

Original Paper

Identifying Reddit Users at a High Risk of Suicide and Their Linguistic Features During the COVID-19 Pandemic: Growth-Based Trajectory Model

Yifei Yan¹, BA; Jun Li², MA; Xingyun Liu³, PhD; Qing Li², PhD; Nancy Xiaonan Yu¹, PhD

¹Department of Social and Behavioural Sciences, City University of Hong Kong, Hong Kong, China (Hong Kong)

²Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

³Key Laboratory of Adolescent Cyberpsychology and Behavior, Central China Normal University, Ministry of Education, School of Psychology, Wuhan, China

Corresponding Author:

Nancy Xiaonan Yu, PhD

Department of Social and Behavioural Sciences

City University of Hong Kong

Tat Chee Avenue, Kowloon, HKSAR, P. R. China

Hong Kong, 000

China (Hong Kong)

Phone: 852 34429436

Fax: 852 34420283

Email: nancy.yu@cityu.edu.hk

Abstract

Background: Suicide has emerged as a critical public health concern during the COVID-19 pandemic. With social distancing measures in place, social media has become a significant platform for individuals expressing suicidal thoughts and behaviors. However, existing studies on suicide using social media data often overlook the diversity among users and the temporal dynamics of suicide risk.

Objective: By examining the variations in post volume trajectories among users on the r/SuicideWatch subreddit during the COVID-19 pandemic, this study aims to investigate the heterogeneous patterns of change in suicide risk to help identify social media users at high risk of suicide. We also characterized their linguistic features before and during the pandemic.

Methods: We collected and analyzed post data every 6 months from March 2019 to August 2022 for users on the r/SuicideWatch subreddit (N=6163). A growth-based trajectory model was then used to investigate the trajectories of post volume to identify patterns of change in suicide risk during the pandemic. Trends in linguistic features within posts were also charted and compared, and linguistic markers were identified across the trajectory groups using regression analysis.

Results: We identified 2 distinct trajectories of post volume among r/SuicideWatch subreddit users. A small proportion of users (744/6163, 12.07%) was labeled as having a high risk of suicide, showing a sharp and lasting increase in post volume during the pandemic. By contrast, most users (5419/6163, 87.93%) were categorized as being at low risk of suicide, with a consistently low and mild increase in post volume during the pandemic. In terms of the frequency of most linguistic features, both groups showed increases at the initial stage of the pandemic. Subsequently, the rising trend continued in the high-risk group before declining, while the low-risk group showed an immediate decrease. One year after the pandemic outbreak, the 2 groups exhibited differences in their use of words related to the categories of personal pronouns; affective, social, cognitive, and biological processes; drives; relativity; time orientations; and personal concerns. In particular, the high-risk group was discriminant in using words related to anger (odds ratio [OR] 3.23, $P<.001$), sadness (OR 3.23, $P<.001$), health (OR 2.56, $P=.005$), achievement (OR 1.67, $P=.049$), motion (OR 4.17, $P<.001$), future focus (OR 2.86, $P<.001$), and death (OR 4.35, $P<.001$) during this stage.

Conclusions: Based on the 2 identified trajectories of post volume during the pandemic, this study divided users on the r/SuicideWatch subreddit into suicide high- and low-risk groups. Our findings indicated heterogeneous patterns of change in suicide risk in response to the pandemic. The high-risk group also demonstrated distinct linguistic features. We recommend conducting real-time surveillance of suicide risk using social media data during future public health crises to provide timely support to individuals at potentially high risk of suicide.

KEYWORDS

COVID-19 pandemic; Reddit; suicide risk; trajectory

Introduction

The COVID-19 pandemic has triggered a global increase in mental disorders [1], with heightened concerns about suicide risk both in the short and in the long term [2,3]. While an interrupted time-series study indicated that most countries or regions have not experienced a significant rise in suicide deaths [4], the prevalence of suicidal ideation (10.81%-12.10%) and suicide attempts (4.86%) has notably increased compared with prepandemic levels [5,6]. Our recent meta-analysis, utilizing longitudinal data, has further corroborated an increase in the prevalence of suicidal ideation and suicide attempts among both nonclinical and clinical populations following the onset of the pandemic [7]. Suicide remains a critical public health concern throughout the pandemic, underscoring the need for ongoing monitoring and vigilance as the situation continues to evolve [8,9]. In this study, we utilized posts from a social media forum to examine patterns of change in suicide risk during the COVID-19 pandemic, analyzing their corresponding linguistic features.

Social media and online discussion forums are increasingly recognized as valuable resources for suicide research [10,11]. These virtual platforms offer users open and widely accessible spaces to share experiences, engage in discussions, seek social support, and exchange information anonymously and in real time [12,13]. Compared with traditional clinical data sets, these online data sets offer several advantages: they are publicly and freely available, involve larger sample sizes, provide access to participants who are ordinarily difficult to engage, enable comparisons with historical data, and exhibit high ecological validity through documentation of first-person experiences [14]. Utilizing social media data from platforms such as Twitter/X (X Corp.), Reddit (Reddit, Inc.), and Weibo (Weibo Corporation), previous studies have detected and predicted suicide risk and developed intervention programs [15-21]. As a result of lockdown measures and heightened concerns during the pandemic, social media data have become even more valuable for suicide research, as these online platforms have become the primary means for many people to receive information and stay connected with the outside world [22,23].

The r/SuicideWatch subreddit is a semianonymous forum that provides “peer support for anyone struggling with suicidal thoughts or worried about someone who may be at risk” [24]. This makes it an important resource for suicide research, as it offers high-quality, self-reported suicide data. Dominant suicide risk assessments for online posts, such as machine learning models combined with manual coding, typically approach it as a multiclassification task. These models output the post-level suicide risk (ie, the risk associated with a single post) using ordinal data. For example, they might label a post as being at ideation, behavior, or attempt level based on the probability score for each level [25-27]. However, such data cannot capture user-level suicide risk (ie, the risk of a user based on posts

during a specific period) and may not conform to the distributional assumptions of many statistical analyses, such as growth-based trajectory models, potentially introducing biases [28,29]. To assess users’ overall risk and track temporal changes in risk, researchers can consider both the quantity and quality of posts published, including methods such as topic modeling or linguistic analysis of content [30,31]. Specifically, the quantitative method focuses on the total number of posts (ie, post volume) within a specific period, which indicates users’ posting activity and social engagement level [32,33]. Changes in users’ post volume can reflect changes in their suicide risk. For instance, the diurnal and weekly patterns of post volume on r/SuicideWatch corresponded to temporal fluctuations in suicide risk, including actual suicide attempts or deaths [34]. Similarly, Twitter users exhibited an increased volume of suicide-related posts before their suicide attempts [35]. Furthermore, analyses of both the quantity and quality of posts have demonstrated that higher post volume (quantity) corresponds to active disclosure of suicidal thoughts in post content (quality) [30,31]. Therefore, post volume in online suicide communities can serve as an effective indicator of user-level suicide risk, offering sufficient accessibility and flexibility for statistical analysis.

