain distinguish people with and without suicide ideation, (b) connectedness is protective against
15 suicidal behavior, and (c) capacity for suicide distinguishes people who attempt suicide from
16 people who do not attempt suicide.¹⁵⁻²⁰ The main concepts that make up 3ST - hopelessness,
17 psychological pain, connectedness, and capacity for suicide - are not systematically recorded in
18 structured fields of the VHA EHR, which may explain why they have not previously been
19 included in VHA's suicide risk prediction models. The VA/DoD clinical practice guidelines for
20 the assessment and management of suicidal patients does recommend that these concepts be
21 evaluated during suicide risk assessments,²¹ and our preliminary research has shown that
22 clinicians do record the results of these assessments in clinical progress notes. Unfortunately,
23 clinical progress notes consist of mostly unstructured free texts and their contents are not easily
24 incorporated in suicide prediction models. Information extraction is an area of computer science
25 that involves the development and evaluation of computational techniques to transform
26 unstructured data into structured information that can be used for predictive analytics. Studies
27 using natural language processing (NLP) and other data science methods have shown that
28 suicide-related information can be automatically identified and extracted from large bodies of
29 unstructured text, like clinical progress notes.²²⁻²⁵ However, none of these studies systematically
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 extracted critical drivers of suicidal behavior that are suggested by current suicide theory and
5
6 which are currently missing in suicide prediction models.
7

8
9 Information extraction from clinical progress notes requires NLP as an integral part of
10
11 the data processing pipeline. Natural language refers to language that is written and spoken, and
12
13 NLP typically refers to the branch of computer science that aims for computers to understand
14
15 natural language, as well as to the methods used to achieve that aim. Clinical progress notes are
16
17 highly complex and include both structured (templated) and unstructured information that must
18
19 be woven together to create a digital phenotyping algorithm. In recent years, efforts have been
20
21 devoted to developing NLP methods to extract information from clinical notes, which have
22
23 resulted in many publicly available NLP systems.^{26 27} However, subtle relationships between
24
25 clinical concepts remain difficult to extract due to the complexity and heterogeneity of the
26
27 language used and the lack of explicit semantic resources codifying those relationships. To
28
29 extract this type of specialized information, a combination of domain knowledge and deeper
30
31 semantic and syntactic analysis of clinical progress notes is required. Use of NLP requires that
32
33 relevant terms and phrases from the progress notes be mapped to a common language, or
34
35 ontology. An ontology is a description of concepts and their relationships pertaining to a certain
36
37 domain, in this case the domain of suicide. Even the well-known and widely used medical
38
39 ontology Systemized Nomenclature of Medicine (SNOMED)²⁸ does not capture the full
40
41 complexity of clinical progress notes, and an ontology that captures information related to the
42
43 concepts of psychological pain, hopelessness, connectedness, and capacity for suicide does not
44
45 exist. Without a domain-specific ontology that describes how language in clinical notes maps to
46
47 these concepts, NLP cannot reliably classify the potentially potent signals of suicide risk
48
49 suggested by suicide theory.
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Given the aforementioned absence of critical drivers of suicidal behaviors from current VHA suicide risk prediction models and lack of a domain specific ontology, our mixed-methods 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 and REACHVET, which VHA suicide prevention coordinators use daily.^{12 29} Enhancing these risk prediction models with potentially potent predictors of suicidal behaviors 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.

51 52 53 54 55 56 57 58 59 60

METHODS AND ANALYSIS

Our mixed-methods 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

1
2
3
4 Corporate Data Warehouse (CDW) will be used to address the aims. Our study will not require
5
6 recruitment of veterans to collect data. Total duration of the study is 3.5 years (June 2021 –
7
8 December 2024).
9
10

11 12 **Aim 1**

13
14 The goal of Aim 1 is to develop a domain-specific ontology for the 3ST concepts
15
16 hopelessness, psychological pain, connectedness, and capacity for suicide, using progress notes
17
18 of clinical encounters with Veterans who attempted or died by suicide. The deliverables for this
19
20 aim will be 1) terms-to-concept maps containing instances of the 3ST concepts, and (2) an
21
22 associated schema that can be used to machine-annotate and label instances of 3ST concepts in
23
24 progress notes. Machine-annotation, also known as tagging, is the process of machine-
25
26 identifying all instances of the concepts (i.e., the exact string of text, its location within the text,
27
28 as well as its classification) in a corpus of text. Labelling is the process of machine-interpreting
29
30 the tagged strings of text as positive or negative for the presence of the concepts they represent.
31
32
33
34
35

36 Aim 1 Annotation Overview and Progress Note Selection

37
38 A team of trained annotators, including a subject matter expert (first author EM), will
39
40 manually annotate VHA clinical progress notes of Veterans who attempted or died by suicide for
41
42 instances of the target concepts. A ‘progress note’ is any text that clinicians enter into the EHR
43
44 during or after their encounter with a Veteran. There can be multiple notes per day, if the
45
46 Veteran sees more than one clinician during the same visit, and multiple notes per clinician
47
48 encounter, for example if a clinician enters an addendum after entering the initial note. VHA
49
50 clinical progress notes are available electronically through the CDW. An instance of a 3ST
51
52 concept can be a single word, abbreviation, or a phrase that is indicative of the concept in
53
54
55
56
57
58
59
60

1
2
3
4 question. We do not need to identify all instances across the entire universe of VHA progress
5 notes, only the set of unique terms and phrases associated with these concepts. We anticipate that
6
7 it may take progress notes from up to 1000 Veterans before new and unique instances of the
8
9 concepts in question are no longer identified (saturation). If saturation is reached earlier, that is
10
11 with notes of fewer Veterans, we will stop manually annotating clinical notes.
12
13
14
15

16 We will employ purposeful sampling to select 40 stations (regionalized facilities or
17 groups of facilities within the VA's network of healthcare localities) among VHA's 130 stations
18 nationwide and then obtain clinical progress notes for 25 randomly selected Veterans who
19 attempted suicide or died by suicide per station. Stations will be selected based on an adequate
20 number of suicide attempts or deaths to ensure data availability, and on coverage across the U.S.
21 to maximize heterogeneity, as words, abbreviations, and phrases may vary across the VHA
22 system. Veterans who attempted suicide will be identified in the CDW by means of the ICD-9/10
23 code for suicide attempt and through their being listed in the Suicide Prevention Action Network
24 (SPAN) database¹ and VA Suicide Behavior Overdose Reports. We will use cause-of-death data
25 from the VA/DoD Suicide Data Repository to identify Veterans who died by suicide.³⁰ We will
26 employ a blocking variable for gender and racial minorities, so that women and non-white
27 Veterans will be adequately represented. To ensure that justice-involved Veterans are adequately
28 represented we will employ a similar blocking variable. To focus our annotation efforts on
29 information-rich clinical notes, we will annotate progress notes entered for encounters in mental
30 health care (including substance use treatment and homelessness), primary care, general (acute)
31 medicine, and urgent and emergency care, as pilot research suggested that clinical progress notes
32 generated in these settings are most likely to contain language related to suicidal thoughts and
33 behaviors. We will focus on inpatient and outpatient progress notes during a six-month period
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 before and after the suicide attempt or death by suicide, and we will limit the corpus to notes
5 entered after 2013, as the initial VA/DoD clinical practice guidelines for the assessment and
6 management of suicidal patients was released in 2013 and may have affected how suicide
7 assessments are recorded in VHA clinical progress notes.²¹
8
9
10
11
12
13

14 Aim 1 Annotation Process

15
16 For manual annotation, we will use the extensible Human Oracle Suite of Tools
17 (eHOST).³¹ eHOST is an open-source annotation tool that was developed in collaboration
18 between the University of Utah and the VA Consortium for Healthcare Informatics Research.
19
20
21
22

23
24 The annotation team will work in an iterative fashion and continually discuss results
25 and evaluate new annotations against a growing list of annotations that were made before.
26
27

28 Annotators will initially annotate and discuss the same notes by way of training. When having
29 demonstrated an acceptable level of understanding of the annotation process, each annotator will
30 annotate different note sets. All annotations will be discussed by the annotation team. Results
31 from our pilot research and our theoretical framework provide a rudimentary ontology that will
32 serve as an initial annotation schema for the annotation of the target concepts. Fig. 1 shows the
33 initial schema for Capacity for Suicide, and similar schema will be developed for hopelessness,
34 psychological pain, and connectedness.
35
36
37
38
39
40
41
42
43

44 While the target concepts of 3ST are fixed and well-defined, we anticipate that child
45 concepts will emerge from the annotated text that provide more granularity for the target concept
46 in question. For each child concept, we will define its relationship to the annotated text. For
47 example, it can be derived from our theoretical framework that the target concept Capacity for
48 Suicide has three child-concepts: 1) Acquired Capacity, 2) Practical Capacity, and 3)
49 Dispositional Capacity (see Figure 1). Note text such as '*Slit his wrists*' and '*OD'd on lithium*'
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 would apply to Acquired Capacity with a ‘has_attempted_suicide’ relationship, whereas text
5
6 such as ‘*Had a firearm with him*’ and ‘*Has access to weapons*’ would apply to Practical Capacity
7
8 with an ‘access_to_means’ relationship. Whenever it is not immediately clear which relationship
9
10 or child concept should be used for annotation, the team will discuss until consensus is reached.
11
12
13 The study team has two suicide experts who can guide this discussion (EM and MI). For each
14
15 annotation, we will indicate whether it evidences the presence or absence of that concept. Each
16
17 annotator’s annotations and relationships will be combined in terms-to-concept maps, one for
18
19 each 3ST concept, which will be discussed in the project team. The target concepts, child
20
21 concepts and their relationship to the annotated text constitute a practical ontology that will
22
23 facilitate machine annotation for 3ST concepts.
24
25
26
27
28
29

30 Figure 1, full width, about here.
31
32

33 Aim 1 Embedding Models 34

35 Using a large corpus of VA clinical progress notes, we will train a word embedding
36
37 model that can be queried with each of the terms in the initial terms-to-concept maps to identify
38
39 statistically similar terms that are not in those maps. These new candidate terms may include
40
41 abbreviations, phrases, and lexical variants, including misspellings of the initial terms. The new
42
43 candidate terms that are closely related, as determined by subject matter experts, will be
44
45 integrated into the final terms-to-concept maps. For example, our manual annotation process may
46
47 identify ‘divorce’ as a term relevant to connectedness. The embedding model may indicate a
48
49 strong correlation between ‘divorce’ and ‘dvorce’ and between ‘divorce’ and ‘divorced’.
50
51
52 Recognizing ‘dvorce’ as a potential misspelling of ‘divorce’ and ‘divorced’ as past tense of
53
54 ‘divorce’, we can then add those terms to our terms-to-concept map for connectedness, so they
55
56
57
58
59
60

