



Published in final edited form as:

Psychiatry Res. 2024 September ; 339: 116097. doi:10.1016/j.psychres.2024.116097.

Using natural language processing to evaluate temporal patterns in suicide risk variation among high-risk Veterans

Maxwell Levis^{a,b,1,*}, Joshua Levy^{b,1}, Monica Dimambro^a, Vincent Dufort^a, Dana J. Ludmer^c, Matan Goldberg^d, Brian Shiner^{a,b,e}

^aWhite River Junction VA Medical Center, White River Junction, VT, USA

^bGeisel School of Medicine at Dartmouth, Hanover, NH, USA

^cNational Institute for the Psychotherapies, New York, NY, USA

^dJerusalem College of Technology, Jerusalem, IL, Israel

^eNational Center for PTSD Executive Division, White River Junction, VT, USA

Abstract

Measuring suicide risk fluctuation remains difficult, especially for high-suicide risk patients. Our study addressed this issue by leveraging Dynamic Topic Modeling, a natural language processing method that evaluates topic changes over time, to analyze high-suicide risk Veterans Affairs patients' unstructured electronic health records. Our sample included all high-risk patients that died (cases) or did not (controls) by suicide in 2017 and 2018. Cases and controls shared the same risk, location, and treatment intervals and received nine months of mental health care during the year before the relevant end date. Each case was matched with five controls. We analyzed case records from diagnosis until death and control records from diagnosis until matched case's death date. Our final sample included 218 cases and 943 controls. We analyzed the

*Corresponding author at: White River Junction VA Medical Center, 215 North Main Street, White River Junction VT, USA. maxwelle.levis@va.gov (M. Levis).

¹Denotes equal contribution as co-first authors.

Author Agreement Statement

On behalf of all the authors, I, as the corresponding author, declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere. I confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. I further confirm that the order of authors listed in the manuscript has been approved by all of us. As the Corresponding Author, I am the sole contact for the Editorial process and will be responsible for communicating with the other authors about progress, submissions of revisions, and final approval of proofs.

CRedit authorship contribution statement

Maxwell Levis: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Joshua Levy:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation, Conceptualization. **Monica Dimambro:** Visualization, Validation, Software, Formal analysis, Data curation. **Vincent Dufort:** Software, Investigation, Formal analysis, Data curation. **Dana J. Ludmer:** **Matan Goldberg:** Writing – review & editing. **Brian Shiner:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2024.116097.

corpus using a Python-based Dynamic Topic Modeling algorithm. We identified five distinct topics, “Medication,” “Intervention,” “Treatment Goals,” “Suicide,” and “Treatment Focus.” We observed divergent change patterns over time, with pathology-focused care increasing for cases and supportive care increasing for controls. The case topics tended to fluctuate more than the control topics, suggesting the importance of monitoring lability. Our study provides a method for monitoring risk fluctuation and strengthens the groundwork for time-sensitive risk measurement.

Keywords

Suicide prediction; Electronic medical records; Natural language processing; Dynamic topic models

1. Introduction

Death by suicide is a leading cause of premature mortality in the United States (Hedegaard et al., 2020; Nock et al., 2022), ranking as the second most common cause of death among individuals 10 to 34 years old and fourth among individuals 35 to 44 years old (NIMH, 2022). The rate of suicide among military Veterans is particularly pronounced, accounting for up to 57 % increased likelihood of death when compared to civilians (VA, 2022). While predicting suicide risk remains difficult due to the relative infrequency of suicide (Kessler, 2019), this challenge is exacerbated by the labile nature of suicide risk (Bayramli et al., 2021). Machine learning tools trained to predict rare events have been increasingly utilized to surmount these challenges (Bayramli et al., 2021; Kessler et al., 2019; Levis et al., 2022).

Although recent innovations have raised accuracy standards, precise risk appraisal remains constrained for a variety of reasons (Kessler et al., 2020), including risk fluctuations over time. Indeed, suicidal ideation and psychosocial suicide risk factors, including hopelessness, burdensomeness, loneliness, and psychotherapeutic alliance have been shown to shift dramatically in the interval leading up to suicide (Bryan et al., 2012; Dunster-Page et al., 2017; Kleiman et al., 2017). Accounting for these variations is made even more arduous due to a lack of reliable suicide risk screening instruments, screening availability, and patient concerns associated with stigma (Kessler et al., 2020).

In a sequence of related studies (Levis et al., 2023b, 2022), we suggested that leveraging more expansive data resources may improve risk evaluation. Our prior research explored the utility of using natural language processing (NLP) to transform unstructured electronic health record (EHR) data, such as clinician notes, into analyzable formats that can be used to enhance suicide risk modeling. We found that this method bolsters predictive accuracy, aids clinical insight, and helps monitor subtle changes over time, especially for high-suicide risk patients. We recently found that this method, when leveraged alongside Recovery Engagement and Coordination for Health–Veterans Enhanced Treatment (REACH-VET; VA, 2017), the VA’s leading suicide prediction algorithm, led to up to 6X improved predictive accuracy for high-suicide risk patients, when compared to singly using REACH-VET (Levis et al., 2023a). We also previously established the feasibility of deploying a novel machine learning method to identify risk fluctuations in a small sample of closely matched patients with Posttraumatic Stress Disorder (PTSD) that died and did not die by suicide

(Levis et al., 2023b). Expanding beyond our prior studies, in the present paper, we utilize a national VA dataset that includes all high-suicide-risk patients that received mental health care. We selected this population based on their elevated mental health burden (Kessler et al., 2017), likelihood of suicide (McCarthy et al., 2015), and because our previous NLP study offered pronounced impact for related groups (Levis et al., 2023a). We defined high-risk status based on patients scoring within the top 1 % of REACH-VET risk scores as this group includes a very high-concentration of suicide decedents and a high proportion of clinically flagged high suicide risk patients (McCarthy et al., 2015). We used this sample to derive targeted psychosocial variables and map changes in the year before death by suicide.

1.1. Background

NLP, a branch of machine learning that evaluates textual patterns, has been shown to be an effective means to convert text into data formats that can be used to identify and track psychotherapy and suicide-related variables (Ben-Ari and Hammond, 2015; Dreisbach et al., 2019; Fernandes et al., 2018; Levis et al., 2020). An array of previous studies have explored using NLP to analyze unstructured EHR data for suicide prediction (Leonard Westgate et al., 2015; Poulin et al., 2014; Rumshisky et al., 2016). NLP may offer improved access to information about patients' interpersonal patterns and relationships between the clinicians and patients (Levis et al., 2022; 2020) and how these change in the intervals leading up to death by suicide (Levis et al., 2023b). Our prior findings suggest the benefit of using NLP to develop EHR-note-derived variables to enhance predictive accuracy, as evidenced by the ability to outperform leading structured EHR suicide risk metrics (Levis et al., 2023a).

