

RESEARCH

Open Access



Acute suicidal ideation in context: highlighting sentiment-based markers through the diary entries of a clinically depressed sample

Damien Lekkas^{1,2*} , Amanda C. Collins^{1,3} , Michael V. Heinz^{1,4} , Tess Z Griffin¹ , Arvind Pillai⁵ , Subigya K. Nepal⁵ , Daniel M. Mackin^{1,3} , Andrew T. Campbell^{1,5} and Nicholas C. Jacobson^{1,2,3,4,5}

Abstract

Background Despite major strides in conceptualizing and modeling the multifaceted nature of suicidal thought and behavior (STB) over the past few decades, the overall predictability of STB has not improved. This may be partly due to the dynamic nature of suicidal ideation (SI), which often fluctuates over hours, yet is largely overlooked in studies. Bolstered by the application and promise of natural language processing (NLP) across the mental health field, efforts toward richer operationalization of acute SI may include analyses on written data that occur alongside changes in SI, thus offering a better understanding of STB as it unfolds.

Methods Ecological momentary assessment (EMA) data from 268 participants with major depressive disorder (MDD) were utilized to investigate acute changes in SI. Data consisted of thrice-daily SI severity scores measured through self-report responses to item 9 of the Patient Health Questionnaire mobile version (MPHQ-9) as well as free-form diary text. Using difference scores and probability of acute change thresholds, eleven acute SI phase trajectory types were defined to label change in SI over three consecutive EMAs. In total, 5,938 acute SI trajectories were paired with the temporally centered diary entries. The Sentiment Analysis and Cognition Engine (SEANCE) tool was applied to quantify the written content of each diary entry across eight established lexica. Entry results were grouped based on phase trajectory type, and the Kruskal-Wallis test was employed with post-hoc multiple hypothesis correction to statistically compare SEANCE features between all group pairs.

Results There were 131 statistically significant (adjusted p -value < 0.05) pairwise differences between acute SI phase trajectory groups, implicating 31 NLP features. Consistent with the literature, results highlighted qualities of writing that are generally associated with heightened SI, including personal pronoun usage, passivity, and negative valence. Patterns of significance also uncovered novel contextual nuance in terms of how characteristics such as verbosity, hostility, anger, and pleasantness present in relation to SI over short change trajectories.

*Correspondence:

Damien Lekkas
Damien.Lekkas.GR@dartmouth.edu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions This work provides an accessible exploratory framework that capitalizes on the benefits of dense EMA sampling and NLP to profile and quantify acute SI trajectories. The use of the MPHQ's item 9 to quantify SI is an important limitation as it is designed to also capture precursory SI, passive SI, and SI-adjacent behaviors, potentially overestimating the SI expressed by participants. Nonetheless, future research should continue to focus on short timeframes as there are likely important signals and interpretative nuances to SI expression that have yet to be fully detailed.

Keywords Acute suicidal ideation, Natural Language processing, Lexicon, Sentiment analysis, Ecological momentary assessment, Free-text response

Background

With 135 suicide deaths per day on average and approximately 1.6 million suicide attempts in 2022 alone, suicide is currently the eleventh leading cause of death in the United States [1]. Despite these numbers, 95% of Americans believe that suicide can be prevented and 75% believe that those who die by suicide provide signs of their intent beforehand [2]. Since Durkheim's seminal study of suicide in 1897 [3], countless theories and models for suicidal thought and behavior (STB) have been proposed, incorporating biological [4], sociological [5], and psychological [6] mechanisms. In recent years, ideation-to-action-based frameworks have become among the most comprehensive and practical, drawing from over a century of ideas and empirical data [7]. These most prominently include the Interpersonal-Psychological Theory of Suicide (IPTS; [8]), the Integrated Motivational-Volitional model (IMV; [9]), and the Three-Step Theory of suicide (3ST; [10]) which have served as the basis for investigations that use cutting edge quantitative and analytical techniques to model STB [11–16].

Despite these theories providing utility to understand STB within research and clinical contexts [17–21], a recent meta-analysis has underlined an overall weaker-than-expected ability to detect and predict STB [22]. Importantly, this extends beyond automated models, where clinician assessments have been found to be slightly better than chance at predicting suicide attempts in at-risk individuals [23]. One reason for this finding may be due to the highly dynamic and heterogeneous nature of suicidal ideation (SI) [11]. While the field of suicidology acknowledges the temporal complexity of STB presentation, Fluid Vulnerability Theory (FVT; [24]) is the only theoretical model of STB that explicitly supports the conceptualization of SI as a dynamic change process with fluctuations over short time scales (i.e., hours or days). FVT also provides a strong precedent for study designs with shorter sampling intervals to be more aligned with the evolving phenomenology of STB. By extension, FVT highlights why so many longitudinal and sparsely sampled investigations have been unsuccessful in their ability to predict STB risk [25]. Indeed, research and modeling of SI is largely cross-sectional or consists of measures over coarse-grained periods of months or

years, and it is only within the past decade that works have begun to focus on acute SI trajectories [11, 26]. Such efforts have documented within-person variability in SI over hours [11] and days [27], and have linked transitory SI with key constructs from IPTS, IMV, and 3ST [11]. Moreover, within clinical contexts, temporal patterns and acute fluctuations in SI presentation have been shown to better characterize vulnerability to recurrent attempts among those receiving outpatient treatment [28].

Notions of STB as a constantly evolving system motivate more precise work aimed at characterizing the array of different risk trajectories and qualitative states that may define acute SI as it unfolds. For example, an individual at the height of an SI escalation over the past several days may express themselves differently from an individual who has rapidly transitioned into a severe ideation state within the past few hours. Despite these two hypothetical individuals expressing high momentary SI, they may possess different risk profiles. In other words, an identical set of observable factors may ultimately lead to different capabilities or patterns in these factors to quantify risk—a result of capturing SI at phenomenologically and contextually different stages along its trajectory. Much of the literature supports this, as efforts in testing the utility of certain risk factors as predictive tools, including time-invariant operationalizations of SI, have been frequently met with inconsistent or contradictory results [29]. Where one behavioral proxy may reliably flag for short-lived bouts of heightened SI, another may be no better than chance in this regard, yet excel at predicting an impending decline in SI from a heightened and sustained state. A 50-year meta-analysis on STB modeling [25] offers insightful direction related to this—the application of advanced technology to more densely sample SI in real time will lead to a more complete understanding of how SI presents within finer temporal intervals.

The rates of smartphone ownership have sharply risen over the past few years with 90% of the United States population owning a smartphone [30]. As a result, smartphones have become a burgeoning method for the unobtrusive, dense, and ecologically valid collection of behavioral and mental health-related data, including SI and STB [11–13]. Smartphones allow for data collection within the context of day-to-day life rather than

within the traditional laboratory setting, which tends to offer limited generalizability and temporal granularity. Prior work within the STB space has leveraged ecological momentary assessment (EMA)—the repeated sampling of current behaviors and experiences in real time and within a subject's natural environment—to study passive SI [31–35], active SI [32–39], suicide desire [37], capability for suicide [40], intent to die by suicide [37, 39], and ability to keep oneself safe or resist the urge to die by suicide [37, 39]. More broadly, the dense and longitudinal nature of EMA data has provided novel insights for highly dynamic psychiatric constructs such as person-specific depression, including person-specific depression symptom profiles [41]. Given the acute nature of SI, further exploration and development of SI markers may benefit from novel and creative operationalizations of EMA-assessed SI that take full advantage of the dense and ecologically valid sampling approach. Doing so may ultimately support useful and novel models for SI which are powered to detect or predict temporally nuanced changes in SI.

One way to enhance the contextual utility of EMA data is to include prompts that invite free-form qualitative responses (e.g., diary entries) alongside more structured, quantitative assessments of state (e.g., SI severity). While both qualitative and quantitative responses are common among EMA-based studies, they are less frequently modeled in combination, especially within suicide research. The further development of, and ease of access to, the vast analytical toolkit of natural language processing (NLP) provides a rich array of means with which to extract and dimensionalize qualitative text for integration into downstream quantitative applications. Accordingly, the use of NLP in mental health research has grown exponentially over the past decade and has been applied in both clinical and non-clinical research [42, 43], including constructs and outcomes within STB [14, 44–48].

While there are many techniques that subsume NLP, sentiment analysis—the assessment of the affective qualities of free-form text—demonstrates strong utility for mental health research. Briefly, there are three overarching approaches to deploying sentiment analysis: (i) lexicon-based, (ii) machine/deep learning, and (iii) hybrid approaches. Where lexicon-based approaches typically rely on manual, expert curation to develop lists of mapped word-value assignments, automated approaches involve the training of machine learning models with labeled or unlabeled text data to automatically recognize key properties of words [49]. While lexicon-based approaches offer greater interpretability, customization, and relative simplicity of deployment/application, automated approaches provide scalability, an ability to detect nuanced patterns, and higher task-specific performance [49]. As the name would suggest, hybrid approaches

combine elements of both lexicon-based and automated approaches, such as first preprocessing raw text through lexica to derive features that are then fed through a deep learning model architecture [50]. No approach is superior to another, and choice may depend on the priorities of the research endeavor (e.g., interpretability over performance, exploration over hypothesis testing, or domain independence over domain dependence).

The current work aimed to identify and interpret novel semantic markers and patterns in acute SI by methodologically focusing on the deployment of lexicon-based sentiment analysis, which has a broad suite of resource databases to explore [51, 52]. These include lexica related to emotions and emotional intensity [53, 54], affective norms [55], polarity [56], broad topical content [57], and the flow of communication [58]. Moreover, lexicon-based approaches have been frequently employed within mental health and STB-related research. For example, the Emotion Lexicon (EmoLex) [53] was used to model and predict emotional responses to public events on social media [59]; the Affective Norms for English Words (ANEW) lexicon [55] served to help predict the emergence of depression and post-traumatic stress disorder online [60]; and Valence Aware Dictionary and Sentiment Reasoner (VADER) [52] features have been applied to machine learning-based pipelines to detect suicide risk in Twitter posts [61]. In each of these applications, the models performed well, with areas under the receiver operating characteristic curve (AUC; a metric that reflects a model's ability to discriminate between two classes and is derived from a plot of the trade-off between the true positive rate and false positive rate) ranging from 0.87 to 0.96.