1
2
3
4 will be appropriately identified when tagging clinical progress notes in Aim 2. Similarly for the
5
6 concept of psychological pain, the embedding model may indicate a strong correlation between
7
8 the phrase ‘emotional pain’ and ‘psychological pain’. Knowing that emotional pain is often used
9
10 as a synonym of psychological pain,³² we can then add emotional pain as a new term to the
11
12 terms-to-concept map for psychological pain, so ‘emotional pain’ will be identified as an
13
14 instance of psychological pain. Embedding models are essential for solving most modern NLP
15
16 problems, and we have applied them in prior work to enhance terms-to-concept maps for other
17
18 clinical areas.³³⁻³⁵ The process of creating embedding models, that is determining the
19
20 misspellings and words or phrases that are statistically similar, is based on the proximity of
21
22 words in clinical progress notes and is relatively straight forward.³⁶ Because embedding models
23
24 can contain less relevant terms, our subject matter experts will review the terms that are
25
26 machine-generated for relevancy, before including them in our 3ST terms-to-concept maps. The
27
28 updated terms-to-concept maps will then be translated into a machine-readable annotation
29
30 schema that will allow NLP tools to tag and label all instances of the 3ST concepts.
31
32
33
34
35

36
37 The use of embedding models aims to enhance generalizability of the terms-to-concept
38
39 maps to the larger corpus of VHA progress notes. To assess the distribution of terms and 3ST
40
41 concepts in the corpus of VHA notes, we will determine term frequencies of machine-annotated
42
43 concepts relative to the number of patients and progress notes for all 130 VA stations.
44
45

46 47 **Aim 2**

48
49 The goal of Aim 2 is to derive longitudinal representations for each patient with respect
50
51 to the presence and intensity of hopelessness, psychological pain, connectedness, and capacity
52
53 for suicide and describe change in these concepts in temporal proximity of suicide attempts and
54
55
56
57
58
59
60

1
2
3
4 deaths by suicide. The ability of these longitudinal representations to improve accuracy of
5
6 suicide prediction models will be evaluated in Aim 3.
7
8
9

10 Aim 2 Methods

11
12 We will determine the longitudinal trajectories of the target concepts in VHA clinical
13
14 progress notes across 6 months leading up to a suicide attempt or death by suicide. Prior research
15
16 has shown that the risk of suicide among Veterans with VHA utilization increases during this
17
18 time frame.¹⁴ Like Aim 1, our primary data source will be clinical progress notes from
19
20 encounters with Veterans who attempted or died by suicide after 2013. Based on recent data, we
21
22 anticipate that this may involve clinical progress notes of about 60,000 Veterans.^{1 37} We will
23
24 exclude Veterans whose progress notes were used for Aim 1, as this will provide a measure of
25
26 the generalizability of the new ontology, independent from the notes that were used to develop
27
28 that ontology. For descriptive and comparative analysis, we will also annotate six months of
29
30 clinical notes for Veterans without a history of suicide attempt. These Veterans will be selected
31
32 to match the Veterans who died by suicide or attempted suicide, in terms of age, gender, race,
33
34 VA facility, and the six-month time frame.
35
36
37
38
39

40 Using the machine-annotation schema developed for Aim 1, we will tag and label the
41
42 clinical progress notes for the target concepts with a text analysis tool called CLEVER (CL-
43
44 inical EVE-nt R-ecognizer).³³ Although CLEVER has not been used to detect suicide-related
45
46 concepts, its applicability to ontology-driven representations and associated expert-rule sets for
47
48 clinical texts has been demonstrated in several studies.^{33 34 38 39} CLEVER was designed to be
49
50 efficient and flexible for extension by computational scientists or subject matter experts with
51
52 some coding expertise and a desire to build custom NLP extractors. For each tag, CLEVER will
53
54 record the applicable semantic relationship, the child and target concept, the surrounding context
55
56
57
58
59
60

1
2
3
4 or ‘snippet’ entailing the annotation, as well as relevant note meta-data such as the clinical note
5
6 type (e.g. mental health note, suicide risk assessment note), entry time, provider type (e.g.
7
8 physician, nurse, social worker), and the patient’s ID. CLEVER will also label each tag,
9
10 indicating whether it is positive or negative for the associated concept, taking into account other
11
12 common semantic modifiers (e.g., history, experiencer).
13
14

15
16 After all concept information is populated into CLEVER, we will translate each
17
18 Veteran’s labelled annotations into a numerical score for each target concept that will be used to
19
20 test the hypotheses of Aim 2. Exploring how to determine concept scores so that they best
21
22 represent the data over time is part of the proposed study. The scores will be determined for
23
24 different time windows (e.g., week-by-week, month-by-month, going back six months prior to
25
26 the Veteran’s suicide attempt or death). In determining scores for sequential windows of varying
27
28 size, we allow for the fact that suicide risk is not necessarily a linear phenomenon. Suicide risk is
29
30 a dynamic property that may wax and wane depending on internal and external factors.^{40 41} The
31
32 volume of annotations related to each concept depends on the window size; longer windows will
33
34 capture more clinical encounters and hence more clinical notes and annotations. Therefore, we
35
36 define concept scores as the number of concept-related annotations over the number of progress
37
38 notes within a window. Figure 2 visualizes this with an example of the score for one Veteran’s
39
40 hopelessness during the first month prior to their suicide attempt or death.
41
42
43
44
45
46
47

48 Figure 2, full width, about here.
49
50
51

52
53 Should a window not contain any progress notes, we will carry forward the score
54
55 determined one window before the current window (i.e., going back in time). As 3ST concepts
56
57
58
59
60

1
2
3
4 contain child-concepts and relationships (see Figure 2), we will also explore scores at the child-
5
6 concept level, defined as the number of child-concept-related annotations over the number of
7
8 progress notes within a window, and at the relationship level, defined as the number of
9
10 relationship-related annotations over the number of progress notes within a window.
11
12

13 Determining the relationships between annotations and child-concepts is done as part of Aim 1,
14
15 and it is unknown beforehand how many will be identified. However, based on our expertise
16
17 with NLP of clinical notes, we expect the number of relationships and child-concepts per 3ST
18
19 concept to be quite manageable ($\ll 100$). The reason to evaluate multiple approaches to quantify
20
21 3ST concepts is that an approach that works for one 3ST concept may not be the best option for
22
23 another.
24
25
26
27

28 Aim 2 Hypotheses and Analysis

29
30 Hierarchical (mixed effects) linear modelling will be used to analyze our repeated
31
32 measures data across the six months leading up to the suicide attempts and deaths. We will
33
34 address the following hypotheses related to our theoretical framework:
35
36
37
38
39

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

41
42 H2b: Psychological pain will increase during the six months prior to a suicide attempt or death.

43
44 H2c: Connectedness will decrease during the six months prior to a suicide attempt or death.

45
46
47 H2d: Capacity for suicide will increase during six months prior to a suicide attempt or death.
48
49
50

51 To test these hypotheses, the fixed-effect part of the models will have the following
52
53 basic form:
54
55

$$56 Y_{ij} = \text{intercept}_i + \beta_i * \text{time}_{ij}$$

57
58
59
60

1
2
3
4
5
6
7 to which random effects will be added for VHA facility and Veterans within those facilities. Y_{ij}
8 represents the outcome (e.g., score on Hopelessness shown in Figure 2) for Veteran i , at time j
9 before the suicide or suicide attempt. The time parameter β will be tested for the Aim 2
10 hypotheses. We will test time expressed in weeks and, separately, months. In secondary analyses,
11 we will examine if these relationships differ for Veterans without a history of suicide attempt by
12 testing the interaction between time and history of suicide attempt (yes/no).
13
14
15
16
17
18
19

20 In descriptive analysis of the target concepts (means, volume of annotations, total
21 number of encounters etc.), we will compare Veteran groups that are of particular interest to
22 VA.⁴² Specifically, we will compare data by race, gender, age category, justice-involvement
23 (yes/no), access to means (yes/no), and history of suicide attempt (yes/no).
24
25
26
27
28
29

30 **Aim 3**

31
32
33 Traditional approaches to developing suicide prediction models rely in large part on
34 known risk and protective factors.^{6 43 44} Suicide prediction models currently used in VHA clinical
35 practice, STORM and REACHVET described elsewhere,¹²⁻¹⁴ extract these factors from
36 structured data in the EHR and do not capture the key drivers of suicidal behavior known from
37 3ST. We expect that enhancing these suicide prediction models with our longitudinal
38 representations of the 3ST concepts will improve the models' accuracy and will better
39 distinguish who will and who will not act on their suicidal thoughts and potentially die by
40 suicide. The goal of Aim 3 is to determine the predictive validity of the 3ST concepts regarding
41 Veteran suicide and suicide attempts.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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-nonserious attempts.⁴⁵ Briefly, medically-serious suicide attempts are defined as suicide attempts that require hospitalization for more than 24 hours in combination with one of the following criteria: a) a highly lethal method was used, like hanging or gunshot, or b) treatment in specialized unit, like an ICU or burns unit, was required, or c) surgery under general anesthesia 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 (e.g., 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

1
2
3
4 existing set of predictors, to evaluate whether their accuracy can be improved.¹² This will enable
5
6 us to evaluate our methods in a general Veteran population (REACHVET) and in a high-risk
7
8 Veteran population (STORM: Veterans with active opioid prescriptions). It also enables us to
9
10 evaluate different outcomes and prediction horizons, as REACHVET predicts the risk of death
11
12 by suicide within 30 days and STORM predicts the risk of suicide attempt or death by suicide
13
14 within 1 year.^{12 13}
15
16

17
18 For each risk prediction model, we will create two feature sets that can be explored in
19
20 predictive analytics. Feature Set 1 will contain the predictors that are currently used in STORM,
21
22 or REACHVET, and will function as our baseline. Feature Set 2 will contain the features derived
23
24 from the 3ST concepts as well as the features from Feature Set 1. For each set of features, we
25
26 will construct machine learning models for the respective outcome of STORM and REACHVET,
27
28 using penalized regression with L1/L2 (elastic net) regularization, support vector machines, and
29
30 random forests. A recent systematic review supports the use of various machine learning models
31
32 to further suicide risk prediction.⁴⁴
33
34
35
36