In this study, we monitored users’ post volume on the r/SuicideWatch subreddit before and during the pandemic to observe changes in their suicide risk. Among active adolescent users during the pandemic (April 2020-September 2021), post volume remained stable compared with prepandemic periods [36]. However, when examining post volume over shorter intervals, it fluctuated and demonstrated an overall decrease until December 2020 [37,38]. There is currently no documented study on how post volume has evolved beyond 2020. It is important to note that users of r/SuicideWatch may vary in their levels of suicide risk, and the findings mentioned above could be ambiguous without accounting for this heterogeneity among users. Through expert annotation, active r/SuicideWatch users (ie, those with at least 10 total posts) were categorized into 4 risk levels: no risk (36/245, 15%), low risk (50/245, 20%), moderate risk (115/245, 47%), and severe risk (44/245, 18%) of suicide [39]. However, each user’s risk label was determined based on the highest risk observed in their posts during a 7-year period before the pandemic. As suggested by the fluid vulnerability theory [40], individual suicide risk is best understood as a temporal process influenced by both baseline and acute risk factors. Environmental stressors or contexts such as sudden outbreaks of infectious disease epidemics or pandemics, social isolation, and fear [3,7] can easily trigger acute suicide risk in individuals predisposed to underlying vulnerabilities (ie, those with a higher baseline risk). Cognitive, emotional, behavioral, and physiological factors interact to either sustain or alleviate their suicide risk [41]. Considering both stability and dynamism, some individuals exhibit fluctuations in suicide risk, moving between low and high states where suicidal behavior may be more or less likely to emerge,

while others display a stable pattern. To address the heterogeneity among users and understand the temporal nature of suicide risk, using trajectory modeling techniques such as group-based trajectory modeling (GBTM) and growth curve modeling can be beneficial. These methods identify subgroups within a population that share similarities in outcomes over time [42].

The linguistic or language styles used in posts or comments can offer valuable insights into the experiences and perspectives of suicidal individuals. This can assist researchers in understanding the underlying thoughts, emotions, and behaviors that individuals may be unwilling or unable to express explicitly [43]. Importantly, linguistic features can act as markers that distinguish suicide-related posts from general posts, aiding in the identification of potential high-risk users who may need support [17,44,45]. Suicide-related social media posts often exhibit characteristics such as simplicity in words and short sentences, reduced lexical diversity, and language disorganization [46]. They may also include more statements related to self-destruction, commands, and conflicts [47], along with increased use of first-person pronouns, adverbs, and multifunctional words. These posts frequently reference death, anger, and the present moment, while showing fewer occurrences of second- and third-person pronouns, nouns, and references to causes and differentiation [17,48,49]. However, it remains unclear whether high-risk users have exhibited these linguistic features during the COVID-19 pandemic, which has introduced different stressors potentially influencing suicide risk. Existing studies during the pandemic that utilized psycholinguistic analysis of r/SuicideWatch posts have primarily concentrated on monitoring temporal shifts in these linguistic features. Specifically, studies have identified increased use of words associated with negative emotions and a focus on the past, along with fewer references to positive emotions, social interactions, and leisure activities. References to death and first-person pronouns remained stable [36,37]. However, these findings only encompassed a limited time frame during the pandemic (until September 2021) and did not account for the heterogeneity among users in terms of suicide risk. Different users may exhibit varying linguistic characteristics and trends in linguistic changes over time. Therefore, it is crucial to first identify groups of users exhibiting similar patterns of suicidal behavior throughout the pandemic. Subsequently, analyzing their respective linguistic trends and markers can yield valuable insights for suicide surveillance and targeted interventions among high-risk users.

Utilizing posts from the r/SuicideWatch subreddit before and during the COVID-19 pandemic (March 2019–September 2022), this study aimed to investigate the following: (1) the potential for distinct patterns of change in suicide risk using GBTM of users' post volumes, (2) the trends and characteristics of linguistic features within posts across each trajectory group, and (3) the linguistic markers associated with each trajectory group. We anticipated that identifying trajectories of post volume on the r/SuicideWatch subreddit would uncover users' diversity by accounting for the temporal dynamics of suicide risk. Analyzing their associated linguistic features could also help identify users potentially at high risk of suicide. These

findings could have significant implications for enhancing suicide screening, monitoring, and interventions during future public health crises.

Methods

Data Set and Participants

For this study, we gathered the longitudinal data set using the Reddit application programming interface [50]. Following a previously established method [14], we crawled posts from users who contributed to the r/SuicideWatch subreddit between March 1, 2020, and August 31, 2022, resulting in a total of 603,802 posts from 6943 users. We expanded our data set by retrieving historical posts dating back to March 1, 2019, to analyze the trajectory of post volume before and during the pandemic. To streamline data usage in subsequent analyses, we excluded accounts that were canceled and posts with deleted content. The final data set comprised 6163 users and their posts from the r/SuicideWatch subreddit (N=33,714) spanning the period from March 1, 2019, to August 31, 2022, encompassing the COVID-19 pandemic period. As a result of the onset of the COVID-19 pandemic around March 2020, there was a notable increase in discussions related to the pandemic on Reddit [37]. Therefore, we used March 1, 2020, as a cutoff point and defined 2 prepandemic periods (T1: March 1, 2019–August 31, 2019 and T2: September 1, 2019–February 29, 2020) and 5 peripandemic periods (T3: March 1, 2020–August 31, 2020; T4: September 1, 2020–February 28, 2021; T5: March 1, 2021–August 31, 2021; T6: September 1, 2021–February 28, 2022; and T7: March 1, 2022–August 31, 2022) to track post volume trajectories across these time frames.

Trajectory Variable: Post Volume

To assess changes in suicide risk, we used post volume from the r/SuicideWatch subreddit within each pre- and peripandemic period for each user as a proxy for the trajectory variable [30,31,34,35,51]. We quantified the number of posts made by each user during each specific period, with periods where no posts were made recorded as 0. For users who joined r/SuicideWatch after March 1, 2020 (ie, those who began posting suicide-related content on r/SuicideWatch only after the pandemic outbreak; 5759/6163, 93.44%) [48], their post counts on r/SuicideWatch during the 2 prepandemic periods and the peripandemic periods before their initial post were recorded as 0. As noted by De Choudhury et al [48], there are a few suicide-related posts found on subreddits outside of r/SuicideWatch. The transition of these users to r/SuicideWatch may indicate the onset and progression of their suicidal concerns following the pandemic outbreak. Therefore, their 0 post volume during the pre- and peripandemic periods can serve as a proxy for their respective suicide risk trajectories, reflecting changes in their behavioral patterns and suicide risk following the pandemic. According to our eligibility criteria, all users included in the study had posted at least once across all periods.

There might be a concern that some of our collected posts discussed the suicide risk of others rather than the user's own risk, although previous studies have indicated that posts on r/SuicideWatch primarily focus on self-directed concerns [30,48]. To address this issue in our data set, we randomly

selected 10% (3372/33,714) of the total posts and manually screened the content to determine the subject of the posts. Among the 3372 posts, only 18 (0.53%) were found not to be about the user's own suicide-related issues: 6 discussed others' suicide risk, 5 provided help to others, 3 were about irrelevant topics, and 4 were unclassified. Thus, the majority of our collected posts accurately reflected users' own experiences and concerns related to suicide risk.