Prior research suggests that NLP methods can be utilized to monitor real-time changes in suicide risk (Fine et al., 2020). Unlike demographic, diagnostic, and service usage variables, which tend to remain fairly stable over time, NLP approaches that account for temporal changes may aid evaluation of the dynamic nature of suicide risk (Bayramli et al., 2021). In a prior study, (Levis et al., 2023b) we used Dynamic Topic Modeling, a method that automatically detects topics distributed throughout text corpora (Blei et al., 2003), to better uncover patterns over time within EHR notes. This approach is conceptualized as being "latent" because it discerns topics that can be hidden from sight, mimicking qualitative pattern detection and characteristic syntheses (Glaser and Strauss, 2017; Knowles et al., 2014). As opposed to qualitative methods, Dynamic Topic Modeling is probabilistic and independent of human consultation, utilizing an unsupervised machine learning method to rank, sort, and synthesize text themes. In contrast to conventional topic modeling, which evaluates word pattern distribution without consideration for time, Dynamic Topic Modeling accounts for changes over time, charting how topic-associated words fluctuate over defined intervals. Dynamic Topic Modeling offers the ability to longitudinally trace risk fluctuations and pinpoint intervals of concern.

Within our pilot study, we identified central topics for patients who died and suicide risk-matched patients who did not die by suicide, including "Treatment", "Medication", "Engagement", "Expressivity", and "Symptomology". We found that patients who did not die showed greater fluctuation over time relative to those who died by suicide for all topics except Medication, which remained consistent across time for both groups. Our findings

suggested that over time patients who did not die showed increased psychotherapeutic benefits, including improved engagement and expressivity. This prior work, however, was constrained by its small sample size, a lack of focus on the prevalence and relevance of identified topics, and a lack of population specificity. In the present study, we built on our past work by 1) including a much larger sample, 2) monitoring changes in topic relevance and prevalence over time, and 3) only including patients that were predicted to be at high suicide risk.

Our choice to focus on high-risk patients is predicated on this community's pronounced suicide risk concentration (McCarthy et al., 2015), elevated mental health burden (Kessler et al., 2017), and active usage of mental health services, including targeted suicide prevention services (McCarthy et al., 2021). Based on this prior study, we anticipated that patients who died by suicide and risk-matched patients who did not die would show divergent topic fluctuations over time and that monitoring these changes could aid monitoring clinical risk patterns. This study advances new risk detection methods and expands clinical understanding of high suicide-risk patient experience.

2. Methods

The VA is the nation's largest integrated health system and provider of psychotherapy and mental health services (OMHSP, 2021). It also maintains a central data repository, the Corporate Data Warehouse (CDW), which contains all patient EHR data. Responding to the increased rates of death by suicide in the Veteran community (OMHSP, 2021; Rubin, 2019), the VA has prioritized advancing innovative methods to monitor suicide risk, data that is now accessible via EHR (Kessler et al., 2017). As such, the VA provides a unique resource for data-focused suicide research. This study utilizes a recent representative sample of Veterans engaged in VA mental health care who were identified as high suicide risk that did (cases) and did not die (controls) by suicide. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

2.1. Sample

We obtained information on patient characteristics and service use, as well as clinical note text associated with mental health encounters from the CDW. We obtained ethical approval for all protocols from the local institutional review board, confirming that the study meets national and international guidelines for research on humans. We also received a waiver of informed consent due to study's retrospective nature, sample size, data protections, and lack of ability to contact patients and their families due to patient death. Following VA privacy standards, analysis was restricted to a VA study-specific airlocked research server that could only be accessed by approved VA personnel.

One of the VA's leading innovations was the development of REACH-VET, an algorithm that identifies VA users with the highest suicide risk based upon structured variables derived from EHR data associated with increased suicide risk (Matarazzo et al., 2019; McCarthy et al., 2015; VA, 2017). REACH-VET's algorithm automatically evaluates 61 structured variables that are strongly associated with suicide from the EHR including healthcare usage, psychotropic medications, and socio-demographics (Supplementary Table 1).

To develop the study sample, we linked EHR data from the CDW with cause of death data from the VA-Department of Defense Mortality Data Repository (MDR;DoD, 2020) to identify all patients who died by suicide that had at least one VA healthcare encounter in either 2017 or 2018 (cases = 4584). We then restricted cases to patients that 1) received mental health care during treatment interval ($n = 2279$), 2) were detected by REACH-VET as being in the top 1 % of suicide risk in the month before death by suicide ($n = 437$), and 3) had at least 9 months of mental health care in the year before suicide ($n = 218$). Although the REACH-VET model is trained to predict the top 0.1 % of risk, we chose to include patients within the top 1 % of predicted risk due to this group's very high suicide risk concentration, accounting for nearly 10 % of VA suicide deaths (Kessler et al., 2017), their elevated level of clinically identified suicide risk (McCarthy et al., 2015), and because we had previously found that NLP risk modeling aided predictive modeling for this group (Levis et al., 2023a). As the study evaluates suicide risk changes over time, we retained patients who received care for at least 9 months to facilitate time-dependent observations. Due to the project's clinical nature, and access restrictions associated with VA data, our project data is not publicly available.

2.2. Matching

Following guidance about rare event matched case-control methods (Lacy, 1997) and with support from the Office of Mental Health and Suicide Prevention, we matched each case with five controls that had the same REACH-VET score, a comparable amount and duration of mental health care at the same VA facility, and were alive at the time of the cases' death. We identified risk matched controls that had mental health care during the select interval ($n = 10,729$), sub-selected those with REACH-VET indicated high-suicide risk ($n = 2095$), and then identified those who had at least 9 months of mental health care in the year before relevant end date ($n = 943$). Given our multistep sample derivation, our final matching ratio was slightly reduced such that for every case there were 4.32 controls, on average. For descriptive purposes, we assessed demographic characteristics for the sample, including age, race and ethnicity, marital status, military service era, and level of VA service-connected disability from the month before matched-cases' death date, and calculated standardized mean differences between case and control differences.

2.3. Dynamic topic modeling

We used LDASeq, a widely used Dynamic Topic Modeling method (Blei and Lafferty, 2006). LDASeq is based on Latent Dirichlet Allocation (LDA), where each document, or EHR note in our case, is premised to contain words that represent a range of topics. The word frequency within a document is assumed to follow a multinomial distribution (Levy et al., 2015). Topics are represented using Dirichlet distributions, a process that ascribes select words to a topic by assigning them higher weight via simplex parameters ("proportions" summing to 1 to indicate relative presence). The breakdown of topics within a document (proportion of topic occurrence based on word content) is also given by a Dirichlet distribution. The multinomial distribution is modeled using a generative modeling approach which samples topics from documents per the Dirichlet distribution, and words from topics per another Dirichlet distribution. Instead of using conventional Dirichlet distributions, LDASeq uses logistic transformations of multivariate normal distributions. This adaptation

preserves the distribution of topics and words within simplex parameters but also allows more flexible modeling of topic complexities. LDASeq model parameters are optimized using variational Kalman filtering and variational wavelet regression (Theodoridis, 2020) to estimate topic distributions.

In Dynamic Topic Modeling, the importance of words within a topic (i.e., prominence) is temporally modeled across time by inducing conditional dependence linking the current topic with the previous topic. This is accomplished by sampling the current topic from a multivariate normal distribution, centered on the previous topic, before application of the logistic distribution that generates the simplex parameter. This generative method mimics qualitative research design, allowing innate lexical patterns to come forward independent of theoretical assumptions. To isolate and compare change patterns within the respective groups, we conducted analysis on a month-by-month basis separately for cases and controls.