37 Aim 3 Hypotheses and Analysis

38
39 We will use 60% of the sample for model training, 20% of the sample for model
40
41 selection/tuning, and the final 20% of the data for model evaluation. To increase the number of
42
43 training events for REACHVET, we will train the models on death by suicide or medically-
44
45 serious suicide attempt, instead of death by suicide alone. Standard classification performance
46
47 measures will be used to compare the alternative models (e.g., precision, recall (sensitivity),
48
49 specificity, area-under-the-curve). We will address the following hypotheses:
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 H3a: Models that contain STORM Feature Set 2 will have better classification accuracy of
5
6 suicide attempt or death by suicide than models that contain STORM Feature Set 1.
7

8
9 H3b: Models that contain REACHVET Feature Set 2 will have better classification accuracy of
10
11 death by suicide than models that contain REACHVET Feature Set 1.
12
13

14
15
16 In a secondary regression analysis, we will address the following hypothesis underlying
17
18 3ST by testing the three-way interaction between connectedness, psychological pain, and
19
20 capacity for suicide:
21
22

23
24
25 H3c: Capacity for suicide will moderate the likelihood of death by suicide, such that among
26
27 Veterans with high psychological pain and a low level of connectedness, those with low
28
29 capacity for suicide are less likely to die by suicide.
30
31

32
33
34 As a secondary analysis, we will assess algorithmic bias of the suicide prediction
35
36 models with Feature Set 2 for age, race, or gender. Note that variables for age, race, and gender
37
38 are already included in Feature Set 2, and that support vector machines, and random forests will
39
40 automatically identify and include interactions for these subgroups if they are significant.⁴⁷ For
41
42 models based on penalized regression, we will manually add interaction terms with age, race,
43
44 and gender to assess dependency on these variables. If evidence of interaction is found, we will
45
46 develop separate models for these groups and assess their performance compared to the
47
48 combined models, as we seek to develop software tools that generalize to the entire population of
49
50 Veterans who receive care through VHA.
51
52
53
54
55
56
57
58
59
60

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 analyzing 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 (e.g., 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 behaviors. The key methodological innovation of our study is to pair a state-of-the-art

1
2
3
4 theoretical framework (3-step theory of suicide) that predicts who is most likely to act on their
5
6 suicidal thoughts with state-of-the-art methods to extract information from unstructured clinical
7
8 progress notes. As 3ST concepts are not currently represented in structured VHA patient data, we
9
10 will develop novel NLP and information extraction tools and apply them to clinical progress
11
12 notes, the potential of which has not been explored to improve suicide prediction models. While
13
14 our study uses VHA data, the ontology for 3ST concepts that will be developed may generalize
15
16 to populations other than veterans (e.g., active-duty military) and will be made available to other
17
18 healthcare systems.
19
20
21
22
23

24 **Availability of data and materials**

25
26 The datasets analyzed during the current study will not be publicly available due to the
27
28 potential for exposure of personal health information. The ontology of terms related to the 3-Step
29
30 Theory of Suicide will be made available from the corresponding author on reasonable request.
31
32
33
34
35

36 **Competing interests**

37
38 The authors declare that they have no competing interests.
39
40
41

42 **Funding**

43
44 This study is funded by U.S. Veterans Health Administration Research & Development
45
46 grant SDR 19-345. The funding agency had no role in the design of the study nor in collection,
47
48 analysis, and interpretation of data or in writing of the manuscript.
49
50
51
52
53
54
55
56
57
58
59
60

Author contributions

EM, ST, and AH 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.

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(6):816-21. doi: 10.1177/0033354916670133 [published Online First: 2017/01/27]
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):1-5. doi: 10.1111/sltb.12068
5. Turecki G, Brent DA. Suicide and suicidal behaviour. *Lancet* 2016;387(10024):1227-39. doi: 10.1016/S0140-6736(15)00234-2

1
2
3
4 6. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and
5 behaviors: A meta-analysis of 50 years of research. *Psychol Bull* 2017;143(2):187-232. doi:
6
7 10.1037/bul0000084
8
9

10
11
12 7. Van Orden KA, Witte TK, Cukrowicz KC, et al. The interpersonal theory of suicide.
13
14 *Psychol Rev* 2010;117(2):575-600. doi: 10.1037/a0018697
15
16

17
18 8. Maltzberger JT. The descent into suicide. *Int J Psychoanal* 2004;85(Pt 3):653-67.
19
20 doi: 10.1516/002075704774200799 [published Online First: 2004/07/02]
21
22

23
24 9. Leenaars AA. Suicide: a multidimensional malaise. *Suicide Life Threat Behav*
25
26 1996;26(3):221-36. [published Online First: 1996/01/01]
27
28

29
30 10. Baumeister RF. Suicide as escape from self. *Psychol Rev* 1990;97(1):90-113. doi:
31
32 10.1037/0033-295x.97.1.90 [published Online First: 1990/01/01]
33
34

35
36 11. Shneidman ES. Perspectives on suicidology. Further reflections on suicide and
37
38 psychache. *Suicide Life Threat Behav* 1998;28(3):245-50. [published Online First: 1998/11/10]
39

40
41 12. Oliva EM, Bowe T, Tavakoli S, et al. Development and applications of the
42
43 Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation (STORM) to
44
45 improve opioid safety and prevent overdose and suicide. *Psychol Serv* 2017;14(1):34-49. doi:
46
47 10.1037/ser0000099
48
49

50
51 13. Kessler RC, Hwang I, Hoffmire CA, et al. Developing a practical suicide risk
52
53 prediction model for targeting high-risk patients in the Veterans health Administration. *Int J*
54
55 *Methods Psychiatr Res* 2017;26(3) doi: 10.1002/mpr.1575
56
57

1
2
3
4 14. McCarthy JF, Bossarte RM, Katz IR, et al. Predictive Modeling and Concentration
5 of the Risk of Suicide: Implications for Preventive Interventions in the US Department of
6
7 Veterans Affairs. *Am J Public Health* 2015;105(9):1935-42. doi: 10.2105/AJPH.2015.302737
8
9 [published Online First: 2015/06/13]
10
11
12

13
14
15 15. Klonsky ED, May AM. The Three-Step Theory (3ST): A New Theory of Suicide
16 Rooted in the "Ideation-to-Action" Framework. *International Journal of Cognitive Theory*
17
18 2015;8(2):114-29. doi: 10.1521/ijct.2015.8.2.114
19
20
21

22
23 16. Yang L, Liu X, Chen W, et al. A Test of the Three-Step Theory of Suicide among
24 Chinese People: A Study Based on the Ideation-to-Action Framework. *Arch Suicide Res*
25
26 2019;23(4):648-61. doi: 10.1080/13811118.2018.1497563 [published Online First: 2018/07/20]
27
28
29

30
31 17. Dhingra K, Klonsky ED, Tapola V. An Empirical Test of the Three-Step Theory of
32 Suicide in U.K. University Students. *Suicide Life Threat Behav* 2019;49(2):478-87. doi:
33
34 10.1111/sltb.12437 [published Online First: 2018/04/01]
35
36
37

38
39 18. Hagan CR, Muehlenkamp JJ. The Three-Step Theory of Suicide: An Independent
40 Replication and Conceptual Extension. *Suicide Life Threat Behav* 2019 doi: 10.1111/sltb.12590
41
42 [published Online First: 2019/10/01]
43
44
45

46
47 19. Wolford-Clevenger C, Flores LY, Stuart GL. Proximal correlates of suicidal
48 ideation among transgender and gender diverse people: A preliminary test of the three-step
49 theory. *Suicide Life Threat Behav* 2021 doi: 10.1111/sltb.12790 [published Online First:
50
51 2021/07/14]
52
53
54
55
56
57
58
59
60

1
2
3
4 20. Pachkowski MC, Hewitt PL, Klonsky ED. Examining suicidal desire through the
5 lens of the three-step theory: A cross-sectional and longitudinal investigation in a community
6 sample. *J Consult Clin Psychol* 2021;89(1):1-10. doi: 10.1037/ccp0000546 [published Online
7 First: 2021/01/29]
8
9
10
11
12

13
14 21. U.S. Departments of Veterans Affairs & Defense. VA/DoD Clinical Practice
15 Guideline for the Assessment and Management of Patients at Risk for Suicide. version 2.0 ed:
16 The Assessment and Management of Suicide Risk Work Group, 2019.
17
18
19
20
21

22 22. Leonard Westgate C, Shiner B, Thompson P, et al. Evaluation of Veterans' Suicide
23 Risk With the Use of Linguistic Detection Methods. *Psychiatr Serv* 2015;66(10):1051-6. doi:
24 10.1176/appi.ps.201400283
25
26
27
28
29

30 23. Hammond KW, Ben-Ari AY, Laundry RJ, et al. The Feasibility of Using Large-
31 Scale Text Mining to Detect Adverse Childhood Experiences in a VA-Treated Population. *J*
32 *Trauma Stress* 2015;28(6):505-14. doi: 10.1002/jts.22058
33
34
35
36
37

38 24. Poulin C, Shiner B, Thompson P, et al. Predicting the risk of suicide by analyzing
39 the text of clinical notes. *PLoS One* 2014;9(1):e85733. doi: 10.1371/journal.pone.0085733
40
41
42
43

44 25. Levis M, Leonard Westgate C, Gui J, et al. Natural language processing of clinical
45 mental health notes may add predictive value to existing suicide risk models. *Psychol Med*
46 2020:1-10. doi: 10.1017/S0033291720000173 [published Online First: 2020/02/18]
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

26. Reategui R, Ratte S. Comparison of MetaMap and cTAKES for entity extraction in clinical notes. *BMC Med Inform Decis Mak* 2018;18(Suppl 3):74. doi: 10.1186/s12911-018-0654-2

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 [Available from: https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html accessed April 12, 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(10):e2129900. doi: 10.1001/jamanetworkopen.2021.29900 [published Online First: 2021/10/19]

30. U.S. Defense Suicide Prevention Office. Suicide Data Repository [Available from: <https://www.dspo.mil/About-Suicide/Suicide-Data-Repository/> accessed April 3, 2019].

31. Leng C. Annotation Tool: The extensible Human Oracle Suite of Tools (eHOST) [Available from: github.com/chrisleng/ehost].