Linguistic Features

To analyze the linguistic features of posts, we used Linguistic Inquiry and Word Count (LIWC) 2015 [52], a widely used tool for language analysis. LIWC encompasses more than 80 word categories, each containing hundreds of dictionary words for the identification and analysis of word use patterns related to suicide risk. The primary psycholinguistic categories in LIWC are personal pronouns and words related to affective, social, cognitive, perceptual, and biological processes; drives; time orientations; relativity; and personal concerns. For each Reddit user included in our study, we calculated LIWC measures for these 10 major psycholinguistic categories during each period [53]. First, we tallied the occurrences of each word within a specific post alongside the post's length (ie, the total number of words used). Second, we summed the occurrences of each word and the total length of posts for each period. Finally, for each period, we calculated the normalized frequency of word use in each LIWC category by dividing the total count of the LIWC category by the total length of posts in that period. Therefore, each LIWC measure represents the normalized frequency of word use within a specific LIWC category during each period analyzed.

Statistical Analysis

GBTM was used to define trajectory groups based on post volume across the COVID-19 pandemic. As a finite mixture model, GBTM is capable of identifying distinct groups of individuals with similar developmental trajectories in a particular outcome or behavior within a population. It accommodates trajectory variables that adhere to distributions such as censored normal (CNORM), zero-inflated Poisson (ZIP), beta, and Bernoulli distributions [54]. Within each period, the exploratory analysis revealed that a large number of eligible users had 0 posts, resulting in a skewed and zero-inflated distribution of the trajectory variable (ie, post volume within each period). Among the available models, the ZIP model was selected to fit our data because of its capability to address excessive zeros. The ZIP model combines an inflation model for zeros with a count model for nonzero values, making it suitable for our data set [54,55].

To identify the best-fitting model with the optimal number of trajectory groups, we followed 3 steps [56]. Initially, we incrementally increased the number of group specifications from 2 to 5 to pinpoint the optimal number of trajectories. Specifically, we selected the model based on 4 commonly used fit statistics in GBTM analysis [56-58]: the Bayesian information criterion (BIC), the Akaike information criterion (AIC), entropy, and group composition. AIC relies on information theory to assess the relative information value of the model by considering the maximum likelihood estimate and the number of parameters within the model [59]. Similar to AIC, BIC originates from the

Bayesian framework and can be interpreted as the posterior probability of a model based on the observed data [60]. Both statistics aim to identify the most informative model by balancing between goodness-of-fit and model complexity. However, BIC imposes a stronger penalty for model complexity compared with AIC, taking into account the sample size [61]. The goodness-of-fit and penalty terms are summed to compute AIC and BIC values, where smaller values indicate better-fitting models [62-64]. Additionally, entropy assesses the classification accuracy of the model by summarizing the likelihood of each participant being correctly classified [57]. With values ranging from 0 to 1, higher entropy values indicate more precise classification, typically considered satisfactory when exceeding 0.8 [65]. We also analyzed the group composition (ie, the percentage of the population represented in each subgroup), ensuring that each subgroup represented at least 5% of the total sample [58]. Second, we determined the shapes of each trajectory by specifying their functional forms (eg, linear and cubic). Starting with a cubic specification (up to 3 degrees), we iteratively dropped nonsignificant ($P > .05$) polynomial terms until only significant ones remained [56]. In the count model part, linear terms were retained regardless of their statistical significance. Third, after identifying the optimal number of trajectories and their shapes, we used the average posterior probabilities (APPs) of group membership to validate the selected model. The APP measures the average probability of each participant belonging to their assigned group and should ideally be at least 0.7 for each group to ensure robustness [58].

After identifying the best-fitting model, we assigned users to their respective trajectory groups. Subsequently, summary descriptive statistics of linguistic features during each period were computed and graphed for each user group. Specifically, we compared linguistic frequencies between groups across different periods. As a result of the skewed and excessively zero-inflated distribution of LIWC frequency, parametric tests such as *t* tests or ANOVA may not be appropriate, as they violate their assumptions and can reduce the robustness of nonparametric tests such as the median or Wilcoxon-Mann-Whitney test. Therefore, Poisson regression modeling was recommended for its improved interpretability of data and comparability among potential models [66]. In the Poisson regression model, group membership was included as an independent variable, while the frequency of each linguistic feature in each period served as the dependent variable. The results of the Poisson model provided rate ratios (RRs) along with SEs, indicating the relative changes in counts of the outcomes between the groups.

In the final analysis, our goal was to identify linguistic markers that could differentiate between groups of users exhibiting different trajectories. In the context of suicidal text analysis, linguistic markers, or linguistic distinguishers, are language features (eg, LIWC categories) extracted from texts that substantially distinguish users or posts with varying risk statuses [17,49]. To identify these features, we analyzed the linguistic profiles of high- and low-risk users and examined which words could indicate their respective suicide risk levels. As the markers were not intended to predict users' group membership, a temporal sequence (eg, baseline linguistic data) was not

necessary. We utilized data from the 7 periods and followed these steps to identify potential linguistic markers. First, we identified linguistic features from the periods that exhibited significant ($P \leq .05$) between-group differences based on the results of Poisson regression, considering them as potential markers. Second, to prevent duplication, we excluded high-level LIWC categories that have hierarchical relationships with each other (eg, pron includes I, we, you, she/he, they). Third, we used Lasso logistic regression with cross-validation to determine the optimal penalty parameter, aiming to mitigate collinearity among the remaining linguistic measures [67]. Given the lack of a theoretical basis for estimating post-selection coefficients with nonlinear Lasso models [67], we utilized the Lasso model solely for model selection purposes to filter out redundant variables. The remaining variables were then integrated into the best-fitting GBTM, and their associations with group membership were assessed using multivariate logistic regression [68]. In this analysis, we computed odds ratios (ORs), where potential linguistic features served as independent variables and group memberships as the dependent variable. Significance levels were determined at a P value $\leq .05$. Data extraction was performed using Python (Python Foundation), while data analysis was conducted using Stata/SE 16.1 (StataCorp LLC) along with the Traj plugin for trajectory analysis.

Ethical Considerations

The data used in this study were obtained from publicly accessible posts on the r/SuicideWatch subreddit through purely observational and nonintrusive means. The raw data did not contain personally identifiable information. To uphold user privacy and confidentiality, selected posts were deidentified before analysis. This involved removing any identifying

information such as names, genders, ages, addresses, and links from the post content. We maintained annotated user data separately from the raw data and stored them on secure servers, linked only through anonymous IDs. Furthermore, all examples presented in [Multimedia Appendix 1](#) were anonymized and paraphrased to safeguard user privacy, following the framework outlined by Bruckman [69]. As publicly available data were utilized, this study fell outside the purview of ethical review by the City University of Hong Kong Research Committee, for which an exemption was obtained.

Results

Post Trajectories on the r/SuicideWatch Subreddit Throughout the COVID-19 Pandemic

To determine post trajectories, we evaluated the model fit statistics for 2- to 5-group solutions of the GBTM to identify the optimal number of trajectory groups ([Table 1](#)). As the number of groups increased from 2, we noted that both the AIC and BIC values tended to increase, while entropy decreased. Additionally, starting from the 3-group model, some group compositions did not meet the 5% threshold. Therefore, we selected the 2-group model as the optimal choice, with AIC and BIC values of $-39,605.31$ and $-39,659.12$, respectively, and an entropy of 0.96. Further analysis indicated that the 2-group solution, using cubic and quadratic functions in the count model and 2 cubic functions in the inflation model, resulted in all polynomial terms being statistically significant ([Table 2](#)). The APPs for groups 1 and 2 were 0.95 and 0.99, respectively, indicating strong alignment between users and their assigned groups within this 2-group ZIP model.