Given the demonstrated need to develop more temporally appropriate and sensitive models of acute SI, the current work aimed to leverage densely collected EMA and free-text response data. Specifically, this research aimed to statistically explore how written language may be used to detect acute shifts in SI severity. Importantly, this work sought to address the current lack of temporal context within the risk factor literature by first developing a classification scheme for acute SI trajectories based on EMA. At baseline, the aim of the scheme was to provide an objective and easy-to-interpret framework for distinguishing among equivalently or similarly severe momentary SI states. Based on SI measures that reflect transitions among current, immediate past, and immediate future states (i.e., 4–8 h timeframes), this profiling of acute SI served to support the discovery of NLP-based features that (i) are associated with near-future increases or decreases in SI severity, (ii) signal for current heightened SI severity states, and (iii) empirically highlight the importance of understanding the acute context in which emotions co-occur with SI. In this manner, these efforts

can be used to guide and support future research that aims to more fully appreciate the dynamics of SI, ultimately working toward improved detection, prediction, and prevention of STB-related outcomes.

Methods

Human subjects protection

The current study received approval from the Committee for the Protection of Human Subjects at Dartmouth College (STUDY00032081). Participants provided written informed consent prior to the collection of information and verbal informed consent prior to the start of the study. This consent procedure was approved by the Committee.

Participant recruitment and screening

Recruitment and screening

Participants across the United States were recruited online via Google Ads and Meta Ads from February 17, 2021 to December 1, 2023. The ads were directed at individuals with depression, and those interested in the study clicked on the ad to complete a screener questionnaire. An initial consent form was completed by 15,507 participants, and 4,625 of these participants completed the screener. Participants who were at least 18 years of age, owned an Android as their primary smartphone device, and self-reported elevated Major Depressive Disorder (MDD) symptoms on the Patient Health Questionnaire-9 (i.e., sum score ≥ 10) were invited to complete a more detailed screening procedure. Conducted on 1,305 participants, the procedure consisted of a more comprehensive battery of assessments and a 20-minute semi-structured screening interview via Zoom which reviewed participant depressive symptoms over the past 30 days. Participants who were still deemed as eligible due to self-report of an MDD-defined depressive episode over the past 30 days were invited to participate in a final clinical interview. This involved the Structured Clinical Interview for DSM-5 (SCID), which was administered via Zoom by either a psychiatrist or a postdoctoral research fellow in clinical psychology. Of the 431 participants interviewed, 312 met criteria for a current MDD episode and did not report a history of (i) mania, (ii) psychoses, or (iii) currently endorse moderate to high acute suicide risk and therefore were invited to participate in the main study.

Moderate to high suicide risk was determined based on self-reported endorsement of items 1 (a wish to be dead) through 5 (planning and intent to carry out a plan) on the screen version of the Columbia-Suicide Severity Rating Scale (C-SSRS) [62]. The C-SSRS was automatically administered as part of the initial screening process to any participant who provided a “non-zero” response to item 9 on the Patient Health Questionnaire-9 (PHQ-9) [63]. Items were presented within the context of the

past month. If any participant endorsed item 3 (thoughts about how to attempt), item 4 (thoughts with some intention to act), or item 5, reflecting moderate or high risk, the participant was considered at-risk of indicating active suicidality during the study and was excluded from enrollment. The C-SSRS was developed across multiple institutions and with support from National Institute of Mental Health (NIMH) [62], and its validity in assessing suicide severity and immediacy has been demonstrated across a wide range of settings and populations [64].

Participants

The current investigation serves as a secondary analysis of Tracking Depression—an NIMH-funded research effort to monitor and study depression across 90 days [41, 65]. At the conclusion of recruitment (see above), 300 individuals met criteria for current MDD and were included in the 90-day study. Of these, $N=268$ participants had sufficient ecological momentary assessment (EMA) and free-text diary entry data available for use in the current analysis (see “Data preprocessing and sub-selection” below). The overall research effort sought to recruit a nationally representative sample of individuals with MDD. Accordingly, participants in the present investigation were primarily White ($n=212$; 79.1%), non-Hispanic ($n=236$; 88.1%), and women ($n=225$; 84%). In addition, 99.3% of participants ($n=266$) endorsed some level of SI during the 90-day study. See Table 1 for additional information on cohort sociodemographics.

Data collection and procedure

Participants meeting eligibility requirements were onboarded on a rolling basis by a research coordinator. The onboarding process included downloading the Android-based MLife app [66]. Following onboarding, participants began the study and completed EMA prompts via the MLife app three times per day for 90 days. This resulted in up to 270 total possible daily EMAs per participant. Each EMA included the PHQ-9, an optional free-text response, as well as questions beyond the scope of the current investigation (e.g., in regards to anxiety and somatic symptoms). Prompts to complete the EMAs began four hours after a participant’s self-reported wake time (e.g., the first prompt of the day would be at 11 AM if the reported wake time was 7 AM). The next two prompts were delivered in four-hour intervals following the first prompt. Additionally, participants received a prompt every Monday morning to complete a weekly assessment, which included the PHQ-9, optional free-text response, and other questions not relevant to the current investigation. On average, it took participants 54.0 s ($\sigma=70.5$) to complete an EMA. Participants were compensated \$1 at the end of the study for every completed daily and weekly EMA. Participants additionally

Table 1 Cohort sociodemographic summary

Characteristic	Attribute	mean (sd) or count (%)	Characteristic	Attribute	count (%)
Age	—	40.72 (11.7)	Hispanic / Latino Ethnicity	No	236 (88.1)
Gender	Women	225 (84.0)	Highest Level of Education	Yes	32 (11.9)
	Men	28 (10.4)		Less than high school	2 (0.7)
Transgender	Non-binary	11 (4.1)		High school	15 (5.6)
	Other (Prefer to self-describe)	4 (1.5)		Trade/technical school	3 (1.1)
Sexual Orientation	No	258 (96.3)	Current Student Status	Some college, no degree	64 (23.9)
	Yes	10 (3.7)		Associate degree	23 (8.6)
Sexual Orientation	Heterosexual	177 (66.0)		Bachelor's degree	88 (32.8)
	Homosexual	14 (5.2)		Master's degree	60 (22.4)
	Bisexual	35 (13.1)		Doctoral degree	13 (4.9)
	Bicurious	6 (2.2)	Income	Not a student	220 (82.1)
	Pansexual	18 (6.7)		Part-time	14 (5.2)
	Asexual	13 (4.9)		Full-time	34 (12.7)
Race	Other (Prefer to self-describe)	5 (1.9)		Less than \$20,000	46 (17.2)
	White	212 (79.1)	Income	\$20,000 to \$39,999	49 (18.3)
	Black or African American	18 (6.7)		\$40,000 to \$59,999	49 (18.3)
	Asian	12 (4.5)		\$60,000 to \$79,999	39 (14.6)
	American Indian or Alaskan Native	1 (0.4)		\$80,000 to \$99,999	26 (9.7)
	More than one race	19 (7.1)		\$100,000 to \$149,999	38 (14.2)
	Other (Prefer to self-describe)	6 (2.2)		\$150,000 or more	21 (7.8)

Note. N = 268

Table 2 Statistical summary of EMA data

Overall Cohort Statistics	Count (%)
Total participants	268
Participants with SI reported	266 (99.3)
Total EMAs	62,243
EMAs with SI reported	28,236 (45.4)
Total diary entries	9,586
Diary entries with representative EMA triplicates [†]	5,938 (61.9)
Per Participant Statistics	Mean [min, max]
EMA count	232.3 [6, 419*]
Diary entry count	35.8 [1, 236]
Percentage of EMAs with diary entry	15.9 [0.3, 86.4]
EMAs with SI reported	105.4 [0, 346]
EMAs with no SI reported	126.9 [0, 305]
Average SI severity	9.4 [0, 95.4]
Maximum SI severity	55.2 [0, 100]
Minimum SI severity	0.9 [0, 40]

Note. SI severity ranged on a sliding scale from 0 to 100. [†]See Data Preprocessing and Subselection within the Data Analytic Plan as well as Fig. 1 below for more information. *While the maximum possible number of EMAs (and free-text responses) for the 90-day study period was 270, participants could continue to provide EMA and free-text response data through the MLife app past this window. Only one participant opted to do so. Moreover, participants could self-initiate EMAs at any time; however, only the most recent and valid EMA completed within each time interval was used to represent that data point (thus, a maximum of three representative EMAs per day)

received a bonus of \$50 for completing at least 90% of EMAs. Evaluation of compliance rates indicated a high level of compliance for participants, with a median of 256 (94.8%) EMAs completed per participant. Of those, 105 (45.4%) of EMAs (on average per participant), included

an endorsement of SI. See Table 2 for more detailed statistics on participant EMA compliance.