32. Meerwijk EL, Weiss SJ. Toward a Unifying Definition of Psychological Pain. *Journal of Loss and Trauma* 2011;16(5):402-12. doi: 10.1080/15325024.2011.572044

1
2
3
4 33. Ling AY, Kurian AW, Caswell-Jin JL, et al. A Semi-Supervised Machine Learning
5 Approach to Detecting Recurrent Metastatic Breast Cancer Cases Using Linked Cancer Registry
6 and Electronic Medical Record Data. *ArXiv:190105958 Cs Q-Bio Stat*, 2019.
7
8
9

10
11
12 34. Tamang SR, Hernandez-Boussard T, Ross EG, et al. Enhanced Quality
13 Measurement Event Detection: An Application to Physician Reporting. *EGEMS (Wash DC)*
14 2017;5(1):5. doi: 10.13063/2327-9214.1270
15
16
17

18
19
20 35. Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations
21 in Vector Space. *arXiv:13013781* 2013.
22
23
24

25
26 36. Chen PH. Essential Elements of Natural Language Processing: What the
27 Radiologist Should Know. *Acad Radiol* 2019 doi: 10.1016/j.acra.2019.08.010 [published Online
28 First: 2019/09/21]
29
30
31

32
33
34 37. U.S. Department of Veterans Affairs - Office of Mental Health and Suicide
35 Prevention. 2019 National Veteran Suicide Prevention Annual Report, 2019.
36
37
38

39 38. 187 Application of text mining methods to identify lupus nephritis from electronic
40 health records. 13th International Conference on Systemic Lupus Erythematosis; 2019; San
41 Francisco, CA.
42
43
44

45
46
47 39. Tamang S, Patel MI, Blayney DW, et al. Detecting unplanned care from clinician
48 notes in electronic health records. *J Oncol Pract* 2015;11(3):e313-9. doi:
49 10.1200/JOP.2014.002741
50
51
52

1
2
3
4 40. Klonsky ED, Saffer BY, Bryan CJ. Ideation-to-action theories of suicide: a
5 conceptual and empirical update. *Curr Opin Psychol* 2018;22:38-43. doi:
6
7 10.1016/j.copsyc.2017.07.020 [published Online First: 2018/08/21]
8
9

10
11
12 41. Bryan CJ, Rudd MD. The Importance of Temporal Dynamics in the Transition
13 From Suicidal Thought to Behavior. *Clinical Psychology Science and Practice* 2016;23(1):21-
14
15 25. doi: <https://doi.org/10.1111/cpsp.12135>
16
17

18
19
20 42. VA Health Services Research and Development Service. Health Services Research
21 and Development Updated Research Priorities.
22
23 <https://www.hsrd.research.va.gov/funding/PriorityDomains2019.pdf> (accessed 3/18/2019).
24
25

26
27
28 43. Belsher BE, Smolenski DJ, Pruitt LD, et al. Prediction Models for Suicide Attempts
29 and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry* 2019 doi:
30
31 10.1001/jamapsychiatry.2019.0174
32
33

34
35
36 44. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study
37 of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *J Affect*
38
39 *Disord* 2019;245:869-84. doi: 10.1016/j.jad.2018.11.073
40
41

42
43
44 45. Gvion Y, Levi-Belz Y. Serious Suicide Attempts: Systematic Review of
45 Psychological Risk Factors. *Front Psychiatry* 2018;9:56. doi: 10.3389/fpsyt.2018.00056
46
47 [published Online First: 2018/03/23]
48
49

1
2
3
4 46. Beautrais AL. Suicides and serious suicide attempts: two populations or one?
5

6
7 *Psychol Med* 2001;31(5):837-45. doi: 10.1017/s0033291701003889 [published Online First:
8
9 2001/07/19]
10

11
12 47. Kuhn M, Johnson K. Feature Engineering and Selection: A Practical Approach for
13
14 Predictive Models: CRC 2019.
15
16

17 18 19 20 **Figure Legends** 21

22
23
24 Figure 1 Initial annotation schema for Capacity for Suicide (based on our pilot
25 research), showing annotations, relationships, and child-concepts.
26
27

28 Figure 2 Translation of concept annotations into scores across time, with example of
29 the hopelessness score for one Veteran during the month before their suicide or suicide attempt.
30 The numbers indicate months looking back in to the past, with zero (not shown) being the
31 moment of the attempt. Similarly, we will determine scores on a week-by-week basis.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

BMJ Open

...

Capacity for Suicide

Acquired Capacity

attempted_suicide

No past suicide attempts

Slit his wrists

Od'd on lithium

...

Dry firing of a gun

...

practised_a_plan

Dispositional Capacity

exposed_to_suicide

Mother committed suicide

...

Practical Capacity

has_access_to_means

No means to carry out plan

Guns/ammo were moved to detached garage

Has access to weapons

Had a firearm with him

...

has_a_plan

Drive off a mountain/cliff

Jump off cliff

Shoot myself

Death by police

Overdose on his medications

Walk into traffic

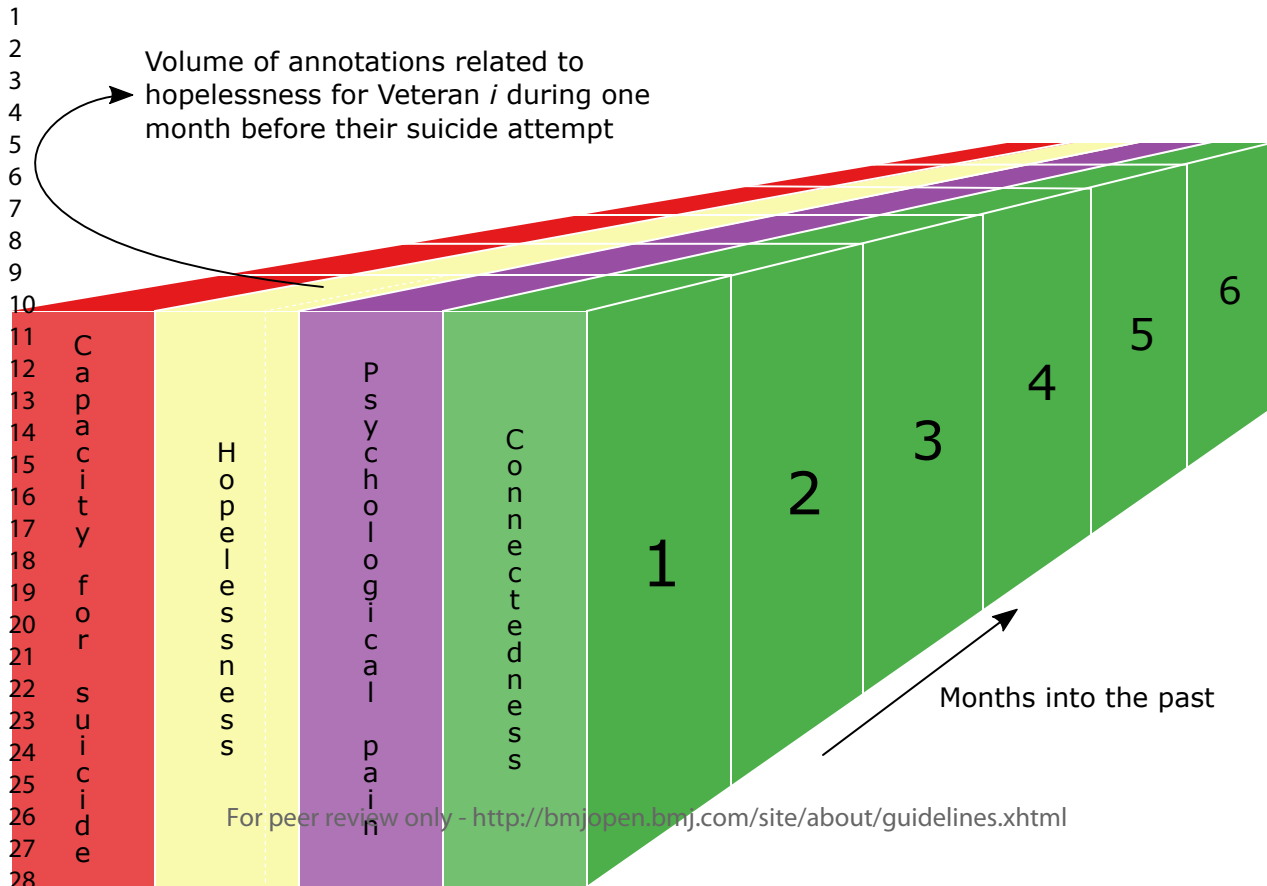
...

...

1
2
3
4
5
6
7
8
9
10
11
12
13
14

Hopelessness score for Veteran i during month 1 defined as:

$$H_{i,1} = \frac{\text{Number of hopelessness annotations for Veteran } i \text{ in month 1}}{\text{Number of progress notes for Veteran } i \text{ in month 1}}$$



BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-065088.R1
Article Type:	Protocol
Date Submitted by the Author:	22-Jul-2022
Complete List of Authors:	Meerwijk, Esther; VA Palo Alto Health Care System, Center for Innovation to Implementation (Ci2i) Tamang, Suzanne R.; VA Palo Alto Health Care System, Center for Innovation to Implementation (Ci2i); Stanford University, Department of Biomedical Data Science Finlay, Andrea K.; VA Palo Alto Health Care System, Center for Innovation to Implementation (Ci2i); George Mason University, Schar School of Policy and Government Ilgen, Mark A.; VA Ann Arbor Healthcare System, Center for Clinical Management Research; University of Michigan, Department of Psychiatry Reeves, Ruth M.; VA Tennessee Valley Healthcare System; Vanderbilt University Medical Center, Department of Biomedical Informatics Harris, Alex H. S.; VA Palo Alto Health Care System, Center for Innovation to Implementation (Ci2i); Stanford University School of Medicine, Stanford–Surgical Policy Improvement Research and Education Center
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Health services research, Health informatics, Mental health
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MENTAL HEALTH, PREVENTIVE MEDICINE, Suicide & self-harm < PSYCHIATRY

SCHOLARONE™
Manuscripts

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

Esther L. 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}

¹ VA Health Services Research & Development, Center for Innovation to Implementation (Ci2i), VA Palo Alto Health Care System, Menlo Park, CA, USA.

² Department of Biomedical Data Science, Stanford University, Stanford, CA, USA.

³ VA National Center on Homelessness Among Veterans, USA.

⁴ Schar School of Policy and Government, George Mason University, Arlington, VA, USA.

⁵ VA Health Services Research & Development, Center for Clinical Management Research, Ann Arbor, MI, USA.