As our goal was to evaluate changes over time, we specifically looked at month intervals. This unit of analysis allowed the inclusion of intermittent mental health appointments. When compared to shorter intervals, this approach required less computational resources, a concern given the complexity of Dynamic Topic Modeling analyses. Additionally, months, as discrete and digestible time intervals, are frequently used within VA suicide prevention efforts, such as the calculation of REACH-VET percentile scores. LDAseq requires manual investigation to validate topic validity and justify topic coherence (AlSumait et al., 2009). We conducted analyses in Python 3.8 using Gensim (Models.Ldaseqmodel, 2021), The Natural Language Toolkit (NLTK; Bird et al., 2009), and Scikit-learn packages (Pedregosa et al., 2011).

2.4. Data preparation

To prepare data for analysis, we performed tokenization, a process that splits text data into consistent units, removed very common words (“stop-words”) that were premised to offer little added meaning to our model (words like “and” or “the”, as well as ubiquitous terms like “vet”, “veteran”, and “patient”), eliminated symbols and punctuation, and made all words lowercase. Following our previous method, we removed notes from five days before death to avoid potential endogeneity as date of death can be delayed within EHR records and we did not want to bias our sample with the possible inclusion of post-mortem notes.

2.5. Analysis

After preliminary analysis selecting different numbers of topics, we constrained analysis to five topics, a number that maximized interpretability without overtaxing computational resources. As cases and controls were closely matched on REACH-VET, derived topics and topic fluctuations were expected to be closely related. Our analysis included the following steps:

2.5.1. Grouping case and control topics—Case and control topics were grouped together based on degree of agreement at the first timepoint, using the rank-biased overlap metric (RBO; Amigó et al., 2018), between topics. RBO is a score between 0 (no agreement) to 1 (perfect agreement), representing the overlap between two sets of words

and weighting words that are ranked more highly in each list (Bonett and Wright, 2000). We calculated 95 % confidence intervals based on a parametric assumption that the RBO would be distributed similarly to the Kendall Tau, a related measure of agreement (Lapata, 2006). These confidence intervals were used to map the topics' degree of association. To conceptualize each derived topic, we identified the 20 most prominent words from each topic by averaging term prominence scores across all time intervals. Topic correlations and prominent words were further evaluated by clinical experts (including a psychiatrist, clinical social worker, and psychologist) to evaluate optimal topic grouping and topic name. Topic names were selected by a process in which each clinician indicated naming preferences. The most consistently selected names were chosen to represent each respective topic.

2.5.2. Characterizing topic evolution—To monitor changes over time, we evaluated case and control topic changes, respectively, from the start to the end of the treatment interval. We also used Dynamic Time Warping (DTW: Berndt and Clifford, 1994) and Latent Class Mixture Models (LCMM: Marcoulides and Schumacker, 2001) to perform a trajectory analysis and identify sub-topics (i.e., groups of words whose prominence co-varies similarly across time intervals). As both dynamic time warping and latent class mixture models yielded similar results, we chose to only utilize dynamic time warping as it better captured non-linear time dependencies. Additional details on both trajectory analysis techniques can be found in the supplementary material, “Sub-topic identification” section. We characterized each word's temporal changes relative to other words' temporal changes within sub-topics based on differences in relative prominence at each timepoint. This metric measures words' centrality within each sub-topic.

2.5.3. Relative suicide risk fluctuations—To track topic association and changes over time between groups, we calculated the prevalence of control-derived topics for cases and controls. This measurement served as an indicator of how the presence of these topics for cases deviated from what would be considered normal expectations. Using multivariable regression models, we identified specific points in time where the presence of specific topics in the corpus differed between cases and controls (see supplementary section “Association between derived topics and subsequent suicide risk at specific timepoints” for additional details on the statistical modeling).

3. Results

Within our final sample, we identified 218 cases and 943 controls that met all selection criteria and were classified as high-risk based on their REACH-VET scores. Matches had similar demographics and service usage trends, as evidenced by very low SMDs across all assessed variables, as presented in Table 1. Small differences, however, were detected between cases and controls' racial background; cases tended to include slightly more non-Hispanic White patients and controls tended to include more non-Hispanic Black patients. Controls also tended to include slightly more members with no service-connected disability and more cases tending to have high levels of service-connected disability. Cases had a total of 13,504 notes, or on average 62 notes per patient, while controls had a total of 55,369 notes, or on average 59 notes per patient.

3.1. Grouping case and control topics

Both groups' Dynamic Topic models identified potentially clinically meaningful topics, showcasing group similarities and differences that shifted over time. After consulting with clinical experts, we grouped the topics with the highest case and control correlations at baseline, naming them as follows: "Medication", "Intervention", "Treatment Goals", "Suicide", and "Treatment Focus". Case and control RBO agreement identified strong associations across most topics. Each topic's RBO and ranked prominent words are presented in Table 2.

3.2. Characterizing topic evolution

Within-group topic changes from the beginning to the end of the treatment interval showcased case and controls' relative rates of change over time. Case topics tended to change more over time than control topics, indicating increased relative fluctuation over time, as presented in Table 3. Only the Intervention and Treatment Goals topics showed statistical differences when comparing group RBO scores over time, suggesting that these topics contained the most pronounced group fluctuation. Case and control sub-topics relative prevalence also fluctuated, as presented in Fig. 1. For instance, we noticed that within the Suicide topic, "attempt" decreased in relative prominence within its sub-topic over time for both groups, whereas "family" decreased for cases and increased for controls over time. Similarly, we noticed that within the Intervention topic, "pain" decreased in relative prominence within its sub-topic over time for cases and increased for controls over time. Additional trajectory analysis results are presented in the Appendix (Supplementary Figure 1 replicates sub-topic analysis using LCMM, Supplemental Figure 2 presents DTW sub-topic grouping).

3.3. Relative suicide risk fluctuations

While case and control topic prevalence remained similar for most intervals, we identified timepoints where case and control topic prevalence differed considerably, as presented in Fig. 2. Cases and controls demonstrated parallel fluctuations in the Medication and Suicide topics, divergences during the middle of treatment in the Intervention topic, and opposite tendencies during the beginning and end of treatment in Treatment focus and Treatment goals topics. Within the Treatment Focus and Treatment Goals topics, significant differences in topic prevalence between groups were apparent during the first 100 days and last 50 days of mental health utilization. Count-based models, as opposed to derived statistical models, are presented in Supplementary Figure 3. Topic prevalence models and confidence intervals on a weekly basis over the treatment year are presented in Supplementary Table 2.

4. Discussion

Even with the increased utilization of targeted machine learning methods, predicting suicide remains beset with challenges (Nock et al., 2022). A central aspect of this challenge stems from suicide risk's temporal variability. In this study, we utilized Dynamic Topic Modeling to measure changes over time among patients with high suicide risk. Our EHR note analysis allowed nuanced evaluation of how cases and controls received and responded to care over time, highlighting group similarities and differences. Although small

demographic differences between cases and controls were detected, these differences were minor (Andrade, 2020). This degree of similarity allowed us the unique ability to focus on novel risk domains and areas to improve care. Each detected topic is reviewed as follows.

4.1. Topic analysis

4.1.1. Medication—Case and control “Medication” topics were very highly correlated across treatment intervals, with overlapping prominent words centering on pharmaceutical regimens (“mouth”, “mg²”). Case and control topics’ high correlations affirm the centrality and consistency of pharmacotherapy within VA mental health care (Krystal et al., 2017).