Measures

Patient Health Questionnaire-9

The PHQ-9 is a nine-item self-report measure that assesses the frequency of depressive symptoms over the past two weeks [63]. For the current study, we utilized a previously developed mobile-friendly version of the PHQ-9 (MPHQ-9) that prompted participants to respond with how they felt over the prior four hours [35, 67]. This version has demonstrated high correlations with the original version ($r=0.84$) and good adherence rates (77.78%) [35]. In a recent analysis on the Tracking Depression study data, the MPHQ-9 demonstrated excellent internal consistency ($\alpha = 0.91$), good longitudinal stability ($r=0.69$), and good convergent validity with the traditional PHQ-9 ($r=0.71$) and the Inventory of Depression and Anxiety Symptoms – Expanded Version ($r=0.65$) [68]. Moreover, it has been used to good effect in other investigations within the larger Tracking Depression study [41, 69–71]. Participants responded to the nine questions, each representing an individual depressive symptom, via a sliding scale from 0 (not at all) to 100 (constantly). To provide participants with more context and to minimize any potential floor or ceiling effects in what constitutes “experiencing a symptom,” participants were asked to reflect on the best and worst they had ever experienced each symptom in their lives when responding to the EMA prompt on frequency over the prior four

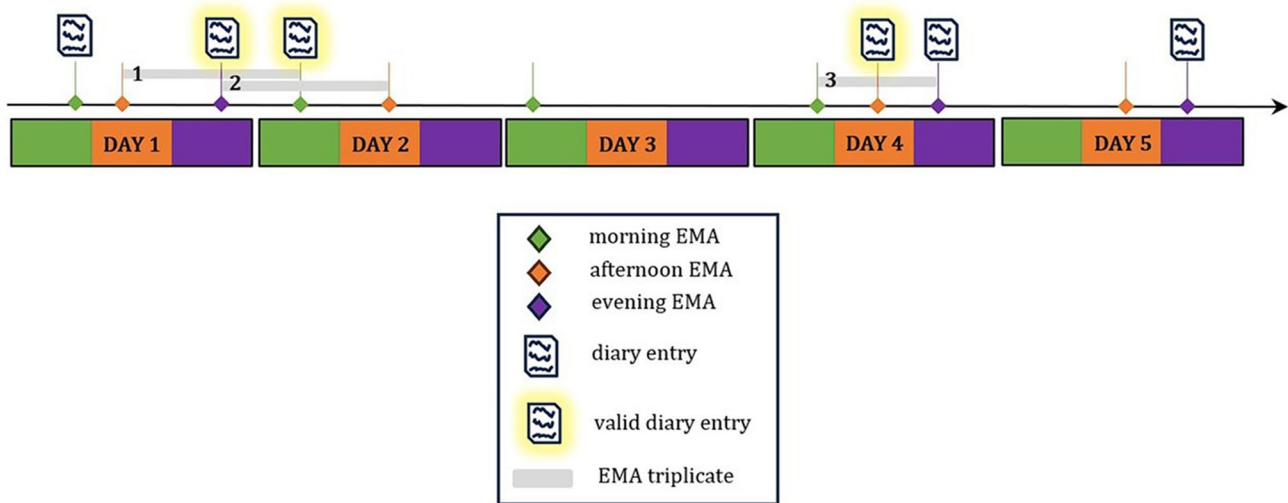


Fig. 1 Hypothetical Example of Valid Participant EMA Triplicates *Note.* EMAs and diary entries completed over the course of the first five days for a hypothetical participant are shown. Data used for analysis was selected based on the requirement that a diary entry was completed as the second of three sequentially completed EMA. Each valid diary entry is illustrated (highlighted) alongside its respective EMA triplicate (gray bar that spans three consecutive EMAs). In the example shown, there are three valid data points (labeled “1,” “2,” and “3”)

hours. Importantly, participants did not have access to their prior EMAs and could not refer to their responses earlier in the day or week. The original version of the PHQ-9 has demonstrated consistently good sensitivity, specificity, internal reliability, and sensitivity to change [72, 73].

Free-Text responses

At the end of each EMA, participants were given the option to provide a free-text response to the prompt, “Please describe how you’ve been feeling over the past four hours and why.” Within a specified 2,500-character limit (increased from a 500-character limit approximately halfway through the study), participants were able to write about events that had transpired since their last prompt, their current thoughts or emotions, or anything else that they wished to write. Importantly, this was an optional component of the EMA with an option to skip this prompt each time if a participant did not wish to provide a text entry.

Data analytic plan

Data preprocessing and subselection

For each of the 300 participants, EMA responses that included a completed optional free-text response (hereafter, “diary entry”) were selected, resulting in 9,586 diary entries across $N=268$ participants. Thirty-two participants did not engage with the diary entry component of the EMA and thus were excluded from analysis. Next, these 9,586 diary entries were further subset to only include those that had a completed EMA (themselves, either with or without an associated diary entry)

immediately before and immediately after. For example, if a participant on a given day completed an afternoon diary entry, and it was followed by an evening EMA but not preceded by a morning EMA, then that diary entry was removed from the dataset. See Fig. 1 for an illustrated example of this rule. This procedure resulted in 5,938 diary entries with temporally adjacent EMAs for analysis. Of particular interest to the current study was item 9 of the PHQ-9 which asks, “In the past four hours, have you had thoughts that you would be better off dead, or of hurting yourself?” Thus, the final dataset for modeling consisted of 5,938 diary entry-SI score triplicates across the cohort. The text of each diary entry served to contextualize an acute SI trajectory for a given participant, herein quantified by the self-report response scores to item 9 of the PHQ-9 before, coincident, and after the diary entry.

It is important to note that item 9 is a broad construct that not only measures SI, but also may capture precursory SI through “thoughts of being better off dead” as well as SI-adjacent behaviors through (“thoughts of hurting yourself”). Generally speaking, the literature is mixed in terms of the PHQ-9’s usefulness as a measure of STB; however, it has the advantage in this exploratory work of being a wide net with which to capture subtle changes in SI and SI-related expressions against the backdrop of those suffering from clinical depression. While it would be preferable to use alternative measures of SI, the use of the PHQ-9 within the context of temporally dense EMA was not only minimally burdensome for participants, but was bolstered by previous work which has begun to establish its deployment within this setting [35, 68].

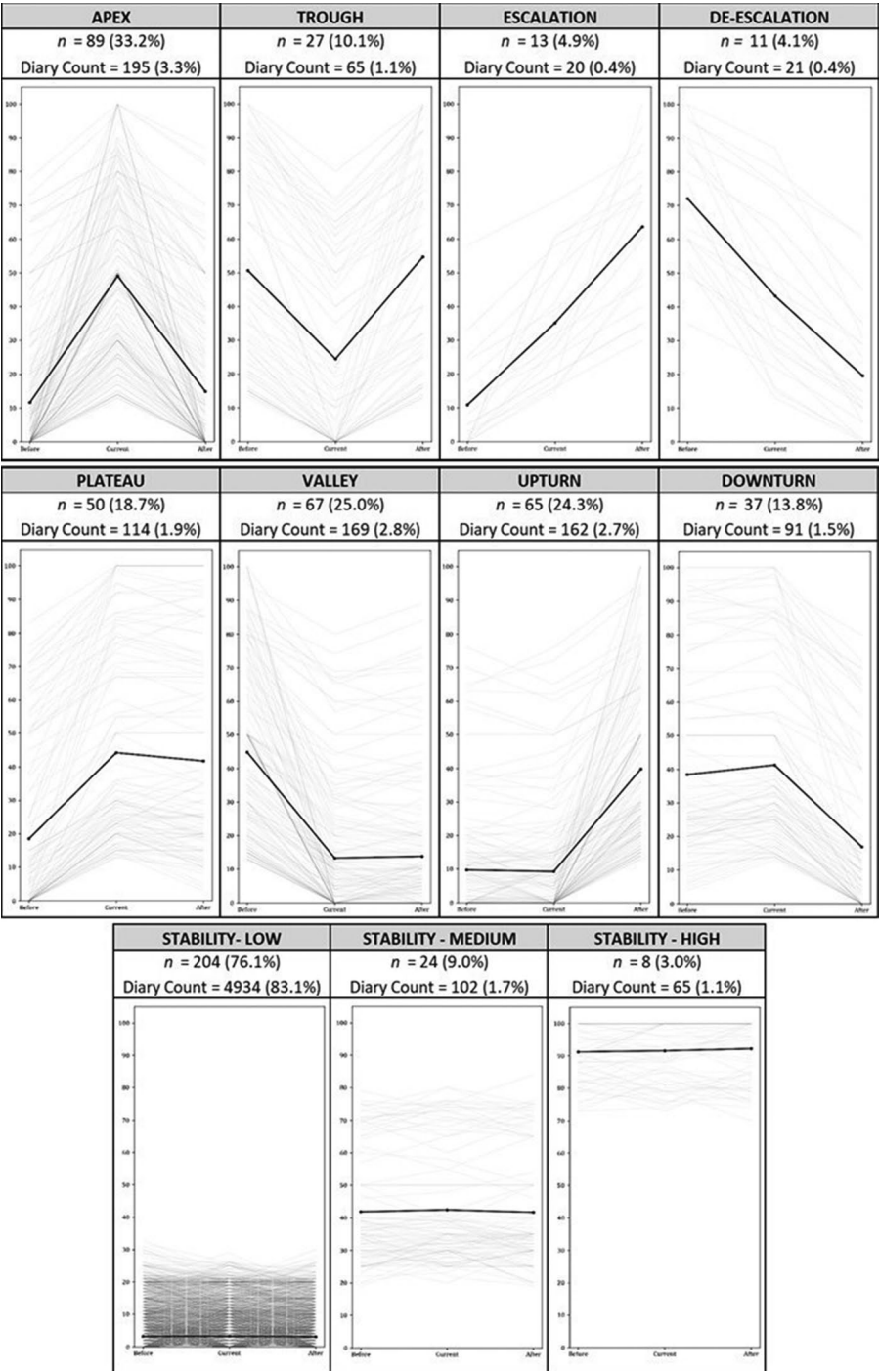


Fig. 2 Summary of Acute SI Phase Trajectories *Note.* Each phase (panel) is illustrated and exemplified by the SI severity trajectories endorsed in the data. A dark line represents the average trajectory for a given phase type, while lighter gray lines represent the individual trajectories. Each panel also reports the total number (*n*) and percentage of participants with at least one instance of the respective trajectory as well as the total number and percentage of diary entries represented by the respective trajectory