⁶ Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA.

⁷ VA Tennessee Valley Healthcare System, Nashville, TN, USA.

⁸ Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA.

⁹ Department of Surgery, Stanford University, Stanford, CA, USA.

Corresponding Author: Esther L. Meerwijk, VA Health Services Research & Development, Center for Innovation to Implementation (Ci2i), VA Palo Alto Health Care System, 795 Willow Road, Menlo Park, CA 94025, USA. E-mail: esther.meerwijk@va.gov, phone: (650) 614-9997 ext. 27275, fax: (650) 617-2736.

1
2
3
4 Word count: 4991
5
6
7
8
9

10 ABSTRACT

11
12 **Introduction** The state-of-the-art 3-step Theory of Suicide (3ST) describes why people
13 consider suicide and who will act on their suicidal thoughts and attempt suicide. The central
14 concepts of 3ST - psychological pain, hopelessness, connectedness, and capacity for suicide - are
15 among the most important drivers of suicidal behavior but they are missing from clinical suicide
16 risk prediction models in use at the US Veterans Health Administration (VHA). These four
17 concepts are not systematically recorded in structured fields of VHA's electronic healthcare
18 records. Therefore, this study will develop a domain-specific ontology that will enable automated
19 extraction of these concepts from clinical progress notes using natural language processing
20 (NLP), and test whether NLP-based predictors for these concepts improve accuracy of existing
21 VHA suicide risk prediction models.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 **Methods and analysis** Our mixed-methods study has an exploratory sequential design
39 where a qualitative component (Aim 1) will inform quantitative analyses (Aims 2 and 3). For
40 Aim 1, subject matter experts will manually annotate progress notes of clinical encounters with
41 Veterans who attempted or died by suicide to develop a domain-specific ontology for the 3ST
42 concepts. During Aim 2 we will use NLP to machine-annotate clinical progress notes and derive
43 longitudinal representations for each patient with respect to the presence and intensity of
44 hopelessness, psychological pain, connectedness, and capacity for suicide in temporal proximity
45 of suicide attempts and deaths by suicide. These longitudinal representations will be evaluated
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 during Aim 3 for their ability to improve existing VHA prediction models of suicide and suicide
5 attempts, STORM and REACHVET.
6
7
8
9
10

11 **Ethics and dissemination** Ethics approval for this study was granted by the Stanford
12 University Institutional Review Board and the Research and Development Committee of the VA
13 Palo Alto Health Care System. Results of the study will be disseminated through several outlets,
14 including peer-reviewed publications and presentations at national conferences.
15
16
17
18
19
20

21 **STRENGTHS AND LIMITATIONS**

- 22 • Our study pairs state-of-the-art Natural Language Processing methods to extract suicide-
23 relevant information from unstructured clinical progress notes with a state-of-the-art
24 theoretical framework – the 3-Step Theory of Suicide - that will guide interpretation of that
25 information regarding who is most likely to act on their suicidal thoughts.
26
27
- 28 • Using expert-derived annotation schema in combination with machine-generated embedding
29 models will help us to capture diverse language forms that correspond with the concepts of
30 interest - psychological pain, hopelessness, connectedness, and capacity for suicide.
31
32
- 33 • By implementing multiple approaches to quantify the longitudinal nature of the concepts of
34 interest, we will be able to evaluate which approach is most sensitive to change in proximity
35 of a suicide attempt or death by suicide.
36
37
- 38 • It is unknown if the measures we take to expand annotations for the concepts of interest from
39 selected VHA stations will suffice to capture the variation and complexity of clinical
40 language used across all VHA.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 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.

Keywords

Natural language processing, clinical progress notes, veterans, 3-step theory of suicide, information extraction, veterans health services.

INTRODUCTION

Reducing suicide and suicide attempts among veterans is a priority in the United States, 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 behavior.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 behaviors 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 (e.g., opioids, antidepressants), health services provided (e.g., hospital discharges, emergency room visits,

1
2
3
4 visits for mental health and substance use services), and prior severe adverse events (e.g., suicide
5 attempts, accidental overdose).
6
7

8
9 Building on work by Durkheim, Shneidman and Joiner, the state-of-the-art 3-step
10 Theory of Suicide (3ST) describes why people consider suicide, and it also distinguishes
11 individuals who will and who will not act on their suicidal thoughts and attempt suicide.¹⁵
12
13 Empirical studies corroborate the central tenets of 3ST that (a) hopelessness and psychological
14 pain distinguish people with and without suicide ideation, (b) connectedness is protective against
15 suicidal behavior, and (c) capacity for suicide distinguishes people who attempt suicide from
16 people who do not attempt suicide.¹⁵⁻²⁰ The main concepts that make up 3ST - hopelessness,
17 psychological pain, connectedness, and capacity for suicide - are not systematically recorded in
18 structured fields of the VHA EHR, which may explain why they have not previously been
19 included in VHA's suicide risk prediction models. The VA/DoD clinical practice guidelines for
20 the assessment and management of suicidal patients does recommend that these concepts be
21 evaluated during suicide risk assessments,²¹ and our preliminary research has shown that
22 clinicians do record the results of these assessments in clinical progress notes. Unfortunately,
23 clinical progress notes consist of mostly unstructured free texts and their contents are not easily
24 incorporated in suicide prediction models. Information extraction is an area of computer science
25 that involves the development and evaluation of computational techniques to transform
26 unstructured data into structured information that can be used for predictive analytics. Studies
27 using natural language processing (NLP) and other data science methods have shown that
28 suicide-related information can be automatically identified and extracted from large bodies of
29 unstructured text, like clinical progress notes.²²⁻²⁵ However, none of these studies systematically
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 extracted critical drivers of suicidal behavior that are suggested by current suicide theory and
5
6 which are currently missing in suicide prediction models.
7

8
9 Information extraction from clinical progress notes requires NLP as an integral part of
10
11 the data processing pipeline. Natural language refers to language that is written and spoken, and
12
13 NLP typically refers to the branch of computer science that aims for computers to understand
14
15 natural language, as well as to the methods used to achieve that aim. Clinical progress notes are
16
17 highly complex and include both structured (templated) and unstructured information that must
18
19 be woven together to create a digital phenotyping algorithm. In recent years, efforts have been
20
21 devoted to developing NLP methods to extract information from clinical notes, which have
22
23 resulted in many publicly available NLP systems.^{26 27} However, subtle relationships between
24
25 clinical concepts remain difficult to extract due to the complexity and heterogeneity of the
26
27 language used and the lack of explicit semantic resources codifying those relationships. To
28
29 extract this type of specialized information, a combination of domain knowledge and deeper
30
31 semantic and syntactic analysis of clinical progress notes is required. Use of NLP requires that
32
33 relevant terms and phrases from the progress notes be mapped to a common language, or
34
35 ontology. An ontology is a description of concepts and their relationships pertaining to a certain
36
37 domain, in this case the domain of suicide. Even the well-known and widely used medical
38
39 ontology Systemized Nomenclature of Medicine (SNOMED)²⁸ does not capture the full
40
41 complexity of clinical progress notes, and an ontology that captures information related to the
42
43 concepts of psychological pain, hopelessness, connectedness, and capacity for suicide does not
44
45 exist. Without a domain-specific ontology that describes how language in clinical notes maps to
46
47 these concepts, NLP cannot reliably classify the potentially potent signals of suicide risk
48
49 suggested by suicide theory.
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Given the aforementioned absence of critical drivers of suicidal behaviors from current VHA suicide risk prediction models and lack of a domain specific ontology, our mixed-methods 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 and REACHVET, which VHA suicide prevention coordinators use daily.^{12 29} Enhancing these risk prediction models with potentially potent predictors of suicidal behaviors 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.

51 52 53 54 55 56 57 58 59 60

METHODS AND ANALYSIS

Our mixed-methods 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

1
2
3
4 Corporate Data Warehouse (CDW) will be used to address the aims. Our study will not require
5
6 recruitment of veterans to collect data. Total duration of the study is 3.5 years (June 2021 –
7
8 December 2024).
9
10

11 **Aim 1**

12
13
14 The goal of Aim 1 is to develop a domain-specific ontology for the 3ST concepts
15
16 hopelessness, psychological pain, connectedness, and capacity for suicide, using progress notes
17
18 of clinical encounters with Veterans who attempted or died by suicide. The deliverables for this
19
20 aim will be 1) terms-to-concept maps containing instances of the 3ST concepts, and (2) an
21
22 associated schema that can be used to machine-annotate and label instances of 3ST concepts in
23
24 progress notes. Machine-annotation, also known as tagging, is the process of machine-
25
26 identifying all instances of the concepts (i.e., the exact string of text, its location within the text,
27
28 as well as its classification) in a corpus of text. Labelling is the process of machine-interpreting
29
30 the tagged strings of text as positive or negative for the presence of the concepts they represent.
31
32
33
34
35

36 Aim 1 Annotation Overview and Progress Note Selection

37
38 A team of trained annotators, including a subject matter expert (first author EM), will
39
40 manually annotate VHA clinical progress notes of Veterans who attempted or died by suicide for
41
42 instances of the target concepts. A ‘progress note’ is any text that clinicians enter into the EHR
43
44 during or after their encounter with a Veteran. There can be multiple notes per day, if the
45
46 Veteran sees more than one clinician during the same visit, and multiple notes per clinician
47
48 encounter, for example if a clinician enters an addendum after entering the initial note. VHA
49
50 clinical progress notes are available electronically through the CDW. An instance of a 3ST
51
52 concept can be a single word, abbreviation, or a phrase that is indicative of the concept in
53
54
55
56
57
58
59
60

1
2
3
4 question. We do not need to identify all instances across the entire universe of VHA progress
5 notes, only the set of unique terms and phrases associated with these concepts. We anticipate that
6
7 it may take progress notes from up to 1000 Veterans before new and unique instances of the
8
9 concepts in question are no longer identified (saturation). If saturation is reached earlier, that is
10
11 with notes of fewer Veterans, we will stop manually annotating clinical notes.
12
13
14
15