When comparing within-group changes over time, cases showed higher change levels, when compared to controls. The case sub-topic dealing with medication frequency (“capsule”, “daily”, “hour”) gained prominence over time, suggesting increased utilization of inpatient pharmacotherapy. This contrasts with the control sub-topic dealing with outpatient medication (“medication”, “outpatient”, “pain”) that also increased over time. This divergence hints toward reliance on inpatient and medical-focused care for cases, compared with a shift towards outpatient, psychological, and psychosocial care for controls. When comparing topic prevalence, rates similarly declined for both groups. Cases, however, showed greater fluctuation and a steeper rate of decline. This finding acknowledges the centrality of pharmacotherapy, but also suggests that, as cases approach death by suicide, there may be lessened engagement or treatment fidelity. Alternatively, it is conceivable that, as cases approach death by suicide, their providers may be more concerned about issues other than medication, prioritizing medical, legal, or psychosocial issues.

4.1.2. Intervention—Case and control “Intervention” topics were highly correlated across the treatment intervals, with overlapping prominent words centering on pharmacotherapy (“tab”, “medication”). Non-overlapping prominent words centered on medication for cases (“qhs³”, “refill”) and personal experience and expression for controls (“denies”, “mood”).

Case sub-topics dealing with inpatient care (“daily”, “inpatient”) and pharmacotherapy (“dose”, “mg daily”) increased over time while those dealing with pain treatment (“pain”, “tablet”) decreased. Control sub-topics dealing with symptoms (“anxiety”, “pain”) and inpatient care (“inpatient”, “urine”, “qh⁴ prn⁵”) increased over time, while the sub-topic dealing with improvement (“discharge”, “good”) decreased, especially towards the end of care. Related shifts may be indicative of controls transitioning out of inpatient care and showing fewer discharge references. Over time, sub-topic differences suggest cases received more medical and inpatient care, while controls received more symptom-specific psychological care. When comparing topic prevalence, we may note that cases showed higher Intervention prevalence, but that this declined over time, relative to controls, whose rates shifted only slightly over time. This trend suggests that cases may have received higher

²Milligram.

³Every night at bedtime.

⁴Every hour.

⁵As often as necessary.

levels of targeted interventions, but that this level diminishes over time. Controls' level of care remained more consistent, albeit at a comparatively lower level.

It is of note that this topic shared core similarities with the Medication topic, corroborating the centrality of pharmaceutical care for VA patients at high suicide risk. That being said, within the "Interventions" topic, control notes showed a network of related non-medication words, such as "denies", "mood", and "thought". This pattern suggests that controls may encounter broader intervention resources and that, potentially, providers may be more sensitive to controls' experiences and expressions. This dynamic could indicate providers' comparative commitment to controls' unique experiences and, potentially, reference these patients' increased engagement in care over time.

4.1.3. Treatment goals—Case and control "Treatment Goals" topics were moderately correlated at the beginning and end of the treatment interval, with overlapping prominent words centering on engagement ("goal", "recovery", "treatment"). Non-overlapping prominent words centered on treatment for cases ("diagnosis", "therapy") and accomplishment for controls ("skills", "resolved").

Cases showed decreased prominence over time for sub-topics dealing with engagement ("discussed", "member", "behavior") and intervention ("intervention", "problem", "program"), and increased prominence for sub-topics dealing with medical scheduling ("appointment", "care", "provider"). Controls showed increased prominence over time for sub-topics dealing with participation ("participated", "plan", "process"), engagement ("discussion", "treatment"), and achievement ("change", "objective", "relapse"). Case and control sub-topic differences suggest that cases shifted away from psychotherapeutic engagement and towards more routine scheduling over time. Controls, in contrast, appear to shift towards increased focus on engagement, achievement, and dedication to the psychotherapeutic process. When comparing within-group changes over time, cases showed higher change levels, when compared to controls.

When comparing topic prevalence, significant differences were apparent during the start and end of care. While both groups began at a similar level, cases showed a higher increase rate relative to cases. Although both groups again return to similar rates during the middle of the treatment year, controls again diverge, significantly increasing towards the end of the interval. This relationship may indicate that the prevalence of control Treatment Goals topics is more fluid, relative to cases.

4.1.4. Suicide—Case and control "Suicide" topics were modestly correlated across treatment intervals, with overlapping prominent words centering on risk ("suicide", "risk"), and non-overlapping terms centering on evaluation and mitigation for cases ("factor", "mental health") and targeted treatment for controls ("alcohol", "disorder"). When comparing within-group changes over time, cases showed higher change levels, when compared to controls.

Cases showed increases over time in sub-topics dealing with suicide ("suicide", "abuse", "history") and medical need ("medical", "need", "problem"), and decreases in sub-

topics dealing with service usage (“call”, “contact”, “mental health”) and targeted care (“assessment”, “care”, “support”). Control sub-topics showed increases in sub-topics dealing with symptoms (“alcohol”, “anxiety”, “symptom”) and personal experience (“family”, “history”, “pain”) and decreases in the suicide risk (“attempt”, “ideation”) sub-topic. These contrasts further characterize case medical-focused care, compared to controls, who over time, received more supportive and targeted psychological interventions. When comparing topic prevalence, a nearly identical pattern of decreasing prevalence is apparent for both groups, suggesting that less attention is dedicated to suicide risk over time.

4.1.5. Treatment focus—Case and control “Treatment Focus” topics were not correlated across treatment intervals, with non-overlapping prominent words centering on symptoms and psychopathology for cases (“disorder”, “anxiety”) and connecting resources for controls (“contact”, “discharge”, “housing”). This topic was nonetheless grouped together as it offered a potential method of identifying clinical differences between cases and controls. Although both groups showed similar levels of change over time, controls had a slightly higher level of fluctuation. Case sub-topics showed increased symptoms (“alcohol”, “anxiety”, “depression”) alongside decreased mental health care usage (“medication”, “mental”, “symptom”) over time. Control sub-topics showed increased supportive care (“home”, “information”, “program”) alongside decreased mental health care usage (“management”, “medication”, “provider”). When comparing topic prevalence, the trends were the opposite of those described in the Treatment Goals topic. In this context, controls prevalence rates sharply declined early in care and again at the end of the treatment interval, relative to cases who appeared to have less fluctuation. While case fluctuations suggest increased pathology, which could signal potential emotional distancing from providers, controls’ fluctuations may have associations with increased engagement and psychotherapeutic growth.

4.2. Implications

This study affirms the potential utility of leveraging NLP to monitor time-sensitive population-specific suicide risk variables. In contrast to REACH-VET, a prediction model that relies on structured EHR variables including demographics and service usage (McCarthy et al., 2015), which are relatively static domains, our method relies on unstructured EHR notes, which may be more responsive to real-time changes. It is of note that cases appeared to change more over the course of treatment, when compared to controls, suggesting the potential of monitoring lability as a risk factor. These changes may be indicative of increased symptomology, declining engagement, or worsened therapeutic alliance.

Our findings about relative change over time stand in contrast to our prior Dynamic Topic Modeling study which identified the opposite pattern (Levis et al., 2023b). This difference could be associated with the samples’ distinct risk concentrations: whereas the current study only included patients classified as high-risk by REACH-VET, a sub-group that accounts for a high concentration of patient suicide deaths (Kessler et al., 2017) in the prior study, we did not impose risk-tier restrictions. As the high-risk sample was flagged by REACH-VET, it is likely that they had overt suicide-associated symptoms and received targeted suicide

prevention services. It is similarly likely that the non-high-risk group had fewer overt symptoms and received fewer suicide prevention services.