Defining acute SI trajectories

Because self-report SI was quantified in terms of a sliding scale ranging from 0 to 100 (see “Measures” above), it was important to first define the limits of what would be considered “no significant change” between adjacent

timepoints. Accordingly, this work utilized probability of acute change (PAC) to determine the threshold of stationarity [74, 75]. Following recommendations in previous work, the difference scores between all consecutive EMAs across the entire cohort were calculated, and the

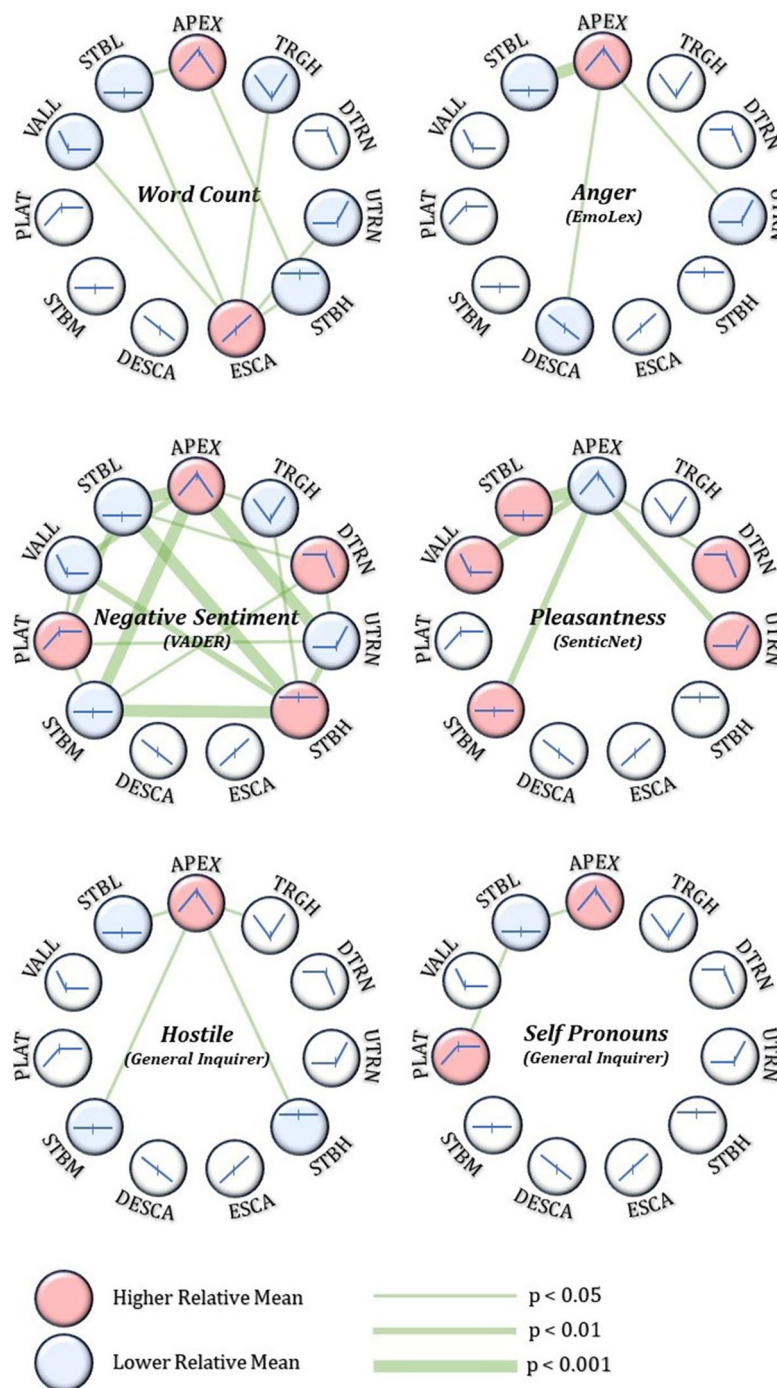


Fig. 3 Selected Significant Acute SI Phase Group Differences by Feature Note. Each network graphlet represents a SEANCE feature with statistically significant differences between acute SI phase trajectory groups. Each edge is a pairwise significant difference between two groups (nodes). Edge thickness is based on p -value significance threshold. APEX = Apex; TRGH = Trough; DTRN = Downturn; UTRN = Upturn; STBH = Stability—High; ESCA = Escalation; DESCA = De-escalation; STBM = Stability—Medium; PLAT = Plateau; VALL = Valley; STBL = Stability—Low

90th percentile value of difference scores was used as the minimum required to define “change” [74, 75]. As a result, any difference below 12.0 units was not considered a change for the purposes of trajectory profiling as described in the following paragraph. To operationalize

acute changes in SI, each of the score triplicates derived in “Data preprocessing and subselection” were assigned a trajectory class. Each trajectory class defines an acute “phase” of SI that is based on the context of endorsed SI severity surrounding a diary entry. Using the

PAC-derived threshold of change, eleven phases were defined to capture an exhaustive list of every possible three-point trajectory of change:

1. *Apex*. Endorsed acute SI severity is higher at the time of the diary entry in comparison to both the EMA immediately before and after the diary entry.
2. *Trough*. Endorsed acute SI severity at the time of the diary entry is lower in comparison to both the EMA immediately before and after the diary entry.
3. *Escalation*. Endorsed acute SI severity at the time of the diary entry is higher in comparison to the EMA immediately before the diary entry and lower in comparison to the EMA immediately after the diary entry.
4. *De-escalation*. Endorsed acute SI severity at the time of the diary entry is lower in comparison to the EMA immediately before the diary entry and higher in comparison to the EMA immediately after the diary entry.
5. *Plateau*. Endorsed acute SI severity at the time of the diary entry is higher in comparison to the EMA immediately before the diary entry and equivalent to the EMA immediately after.
6. *Valley*. Endorsed acute SI severity at the time of the diary entry is lower in comparison to the EMA immediately before the diary entry and equivalent to the EMA immediately after.
7. *Upturn*. Endorsed acute SI severity at the time of the diary entry is equivalent to the EMA immediately before the diary entry and lower in comparison to the EMA immediately after.
8. *Downturn*. Endorsed acute SI severity at the time of the diary entry is equivalent to the EMA immediately before the diary entry and higher in comparison to the EMA immediately after.
9. *Stability—Low*. Endorsed acute SI severity at the time of the diary entry is equivalent to both the EMA immediately before and after the diary entry while also ranging in severity from 0 to 24.
10. *Stability—Medium*. Endorsed acute SI severity at the time of the diary entry is equivalent to both the EMA immediately before and after the diary entry while also ranging in severity from 25 to 75.
11. *Stability—High*. Endorsed acute SI severity at the time of the diary entry is equivalent to both the EMA immediately before and after the diary entry while also ranging in severity from 76 to 100.

A visualization of each phase class is provided in Fig. 2. Additionally, Fig. 2 provides the number of participants who had endorsed a given phase at least once, as well as the total number of diary entries (and their unique trajectories) defined by the phase across the dataset.

Natural Language processing with SEANCE

To quantify the semantic and sentiment-based properties of each diary entry, this work utilized SEANCE (Sentiment Analysis and Cognition Engine), a highly cited and freely available natural language processing tool that leverages a rich suite of over 250 indices from eight well-known reference lexica [51]. SEANCE is provided as a stand-alone executable that takes in plain text files as input and outputs a comma-delimited file of NLP feature values that correspond with the content of the text. SEANCE allows for some customization at runtime, including which indices (lexica) to utilize, which types of words to analyze (i.e., nouns, verbs, adjectives, adverbs), and whether to turn on negation control. In the current study, all lexica were utilized, all words were considered, and negation control was enabled. Importantly, negation control operates by checking for negation terms (e.g., not) within a three-word window before and after each word in text. This is important for context as the word “happy” in “I’m not very happy” would be interpreted as positively-valenced in the absence of negation control. From the 271 NLP features generated across the 5,938 diary entries (see Supplemental File 1 for the complete dataset used in downstream analyses), 82 features were selected based on relevance and presence of variability across the data. For brevity, the following is a description of each lexicon along with example features and the total number of features utilized in the current work. A complete list of features along with their corresponding distributional statistics is provided in Supplemental File 2. For more details on the utilized lexica, features, and SEANCE, interested readers are encouraged to consult the excellent documentation and references provided by the SEANCE team [76].

1. *General Inquirer (GI)*. A lexicon consisting of 119 lists with over 11,000 words originally developed for content analysis in politics, sociology, and psychology [57]. Example features: self, virtue, work. Total number: 31.
2. *Lasswell*. Drawing directly from the *Dynamics of Culture*, these 63 lists are a subset of the GI which distinguish between substantive goals and the elements and attributes of personal evaluation and social allocation [77]. Example features: power, rectitude, well-being. Total number: 4.
3. *Geneva Affect Label Coder (GALC)*. Contains 38 short lists of emotion-based words [54]. Example features: boredom, envy, relief. Total number: 0.
4. *Affective Norms for English Words (ANEW)*. Measures valence, pleasure, arousal, and dominance of words for around 1,000 English words [55]. Example features: valence, arousal, dominance. Total number: 6.

5. *EmoLex*. Consists of ten emotion-based word lists that were developed using crowdsourced rating efforts in combination with expert curation and references to other established lexica to label the emotional properties of words [53]. Example features: anger, joy, sadness. Total number: 10.
6. *SenticNet*. Measures four emotional dimensions using a semi-supervised algorithmic approach to serve as an extension of WordNet [78, 79]. Example features: pleasantness, attention, sensitivity. Total number: 5.
7. *Valence Aware Dictionary for Sentiment Reasoning (VADER)*. Rule-based, crowd-sourced valence ratings developed for shorter texts that take syntactic and grammatical context into account [52]. Example features: negative, positive, neutral. Total number utilized: 4.
8. *Hu-Liu polarity*. Positive and negative word lists developed for product reviews and social texts [56]. Example features: proportion positive, number of negative words, proportion of positive to negative. Total number utilized: 5.
9. *SEANCE components*. Principal component analysis-based indices that cluster related features based on frequent co-occurrence [51]. Example features: failure component, certainty component, positive verbs component. Total number utilized: 16.