Table 1. Fit statistics for the 2- to 5-group solution-based trajectory modeling of post volume among r/SuicideWatch users throughout the COVID-19 pandemic.

Model	AIC ^a	BIC ^b	Entropy ^c	Composition ^d , n/N (%)
2-Group model	-39,605.31	-39,659.12	0.96	744/6163 (12.07)/5419/6163 (87.93)
3-Group model	-36,862.86	-36,940.22	0.892	1562/6163 (25.34)/4379/6163 (71.05)/222/6163 (3.60)
4-Group model	-35,890.32	-35,981.12	0.836	625/6163 (10.14)/1502/6163 (24.37)/3936/6163 (63.87)/100/6163 (1.62)
5-Group model	-36,099.67	-36,190.48	0.873	293/6163 (4.75)/1699/6163 (27.57)/91/6163 (1.48)/4003/6163 (64.95)/77/6163 (1.25)

^aAIC: Akaike information criterion (a lower value is better).

^bBIC: Bayesian information criterion (a lower value is better).

^cEntropy (a value >0.8 is better).

^dGroup composition (the percentage of the population represented in each subgroup should exceed 5%).

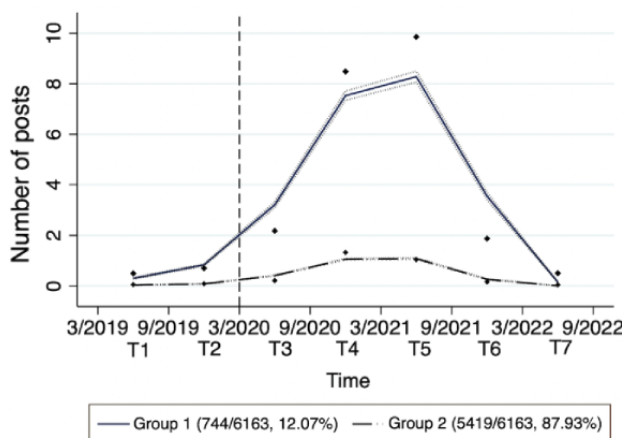
Table 2. Parameter estimates for the 2-group zero-inflated group-based trajectory modeling of post volume among r/SuicideWatch users throughout the COVID-19 pandemic.

Group and parameter	Estimate	SE	t value (df)	P value
1 (count)				
Intercept	2.62	0.01	214.7 (743)	<.001
Linear	0.1	0.03	3.63 (743)	<.001
Quadratic	-0.44	0.02	-24.52 (743)	<.001
Cubic	-0.22	0.02	-10.56 (743)	<.001
2 (count)				
Intercept	0.59	0.01	42.46 (5418)	<.001
Linear	-0.14	0.02	-5.99 (5418)	<.001
Quadratic	-0.12	0.02	-5.07 (5418)	<.001
1 (inflation)				
Alpha0	-0.19	0.05	-3.78 (743)	<.001
Alpha1	-1.85	0.1	-18.4 (743)	<.001
Alpha2	1.49	0.06	24.17 (743)	<.001
Alpha3	0.75	0.07	10.8 (743)	<.001
2 (inflation)				
Alpha0	-0.34	0.03	-13.16 (5418)	<.001
Alpha1	-2.31	0.06	-37.54 (5418)	<.001
Alpha2	2.63	0.04	58.83 (5418)	<.001
Alpha3	1.49	0.05	27.77 (5418)	<.001

Figure 1 depicts the post volume trajectories across the COVID-19 pandemic for the 2 identified groups of users. Group 1, designated as the “high risk of suicide” group, consisted of 744 (12.07%) users. Their post volume on r/SuicideWatch showed a gradual increase during the 2 prepandemic periods, followed by a rapid acceleration after the pandemic began. This trend peaked approximately 1 year after the pandemic outbreak and subsequently declined, returning to its initial level during

the second year of the pandemic. Group 2, identified as the “low risk of suicide” group, comprised the majority of users (5419/6163, 87.93%). This group exhibited a slight increase in post volume on r/SuicideWatch following the pandemic outbreak, followed by stabilization and eventual recovery. Throughout the pandemic, group 2 maintained a relatively low post volume on the subreddit.

Figure 1. Trajectories of post volume among r/SuicideWatch users throughout the COVID-19 pandemic.



Linguistic Feature Analysis

The summary distribution of frequency of use in LIWC for the 2 groups can be found in Multimedia Appendix 2. Similar to the distribution of post volume, we observed a zero-inflated

phenomenon for these linguistic features across periods. Therefore, descriptive statistics including the median, first quartile, and third quartile were used. By plotting the median frequency trend for each included LIWC feature throughout the pandemic for the 2 groups, trends by category were illustrated

(Multimedia Appendix 3). We observed the following: (1) During the year before the pandemic (T1: March 2019-September 2019 and T2: September 2019-March 2020), word frequency was generally low for both groups. (2) Throughout the pandemic (T3-T6: March 2020-March 2022), words related to cognitive processes, perceptual processes, biological processes, and personal concerns showed relatively lower frequency compared with personal pronouns, affective processes, social processes, drives, relativity, and time orientations. (3) During the first year of the pandemic (T3-T4: March 2020-March 2021), both groups exhibited sharp increases in word frequency. (4) During the second year of the pandemic (T5-T6: March 2021-March 2022), the high-risk group continued to experience a slower increase until reaching a peak and subsequent decrease, while the low-risk group's frequency decreased. (5) Moving into the third year of the pandemic (T7: March 2022-September 2022), word frequency returned to prepandemic levels in both groups. Despite both groups showing increased use of most word types during the pandemic, the

high-risk group exhibited a longer-lasting increase with a peak lagging behind that of the low-risk group. This suggests that the pandemic had a more enduring impact on high-risk users.

The results of between-group comparisons using Poisson regression (with the low-risk group as the reference) are depicted in Figure 2. In general, the high-risk group utilized most types of words more frequently than the low-risk group both before and during the initial 6 months of the pandemic (illustrated in red for T1-T3). Later, in the second half of the pandemic, their differences narrowed and even reversed (as shown in green during T4), with both groups demonstrating increased word use. Subsequently, the high-risk group once again surpassed the low-risk group, and these differences grew larger in the subsequent periods (as indicated in deeper red from T5 to T7). This pattern corresponded with the plotted trend, where the high-risk group exhibited a prolonged increase and a delayed peak following the rise during T4, whereas the frequency of the low-risk group quickly decreased and returned to its initial level.

Figure 2. Visualization of Poisson regression results comparing LIWC frequency between r/SuicideWatch users in the high- and low-risk groups throughout the COVID-19 pandemic (low-risk group as a reference; red: higher frequency in the high-risk group; green: lower frequency in the high-risk group). LIWC: Linguistic Inquiry and Word Count. * $P<.05$; ** $P<.01$; *** $P<.001$.