16 We will employ purposeful sampling to select 40 stations (regionalized facilities or
17 groups of facilities within the VA's network of healthcare localities) among VHA's 130 stations
18 nationwide and then obtain clinical progress notes for 25 randomly selected Veterans who
19 attempted suicide or died by suicide per station. Stations will be selected based on an adequate
20 number of suicide attempts or deaths to ensure data availability, and on coverage across the U.S.
21 to maximize heterogeneity, as words, abbreviations, and phrases may vary across the VHA
22 system. Veterans who attempted suicide will be identified in the CDW by means of the ICD-9/10
23 code for suicide attempt and through their being listed in the Suicide Prevention Action Network
24 (SPAN) database¹ and VA Suicide Behavior Overdose Reports. We will use cause-of-death data
25 from the VA/DoD Suicide Data Repository to identify Veterans who died by suicide.³⁰ We will
26 employ a blocking variable for gender and racial minorities, so that women and non-white
27 Veterans will be adequately represented. To ensure that justice-involved Veterans are adequately
28 represented we will employ a similar blocking variable. To focus our annotation efforts on
29 information-rich clinical notes, we will annotate progress notes entered for encounters in mental
30 health care (including substance use treatment and homelessness), primary care, general (acute)
31 medicine, and urgent and emergency care, as pilot research suggested that clinical progress notes
32 generated in these settings are most likely to contain language related to suicidal thoughts and
33 behaviors. We will focus on inpatient and outpatient progress notes during a six-month period
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 before and after the suicide attempt or death by suicide, and we will limit the corpus to notes
5 entered after 2013, as the initial VA/DoD clinical practice guidelines for the assessment and
6 management of suicidal patients was released in 2013 and may have affected how suicide
7 assessments are recorded in VHA clinical progress notes.²¹
8
9
10
11
12
13

14 Aim 1 Annotation Process

15
16 For manual annotation, we will use the extensible Human Oracle Suite of Tools
17 (eHOST).³¹ eHOST is an open-source annotation tool that was developed in collaboration
18 between the University of Utah and the VA Consortium for Healthcare Informatics Research.
19
20
21
22

23 The annotation team will work in an iterative fashion and continually discuss results
24 and evaluate new annotations against a growing list of annotations that were made before.
25
26
27

28 Annotators will initially annotate and discuss the same notes by way of training. When having
29 demonstrated an acceptable level of understanding of the annotation process, each annotator will
30 annotate different note sets. All annotations will be discussed by the annotation team. Results
31 from our pilot research and our theoretical framework provide a rudimentary ontology that will
32 serve as an initial annotation schema for the annotation of the target concepts. Fig. 1 shows the
33 initial schema for Capacity for Suicide, and similar schema will be developed for hopelessness,
34 psychological pain, and connectedness.
35
36
37
38
39
40
41
42
43

44 While the target concepts of 3ST are fixed and well-defined, we anticipate that child
45 concepts will emerge from the annotated text that provide more granularity for the target concept
46 in question. For each child concept, we will define its relationship to the annotated text. For
47 example, it can be derived from our theoretical framework that the target concept Capacity for
48 Suicide has three child-concepts: 1) Acquired Capacity, 2) Practical Capacity, and 3)
49 Dispositional Capacity (see Figure 1). Note text such as '*Slit his wrists*' and '*OD'd on lithium*'
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 would apply to Acquired Capacity with a ‘has_attempted_suicide’ relationship, whereas text
5
6 such as ‘*Had a firearm with him*’ and ‘*Has access to weapons*’ would apply to Practical Capacity
7
8 with an ‘access_to_means’ relationship. Whenever it is not immediately clear which relationship
9
10 or child concept should be used for annotation, the team will discuss until consensus is reached.
11
12
13 The study team has two suicide experts who can guide this discussion (EM and MI). For each
14
15 annotation, we will indicate whether it evidences the presence or absence of that concept. Each
16
17 annotator’s annotations and relationships will be combined in terms-to-concept maps, one for
18
19 each 3ST concept, which will be discussed in the project team. The target concepts, child
20
21 concepts and their relationship to the annotated text constitute a practical ontology that will
22
23 facilitate machine annotation for 3ST concepts.
24
25
26
27
28
29

30 Figure 1, full width, about here.
31
32

33 Aim 1 Embedding Models 34

35 Using a large corpus of VA clinical progress notes, we will train a word embedding
36
37 model that can be queried with each of the terms in the initial terms-to-concept maps to identify
38
39 statistically similar terms that are not in those maps. These new candidate terms may include
40
41 abbreviations, phrases, and lexical variants, including misspellings of the initial terms. The new
42
43 candidate terms that are closely related, as determined by subject matter experts, will be
44
45 integrated into the final terms-to-concept maps. For example, our manual annotation process may
46
47 identify ‘divorce’ as a term relevant to connectedness. The embedding model may indicate a
48
49 strong correlation between ‘divorce’ and ‘dvorce’ and between ‘divorce’ and ‘divorced’.
50
51
52 Recognizing ‘dvorce’ as a potential misspelling of ‘divorce’ and ‘divorced’ as past tense of
53
54 ‘divorce’, we can then add those terms to our terms-to-concept map for connectedness, so they
55
56
57
58
59
60

1
2
3
4 will be appropriately identified when tagging clinical progress notes in Aim 2. Similarly for the
5
6 concept of psychological pain, the embedding model may indicate a strong correlation between
7
8 the phrase ‘emotional pain’ and ‘psychological pain’. Knowing that emotional pain is often used
9
10 as a synonym of psychological pain,³² we can then add emotional pain as a new term to the
11
12 terms-to-concept map for psychological pain, so ‘emotional pain’ will be identified as an
13
14 instance of psychological pain. Embedding models are essential for solving most modern NLP
15
16 problems, and we have applied them in prior work to enhance terms-to-concept maps for other
17
18 clinical areas.³³⁻³⁵ The process of creating embedding models, that is determining the
19
20 misspellings and words or phrases that are statistically similar, is based on the proximity of
21
22 words in clinical progress notes and is relatively straight forward.³⁶ Because embedding models
23
24 can contain less relevant terms, our subject matter experts will review the terms that are
25
26 machine-generated for relevancy, before including them in our 3ST terms-to-concept maps. The
27
28 updated terms-to-concept maps will then be translated into a machine-readable annotation
29
30 schema that will allow NLP tools to tag and label all instances of the 3ST concepts.
31
32
33
34
35

36
37 The use of embedding models aims to enhance generalizability of the terms-to-concept
38
39 maps to the larger corpus of VHA progress notes. To assess the distribution of terms and 3ST
40
41 concepts in the corpus of VHA notes, we will determine term frequencies of machine-annotated
42
43 concepts relative to the number of patients and progress notes for all 130 VA stations.
44
45

46 47 **Aim 2**

48
49 The goal of Aim 2 is to derive longitudinal representations for each patient with respect
50
51 to the presence and intensity of hopelessness, psychological pain, connectedness, and capacity
52
53 for suicide and describe change in these concepts in temporal proximity of suicide attempts and
54
55
56
57
58
59
60

1
2
3
4 deaths by suicide. The ability of these longitudinal representations to improve accuracy of
5
6 suicide prediction models will be evaluated in Aim 3.
7
8
9

10 Aim 2 Methods

11
12 We will determine the longitudinal trajectories of the target concepts in VHA clinical
13
14 progress notes across 6 months leading up to a suicide attempt or death by suicide. Prior research
15
16 has shown that the risk of suicide among Veterans with VHA utilization increases during this
17
18 time frame.¹⁴ Like Aim 1, our primary data source will be clinical progress notes from
19
20 encounters with Veterans who attempted or died by suicide after 2013. Based on recent data, we
21
22 anticipate that this may involve clinical progress notes of about 60,000 Veterans.¹³⁷ We will
23
24 exclude Veterans whose progress notes were used for Aim 1, as this will provide a measure of
25
26 the generalizability of the new ontology, independent from the notes that were used to develop
27
28 that ontology. For descriptive and comparative analysis, we will also annotate six months of
29
30 clinical notes for Veterans without a history of suicide attempt. These Veterans will be selected
31
32 to match the Veterans who died by suicide or attempted suicide, in terms of age, gender, race,
33
34 VA facility, and the six-month time frame.
35
36
37
38
39

40 Using the machine-annotation schema developed for Aim 1, we will tag and label the
41
42 clinical progress notes for the target concepts with a text analysis tool called CLEVER (CL-
43
44 inical EVE-nt R-ecognizer).³³ Although CLEVER has not been used to detect suicide-related
45
46 concepts, its applicability to ontology-driven representations and associated expert-rule sets for
47
48 clinical texts has been demonstrated in several studies.^{33 34 38 39} CLEVER was designed to be
49
50 efficient and flexible for extension by computational scientists or subject matter experts with
51
52 some coding expertise and a desire to build custom NLP extractors. For each tag, CLEVER will
53
54 record the applicable semantic relationship, the child and target concept, the surrounding context
55
56
57
58
59
60

1
2
3
4 or ‘snippet’ entailing the annotation, as well as relevant note meta-data such as the clinical note
5
6 type (e.g. mental health note, suicide risk assessment note), entry time, provider type (e.g.
7
8 physician, nurse, social worker), and the patient’s ID. CLEVER will also label each tag,
9
10 indicating whether it is positive or negative for the associated concept, taking into account other
11
12 common semantic modifiers (e.g., history, experiencer).
13
14

15
16 After all concept information is populated into CLEVER, we will translate each
17
18 Veteran’s labelled annotations into a numerical score for each target concept that will be used to
19
20 test the hypotheses of Aim 2. Exploring how to determine concept scores so that they best
21
22 represent the data over time is part of the proposed study. The scores will be determined for
23
24 different time windows (e.g., week-by-week, month-by-month, going back six months prior to
25
26 the Veteran’s suicide attempt or death). In determining scores for sequential windows of varying
27
28 size, we allow for the fact that suicide risk is not necessarily a linear phenomenon. Suicide risk is
29
30 a dynamic property that may wax and wane depending on internal and external factors.^{40 41} The
31
32 volume of annotations related to each concept depends on the window size; longer windows will
33
34 capture more clinical encounters and hence more clinical notes and annotations. Therefore, we
35
36 define concept scores as the number of concept-related annotations over the number of progress
37
38 notes within a window. Figure 2 visualizes this with an example of the score for one Veteran’s
39
40 hopelessness during the first month prior to their suicide attempt or death.
41
42
43
44
45
46
47

48 Figure 2, full width, about here.
49
50
51

52
53 Should a window not contain any progress notes, we will carry forward the score
54
55 determined one window before the current window (i.e., going back in time). As 3ST concepts
56
57
58
59
60

1
2
3
4 contain child-concepts and relationships (see Figure 2), we will also explore scores at the child-
5
6 concept level, defined as the number of child-concept-related annotations over the number of
7
8 progress notes within a window, and at the relationship level, defined as the number of
9
10 relationship-related annotations over the number of progress notes within a window.
11
12

13 Determining the relationships between annotations and child-concepts is done as part of Aim 1,
14
15 and it is unknown beforehand how many will be identified. However, based on our expertise
16
17 with NLP of clinical notes, we expect the number of relationships and child-concepts per 3ST
18
19 concept to be quite manageable ($\ll 100$). The reason to evaluate multiple approaches to quantify
20
21 3ST concepts is that an approach that works for one 3ST concept may not be the best option for
22
23 another.
24
25
26
27

28 Aim 2 Hypotheses and Analysis

29
30 Hierarchical (mixed effects) linear modelling will be used to analyze our repeated
31
32 measures data across the six months leading up to the suicide attempts and deaths. We will
33
34 address the following hypotheses related to our theoretical framework:
35
36
37
38
39

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

41
42 H2b: Psychological pain will increase during the six months prior to a suicide attempt or death.