Borrowing from the literature about alliance and suicide risk (Dunster-Page et al., 2017; Gysin-Maillart et al., 2017), providers appear to have lower expectations and worse relationships with high-risk patients when compared to lower-risk patients (Barzilay et al., 2020; Dunster-Page et al., 2017). It stands to reason that for high-risk patients, not experiencing symptom worsening over time could be indicative of therapeutic benefit (Luoma et al., 2002). In contrast, providers may have less defined expectations for lower-risk patients. For lower-risk patients, fluctuation over time could be more tied to positive alliance and productive therapeutic relationships (Huggett et al., 2022). Moving forward, Dynamic Topic Modeling could be a useful tool to hone-in on these changes, and in turn, improve risk stratification and risk management over time.

It is also interesting that prominent Treatment Focus and Treatment Goals prevalence differences were identified during the early and late stages of treatment. These intervals may be important points to anchor future temporal investigations (Luoma et al., 2002). It is similarly noteworthy that the Suicide and Medication topics declined in prevalence for both groups, a finding that reiterates the importance of monitoring these domains for high-risk patients.

4.3. Limitations

Given the exploratory nature of our methods and the unique high-risk characteristics of our sample, we are cautious to not imply that our topic labeling represents the only way to interpret or name topic constellations, nor do we suggest that these topics will be pervasive in other analyses. We are additionally cautious about extrapolation based on our analysis, given the inherent challenges in interpreting machine learning derived models (Kaur et al., 2020).

Our study focused on conceptualizing differences between cases and controls, rather than developing predictive models. As we focused on changes in months leading up to death by suicide, we specifically selected patients that lived at least nine months after indexing to model changes over time, thus limiting cohort size and potential relevance. We are cautious to not speculate about this method's utility for patients who died by suicide sooner after indexing. Future research is necessary to evaluate if this method could be usefully applied to notes from non-mental health medical visits. We did not consider clinician training or specialization. Given our exploratory focus, and the computational burden associated with our analysis, we developed case and control models independently, restricting the ability to utilize our model as a predictive method. Due to data privacy regulations, we cannot share our data or post to a repository. Data privacy and regulations, and ability to receive waivers of consent, could be different in other contexts, constraining ability to conduct comparable analysis. For example, the inability to use post-mortem data would make comparable analysis very arduous.

This study lays the groundwork to derive a predictive method using population-specific, psychosocial, and time-sensitive suicide risk variables from unstructured EHR notes. In a

future study, we hope to evaluate changes in structured and unstructured EHR over time and better understand how topic groupings relate to service utilization. Integrating these data formats would allow richer measurement of service utilization and clinical experiences. Subsequent studies could leverage this approach to develop a predictive model that could be implemented alongside or in combination with structured EHR-based predictive models like REACH-VET. Future research would be strengthened by comparing the identified topics and how they change over time with psychotherapy session transcripts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

Dr. Levis' was supported by a VA New England Career Development Award (VICDA-2020-60). Dr. Levy was supported by the Department of Defense Award (HT9425-23-1-0267).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Maxwell Levis reports financial support was provided by VA New England Healthcare System. Joshua Levy reports financial support was provided by the Department of Defense Award (HT9425-23-1-0267). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data access is restricted due to the clinical nature of the dataset and VA privacy protections.

Abbreviations:

NLP	Natural language processing
EHR	Electronic health records
REACH-VET	Recovery Engagement and Coordination for Health–Veterans Enhanced Treatment

References

- AlSumait L, Barbará D, Gentle J, Domeniconi C, Buntine W, Grobelnik M, Mladeníc D, 2009. Topic significance ranking of LDA generative models. In: Shawe-Taylor J (Ed.), Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 67–82. 10.1007/978-3-642-04180-8_22.
- Amigó E, Spina D, Carrillo-de-Albornoz J, 2018. An axiomatic analysis of diversity evaluation metrics: introducing the rank-biased utility metric. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Presented at the SIGIR '18: The 41st International ACM SIGIR conference on research and development in Information Retrieval. ACM, Ann Arbor MI USA, pp. 625–634. 10.1145/3209978.3210024.
- Andrade C, 2020. Mean difference, standardized mean difference (SMD), and their use in meta-analysis: as simple as it gets. *J. Clin. Psychiatry* 81. 10.4088/JCP.20f13681.

- Barzilay S, Schuck A, Bloch-Elkouby S, Yaseen ZS, Hawes M, Rosenfield P, Foster A, Galynker I, 2020. Associations between clinicians' emotional responses, therapeutic alliance, and patient suicidal ideation. *Depress. Anxiety* 37, 214–223. 10.1002/da.22973. [PubMed: 31730737]
- Bayramli I, Castro V, Barak-Corren Y, Madsen EM, Nock MK, Smoller JW, Reis BY, 2021. Temporally-informed random forests for suicide risk prediction (preprint). *Health Inf.* 10.1101/2021.06.01.21258179.
- Ben-Ari A, Hammond K, 2015. Text mining the EMR for modeling and predicting suicidal behavior among US Veterans of the 1991 Persian Gulf War. In: 2015 48th Hawaii International Conference on System Sciences. Presented at the 2015 48th Hawaii International Conference on System Sciences (HICSS) IEEE, HI, USA, pp. 3168–3175. 10.1109/HICSS.2015.382.
- Berndt DJ, Clifford J, 1994. Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. Presented at the AAAIWS'94 AAAI Press, Seattle, WA, USA, pp. 359–370.
- Bird S, Lopper E, Klein E, 2009. *Natural Language Processing with Python*. O'Reilly Media Inc., Stanford, CA.
- Blei DM, Lafferty JD, 2006. Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning - ICML '06. Presented at the the 23rd international conference. ACM Press, Pittsburgh, Pennsylvania, pp. 113–120. 10.1145/1143844.1143859.
- Blei DM, Ng AY, Jordan MI, 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res* 3, 993–1022.
- Bonett DG, Wright TA, 2000. Sample size requirements for estimating Pearson, Kendall and spearman correlations. *Psychometrika* 65, 23–28. 10.1007/BF02294183.
- Bryan CJ, Corso KA, Corso ML, Kanzler KE, Ray-Sannerud B, Morrow CE, 2012. Therapeutic alliance and change in suicidal ideation during treatment in integrated primary care settings. *Arch. Suicide Res* 16, 316–323. 10.1080/13811118.2013.722055. [PubMed: 23137221]
- DoD VA, 2020. Center of Excellence for Suicide Prevention. Joint Department of Veterans Affairs (VA) and Department of Defense (DoD) Mortality Data Repository - National Death Index (NDI). MIRECC.
- Dreisbach C, Koleck TA, Bourne PE, Bakken S, 2019. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *Int. J. Med. Inf* 125, 37–46. 10.1016/j.jmedinf.2019.02.008.
- Dunster-Page C, Haddock G, Wainwright L, Berry K, 2017. The relationship between therapeutic alliance and patient's suicidal thoughts, self-harming behaviours and suicide attempts: a systematic review. *J. Affect. Disord* 223, 165–174. 10.1016/j.jad.2017.07.040. [PubMed: 28755624]
- Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D, 2018. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. *Sci. Rep* 8, 7426. 10.1038/s41598-018-25773-2. [PubMed: 29743531]
- Fine A, Crutchley P, Blase J, Carroll J, Coppersmith G, 2020. Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data. In: Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science. Presented at the Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science Association for Computational Linguistics, pp. 50–54. 10.18653/v1/2020.nlpccs-1.6. Online.
- Glaser BG, Strauss AL, 2017. *The Discovery of Grounded Theory: Strategies for Qualitative Research*, 1st ed. Routledge. 10.4324/9780203793206.
- Gysin-Maillart AC, Soravia LM, Gemperli A, Michel K, 2017. Suicide ideation is related to therapeutic alliance in a brief therapy for attempted suicide. *Arch. Suicide Res* 21, 113–126. 10.1080/13811118.2016.1162242. [PubMed: 26984644]
- Hedegaard H, Curtin S, Warner M, 2020. Increase in Suicide Mortality in the United States, 1999–2018 (No. NCHS Data Brief; Nos. 362; DHHS Publication; 2020–1209) National Center for Health Statistics (U.S.).
- Huggett C, Gooding P, Haddock G, Quigley J, Pratt D, 2022. The relationship between the therapeutic alliance in psychotherapy and suicidal experiences: a systematic review. *Clin. Psychol. Psychother* 29, 1203–1235. 10.1002/cpp.2726. [PubMed: 35168297]