	T1 3/2019-9/2019	T2 9/2019-3/2020	T3 3/2020-9/2020	T4 9/2020-3/2021	T5 3/2021-9/2021	T6 9/2021-3/2022	T7 3/2022-9/2022
Personal pronouns							
Personal pronouns	1.79	2.09	2.09*	0.97	1.51*	3.95***	3.61*
First-person person singular	1.93	2.16	2.13	0.99	1.54*	4.27***	3.75*
First-person person plural	2.39	1.57	1.5	0.89	1.25	3.3	3.46
Second person	1.32	1.98	2.6	1.09	1.57	3.2	3.47
Third-person singular	1.16	2.15	1.79	0.72	1.24	3.16	2.1
Third-person plural	1.74	1.73	1.86	0.95	1.55	2.7	4.14
Affective processes							
Affective processes	2.16	2.01	2.11*	1.01	1.6**	3.91***	3.98*
Positive emotions	1.9	1.75	2.16	1	1.51	3.46*	3.47
Negative emotions	2.32	2.16	2.08	1.02	1.66*	4.18***	4.3*
Anxiety	3.08	1.51	1.6	0.86	1.42	4.05	5.36
Anger	2.3	2.51	2.4	1.13	1.84	4.14*	4.4
Sad	1.98	2.01	2.05	0.99	1.56	4.32*	3.45
Social processes							
Social processes	1.73	1.76	1.89	0.9	1.43	3.1**	3.04
Family	2.37	1.64	1.31	0.75	1.16	2.6	4.24
Friend	2.68	2.18	1.69	0.76	1.22	2.75	3.26
Female	1.83	2.38	1.7	0.8	1.32	2.66	2.12
Male	1.3	1.63	1.74	0.75	1.12	3.46	3.55
Cognitive processes							
Cognitive processes	1.91	1.84	2.02	0.94	1.49	3.58**	3.63
Insight	1.63	1.75	1.86	0.86	1.48	3.33	4
Cause	2.11	2.01	2.04	0.96	1.48	3.43	2.7
Discrepancy	1.89	1.82	2.2	0.99	1.56	3.63	3.79
Tentativeness	1.79	1.66	2.16	0.92	1.48	3.54	3.57
Certainty	2.36	1.99	2.03	1.03	1.53	3.96	3.98
Differentiation	1.89	1.89	1.99	0.94	1.43	3.66	3.8
Perceptual processes							
Perceptual processes	1.54	2.18	2.02	0.95	1.51	3.58	4.09
See	1.7	2.41	2.15	0.97	1.53	3.77	4.86
Hear	1.66	1.7	2.02	0.97	1.47	2.79	4.62
Feel	1.38	2.18	1.92	0.92	1.51	3.76	3.6
Biological processes							
Biological processes	2.23	2.43	2.17	1.01	1.55	3.76**	4.68
Body	2.96	2.61	2.37	1.15	1.75	4.83	5.06
Health	1.87	2.35	2.02	0.95	1.4	3.44*	4.49
Sexual	3.01	2.99	2.71	1.15	1.92	4.14	3.86
Ingest	2.28	2.03	2	0.99	1.36	2.48	9.29
Drives							
Drives	2.08	1.98	1.92	0.93	1.42	3.39***	3.78*
Affiliation	2.24	1.88	1.75	0.85	1.29	2.68	3.06
Achievement	2.13	1.95	2.24	0.92	1.4	3.8*	3.75
Power	1.98	2.25	2.08	0.94	1.47	3.25	3.78
Reward	2.05	1.97	1.85	0.98	1.42	3.3	4.15
Risk	1.78	1.46	1.97	0.98	1.57	3.99	4.15
Relativity							
Relativity	1.76	1.91	2.02*	0.95	1.48**	3.6***	4.01**
Motion	1.91	2.06	2.03	0.98	1.61	3.6*	3.69
Space	1.79	1.79	2.02	0.95	1.46	3.51***	3.88*
Time	1.67	1.95	2.02	0.94	1.47	3.64***	4.33*
Time orientations							
Past focus	1.53	2.02	1.84	0.88	1.37	3.54**	3.41
Present focus	2.04	1.82	2*	0.97	1.53**	3.75***	4.19**
Future focus	2.11	2.11	2.64	1.12	1.73	4.72*	4.56
Personal concerns							
Work	2.7	1.81	2	0.87	1.34	3.73	3.14
Leisure	3.1	2.03	1.76	1.03	1.42	3.07	6.32
Home	1.66	1.54	1.86	0.84	1.37	3.42	2.71
Money	1.89	2.59	1.66	0.76	1.22	3.31	2.42
Religion	1.72	1.46	1.86	1.18	1.54	3.81	5.26
Death	1.74	1.88	2.07	1.07	1.5	3.97*	4.86

Statistical differences in the frequency of word use were primarily observed after the pandemic outbreak (Figure 2). During T3 (March 2020-September 2020), the high-risk group showed significantly more frequent use of words related to personal pronouns (RR 2.09, SE 0.73, $P=.03$), affective processes (RR 2.11, SE 0.73, $P=.03$), relativity (RR 2.02, SE 0.55, $P=.01$), and present focus (RR 2.00, SE 0.54, $P=.01$) compared with the low-risk group. During T5 (March 2021-September 2021), posts in the high-risk group also exhibited higher frequencies of words related to personal pronouns (RR 1.51, SE 0.26, $P=.02$), first-person singular (RR 1.54, SE 0.31, $P=.03$), affective processes (RR 1.60, SE 0.27, $P=.005$), negative emotions (RR 1.66, SE 0.35, $P=.02$), relativity (RR 1.48, SE 0.20, $P=.005$), and present focus (RR 1.53, SE 0.20, $P=.001$). During T6 (September 2021-March 2022), 21 types of words across categories such as personal pronouns; affective, social, cognitive, and biological processes; drives;

relativity; time orientations; and personal concerns showed higher frequencies in the high-risk group compared with the low-risk group. Meanwhile, in T7 (March 2022-September 2022), the high-risk group exhibited higher frequencies of words related to personal pronouns (RR 3.61, SE 2.08, $P=.03$), first-person singular (RR 3.75, SE 2.38, $P=.04$), affective processes (RR 3.98, SE 2.22, $P=.01$), negative emotions (RR 4.3, SE 3.02, $P=.04$), drives (RR 3.78, SE 2.35, $P=.03$), relativity (RR 4.01, SE 1.75, $P=.001$), space (RR 3.88, SE 2.66, $P=.04$), time (RR 4.33, SE 2.76, $P=.02$), and present focus (RR 4.19, SE 1.8, $P=.001$).

To investigate the linguistic markers that could distinguish group membership, we identified 21 word types that significantly ($P\leq.05$) differed between the 2 groups during the last 3 periods (T5, T6, and T7). These word types include personal pronouns, first-person singular, affective processes, positive emotions,

negative emotions, anger, sadness, social processes, cognitive processes, biological processes, health, drives, achievement, relativity, motion, space, time, past focus, present focus, future focus, and death. The word types with the most observed differences were selected as potential linguistic markers for further examination. Then, we omitted 6 word types (ie, personal pronouns, affective processes, negative emotions, biological processes, drives, and relativity) due to their hierarchical relationship with their subcategory words to avoid duplication. To better fit the multivariate logistic regression, we calculated a binary measure for each of the remaining 15 potential markers, indicating no use (0) or use (1) of the word. We calculated the average frequency of each word across T5, T6, and T7, and then dichotomized these averages. Averaged values of 0 were retained as 0, indicating no use of the word during T5, T6, and T7. For averaged values greater than 0, we recoded the value as 1, indicating that the word was used at least once during T5, T6, and T7, regardless of the actual frequency. To mitigate collinearity among the 15 words, we used lasso regression for variable selection. Ultimately, we omitted 3 word types—specifically, the first-person singular, space, and time—leaving us with 12 linguistic features: positive emotions, anger, sadness, social processes, cognitive processes, health, achievement, motion, past focus, present focus, future focus, and death.