43
44 H2c: Connectedness will decrease during the six months prior to a suicide attempt or death.

45
46
47 H2d: Capacity for suicide will increase during six months prior to a suicide attempt or death.
48
49
50

51 To test these hypotheses, the fixed-effect part of the models will have the following
52
53 basic form:
54
55

$$56 Y_{ij} = \text{intercept}_i + \beta_i * \text{time}_{ij}$$

57
58
59
60

1
2
3
4
5
6
7 to which random effects will be added for VHA facility and Veterans within those facilities. Y_{ij}
8 represents the outcome (e.g., score on Hopelessness shown in Figure 2) for Veteran i , at time j
9 before the suicide or suicide attempt. The time parameter β will be tested for the Aim 2
10 hypotheses. We will test time expressed in weeks and, separately, months. In secondary analyses,
11 we will examine if these relationships differ for Veterans without a history of suicide attempt by
12 testing the interaction between time and history of suicide attempt (yes/no).
13
14
15
16
17
18
19

20 In descriptive analysis of the target concepts (means, volume of annotations, total
21 number of encounters etc.), we will compare Veteran groups that are of particular interest to
22 VA.⁴² Specifically, we will compare data by race, gender, age category, justice-involvement
23 (yes/no), access to means (yes/no), and history of suicide attempt (yes/no).
24
25
26
27
28
29

30 **Aim 3**

31
32
33 Traditional approaches to developing suicide prediction models rely in large part on
34 known risk and protective factors.^{6 43 44} Suicide prediction models currently used in VHA clinical
35 practice, STORM and REACHVET described elsewhere,¹²⁻¹⁴ extract these factors from
36 structured data in the EHR and do not capture the key drivers of suicidal behavior known from
37 3ST. We expect that enhancing these suicide prediction models with our longitudinal
38 representations of the 3ST concepts will improve the models' accuracy and will better
39 distinguish who will and who will not act on their suicidal thoughts and potentially die by
40 suicide. The goal of Aim 3 is to determine the predictive validity of the 3ST concepts regarding
41 Veteran suicide and suicide attempts.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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-nonserious attempts.⁴⁵ Briefly, medically-serious suicide attempts are defined as suicide attempts that require hospitalization for more than 24 hours in combination with one of the following criteria: a) a highly lethal method was used, like hanging or gunshot, or b) treatment in specialized unit, like an ICU or burns unit, was required, or c) surgery under general anesthesia 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 (e.g., 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

1
2
3
4 existing set of predictors, to evaluate whether their accuracy can be improved.¹² This will enable
5
6 us to evaluate our methods in a general Veteran population (REACHVET) and in a high-risk
7
8 Veteran population (STORM: Veterans with active opioid prescriptions). It also enables us to
9
10 evaluate different outcomes and prediction horizons, as REACHVET predicts the risk of death
11
12 by suicide within 30 days and STORM predicts the risk of suicide attempt or death by suicide
13
14 within 1 year.^{12 13}
15
16

17
18 For each risk prediction model, we will create two feature sets that can be explored in
19
20 predictive analytics. Feature Set 1 will contain the predictors that are currently used in STORM,
21
22 or REACHVET, and will function as our baseline. Feature Set 2 will contain the features derived
23
24 from the 3ST concepts as well as the features from Feature Set 1. For each set of features, we
25
26 will construct machine learning models for the respective outcome of STORM and REACHVET,
27
28 using penalized regression with L1/L2 (elastic net) regularization, support vector machines, and
29
30 random forests. A recent systematic review supports the use of various machine learning models
31
32 to further suicide risk prediction.⁴⁴
33
34
35
36

37 Aim 3 Hypotheses and Analysis

38
39 We will use 60% of the sample for model training, 20% of the sample for model
40
41 selection/tuning, and the final 20% of the data for model evaluation. To increase the number of
42
43 training events for REACHVET, we will train the models on death by suicide or medically-
44
45 serious suicide attempt, instead of death by suicide alone. Standard classification performance
46
47 measures will be used to compare the alternative models (e.g., precision, recall (sensitivity),
48
49 specificity, area-under-the-curve). We will address the following hypotheses:
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 H3a: Models that contain STORM Feature Set 2 will have better classification accuracy of
5
6
7 suicide attempt or death by suicide than models that contain STORM Feature Set 1.

8
9 H3b: Models that contain REACHVET Feature Set 2 will have better classification accuracy of
10
11
12 death by suicide than models that contain REACHVET Feature Set 1.

13
14
15
16 In a secondary regression analysis, we will address the following hypothesis underlying
17
18 3ST by testing the three-way interaction between connectedness, psychological pain, and
19
20 capacity for suicide:
21

22
23
24
25 H3c: Capacity for suicide will moderate the likelihood of death by suicide, such that among
26
27
28 Veterans with high psychological pain and a low level of connectedness, those with low
29
30 capacity for suicide are less likely to die by suicide.

31
32
33
34 As a secondary analysis, we will assess algorithmic bias of the suicide prediction
35
36 models with Feature Set 2 for age, race, or gender. Note that variables for age, race, and gender
37
38 are already included in Feature Set 2, and that support vector machines, and random forests will
39
40 automatically identify and include interactions for these subgroups if they are significant.⁴⁷ For
41
42 models based on penalized regression, we will manually add interaction terms with age, race,
43
44 and gender to assess dependency on these variables. If evidence of interaction is found, we will
45
46 develop separate models for these groups and assess their performance compared to the
47
48 combined models, as we seek to develop software tools that generalize to the entire population of
49
50 Veterans who receive care through VHA.
51
52
53
54
55
56
57
58
59
60

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 analyzing 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 (e.g., 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 behaviors. The key methodological innovation of our study is to pair a state-of-the-art

1
2
3
4 theoretical framework (3-step theory of suicide) that predicts who is most likely to act on their
5
6 suicidal thoughts with state-of-the-art methods to extract information from unstructured clinical
7
8 progress notes. As 3ST concepts are not currently represented in structured VHA patient data, we
9
10 will develop novel NLP and information extraction tools and apply them to clinical progress
11
12 notes, the potential of which has not been explored to improve suicide prediction models. While
13
14 our study uses VHA data, the ontology for 3ST concepts that will be developed may generalize
15
16 to populations other than veterans (e.g., active-duty military) and will be made available to other
17
18 healthcare systems.
19
20
21
22
23

24 **Availability of data and materials**

25
26 The datasets analyzed during the current study will not be publicly available due to the
27
28 potential for exposure of personal health information. The ontology of terms related to the 3-Step
29
30 Theory of Suicide will be made available from the corresponding author on reasonable request.
31
32
33
34
35

36 **Competing interests**

37
38 The authors declare that they have no competing interests.
39
40

41 **Funding**

42
43 This study is funded by U.S. Veterans Health Administration Research & Development
44
45 grant SDR 19-345. The funding agency had no role in the design of the study nor in collection,
46
47 analysis, and interpretation of data or in writing of the manuscript.
48
49
50
51
52
53
54
55
56
57
58
59
60

Author contributions

EM, ST, and AH made substantial contributions to the conception and design of the study. EM, ST, AH, AF, RR, and MI significantly contributed to the development of the protocol. EM, ST, AH, AF, RR, and MI authors read, revised, and approved the final manuscript.

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(6):816-21. doi: 10.1177/0033354916670133 [published Online First: 2017/01/27]
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):1-5. doi: 10.1111/sltb.12068
5. Turecki G, Brent DA. Suicide and suicidal behaviour. *Lancet* 2016;387(10024):1227-39. doi: 10.1016/S0140-6736(15)00234-2

1
2
3
4 6. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and
5 behaviors: A meta-analysis of 50 years of research. *Psychol Bull* 2017;143(2):187-232. doi:
6
7 10.1037/bul0000084
8
9

10
11
12 7. Van Orden KA, Witte TK, Cukrowicz KC, et al. The interpersonal theory of suicide.
13
14 *Psychol Rev* 2010;117(2):575-600. doi: 10.1037/a0018697
15
16

17
18 8. Maltzberger JT. The descent into suicide. *Int J Psychoanal* 2004;85(Pt 3):653-67.
19
20 doi: 10.1516/002075704774200799 [published Online First: 2004/07/02]
21
22

23
24 9. Leenaars AA. Suicide: a multidimensional malaise. *Suicide Life Threat Behav*
25
26 1996;26(3):221-36. [published Online First: 1996/01/01]
27
28

29
30 10. Baumeister RF. Suicide as escape from self. *Psychol Rev* 1990;97(1):90-113. doi:
31
32 10.1037/0033-295x.97.1.90 [published Online First: 1990/01/01]
33
34

35
36 11. Shneidman ES. Perspectives on suicidology. Further reflections on suicide and
37
38 psychache. *Suicide Life Threat Behav* 1998;28(3):245-50. [published Online First: 1998/11/10]
39
40

41
42 12. Oliva EM, Bowe T, Tavakoli S, et al. Development and applications of the
43
44 Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation (STORM) to
45
46 improve opioid safety and prevent overdose and suicide. *Psychol Serv* 2017;14(1):34-49. doi:
47
48 10.1037/ser0000099
49

50
51 13. Kessler RC, Hwang I, Hoffmire CA, et al. Developing a practical suicide risk
52
53 prediction model for targeting high-risk patients in the Veterans health Administration. *Int J*
54
55 *Methods Psychiatr Res* 2017;26(3) doi: 10.1002/mpr.1575
56
57