Group comparisons of natural Language features by acute SI phase

Stratifying the 5,938 diary entries and their associated NLP feature values into the 11 acute SI trajectory phase groups defined above (see Fig. 2), the Kruskal-Wallis test for multiple group comparisons (non-parametric equivalent to ANOVA) was performed on each of the 82 features (a total of 82 Kruskal-Wallis tests). Aligning with current recommendations [80], a minimum of five diary entries (each associated with three consecutive EMAs; 15 EMAs total) is required for each group to ensure robust approximation of the test statistic. As shown in Fig. 2, the smallest group (Escalation) exceeded this requirement (20 diary entries) and most groups (9 of 11) were more than an order of magnitude larger than the minimum recommended. The post-hoc Dunn test was then employed with the Benjamini-Hochberg correction procedure to control for multiple hypotheses. Given the exploratory nature of the work, control of the false discovery rate was selected over family-wise error control to ultimately highlight statistically significant feature differences between acute SI phase groups. This procedure was carried out using the *sci-kit posthocs* (v0.11.2) package in Python [81].

Results

There were 131 significant (adjusted p -value < 0.05) pairwise acute SI group differences across 31 NLP features. Most generally, these differences highlighted several semantic and sentiment-based differences in written content produced during heightened and diminished states of SI. These differences both recapitulate previous findings in the literature and offer promising support for further investigation into leveraging sentiment-based lexica (especially, VADER) in acute SI-based detection and prediction models. More specifically, findings drew attention to the contextual nuance of verbosity, hostility and anger (via the General Inquirer), as well as the emotional dimensions of pleasantness and aptitude (via SenticNet), that characterize the diary entries of individuals throughout various points in their SI trajectories. Additionally, features of writing such as increased personal pronoun usage, lower dominance, higher passivity, and a more negative outlook emerged as markers for more severe SI states. Throughout these group comparisons, the Apex contextual state of acute SI most frequently emerged as semantically unique and distinct from other states (78 of 131 significant associations; 60%). Figure 3 illustrates significant acute SI phase group differences across a selected group of NLP features. A complete account of statistics associated with these group differences is provided in Supplemental File 3.

From the helpful suggestion of an anonymous reviewer, a sensitivity analysis was also conducted after the fact to determine if there were any differences in significance patterns when utilizing a subset of the data derived from participants who did not exceed the 500-character limit as imposed during the first half of the study (see “Free-Text responses” in “Measures”). Of the 5,938 total diary entries, 365 (6.15%) were greater than 500 characters in length. No participant had diary entries that were exclusively more than 500 characters, and 43 participants had at least one entry that was over 500 characters. After re-running statistical analyses on this excluded subset of the data, 135 significant group differences emerged (compared to the 131 significant group differences as reported). The additional group differences involved features from the *EmoLex* lexicon (i.e., negative emotion, anger, and sadness), and served to reinforce specific contextual patterns related to anger and more general, non-contextual signals of heightened SI as described in the Discussion. Moreover, the same 31 identical NLP features were implicated. All major patterns and associations as discussed and presented are supported through the results obtained on this cohort subset.

Discussion

The current work utilized EMA data across a 90-day study period to systematically profile the acute trajectories of self-report SI severity for $N=268$ individuals with MDD. Focusing on a rich suite of NLP-based features, 11 classes of SI change were quantified and compared to explore statistical links between SI context and feature utility. From a starting set of 82 features, 31 were found to be statistically significant ($p < 0.05$), highlighting qualities of writing that were generally associated (and aligned in the literature) with heightened SI (e.g., personal pronoun usage, passivity, and negative valence). Importantly, some of these features were contextually nuanced in terms of how they present in relation to SI over short change trajectories (e.g., verbosity, hostility, and aptitude). Serving as an investigative and methodologically accessible template for future introspection and development within the acute SI space, this study offered a first look into signals and patterns of sentiment on relatively short (and more phenomenologically aligned) timescales of SI change. In doing so, the results draw preliminary attention to how the near-past and near-future trajectories of SI may influence the presentation and interpretation of in-the-moment, NLP-based signals of risk.

One feature that emerged as significantly different between multiple pairs of SI phase trajectory groups was word count (Fig. 3). Diary entries with significantly higher verbosity were found to belong to EMA triplets that reflected short-lived peaks in (APEX), and transitions to (ESCA), heightened states of SI severity. Moreover, these contrasted with states of both consistently low (STBL) and high (STBH) SI severity, where the number of words was significantly fewer. Where a previous linguistic analysis of suicide-related posts on Twitter (now X) generally found that the most strongly concerning tweets were characterized by a higher word count in relation to those that were non-suicide-related [82], there is no empirical work that more specifically distinguishes suicide-related content generated as a consequence of severe SI states with various gradations of permanency. To this end, the current finding suggests that there may in fact be a difference in the writing behavior of individuals with elevated SI that occurs within very short (e.g., 4–8 h) time scales, and that this difference in verbosity is more a function of SI's transiency than its severity. Future work may benefit from investigating how relatively simple structural characteristics of writing such as number of words, sentence length, parallelism, and readability may differ when penned across varying spans of heightened ideation states, as these may be useful signals of near-future increases or decreases in risk.

A contextually interesting trend emerged in both hostility and anger. As shown in Fig. 3, higher hostility signaled for a rapid transition into and out of a more severe

state of SI (APEX), while consistent levels of severe ideation had significantly lower hostility by comparison (STBH). This difference may suggest a capability for hostility to predict near-future decreases in heightened SI or serve as a marker of a short-lived, transient state of heightened SI. Higher anger, like hostility, was characterized by a rapid transition into and out of a severe SI state (APEX), yet unlike hostility, lower anger uniquely signaled for an ongoing or transitional state of SI decrease (DESC). In practical terms, a decrease or lack of anger in writing may serve as a marker for detecting remission of SI even if the individual is currently experiencing some level of appreciable ideation. In other words, a lack of anger need not be solely indicative of low or absent SI as would be expected, but may also define an active period of waning SI.

The literature strongly supports the importance of hostility and anger in profiling SI expression. For example, short-lived emotional hostility is consistent with how those with vulnerable narcissism—a trait for SI risk [83]—experience outbursts of this emotion that are followed in quick succession by shame and depression [84]. As such, the communication of hostility may be a useful marker for detecting spikes and predicting near-future decreases in SI severity. Additionally, broad conceptualizations of anger have been cross-sectionally associated with SI in adolescents [85], and aggression, along with its phenomenological ties to impulsivity, has been shown to be associated with increased suicidality and a likelihood of having a history of both SI and suicide attempts [86]. In support of acute temporal framing, irritability—a decreased threshold for anger—has been implicated as a proximal risk factor for elevated STB with differing escalating trajectories of STB across weeks influencing changes in baseline week-to-week irritability [87]. Moreover, in a study leveraging survey data from a large representative sample of adults in the United States, phasic irritability—episodic aggressive outbursts—was indicated to associate with SI, suicide planning, and suicide attempt [88]. Overall, the literature positively relates anger and associated emotional states with SI [89, 90] and the results herein complement these findings while providing additional temporal insight regarding the relative timing of how anger may decrease alongside SI.

Another interesting pair of significant features came from the SenticNet lexicon, where both higher and lower pleasantness in language signaled for a decrease in SI severity in the immediate future (Fig. 3). At first approximation, this result seems contradictory; however, higher pleasantness uniquely signaled for a drop in SI severity only when preceded by a more prolonged term of elevated SI (DTRN), while higher pleasantness was not indicative of decreased SI severity in the immediate future when the heightened state of SI was more transient (APEX). While

speculative, expressions of pleasantness may serve as a regulatory or coping mechanism to decrease SI following habituation to a higher severity state. The same pattern was observed for aptitude. As the negative emotion of disgust underlies the construct of aptitude (see following paragraph), one speculative thought is that more severe states of SI may accompany or reflect an individual's negative stable perceptions of themselves. Thus, an increase in aptitude (a decrease in loathing or disgust) at the end of a prolonged state of heightened SI may indicate an imminent breakdown in this mental paradigm. More precise research with greater levels of temporal contextualization is needed to properly address the potential explanations for these observed patterns. Nevertheless, the differences observed between short-term (APEX) and longer-term (DTRN) states of heightened SI suggest that how momentary emotion in language may signal for future SI is not consistent because it is in part a function of the immediate past. By extension, this result highlights the limitations of developing markers for SI without proper temporal context of the individual's SI history, a consideration that may be even more pertinent when dealing with predicting SI severity within relatively narrow windows of time. To the authors' knowledge, no work has offered the resolution of temporal change in SI to properly interrogate this dynamic more rigorously. To develop effective risk models of SI in an acute setting, it is important that future research efforts are sensitive to the contexts of an individual's SI state, thereby emphasizing temporal frameworks that quantify and model SI more completely.

Although pleasantness and aptitude were modeled discretely, they are tied to more complex and integrated conceptualizations of emotion. Indeed, SenticNet was developed using the Hourglass Theory of Emotion, which is an affective categorization model that organizes primary emotions around four independent and co-occurring dimensions (i.e., pleasantness, attention, aptitude, and sensitivity) with multiple levels of activation [91]. Thus, based on this framing, the overarching significance of lower pleasantness and lower aptitude dimensions as markers of heightened SI severity more precisely encapsulates associations with the negative emotions of grief, sadness, and pensiveness in the former, as well as loathing, disgust, and boredom in the latter. Most of these more nuanced emotions have been found in the literature to play key roles in STB progression and risk elevation [92–95], thus it may be fruitful to apply one or more components of SenticNet in future NLP-based studies to further explore the temporal qualities of these emotions in those with SI.