- Kaur H, Nori H, Jenkins S, Caruana R, Wallach H, Wortman Vaughan J, 2020. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Presented at the CHI '20: CHI Conference on Human Factors in Computing Systems ACM, Honolulu HI USA, pp. 1–14. 10.1145/3313831.3376219.
- Kessler RC, Hwang I, Hoffmire CA, McCarthy JF, Petukhova MV, Rosellini AJ, Sampson NA, Schneider AL, Bradley PA, Katz IR, Thompson C, Bossarte RM, 2017. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration. *Int. J. Methods Psychiatr. Res* 26 10.1002/mpr.1575.
- Kessler RC, Bernecker SL, Bossarte RM, Luedtke AR, McCarthy JF, Nock MK, Pigeon WR, Petukhova MV, Sadikova E, VanderWeele TJ, Zuromski KL, Zaslavsky AM, 2019. The role of big data analytics in predicting suicide. In: Passos IC, Mwangi B, Kapczinski F (Eds.), *Personalized Psychiatry*. Springer International Publishing, Cham, pp. 77–98. 10.1007/978-3-030-03553-2_5.
- Kessler RC, Bossarte RM, Luedtke A, Zaslavsky AM, Zubizarreta JR, 2020. Suicide prediction models: a critical review of recent research with recommendations for the way forward. *Mol. Psychiatry* 25, 168–179. 10.1038/s41380-019-0531-0. [PubMed: 31570777]
- Kessler RC, 2019. Clinical epidemiological research on suicide-related behaviors—Where we are and where we need to go. *JAMA Psychiatry* 76, 777. 10.1001/jamapsychiatry.2019.1238. [PubMed: 31188420]
- Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK, 2017. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. *J. Abnorm. Psychol* 126, 726–738. 10.1037/abn0000273. [PubMed: 28481571]
- Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, Coyle D, Kennedy CM, Littlewood E, Kessler D, Gilbody S, Bower P, 2014. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. *PLoS ONE* 9, e84323. 10.1371/journal.pone.0084323. [PubMed: 24465404]
- Krystal JH, Davis LL, Neylan TC, Raskind MA, Schnurr PP, Stein MB, Vessicchio J, Shiner B, Gleason TD, Huang GD, 2017. It is time to address the crisis in the pharmacotherapy of posttraumatic stress disorder: a consensus statement of the PTSD Psychopharmacology Working Group. *Biol. Psychiatry* 82, e51–e59. 10.1016/j.biopsych.2017.03.007. [PubMed: 28454621]
- Lacy MG, 1997. Efficiently studying rare events: case-control methods for sociologists. *Sociol. Perspect* 40, 129–154. 10.2307/1389496.
- Lapata M, 2006. Automatic evaluation of information ordering: Kendall's Tau. *Comput. Linguist* 32, 471–484. 10.1162/coli.2006.32.4.471.
- Leonard Westgate C, Shiner B, Thompson P, Watts BV, 2015. Evaluation of veterans' suicide risk with the use of linguistic detection methods. *Psychiatr. Serv* 66, 1051–1056. [PubMed: 26073409]
- Levis M, Leonard Westgate C, Gui J, Watts BV, Shiner B, 2020. Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models. *Psychol. Med* 1–10. 10.1017/S0033291720000173.
- Levis M, Levy J, Dufort V, Gobbel GT, Watts BV, Shiner B, 2022. Leveraging unstructured electronic medical record notes to derive population-specific suicide risk models. *Psychiatry Res.* 315, 114703 10.1016/j.psychres.2022.114703. [PubMed: 35841702]
- Levis M, Levy J, Dent, Kallisse R, Dufort V, Gobbel GT, Watts BV, Shiner B, 2023a. Leveraging natural language processing to improve electronic health record suicide risk prediction for Veterans Health Administration users. *J. Clin. Psychiatry* 84. 10.4088/JCP.22m14568.
- Levis M, Levy J, Dufort V, Russ CJ, Shiner B, 2023b. Dynamic suicide topic modelling: deriving population-specific, psychosocial and time-sensitive suicide risk variables from Electronic Health Record psychotherapy notes. *Clin. Psychol. Psychother* 10.1002/cpp.2842.
- Levy O, Goldberg Y, Dagan I, 2015. Improving distributional similarity with lessons learned from word embeddings. *Trans. Assoc. Comput. Linguist* 3, 211–225. 10.1162/tacl_a_00134.
- Luoma JB, Martin CE, Pearson JL, 2002. Contact with mental health and primary care providers before suicide: a review of the evidence. *Am. J. Psychiatry* 159, 909–916. 10.1176/appi.ajp.159.6.909. [PubMed: 12042175]

- Marcoulides GA, Schumacker RE, 2001. *New Developments and Techniques in Structural Equation Modeling*. Erlbaum L, Mahwah NJ.
- Matarazzo BB, Brenner LA, Reger MA, 2019. Positive predictive values and potential success of suicide prediction models. *JAMA Psychiatry* 76, 869. 10.1001/jamapsychiatry.2019.1519.
- McCarthy JF, Bossarte RM, Katz IR, Thompson C, Kemp J, Hannemann CM, Nielson C, Schoenbaum M, 2015. Predictive modeling and concentration of the risk of suicide: implications for preventive interventions in the US Department of Veterans Affairs. *Am. J. Public Health* 105, 1935–1942. 10.2105/AJPH.2015.302737. [PubMed: 26066914]
- McCarthy JF, Cooper SA, Dent KR, Eagan AE, Matarazzo BB, Hannemann CM, Reger MA, Landes SJ, Trafton JA, Schoenbaum M, Katz IR, 2021. 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* 4, e2129900. 10.1001/jamanetworkopen.2021.29900.
- models.lidaseqmodel – Dynamic Topic Modeling in Python, n.d.
- NIMH, 2022. Suicide is a Leading Cause of Death in the United States [WWW Document]. Suicide. URL <https://www.nimh.nih.gov/health/statistics/suicide>.
- Nock MK, Millner AJ, Ross EL, Kennedy CJ, Al-Suwaidi M, Barak-Corren Y, Castro VM, Castro-Ramirez F, Lauricella T, Murman N, Petukhova M, Bird SA, Reis B, Smoller JW, Kessler RC, 2022. Prediction of suicide attempts using clinician assessment, patient self-report, and electronic health records. *JAMA Netw. Open* 5, e2144373. 10.1001/jamanetworkopen.2021.44373.
- OMHSP, 2021. 2021 National Veteran Suicide Prevention ANNUAL REPORT. VA Office of Mental Health and Suicide Prevention, Washington, DC.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E, 2011. Scikit-Learn: machine learning in Python. *J. Mach. Learn. Res* 12, 2825–2830. 10.5555/1953048.2078195.
- Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B, Watts B, Flashman L, McAllister T, 2014. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE* 9, e85733. [PubMed: 24489669]
- Rubin R, 2019. Task force to prevent Veteran suicides. *JAMA* 322, 295. 10.1001/jama.2019.9854.
- Rumshisky A, Ghassemi M, Naumann T, Szolovits P, Castro VM, McCoy TH, Perlis RH, 2016. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl. Psychiatry* 6, e921. [PubMed: 27754482]
- Theodoridis S, 2020. *Machine Learning: A Bayesian and Optimization Perspective*.
- VA, 2017. REACH VET, Predictive Analytics for Suicide Prevention.
- VA, 2022. 2022 National Veteran Suicide Prevention Annual Report. Washington, DC.