Table 3 presents the results of the multivariate logistic regression, incorporating potential linguistic markers into the 2-group GBTM. This analysis models the odds of being in the high-risk group based on the usage of potential linguistic features, with no use of the word serving as the reference. The final model indicated that 9 linguistic features emerged as significant ($P \leq .05$) markers distinguishing the 2 groups. Notably, using words related to cognitive processes and present focus during the later COVID-19 periods had lower odds of being in the high-risk group compared with not using these words ($OR_{\text{cognitive processes}} 0.06$, $SE 0.85$, $P < .001$; $OR_{\text{present focus}} 0.03$, $SE 0.85$, $P < .001$). This indicates that the use of these words was associated with being in the low-risk group. Contrastingly, the odds of being in the high-risk group were substantially higher when using words related to anger, sadness, health, achievement, motion, future focus, and death, compared with not using these words ($OR_{\text{anger}} 3.23$, $SE 0.29$, $P < .001$; $OR_{\text{sadness}} 3.23$, $SE 0.25$, $P < .001$; $OR_{\text{health}} 2.56$, $SE 0.33$, $P = .005$; $OR_{\text{achievement}} 1.67$, $SE 0.26$, $P = .049$; $OR_{\text{motion}} 4.17$, $SE 0.37$, $P < .001$; $OR_{\text{future focus}} 2.86$, $SE 0.3$, $P < .001$; $OR_{\text{death}} 4.35$, $SE 0.26$, $P < .001$). The results illustrated that these 7 words, used 1 year after the pandemic outbreak, were linguistic markers for being in the high-risk group. [Multimedia Appendix 1](#) provides examples of posts that high-risk users published 1 year after the pandemic outbreak.

Table 3. Odds ratios from multivariate logistic regression to identify linguistic markers that discriminate r/SuicideWatch users in the high- and low-risk groups.

Linguistic markers	Odds ratio	SE	<i>t</i> value (<i>df</i>)	<i>P</i> value
Constant	14.83	0.08	32.37 (6162)	<.001
Positive emotions	1.79	0.44	1.34 (6162)	.18
Anger	3.23	0.29	4.12 (6162)	<.001
Sad	3.23	0.25	4.75 (6162)	<.001
Social processes	0.56	0.53	-1.11 (6162)	.27
Cognitive processes	0.06	0.85	-3.38 (6162)	<.001
Health	2.56	0.33	2.84 (6162)	.005
Achievement	1.67	0.26	1.97 (6162)	.049
Motion	4.17	0.37	3.81 (6162)	<.001
Past focus	1.82	0.40	1.48 (6162)	.14
Present focus	0.03	0.85	-4.27 (6162)	<.001
Future focus	2.86	0.3	3.52 (6162)	<.001
Death	4.35	0.26	5.61 (6162)	<.001

Discussion

Principal Findings

To the best of our knowledge, this work is the first to address heterogeneity in suicide risk among social media users by incorporating the temporal characteristics of suicide. Based on the 2 identified trajectories of post volume throughout the COVID-19 pandemic, users on the r/SuicideWatch subreddit were divided into the “high risk of suicide” group (744/6163, 12.07%), characterized by a sharp and lasting increase in post

volume, and the “low risk of suicide” group (5419/6163, 87.93%), characterized by a consistently low and mild increase in post volume during the pandemic. In terms of linguistic features, the 2 groups exhibited distinct frequency trends throughout the pandemic. The high-risk group demonstrated longer-lasting increases and lagged peaks in most linguistic frequencies. Contrarily, the low-risk group displayed different trends. Notably, the use of words related to anger, sadness, health, achievement, motion, future focus, and death 1 year after the pandemic outbreak emerged as markers for membership in the high-risk group. Conversely, words associated with cognitive

processes and present focus were identified as linguistic markers for the low-risk group.

Across the pre- and peripandemic periods, this study identified 2 distinct patterns of change in suicide risk among r/SuicideWatch users based on trajectory modeling of their post volume. These findings underscore the heterogeneity in suicide risk among r/SuicideWatch users from a longitudinal perspective during the pandemic. Users' participation in subreddits, including posting frequency, commenting habits, and emotional expression, was influenced by significant pandemic events [38], particularly its progression in Western countries such as the US, the UK, Canada, Australia, and Germany, where a majority of Redditors originate [70]. Both groups of users exhibited immediate increases in post volume following the onset of the COVID-19 pandemic in March 2020. However, post volume returned to prepandemic levels in later stages, around September 2021, as many Western countries began to resume normalcy [71]. According to the fluid vulnerability theory [40], an environmental stressor can trigger a suicidal response within individuals who have predispositions to such reactions. While the half-year intervals may not fully capture users' detailed responses to the pandemic or fluctuations in their suicidal episodes, the heightened posting activity observed in both groups following the pandemic's onset suggests an overall increase in their suicide risk. Therefore, the ongoing pandemic and its repercussions may serve as a persistent environmental stressor for users. Importantly, the high-risk group exhibited significantly greater increases in post volume during the pandemic (T3-T5: March 2020-September 2021) compared with the low-risk group. This suggests that the onset of suicidal episodes was more pronounced among the high-risk group than the low-risk group. The finding of users' heterogeneity in suicide risk can be explained by the interaction between one's baseline and acute risk of suicide, as proposed by the fluid vulnerability theory [41]. Individuals in the high-risk group may have a higher baseline risk due to underlying vulnerabilities, making their suicidal tendencies more readily activated compared with those in the low-risk group, who have fewer vulnerabilities and a lower baseline risk. The higher level of predispositions among high-risk users also renders them more vulnerable to the adverse impacts of the pandemic. This vulnerability activates heightened risks in various domains including cognition (eg, hopelessness), emotion (eg, depression), behavior (eg, social withdrawal), and physiology (eg, sleep disturbances), contributing to their increased acute risk. The higher baseline and acute risks motivate high-risk users to express their heightened concerns, seek support, and exchange information online, leading to a significant increase in social media engagement [72]. By contrast, the low-risk group, which showed consistently low and mild increases in post volume, likely represents the majority less predisposed to suicide risk, indicating greater resilience to the pandemic. Therefore, they may perceive the pandemic as less threatening and experience fewer burdens related to cognitive, emotional, behavioral, and physiological factors. With fewer concerns to share, they exhibited only a mild and minimal increase in post volume. Our findings underscore the heterogeneity in patterns of suicide risk change during the pandemic within this population, highlighting the importance

of considering users' individual differences and the temporal dynamics of suicide in future studies using social media data.