With 33 of 131 significant SI phase group differences attributed to VADER-based features of negative (17; Fig. 3), compound (15), and neutral (1) sentiment alone, VADER emerged among the leveraged lexica as most

consistently discerning in terms of highlighting sentiment-based differences between high and low severity SI states. In contrast to the features of SenticNet, which demonstrated a contextual sensitivity to near-past and near-future SI, VADER's features were more attuned to capture momentary associations between sentiment and SI. Further bolstered by previous works that have utilized VADER to promisingly probe the presentation of trauma-related [96], depression-related [97], and STB-related sentiment [98], the findings are nevertheless an endorsement for future work to explore the integration of VADER into the sentiment-based components of STB operationalization tasks.

Several other features, while not exhibiting patterns of significance between phase groups that can be used to inform acute SI context, nonetheless suggested a general utility in discriminating high from low SI severity. In terms of emotion, the current results implicate, lower positive emotion, joy and trust, as well as higher negative emotion and fear, as associated with heightened states of SI across the cohort; however, it is important to recognize that emotional experience in SI tends to be highly variable, dependent on the acute trajectories of co-occurring SI fluctuations, and at times unintuitive [47, 99]. In terms of semantics, writing that was more passive/less dominant and of a more negative outlook was indicative of higher SI. Passivity, through the lens of social problem solving, has been implicated as a contributor to suicidal vulnerability [100]. Additionally, echoing repeated findings in the literature [82, 101, 102], higher personal pronoun usage emerged as a clear marker (Fig. 3). Taken together, a majority of the observed patterns in written language and emotion echo previous findings in the literature, further bolstering the potential utility of these features and the convergent validity of this work.

Strengths and limitations

The current work offers a novel, accessible, and easily reproducible analytical framework for exploring the acute longitudinal contexts of SI written expression, leveraging a robust NLP tool that integrates a wide and rich array of prominent sentiment resources. Moreover, the data used to accomplish this was densely sampled over 90 continuous days, enabling a more temporally-nuanced investigation of acute SI dynamics beyond what has been typically seen in the literature. Despite these strengths, this study has important limitations to mention.

First, and most importantly, the use of item 9 of the MPHQ-9 to serve as a proxy for SI was broad and perhaps oversensitive to SI expression within the cohort (see “Data preprocessing and subselection” in “Data analytic plan”). While inappropriate for modeling within practical or clinical settings, item 9 was a means by which foundational NLP-based feature exploration could be logistically

carried out, not only with minimal burden to participants, but within longitudinal and densely sampled timescales. Such efforts were carried out with the hopes of promoting future work that grounds theories and models in ecologically valid and empirically supported data paradigms. The measure is certainly imperfect, and its validity as a tool within mobile environments is still in the early days of assessment [35, 68]. Naturally, such nascency precludes robust and generalizable psychometric evidence, but there is precedence for further application and testing. For example, findings provide preliminary evidence that the MPHQ-9 is a valid EMA measure of depression and exhibits a strong internal consistency and stability that is superior to the PHQ-9 [68].

Second, the cohort was a representative sample of individuals with MDD, thus affecting the generalizability of the findings to the general population. For example, it is possible that the way in which depressed persons with SI express themselves in writing is different from those without MDD. Relatedly, the exclusion criterion of active suicidality established through the larger Tracking Depression study potentially led to a phenomenologically limited appreciation of NLP-based SI in the current investigation. While the results are still useful in directing future research emphases, it is possible that insights gleaned from these results on a well-educated, majority White, female, and heterosexual cohort are not generalizable to broader populations of suicidal individuals who may fall outside of these sociodemographic designations or may exist at the more extreme end of the risk severity spectrum.

Third, participant engagement with the optional diary entries as well as compliance with SI self-report through EMA completion were variable, thus impacting the total number of representative data points within and between participants in the analysis. As a consequence, some participants were more represented than others which could have affected the resulting patterns of feature significance. This fact also had an influence on how PAC was calculated—necessitating a cohort-level approach to cut-off determination rather than more robust and sensitive person-specific designations. Indeed, approximately 60% of the analyzed cohort had a small number of representative person-specific samples (e.g., less than 30), rendering estimations of within-person PAC unreliable and likely inaccurate. Variable engagement with EMA and diary entries also raises the question regarding whether a participant with more severe SI is less likely to complete a diary entry. Indeed, previous research into STB within a veteran population had shown that individuals who declined to endorse their STB were among a subset at heightened risk for suicide [103]. While this variable engagement with EMA in the current study does not invalidate the findings, it does make their generalizability

less clear. The analytical approach outlined herein would thus ideally be repeated on cohorts with more consistently representative data.

Fourth, some of the acute SI phase trajectories were of low or imbalanced prevalence which may have hindered the ability to uncover novel SI contexts. For example, there were 195 instances of the Apex phase trajectory, but only 20 instances of Escalation (see Fig. 2), ultimately providing less of an opportunity to utilize the trend of SI escalation as an investigative piece of the contextual puzzle.

Last, given the novel methodological aspects of this work, the decision to define MPHQ-9 scores as “low” (0–24), “medium” (25–75), and “high” (76–100) for constructing the Stability phases, while based on a general correspondence with the 27-point PHQ-9 scale, has not been substantiated in the literature. The goal was to roughly capture extremes on the MPHQ-9 sliding scale for the purposes of highlighting heightened and dampened acute SI endorsement; however, future work is needed to examine how the MPHQ-9 may be most effectively translated from the traditional PHQ-9.

Conclusions

The prediction and prevention of STB-related outcomes will benefit from a more comprehensive appreciation of how SI acutely presents and fluctuates—a currently understudied and underappreciated component of the STB behavioral mosaic. The exploration of written expression through the lens of NLP is one approach that has the opportunity to provide useful quantitative markers for SI severity and change, yet such work often requires densely sampled data collection regimes to phenomenologically match the timescales of acute SI. Accordingly, the current research offered a broad and repeatable exploratory approach that married profiles of acute SI change with those of written sentiment, ultimately capitalizing on the affordances of frequent and continuous EMA data sampling—a ubiquitous technique within the behavioral and social sciences. While the findings implicated several “usual suspects” in terms of sentiment and SI associations, some qualities of writing related to verbosity, hostility, and pleasantness, among others, offered novel perspectives regarding how features of writing (and their associated interpretations with SI) may be dependent on the greater context of near-past and near-future SI states. Importantly, future efforts toward profiling acute SI via NLP- and sentiment-based features should strive to analyze data within shorter timescales as there are potentially important signals that have yet to be fully appreciated. To this end, the dynamics and shifting contexts of SI may be more robustly captured, quantified, and leveraged toward the increased sophistication and

deployment of STB risk models within ecologically valid settings.

Abbreviations

3ST	Three Step Theory of suicide
ANEW	Affective Norms for English Words
ANOVA	Analysis of variance
AUC	Area under the curve
EMA	Ecological momentary assessment
GALC	Geneva Affect Label Coder
GI	General Inquirer
IMV	Integrated Motivational-Volitional (model of suicide)
IPTS	Interpersonal-Psychological Theory of Suicide
MDD	Major depressive disorder
NIMH	National Institute of Mental Health
NLP	Natural language processing
PAC	Probability of acute change
PHQ-9	Patient Health Questionnaire-9
SCID	Structured Clinical Interview for DSM-5
SEANCE	Sentiment Analysis and Cognition Engine
SI	Suicidal ideation
STB	Suicidal thought and behavior
VADER	Valence Aware Dictionary for Sentiment Reasoning

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12888-025-07108-4>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

Not applicable.

Author contributions

The following reflects individual contributions per the nomenclature of the Contributor Roles Taxonomy (CRediT): DL: Conceptualization, Data Curation, Methodology, Software, Investigation, Formal analysis, Visualization, Writing—original draft, Writing—review & editing. ACC: Investigation, Writing—original draft, Writing—review & editing. MVH: Investigation, Writing—review & editing. TZG: Investigation, Project administration, Writing—review & editing. AP: Software, Investigation, Writing—review & editing. SN: Software, Investigation, Writing—review & editing. DMM: Writing—review & editing. ATC: Software, Investigation, Writing—review & editing. NCJ: Investigation, Resources, Methodology, Writing—review & editing.

Funding

This work was funded by the National Institute of Mental Health (NIMH) and the National Institute of General Medical Sciences (NIGMS) grant (R01MH123482) as well as by an institutional grant from the National Institute on Drug Abuse (NIDA; 5P30DA02992610).

Data availability

The complete dataset used in analysis, the distributional statistics for all NLP features quantified with the data, as well as all statistical results are available in the supplemental files associated with this work. Supplemental materials are also available on Open Science Framework at <https://osf.io/2uhkg/>.