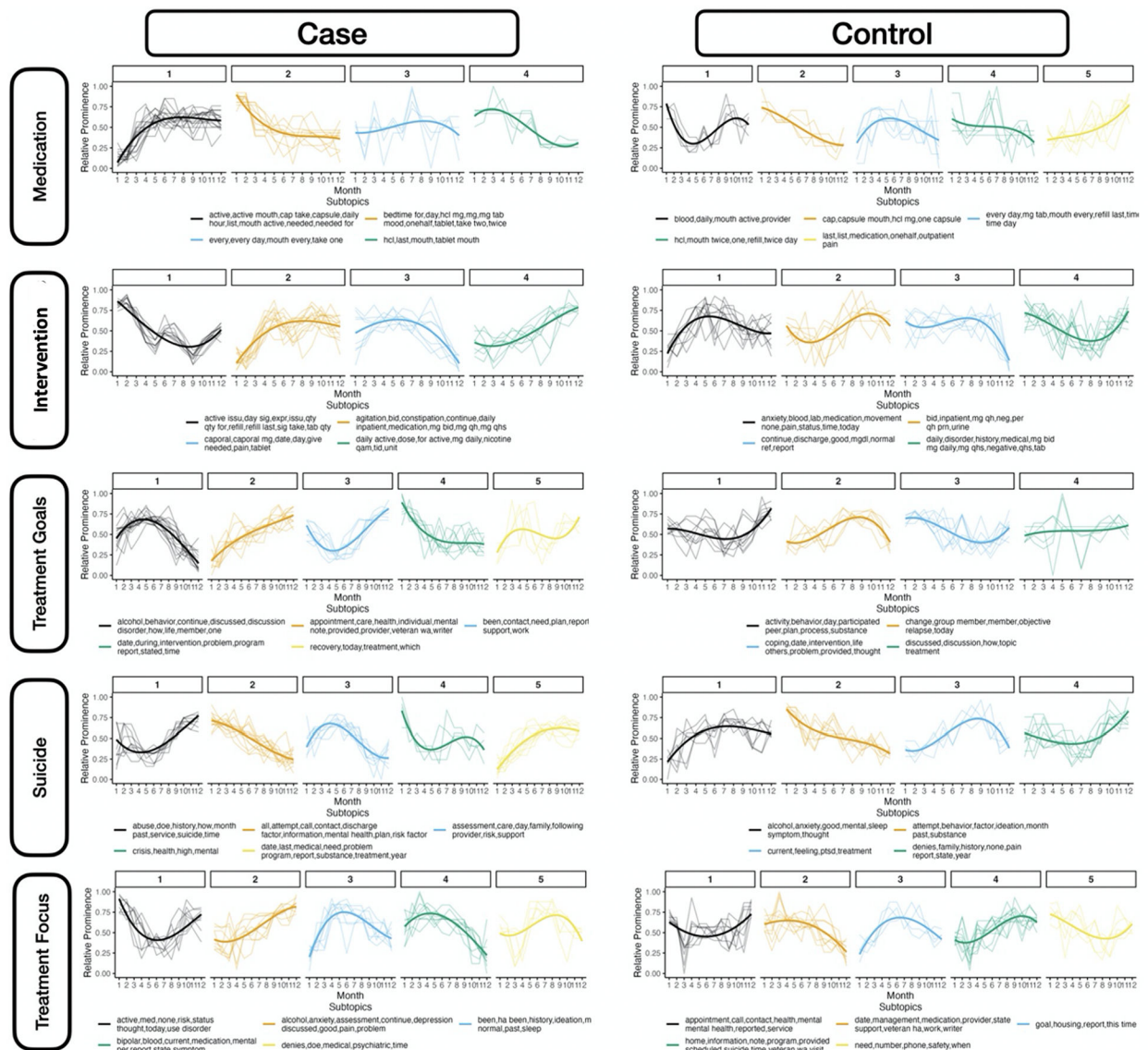


Fig. 1. Dynamic Time Warping Sub-Topic Trajectories.

This figure presents the relative prominence of identified terms (“sub-topics”), clustered from the top 50 most prominent words per topic across time within each topic that had comparable rates of change. Clustering was accomplished using dynamic time warping followed by hierarchical clustering with additional filtration of words and sub-topics. The thick line indicates overall trend per sub-topic, estimated by fitting a third-order polynomial to the data. The X-Axis presents Time (1 is index month, 12 is suicide for cases/end of observation for controls). The Y-axis presents Relative Prominence (higher values are more prominent).

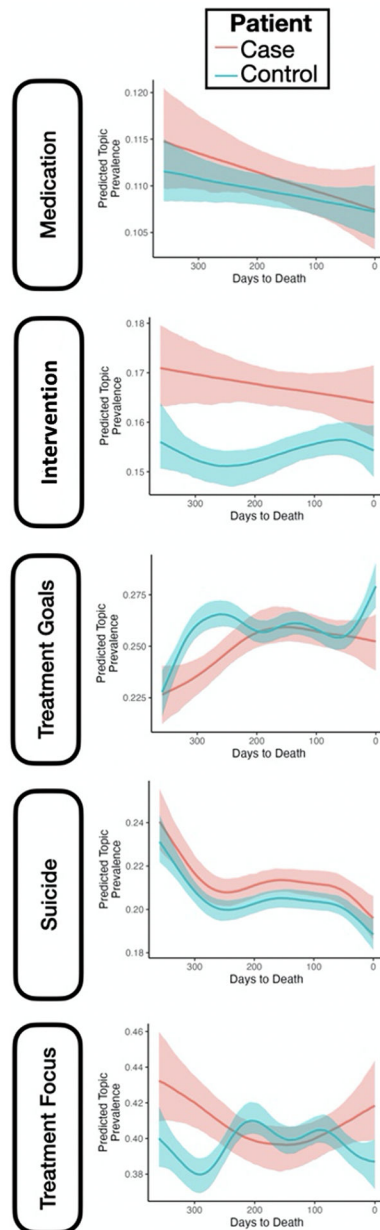


Fig. 2. Differences in topic prevalence over time by suicide status.

This figure presents Dynamic Topic Modeling (Blei et al., 2003) derived topics and monitors changes in topic prevalence over time within the case and control corpus. This figure presents each topic's prevalence among case and control records over time, estimated by using Bayesian hierarchical splines to assess statistical significance (translucent ribbons surrounding line indicates 95 % confidence intervals for the prevalence estimates).