Additionally, this study observed differences in the trends of linguistic features between the high- and low-risk groups. During the first year of the pandemic (T3-T4: March 2020-March 2021), both groups significantly increased their use of words related to personal pronouns, positive and negative emotions, social processes, drives, relativity, and time orientations compared with other word categories, indicating broader topics of interest during this period [73]. However, the increase in linguistic frequency continued at a slower pace in the high-risk group before reaching a peak and returning to its original volume (T5-T7: March 2021-September 2022), whereas the low-risk group experienced an early, mild peak followed by an immediate decrease. This divergent trend highlights that most statistical differences in linguistic frequency between the 2 groups became evident 1 year after the outbreak of the pandemic (T5-T7: March 2021-September 2022), indicating that the impact of the pandemic on the high-risk group was more prolonged and delayed compared with the low-risk group. This finding not only underscores the heterogeneity between the 2 groups but also highlights that high-risk users have experienced prolonged stress and heightened sensitivity during the pandemic.

To better identify users at high risk of suicide and understand their underlying concerns, we examined linguistic markers based on several features that showed between-group differences 1 year into the pandemic. Specifically, words related to anger, sadness, health, achievement, motion, future focus, and death were identified as linguistic markers for the high-risk group, which partially aligns with previous findings [36,49]. We delved deeper into the post content of high-risk users to grasp the context in which these linguistic markers were used. Words related to anger and sadness were used by high-risk users to express agitation and hopelessness concerning the overwhelming impact of the pandemic, emotions strongly linked with an increased risk of suicidal thoughts and behaviors [74-76]. When discussing health and motion, high-risk users conveyed heightened concerns about their physical well-being and limitations in movement due to pandemic-related lockdowns [37]. Additionally, they used achievement-related words to express feelings of failure in meeting their goals and fulfilling their need for social recognition. These users may place high demands on themselves, striving to accomplish difficult tasks and meet high standards, which can increase their vulnerability to depression and suicidal behaviors [17,77]. The widespread economic losses, unemployment, and disruptions in educational settings caused by the pandemic further impeded their ability to achieve success, leading to lowered self-esteem, depressive mood, and heightened suicidal risk [78]. Additionally, we discovered that words related to future focus served as linguistic markers for the high-risk group. While previous studies have noted that suicidal individuals often emphasize present-focused words, reflecting their hopelessness about the future and acute concerns about their current state [49,79], this pattern may differ during the pandemic. High-risk users articulated their apprehensions about an uncertain and uncontrollable future amid the evolving pandemic, as exemplified in the texts (Multimedia Appendix 1). Additionally, the high-risk group

used more words related to death. In addition to referencing suicide or hopelessness, this marker also indicated their perceived threats from virus infections, death cases, or the loss of loved ones during the pandemic [37,80].

Our findings have significant implications for managing suicide issues during future public health crises. By analyzing social media posts, we identified a small percentage of users at high risk of suicide who appear particularly sensitive and vulnerable to pandemic-related events or similar public health crises in the future. Although the majority are at low risk of suicide, these results underscore serious concerns, as high-risk users may be poised to progress to the next stage of suicidal ideation or take action [36]. Therefore, it is crucial to pay particular attention to this subset of users to alleviate their difficulties in such situations. Moreover, the active posting and disclosure by these high-risk users may lead to “suicidal contagion” affecting low-risk users, potentially propagating suicidal tendencies within online communities [81]. Therefore, ongoing surveillance, screening, and timely intervention during public health crises are necessary to prevent this issue. Furthermore, the distinct linguistic patterns observed in the 2 groups in this study can serve as a foundation for understanding the underlying concerns contributing to these users’ suicide risk, thereby aiding in the development of targeted interventions. The identified language markers for the high-risk group can also serve as a basis for screening high-risk individuals in future pandemic-like events.

Additionally, this study has several limitations. First, aside from users disclosing their own suicidal issues, r/SuicideWatch includes posts about others’ suicide risk, providing assistance to those in distress, and disseminating research messages [24]. Although the percentage of these posts was small in our manual screening of selected posts (18/3372, 0.53%, sampled posts), future studies are advised to mitigate this noise or incorporate users’ other online behaviors (eg, commenting frequency and post length) to more accurately assess users’ suicide risk. Moreover, a significant portion of users in our data set transitioned from other subreddits to r/SuicideWatch following the onset of the pandemic, starting with 0 post volume in periods preceding their initial posts (eg, 2 prepandemic periods). Future studies could track users’ earlier psychosocial characteristics on other subreddits to identify indicators that might foreshadow their shift toward actively discussing suicidal concerns on r/SuicideWatch. Second, we utilized seven 6-month intervals as the time frames for capturing post volume and linguistic frequency, which may have been too lengthy to capture specific

fluctuations. Nan et al [82] also utilized 6-month intervals and identified a 2-trajectory model for changes in suicidal ideation throughout the pandemic using scores from multiple-item scales as the trajectory variable. However, using shorter intervals (eg, 2-6 weeks) can reveal more trajectories, as it considers minor but significant differences rather than averaging them in the analysis [83,84]. Given the frequent release of pandemic-related news and information (eg, daily reports), users shared real-time reactions to these updates in their posts, potentially reflecting immediate changes in their suicidal thoughts or behaviors, a nuance that might not have been fully captured in our study [38]. Future studies could benefit from shorter time intervals to capture more nuanced and continuous changes in suicide risk, potentially revealing diverse trajectories of suicidal ideation. Third, due to the anonymity of Reddit data, our access was restricted to users’ demographics (eg, country or region, age, and sex). Consequently, these factors could not be included as potential covariates for modeling trajectory groups or for comparing the demographic compositions between high- and low-risk user groups. We also acknowledge the potential confounding impact of varying pandemic waves and government control policies across different countries, which we were unable to explore due to the lack of geographical information from users. Future studies should aim to investigate these factors while maintaining the integrity of data characterized by high self-disclosure and authenticity [14]. Additionally, our analysis focused exclusively on Reddit data from a Western context [70]. Cross-cultural validation using data from other platforms, such as Weibo, will be crucial to enhance the generalizability of findings and consider cultural and national policy influences.

Conclusions

This study used social media posts to demonstrate the heterogeneous patterns of change in suicide risk during the COVID-19 pandemic. A group of Reddit users at high risk of suicide was identified, characterized by a sharp and sustained increase in post volume. These high-risk users exhibited distinct linguistic patterns, particularly in their use of words related to anger, sadness, health, achievement, motion, future focus, and death during the later stages of the pandemic. Our findings underscore the importance of recognizing users’ heterogeneity in long-term suicide risk. Real-time surveillance of suicide risk using social media data during future public health crises is essential to provide timely support to individuals potentially at high risk of suicide.

Acknowledgments

The study was sponsored by the Research Grants Council of the Hong Kong Special Administrative Region, China (Collaborative Research Fund, Project No. C1031-18G). The sponsors had no further role in study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Data Availability

The data set used in this study is available from the corresponding author upon request.

Authors' Contributions

NXY and QL conceptualized the study. YY, JL, and XL further completed the study design. JL and XL contributed to the data collection. YY was responsible for data analysis and interpretation. YY drafted the manuscript. NXY, QL, JL, and XL reviewed and edited the draft. QL and NXY administrated the project and acquired the funding. We thank the 2 anonymous reviewers for their valuable input.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of linguistic markers for users in the high-risk group.