1
2
3
4 14. McCarthy JF, Bossarte RM, Katz IR, et al. Predictive Modeling and Concentration
5 of the Risk of Suicide: Implications for Preventive Interventions in the US Department of
6
7 Veterans Affairs. *Am J Public Health* 2015;105(9):1935-42. doi: 10.2105/AJPH.2015.302737
8
9 [published Online First: 2015/06/13]
10
11

12
13
14 15. Klonsky ED, May AM. The Three-Step Theory (3ST): A New Theory of Suicide
15 Rooted in the "Ideation-to-Action" Framework. *International Journal of Cognitive Theory*
16
17 2015;8(2):114-29. doi: 10.1521/ijct.2015.8.2.114
18
19

20
21
22 16. Yang L, Liu X, Chen W, et al. A Test of the Three-Step Theory of Suicide among
23 Chinese People: A Study Based on the Ideation-to-Action Framework. *Arch Suicide Res*
24
25 2019;23(4):648-61. doi: 10.1080/13811118.2018.1497563 [published Online First: 2018/07/20]
26
27

28
29
30 17. Dhingra K, Klonsky ED, Tapola V. An Empirical Test of the Three-Step Theory of
31 Suicide in U.K. University Students. *Suicide Life Threat Behav* 2019;49(2):478-87. doi:
32
33 10.1111/sltb.12437 [published Online First: 2018/04/01]
34
35

36
37
38 18. Hagan CR, Muehlenkamp JJ. The Three-Step Theory of Suicide: An Independent
39 Replication and Conceptual Extension. *Suicide Life Threat Behav* 2019 doi: 10.1111/sltb.12590
40
41 [published Online First: 2019/10/01]
42
43

44
45
46 19. Wolford-Clevenger C, Flores LY, Stuart GL. Proximal correlates of suicidal
47 ideation among transgender and gender diverse people: A preliminary test of the three-step
48
49 theory. *Suicide Life Threat Behav* 2021 doi: 10.1111/sltb.12790 [published Online First:
50
51 2021/07/14]
52
53

1
2
3
4 20. Pachkowski MC, Hewitt PL, Klonsky ED. Examining suicidal desire through the
5 lens of the three-step theory: A cross-sectional and longitudinal investigation in a community
6 sample. *J Consult Clin Psychol* 2021;89(1):1-10. doi: 10.1037/ccp0000546 [published Online
7 First: 2021/01/29]
8
9
10
11
12

13
14 21. U.S. Departments of Veterans Affairs & Defense. VA/DoD Clinical Practice
15 Guideline for the Assessment and Management of Patients at Risk for Suicide. version 2.0 ed:
16 The Assessment and Management of Suicide Risk Work Group, 2019.
17
18
19
20
21

22 22. Leonard Westgate C, Shiner B, Thompson P, et al. Evaluation of Veterans' Suicide
23 Risk With the Use of Linguistic Detection Methods. *Psychiatr Serv* 2015;66(10):1051-6. doi:
24 10.1176/appi.ps.201400283
25
26
27
28
29

30 23. Hammond KW, Ben-Ari AY, Laundry RJ, et al. The Feasibility of Using Large-
31 Scale Text Mining to Detect Adverse Childhood Experiences in a VA-Treated Population. *J*
32 *Trauma Stress* 2015;28(6):505-14. doi: 10.1002/jts.22058
33
34
35
36
37

38 24. Poulin C, Shiner B, Thompson P, et al. Predicting the risk of suicide by analyzing
39 the text of clinical notes. *PLoS One* 2014;9(1):e85733. doi: 10.1371/journal.pone.0085733
40
41
42
43

44 25. Levis M, Leonard Westgate C, Gui J, et al. Natural language processing of clinical
45 mental health notes may add predictive value to existing suicide risk models. *Psychol Med*
46 2020:1-10. doi: 10.1017/S0033291720000173 [published Online First: 2020/02/18]
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 26. Reategui R, Ratte S. Comparison of MetaMap and cTAKES for entity extraction in
5 clinical notes. *BMC Med Inform Decis Mak* 2018;18(Suppl 3):74. doi: 10.1186/s12911-018-
6 0654-2
7
8
9

10
11
12 27. Chen ES, Hripesak G, Friedman C. Disseminating natural language processed
13 clinical narratives. *AMIA Annu Symp Proc* 2006:126-30.
14
15

16
17
18 28. National Library of Medicine. Overview of the Systemized Nomenclature of
19 Medicine (SNOMED) CT [Available from:
20 https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html accessed April 12, 2022
21 2022.
22
23
24
25

26
27
28 29. McCarthy JF, Cooper SA, Dent KR, et al. Evaluation of the Recovery Engagement
29 and Coordination for Health-Veterans Enhanced Treatment Suicide Risk Modeling Clinical
30 Program in the Veterans Health Administration. *JAMA Netw Open* 2021;4(10):e2129900. doi:
31 10.1001/jamanetworkopen.2021.29900 [published Online First: 2021/10/19]
32
33
34
35

36
37
38 30. U.S. Defense Suicide Prevention Office. Suicide Data Repository [Available from:
39 <https://www.dspo.mil/About-Suicide/Suicide-Data-Repository/> accessed April 3, 2019.
40
41
42
43

44 31. Leng C. Annotation Tool: The extensible Human Oracle Suite of Tools (eHOST)
45 [Available from: github.com/chrisleng/ehost.
46
47
48

49 32. Meerwijk EL, Weiss SJ. Toward a Unifying Definition of Psychological Pain.
50 *Journal of Loss and Trauma* 2011;16(5):402-12. doi: 10.1080/15325024.2011.572044
51
52
53
54
55
56
57
58
59
60

1
2
3
4 33. Ling AY, Kurian AW, Caswell-Jin JL, et al. A Semi-Supervised Machine Learning
5 Approach to Detecting Recurrent Metastatic Breast Cancer Cases Using Linked Cancer Registry
6 and Electronic Medical Record Data. *ArXiv:190105958 Cs Q-Bio Stat*, 2019.
7
8
9

10
11
12 34. Tamang SR, Hernandez-Boussard T, Ross EG, et al. Enhanced Quality
13 Measurement Event Detection: An Application to Physician Reporting. *EGEMS (Wash DC)*
14 2017;5(1):5. doi: 10.13063/2327-9214.1270
15
16
17

18
19
20 35. Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations
21 in Vector Space. *arXiv:13013781* 2013.
22
23
24

25
26 36. Chen PH. Essential Elements of Natural Language Processing: What the
27 Radiologist Should Know. *Acad Radiol* 2019 doi: 10.1016/j.acra.2019.08.010 [published Online
28 First: 2019/09/21]
29
30
31

32
33
34 37. U.S. Department of Veterans Affairs - Office of Mental Health and Suicide
35 Prevention. 2019 National Veteran Suicide Prevention Annual Report, 2019.
36
37
38

39 38. 187 Application of text mining methods to identify lupus nephritis from electronic
40 health records. 13th International Conference on Systemic Lupus Erythematosis; 2019; San
41 Francisco, CA.
42
43
44

45
46
47 39. Tamang S, Patel MI, Blayney DW, et al. Detecting unplanned care from clinician
48 notes in electronic health records. *J Oncol Pract* 2015;11(3):e313-9. doi:
49 10.1200/JOP.2014.002741
50
51
52

1
2
3
4 40. Klonsky ED, Saffer BY, Bryan CJ. Ideation-to-action theories of suicide: a
5 conceptual and empirical update. *Curr Opin Psychol* 2018;22:38-43. doi:
6
7 10.1016/j.copsyc.2017.07.020 [published Online First: 2018/08/21]
8
9

10
11
12 41. Bryan CJ, Rudd MD. The Importance of Temporal Dynamics in the Transition
13 From Suicidal Thought to Behavior. *Clinical Psychology Science and Practice* 2016;23(1):21-
14
15 25. doi: <https://doi.org/10.1111/cpsp.12135>
16
17
18

19
20 42. VA Health Services Research and Development Service. Health Services Research
21 and Development Updated Research Priorities.
22
23 <https://www.hsrd.research.va.gov/funding/PriorityDomains2019.pdf> (accessed 3/18/2019).
24
25
26

27
28 43. Belsher BE, Smolenski DJ, Pruitt LD, et al. Prediction Models for Suicide Attempts
29 and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry* 2019 doi:
30
31 10.1001/jamapsychiatry.2019.0174
32
33
34

35
36 44. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study
37 of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *J Affect*
38
39 *Disord* 2019;245:869-84. doi: 10.1016/j.jad.2018.11.073
40
41
42

43
44 45. Gvion Y, Levi-Belz Y. Serious Suicide Attempts: Systematic Review of
45 Psychological Risk Factors. *Front Psychiatry* 2018;9:56. doi: 10.3389/fpsy.2018.00056
46
47 [published Online First: 2018/03/23]
48
49
50

1
2
3
4 46. Beautrais AL. Suicides and serious suicide attempts: two populations or one?

5
6
7 *Psychol Med* 2001;31(5):837-45. doi: 10.1017/s0033291701003889 [published Online First:
8
9 2001/07/19]

10
11
12 47. Kuhn M, Johnson K. Feature Engineering and Selection: A Practical Approach for
13
14 Predictive Models: CRC 2019.
15

16 17 18 19 20 **Figure Legends** 21

22
23
24 Figure 1 Initial annotation schema for Capacity for Suicide (based on our pilot
25 research), showing annotations, relationships, and child-concepts.
26

27
28 Figure 2 Translation of concept annotations into scores across time, with example of
29 the hopelessness score for one Veteran during the month before their suicide or suicide attempt.
30 The numbers indicate months looking back in to the past, with zero (not shown) being the
31 moment of the attempt. Similarly, we will determine scores on a week-by-week basis.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14

No means to carry out plan

Guns/ammo were moved to detached garage

Has access to weapons

Had a firearm with him

Drive off a mountain/cliff

Jump off cliff

Shoot myself

Death by police

Overdose on his medications

Walk into traffic

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

...

Capacity for Suicide

Acquired Capacity

attempted_suicide

practised_a_plan

...

No past suicide attempts

Slit his wrists

Od'd on lithium

...

Dry firing of a gun

...

Dispositional Capacity

exposed_to_suicide

...

Mother committed suicide

...

Practical Capacity

has_access_to_means

has_a_plan

...

Hopelessness score for Veteran i during month 1 defined as:

$$H_{i,1} = \frac{\text{Number of hopelessness annotations for Veteran } i \text{ in month 1}}{\text{Number of progress notes for Veteran } i \text{ in month 1}}$$