Table 1

Case and control demographics. The sample was selected based on having the same high suicide risk score on Recovery Engagement and Coordination for Health–Veterans Enhanced Treatment (REACH-VET; VA, 2017), an algorithm that identifies patients with the highest suicide risk based on demographics and service usage variables. To evaluate the success of our matching methodology (cases ($n = 218$) and controls ($n = 943$)), we calculated standardized mean difference (SMD). We considered SMD values of 0.2–0.5 as small, values of 0.5–0.8 as medium, and values > 0.8 as large (Andrade, 2020). Following this metric, differences between cases and controls were very small, a finding that makes sense given that cases and controls were matched on REACH-VET suicide risk score.

	Sub-sample Match		
	Case ($n = 218$)	Control ($n = 943$)	SMD
Patient Demographics			
Female, n (%)	<11	65 (6.9)	0.10
Patient Age, Mean (SD)	52.76 (15.19)	53.46 (14.63)	0.05
Patient Age 18–34, n (%)	<11	133 (14.1)	0.10
Patient Age 35–54, n (%)	73 (33.5)	303 (32.1)	0.03
Patient Age 55–74, n (%)	99 (45.4)	461 (48.9)	0.07
Patient Age 75+, n (%)	<11	46 (4.9)	0.10
Patient Race/Ethnicity			
Non-Hispanic White, n (%)	189 (86.7)	743 (78.8)	0.21
Non-Hispanic Black, n (%)	<11	107 (11.3)	0.20
Hispanic, n (%)	15 (6.9)	69 (7.3)	0.02
Non-Hispanic APAC/AMIND/Unknown, n (%)	<11	22 (2.3)	0.10
Era of Service			
Post-911, n (%)	97 (44.5)	415 (44.0)	0.01
Vietnam, n (%)	72 (33.0)	286 (30.3)	0.06
Marital Status			
Married, n (%)	55 (25.2)	230 (24.4)	0.02
Divorced/Separated n (%)	96 (44.0)	455 (48.3)	0.09
Single n (%)	52 (23.9)	188 (19.9)	0.10
Other, n (%)	15 (6.9)	70 (7.4)	0.02
Service Connected Disability Percent			
None, n (%)	85 (39.0)	497 (52.7)	0.30
0–60 %, n (%)	54 (24.8)	192 (20.4)	0.11
70 %+, n (%)	79 (36.2)	254 (26.9)	0.20
Burden of Physical Illness			
Low, n (%)	74 (33.9)	275 (29.2)	0.10
Mid, n (%)	86 (39.4)	351 (37.2)	0.05
High, n (%)	58 (26.6)	317 (33.6)	0.15
Burden of Mental Illness			
Low, n (%)	12 (5.5)	21 (2.2)	0.17
Mid, n (%)	56 (25.7)	235 (24.9)	0.02

	Sub-sample Match		
	Case (n = 218)	Control (n = 943)	SMD
High, n (%)	150 (68.8)	687 (72.9)	0.09

Notes. Following death-related research protocols, we have suppressed any REACH-VET variables that were found to contain less than 11 individuals or more than 66 individuals.

Table 2

Identified topics.

This table presents the five identified topics. The relationship between associated case and control topics is measured by Rank Biased Overlap (RBO; Amigó et al., 2018) coefficients at baseline using Kendall Tau to derive confidence intervals (CI). The top 20 most frequent words within each topic are listed in order of descending prominence. Topic names were selected based on clinical team, including a psychiatrist, psychologist, and social worker, topic review and ranking. Bold words are indicative of being present in both case and control topics. Abbreviations are noted beneath the table.

Medication	Intervention	Treatment goals	Suicide	Treatment focus
RBO: 0.88 95 % CI [0.85–0.90]	RBO: 0.75 95 % CI [0.70–0.79]	RBO: 0.63 95 % CI [0.56–0.69]	RBO: 0.5 95 % CI [0.41–0.58]	RBO: 0.09 95 % CI [0.0–0.19]
<i>Case</i>	<i>Control</i>	<i>Control</i>	<i>Case</i>	<i>Control</i>
take	active	group	risk	medication
mouth	MG ^a	session	suicide	disorder
MG ^a	tab	treatment	health	denies
active	PRN ^c	recovery	treatment	mood
medication	tab	how	plan	thought
day	MG ^a	goal	history	history
tablet	last	discussed	day	need
tablet	QH ^d	veteran was	factor	care
tab	Daily	writer	time	program
one	QH ^e	disorder	past	provider
MG ^a tab	medication	recovery	mental	health
take one	PRN ^c for	skill	current	report
tablet mouth	QH ^d PRN ^c	comment	assessment	alcohol
every	refill	discussion	care	symptom
one tablet	sig take	disorder	been	call
tab take	MG ^a QHS ^e	date	service	veteran was
needed	BID ^f	resolved	family	service
HCL ^b	day	day	contact	discussed
bedtime	thought	support	mental health	date
needed	treatment	problem	symptom	housing
HCL ^b	expire	therapy	normal	stated
time	pain	blood		

Medication	Intervention	Treatment goals	Suicide	Treatment focus
RBO: 0.88 95 % CI [0.85–0.90]	RBO: 0.75 95 % CI [0.70–0.79]	RBO: 0.63 95 % CI [0.56–0.69]	RBO:0.5 95 % CI [0.41–0.58]	RBO: 0.09 95 % CI [0.0–0.19]
mouth active	mouth every	urine	problem	MG ¹ mental
	continue	reported	program	month

^aMG: Milligram.

^bHCL: Hydrochloric acid (frequently a suffix in generic medication names).

^cPRN: *pro re nata*; “as needed.”

^dQH: *quaque hora*; “every hour.

^eQHS: *quaque hora somni*; “at bedtime.”

^fBID: *bis in die*; “twice a day.”

Table 3
Within group topic changes over time.

This table presents case (top) and control's (bottom) derived Dynamic Topic Modeling (Blei et al., 2003) topic correlations, comparing the beginning and end of the treatment interval. We used Rank Biased Overlap (RBO; Amigó et al., 2018) to evaluate differences over time. RBO is a score between 0 (no agreement) to 1 (perfect agreement), representing the overlap between two sets of words and weighting words that are ranked more highly in each list (Bonett and Wright, 2000). We calculated 95 % confidence intervals based on a parametric assumption that the RBO would be distributed similarly to the Kendall Tau, a related measure of agreement (Lapata, 2006). The case and control topics that have asterisks had non-overlapping confidence intervals and are thus significantly different from each other.

Case Rank Biased Overlap				
Medication	Intervention	Treatment Goals	Suicide	Treatment Focus
RBO: 0.80 95 % CI [0.75–0.84]	RBO: 0.83* 95 % CI [0.78–0.86]	RBO: 0.73* 95 % CI [0.67–0.78]	RBO:0.73 95 % CI [0.66–0.78]	RBO: 0.83 95 % CI [0.79–0.87]
Control Rank Biased Overlap				
Medication	Intervention	Treatment Goals	Suicide	Treatment Focus
RBO: 0.81 95 % CI [0.76–0.85]	RBO: 0.93* 95 % CI [0.91–0.94]	RBO: 0.89* 95 % CI [0.86–0.91]	RBO: 0.79 95 % CI [0.74–0.82]	RBO: 0.88 95 % CI [0.84–0.90]