Declarations

Ethics approval and consent to participate

The current study received approval from the Committee for the Protection of Human Subjects at Dartmouth College (STUDY00032081) and was carried out in compliance with the Declaration of Helsinki. Participants provided written informed consent prior to the collection of information and verbal informed

consent prior to the start of the study. This consent procedure was approved by the Committee.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, 46 Centerra Parkway, Suite 300, Lebanon 03766, NH, USA

²Quantitative Biomedical Sciences Program, Dartmouth College, Hanover, NH, USA

³Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA

⁴Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA

⁵Department of Computer Science, Dartmouth College, Hanover, NH, USA

Received: 16 October 2024 / Accepted: 12 June 2025

Published online: 01 July 2025

References

1. SAVE: Suicide Prevention, Information, and Awareness. Learn Suicide Statistics. 2024. <https://www.save.org/learn/suicide-statistics/>. Accessed 11 Sep 2024.
2. American Foundation for Suicide Prevention, National Action Alliance for Suicide Prevention, Suicide Prevention Resource Center. Suicide Prevention Now. 2022. <https://suicidepreventionnow.org/>. Accessed 20 May 2024.
3. Durkheim É. [Suicide: A study in sociology]. Le suicide: étude de sociologie. Paris, France: F. Alcan; 1897.
4. Lengvenyte A, Conejero I, Courtet P, Olié E. Biological bases of suicidal behaviours: A narrative review. *Eur J Neurosci*. 2021;53:330–51.
5. Mueller AS, Abrutyn S, Pescosolido B, Diefendorf S. The social roots of suicide: theorizing how the external social world matters to suicide and suicide prevention. *Front Psychol*. 2021;12:621569.
6. O'Connor RC, Nock MK. The psychology of suicidal behaviour. *Lancet Psychiatry*. 2014;1:73–85.
7. Klonsky ED, Saffer BY, Bryan CJ. Ideation-to-action theories of suicide: A conceptual and empirical update. *Curr Opin Psychol*. 2018;22:38–43.
8. Chu C, Buchman-Schmitt JM, Stanley IH, Hom MA, Tucker RP, Hagan CR, et al. The interpersonal theory of suicide: A systematic review and meta-analysis of a decade of cross-national research. *Psychol Bull*. 2017;143:1313–45.
9. O'Connor RC, Kirtley OJ. The integrated motivational-volitional model of suicidal behaviour. *Phil Trans R Soc B*. 2018;373:20170268.
10. Klonsky ED, May AM. The Three-Step Theory (3ST): A new theory of suicide rooted in the ideation-to-action framework. *International Journal of Cognitive Therapy*. 2015;8:114–29.
11. Coppersmith DDL, Ryan O, Fortgang RG, Millner AJ, Kleiman EM, Nock MK. Mapping the timescale of suicidal thinking. *Proceedings of the National Academy of Sciences*. 2023;120:e2215434120.
12. Ammerman BA, Law KC. Using intensive time sampling methods to capture daily suicidal ideation: A systematic review. *J Affect Disord*. 2022;299:108–17.
13. Sedano-Capdevila A, Porras-Segovia A, Bello HJ, Baca-García E, Barrigón ML. Use of ecological momentary assessment to study suicidal thoughts and behavior: A systematic review. *Curr Psychiatry Rep*. 2021;23:41.
14. Arowosegbe A, Oyelade T. Application of natural Language processing (NLP) in detecting and preventing suicide ideation: A systematic review. *IJERPH*. 2023;20:1514.
15. Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, et al. Smart-phones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. *Curr Psychiatry Rep*. 2018;20:51.
16. Fazel S, O'Reilly L. Machine learning for suicide research—Can it improve risk factor identification? *JAMA Psychiatry*. 2020;77:13.

17. Horton SE, Hughes JL, King JD, Kennard BD, Westers NJ, Mayes TL, et al. Preliminary examination of the interpersonal psychological theory of suicide in an adolescent clinical sample. *J Abnorm Child Psychol*. 2016;44:1133–44.
18. Pagliaccio D, Bitran A, Kirshenbaum JS, Alqueza KL, Durham K, Chernick LS, et al. Testing the interpersonal theory of suicide in adolescents: A multi-wave longitudinal study. *J Child Psychol Psychiatry*. 2024;65:668–79.
19. Cleare S, Wetherall K, Eschle S, Forrester R, Drummond J, O'Connor RC. Using the integrated motivational-volitional (IMV) model of suicidal behaviour to differentiate those with and without suicidal intent in hospital treated self-harm. *Prev Med*. 2021;152:106592.
20. Pérez S, Marco JH, García-Alandete J. Psychopathological differences between suicide ideators and suicide attempters in patients with mental disorders. *Clin Psychol Psychother*. 2017;24:1002–13.
21. Meerwijk EL, Jones GA, Shotqara AS, Reyes S, Tamang SR, Eddington HS, et al. Development of a 3-step theory of suicide ontology to facilitate 3ST factor extraction from clinical progress notes. *J Biomed Inform*. 2024;150:104582.
22. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: A systematic review and simulation. *JAMA Psychiatry*. 2019;76:642.
23. Nock MK, Millner AJ, Ross EL, Kennedy CJ, Al-Suwaidi M, Barak-Corren Y, et al. Prediction of suicide attempts using clinician assessment, patient self-report, and electronic health records. *JAMA Netw Open*. 2022;5:e2144373.
24. Rudd MD. Fluid vulnerability theory: A cognitive approach to Understanding the process of acute and chronic suicide risk. In: Ellis TE, editor. *Cognition and suicide: theory, research, and therapy*. American Psychological Association; 2006. pp. 355–68.
25. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychol Bull*. 2016;143:187.
26. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. *J Abnorm Psychol*. 2017;126:726–38.
27. Czyz EK, King CA, Al-Dajani N, Zimmermann L, Hong V, Nahum-Shani I. Ecological momentary assessments and passive sensing in the prediction of short-term suicidal ideation in young adults. *JAMA Netw Open*. 2023;6:e2328005.
28. Bryan CJ, Rozek DC, Butner J, Rudd MD. Patterns of change in suicide ideation signal the recurrence of suicide attempts among high-risk psychiatric outpatients. *Behav Res Ther*. 2019;120:103392.
29. Riera-Serra P, Navarra-Ventura G, Castro A, Gili M, Salazar-Cedillo A, Ricci-Cabello I, et al. Clinical predictors of suicidal ideation, suicide attempts and suicide death in depressive disorder: A systematic review and meta-analysis. *Eur Arch Psychiatry Clin Neurosci*. 2024;274:1543–63.
30. Pew Research Center. Mobile Fact Sheet. 2024. <https://www.pewresearch.org/internet/fact-sheet/mobile/#:~:text=and%20its%20methodology%20phone%20ownership%20over%20time,smartphone%20ownership%20conducted%20in%202011>. Accessed 22 May 2024.
31. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, López-Castroman J, Barrigón ML, Pérez-Rodríguez MS, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: A real-world feasibility study. *J Affect Disord*. 2020;274:733–41.
32. Gratch I, Choo T, Galfalvy H, Keilp JG, Itzhaky L, Mann JJ, et al. Detecting suicidal thoughts: the power of ecological momentary assessment. *Depress Anxiety*. 2021;38:8–16.
33. Oquendo MA, Galfalvy HC, Choo T-H, Kandlur R, Burke AK, Sublette ME, et al. Highly variable suicidal ideation: A phenotypic marker for stress induced suicide risk. *Mol Psychiatry*. 2021;26:5079–86.
34. Hallensleben N, Glaesmer H, Forkmann T, Rath D, Strauss M, Kersting A, et al. Predicting suicidal ideation by interpersonal variables, hopelessness and depression in real-time. An ecological momentary assessment study in psychiatric inpatients with depression. *Eur Psychiatr*. 2019;56:43–50.
35. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the patient health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. *JMIR Ment Health*. 2015;2:e8.
36. Vine V, Victor SE, Mohr H, Byrd AL, Stepp SD. Adolescent suicide risk and experiences of dissociation in daily life. *Psychiatry Res*. 2020;287:112870.
37. Glenn CR, Kleiman EM, Kearns JC, Santee AC, Esposito EC, Conwell Y, et al. Feasibility and acceptability of ecological momentary assessment with high-risk suicidal adolescents following acute psychiatric care. *J Clin Child Adolesc Psychol*. 2022;51:32–48.
38. Ben-Zeev D, Scherer EA, Brian RM, Mistler LA, Campbell AT, Wang R. Use of multimodal technology to identify digital correlates of violence among inpatients with serious mental illness: A pilot study. *PS*. 2017;68:1088–92.
39. Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. *Depress Anxiety*. 2018;35:601–8.
40. Spangenberg L, Glaesmer H, Hallensleben N, Rath D, Forkmann T. (In)stability of capability for suicide in psychiatric inpatients: longitudinal assessment using ecological momentary assessments. *Suicide Life Threat Behav*. 2019;49:1560–72.
41. Nemesure MD, Collins AC, Price GD, Griffin TZ, Pillai A, Nepal S, et al. Depressive symptoms as a heterogeneous and constantly evolving dynamical system: idiographic depressive symptom networks of rapid symptom changes among persons with major depressive disorder. *J Psychopathol Clin Sci*. 2024;133:155–66.
42. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural Language processing for mental health interventions: A systematic review and research framework. *Transl Psychiatry*. 2023;13:1–17.
43. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural Language processing in mental health applications using non-clinical texts. *Nat Lang Eng*. 2017;23:649–85.
44. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel use of natural Language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. *Comput Math Methods Med*. 2016;2016:1–8.
45. Diniz EJS, Fontenele JE, De Oliveira AC, Bastos VH, Teixeira S, Rabêlo RL, et al. Boamente: A natural Language processing-based digital phenotyping tool for smart monitoring of suicidal ideation. *Healthcare*. 2022;10:698.
46. Franz PJ, Nook EC, Mair P, Nock MK. Using topic modeling to detect and describe self-injurious and related content on a large-scale digital platform. *Suicide Life Threat Behav*. 2020;50:5–18.
47. Lekkas D, Jacobson NC. The hidden depths of suicidal discourse: network analysis and natural Language processing unmask uncensored expression. *Digit HEALTH*. 2023;9:20552076231210714.
48. Lekkas D, Klein RJ, Jacobson NC. Predicting acute suicidal ideation on Instagram using ensemble machine learning models. *Internet Interventions*. 2021;25:100424.
49. Wankhade M, Rao ACS, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. *Artif Intell Rev*. 2022;55:5731–80.
50. Mohamad Sham N, Mohamed A. Climate change sentiment analysis using lexicon, machine learning and hybrid approaches. *Sustainability*. 2022;14:4723.
51. Crossley SA, Kyle K, McNamara DS. Sentiment analysis and social cognition engine (SEANCE): an automatic tool for sentiment, social cognition, and social-order analysis. *Behav Res*. 2017;49:803–21.
52. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proc Int AAAI Conf Web Social Media*. 2014;8:216–25.
53. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. *Comput Intell*. 2013;29:436–65.
54. Scherer KR. What are emotions? And how can they be measured? *Social Sci Inform*. 2005;44:695–729.
55. Bradley M, Lang P. Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. 1999.
56. Hu M, Liu B. Mining and summarizing customer reviews. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: Association for Computing Machinery; 2004. pp. 168–77.
57. Stone PJ, Dunphy DC, Smith MS. The general inquirer: A computer approach to content analysis. Oxford, England: M.I.T.; 1966.
58. Lasswell HD, Namenwirth JZ. The Lasswell value dictionary. New Haven: Yale University Press; 1969.
59. Ismail L, Shahin N, Materwala H, Hennebel A, Frermann L. ML-NLPEmot: machine learning-natural Language processing event-based emotion detection proactive framework addressing mental health. *IEEE Access*. 2023;11:144126–49.
60. Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. *Sci Rep*. 2017;7:13006.
61. Lim YQ, Loo YL. Characteristics of multi-class suicide risks tweets through feature extraction and machine learning techniques. *JOIV: Int J Inf Visualization*. 2023;7:2297–305.

62. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia–Suicide severity rating scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *AJP*. 2011;168:1266–77.
63. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*. 2002;32:509–15.
64. Nam S, Um B, Jeong J, Rodriguez M, Lardier D. Score reliability generalization of the Columbia–Suicide severity rating scale (C-SSRS): A meta-analysis. *Meas Evaluation Couns Dev*. 2024;57:173–85.
65. Jacobson NC. Personalized Deep Learning Models of Rapid Changes in Major Depressive Disorder Symptoms using Passive Sensor Data from Smartphones and Wearable Devices. NIH RePORTER. 2020. <https://reporter.nih.gov/search/gDrJ5TTrX0aHxQbkPgrq0A/project-details/10629273>. Accessed 17 Apr 2024.
66. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S et al. StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Seattle Washington: ACM; 2014. pp. 3–14.
67. Pillai A, Nepal SK, Wang W, Nemesure M, Heinz M, Price G, et al. Investigating generalizability of speech-based suicidal ideation detection using mobile phones. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. 2023;7:1–38.
68. Haddox D, Mackin D, Griffin T, Heinz MV, Nemesure MD, Collins AC et al. Psychometric properties and validity of a mobile patient health Questionnaire-9 (MPHQ-9) for ecological momentary assessment in depressed adults. Preprint. 2024.
69. Collins AC, Lekkas D, Struble CA, Trudeau BM, Jewett AD, Griffin TZ, et al. From mood to use: using ecological momentary assessments to examine how anhedonia and depressed mood impact cannabis use in a depressed sample. *Psychiatry Res*. 2024;339:116110.
70. Kline EA, Lekkas D, Bryan A, Nemesure MD, Griffin TZ, Collins AC, et al. The role of borderline personality disorder traits in predicting longitudinal variability of major depressive symptoms among a sample of depressed adults. *J Affect Disord*. 2024;363:492–500.
71. Gallagher MR, Collins AC, Lekkas D, Nemesure MD, Griffin TZ, Price GD, et al. Anhedonia in flux: Understanding the associations of emotion regulation and anxiety with anhedonia dynamics in a sample with major depressive disorder. *J Affect Disord*. 2025;372:27–36.
72. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. *J Gen Intern Med*. 2001;16:606–13.
73. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *Gen Hosp Psychiatry*. 2010;32:345–59.
74. Jahng S, Wood PK, Trull TJ. Analysis of affective instability in ecological momentary assessment: indices using successive difference and group comparison via multilevel modeling. *Psychol Methods*. 2008;13:354–75.
75. Santangelo P, Mussgay L, Sawitzki G, Trull TJ, Reinhard I, Steil R, et al. Specificity of affective instability in patients with borderline personality disorder compared to posttraumatic stress disorder, bulimia nervosa, and healthy controls. *J Abnorm Psychol*. 2014;123:258–72.
76. Crossley SA, McNamara DS. SEANCE: Sentiment Analysis and Cognition Engine. NLP TOOLS FOR THE SOCIAL SCIENCES. 2024. <https://www.linguisticsnalysisistools.org/seance.html>
77. Namenwirth JZ, Weber RP. Dynamics of Culture. 1st edition. London: Routledge; 1987.
78. Cambria E, Speer R, Havasi C, Hussain A. SenticNet: A publicly available semantic resource for opinion mining. *AAAI*; 2010. pp. 14–8.
79. Fellbaum C. A semantic network of english: the mother of all WordNets. *Comput Humanit*. 1998;32:209–20.
80. Ostertagová E, Ostertag O, Kováč J. Methodology and application of the Kruskal–Wallis test. *Appl Mech Mater*. 2014;611:115–20.
81. Terpilowski MA. scikit-posthocs: pairwise multiple comparison tests in Python. *J Open Source Softw*. 2019;4:1169.
82. O’Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related Twitter posts. *Crisis*. 2017;38:319–29.
83. Wilson K, Van Doorn D, Dye J. Vulnerable dark traits mediate the association between childhood adversity and suicidal ideation. *Pers Individ Differ*. 2023;202:111959.
84. Ponzoni S, Beomonte Zobel S, Rogier G, Velotti P. Emotion dysregulation acts in the relationship between vulnerable narcissism and suicidal ideation. *Scand J Psychol*. 2021;62:468–75.
85. Musacchio Schafer K, Joiner T. Anger as a correlate of and longitudinal risk factor for suicidal ideation in adolescents. *Archives Suicide Res*. 2024;28(4):1336–49.
86. Moore FR, Allott C, O’Connor R. Impulsivity, aggression, and impulsive aggression in suicidality. *Pers Individ Differ*. 2023;202:111971.
87. Kaurin A, Wright AGC, Porta G, Hamilton E, Poling K, Bero K, Brent D, Goldstein TR. Weekly links among irritability and suicidal thoughts and behaviors in high-risk youth. *J Child Psychol Psychiatry*. 2024;65(11):1429–40.
88. Liu Q, Cole DA. The association of phasic irritability (aggressive outbursts) and tonic irritability (irritable mood) to depression occurrences, symptoms, and subtypes. *J Affect Disord*. 2021;293:9–18.
89. Schafer KM, Daurio A, French JE, Rogers ML, Meltzer AL, Cogle JR, et al. The relationship between anger and suicidal ideation: investigations in two samples. *J Clin Psychol*. 2022;78:1866–77.
90. Gee BL, Han J, Benassi H, Batterham PJ. Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. *Digit Health*. 2020;6:2055207620963958.
91. Cambria E, Livingstone A, Hussain A. The hourglass of emotions. In: Esposito A, Esposito AM, Vinciarelli A, Hoffmann R, Müller VC, editors. Cognitive behavioural systems. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. pp. 144–57.
92. Sandler I, Yun-Tien J, Zhang N, Wolchik S, Thielemann K. Grief as a predictor of long-term risk for suicidal ideation and attempts of parentally bereaved children and adolescents. *J Trauma Stress*. 2021;34:1159–70.
93. Lin Y-C, Wee J, Marks R, O’Connell K, Hassler M, Law K. Shame-proneness and suicidal ideation: the roles of depressive and anger rumination. *J Affect Disorders Rep*. 2022;7:100303.
94. Chu C, Bodell LP, Ribeiro JD, Joiner TE. Eating disorder symptoms and suicidal ideation: the moderating role of disgust. *Eur Eat Disorders Rev*. 2015;23:545–52.
95. Lissak S, Ophir Y, Tikochinski R, Brunstein Klomek A, Sisso I, Fruchter E et al. Bored to death: artificial intelligence research reveals the role of boredom in suicide behavior. *Front Psychiatry*. 2024;15.
96. Sawalha J, Yousefnezhad M, Shah Z, Brown MRG, Greenshaw AJ, Greiner R. Detecting presence of PTSD using sentiment analysis from text data. *Front Psychiatry*. 2022;12.
97. Leiva V, Freire A et al. Towards Suicide Prevention: Early Detection of Depression on Social Media. In: Kompatsiaris I, Cave J, Satsiou A, Carle G, Passani A, Kontopoulos E, editors. International Conference on Internet Science. Cham: Springer; 2017. pp. 428–36.
98. Deas N, Kowalski R, Finnell S, Radovic E, Carroll H, Robbins C, et al. I just want to matter: examining the role of anti-mattering in online suicide support communities using natural Language processing. *Comput Hum Behav*. 2023;139:107499.
99. Yin Q, Hughes CD, Rizvi SL. Using GIMME to model the emotional context of suicidal ideation based on clinical data: from research to clinical practice. *Behav Res Ther*. 2023;171:104427.
100. Linda WP, Marroquin B, Miranda R. Active and passive problem solving as moderators of the relation between negative life event stress and suicidal ideation among suicide attempters and non-attempters. *Archives Suicide Res*. 2012;16:183–97.
101. Egnoto MJ, Griffin DJ. Analyzing Language in suicide notes and legacy tokens. *Crisis*. 2016;37:140–7.
102. Homan S, Gabi M, Klee N, Bachmann S, Moser A-M, Duri’M, et al. Linguistic features of suicidal thoughts and behaviors: A systematic review. *Clin Psychol Rev*. 2022;95:102161.
103. Stanley IH, Marx BP, Nichter B, Pietrzak RH. Non-response to questions about suicide ideation and attempts among veterans: results from the National health and resilience in veterans study. *Suicide Life-threatening Behav*. 2022;52(4):763–72.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.