[\[DOCX File, 19 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Summary distribution of LIWC frequency for r/SuicideWatch users in the high- and low-risk groups during each period (median [Q1, Q3]). LIWC: Linguistic Inquiry and Word Count.

[\[DOCX File, 34 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Trends of LIWC frequency by category for r/SuicideWatch users in the high- and low-risk groups throughout the COVID-19 pandemic (based on the median). LIWC: Linguistic Inquiry and Word Count.

[\[DOCX File, 392 KB-Multimedia Appendix 3\]](#)

References

1. Riedel-Heller S, Richter D. COVID-19 pandemic and mental health of the general public: is there a tsunami of mental disorders? *Psychiatr Prax*. Nov 02, 2020;47(8):452-456. [doi: [10.1055/a-1290-3469](https://doi.org/10.1055/a-1290-3469)] [Medline: [33137827](https://pubmed.ncbi.nlm.nih.gov/33137827/)]
2. Gunnell D, Appleby L, Arensman E, Hawton K, John A, Kapur N, et al. COVID-19 Suicide Prevention Research Collaboration. Suicide risk and prevention during the COVID-19 pandemic. *Lancet Psychiatry*. Jun 2020;7(6):468-471. [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30171-1](https://doi.org/10.1016/S2215-0366(20)30171-1)] [Medline: [32330430](https://pubmed.ncbi.nlm.nih.gov/32330430/)]
3. Sher L. The impact of the COVID-19 pandemic on suicide rates. *QJM*. Oct 01, 2020;113(10):707-712. [FREE Full text] [doi: [10.1093/qjmed/hcaa202](https://doi.org/10.1093/qjmed/hcaa202)] [Medline: [32539153](https://pubmed.ncbi.nlm.nih.gov/32539153/)]
4. Pirkis J, Gunnell D, Shin S, Del Pozo-Banos M, Arya V, Aguilar PA, et al. Suicide numbers during the first 9-15 months of the COVID-19 pandemic compared with pre-existing trends: an interrupted time series analysis in 33 countries. *EClinicalMedicine*. Sep 2022;51:101573. [FREE Full text] [doi: [10.1016/j.eclinm.2022.101573](https://doi.org/10.1016/j.eclinm.2022.101573)] [Medline: [35935344](https://pubmed.ncbi.nlm.nih.gov/35935344/)]
5. Dubé JP, Smith MM, Sherry SB, Hewitt PL, Stewart SH. Suicide behaviors during the COVID-19 pandemic: a meta-analysis of 54 studies. *Psychiatry Res*. Jul 2021;301:113998. [FREE Full text] [doi: [10.1016/j.psychres.2021.113998](https://doi.org/10.1016/j.psychres.2021.113998)] [Medline: [34022657](https://pubmed.ncbi.nlm.nih.gov/34022657/)]
6. Pathirathna ML, Nandasena HMRKG, Atapattu AMMP, Weerasekara I. Impact of the COVID-19 pandemic on suicidal attempts and death rates: a systematic review. *BMC Psychiatry*. Jul 28, 2022;22(1):506. [FREE Full text] [doi: [10.1186/s12888-022-04158-w](https://doi.org/10.1186/s12888-022-04158-w)] [Medline: [35902951](https://pubmed.ncbi.nlm.nih.gov/35902951/)]
7. Yan Y, Hou J, Li Q, Yu NX. Suicide before and during the COVID-19 pandemic: a systematic review with meta-analysis. *Int J Environ Res Public Health*. Feb 14, 2023;20(4):3346. [FREE Full text] [doi: [10.3390/ijerph20043346](https://doi.org/10.3390/ijerph20043346)] [Medline: [36834037](https://pubmed.ncbi.nlm.nih.gov/36834037/)]
8. John A, Pirkis J, Gunnell D, Appleby L, Morrissey J. Trends in suicide during the covid-19 pandemic. *BMJ*. Nov 12, 2020;371:m4352. [doi: [10.1136/bmj.m4352](https://doi.org/10.1136/bmj.m4352)] [Medline: [33184048](https://pubmed.ncbi.nlm.nih.gov/33184048/)]
9. Knipe D, John A, Padmanathan P, Eyles E, Dekel D, Higgins JPT, et al. Suicide and self-harm in low- and middle- income countries during the COVID-19 pandemic: a systematic review. *PLOS Glob Public Health*. 2022;2(6):e0000282. [FREE Full text] [doi: [10.1371/journal.pgph.0000282](https://doi.org/10.1371/journal.pgph.0000282)] [Medline: [36962383](https://pubmed.ncbi.nlm.nih.gov/36962383/)]
10. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. *Early Interv Psychiatry*. Apr 2016;10(2):103-121. [doi: [10.1111/eip.12229](https://doi.org/10.1111/eip.12229)] [Medline: [25702826](https://pubmed.ncbi.nlm.nih.gov/25702826/)]
11. Tadesse MM, Lin H, Xu B, Yang L. Detection of suicide ideation in social media forums using deep learning. *Algorithms*. Dec 24, 2019;13(1):7. [doi: [10.3390/a13010007](https://doi.org/10.3390/a13010007)]
12. Slemmon A, McAuliffe C, Goodyear T, McGuinness L, Shaffer E, Jenkins EK. Reddit users' experiences of suicidal thoughts during the COVID-19 pandemic: a qualitative analysis of r/Covid19_support posts. *Front Public Health*. 2021;9:693153. [FREE Full text] [doi: [10.3389/fpubh.2021.693153](https://doi.org/10.3389/fpubh.2021.693153)] [Medline: [34458223](https://pubmed.ncbi.nlm.nih.gov/34458223/)]

13. Wong K, Chan CS, Chan M, Wong C, Cheng Q, Xiong C, et al. Who seeks help online? Comparing online and offline help-seeking preferences amongst youths with suicidal ideation. *J Affect Disord*. Sep 01, 2021;292:21-29. [[FREE Full text](#)] [doi: [10.1016/j.jad.2021.05.056](https://doi.org/10.1016/j.jad.2021.05.056)] [Medline: [34087633](https://pubmed.ncbi.nlm.nih.gov/34087633/)]
14. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: observational study. *J Med Internet Res*. Oct 12, 2020;22(10):e22635. [[FREE Full text](#)] [doi: [10.2196/22635](https://doi.org/10.2196/22635)] [Medline: [32936777](https://pubmed.ncbi.nlm.nih.gov/32936777/)]
15. Bitew S, Bekoulis G, Deleu J, Sterckx L, Zaporozets K, Demeester T. Predicting suicide risk from online postings in Reddit the UGent-IDLab submission to the CLPsych 2019 shared task A. In: *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*. New York, NY. Association for Computational Linguistics; 2019. Presented at: The Sixth Workshop on Computational Linguistics and Clinical Psychology; June 6, 2019:158-161; Minneapolis, MN. URL: <https://aclanthology.org/W19-3019> [doi: [10.18653/v1/w19-3019](https://doi.org/10.18653/v1/w19-3019)]
16. Chatterjee M, Samanta P, Kumar P, Sarkar D. Suicide ideation detection using multiple feature analysis from Twitter data. New York, NY. IEEE; 2022. Presented at: 2022 IEEE Delhi Section Conference (DELCON); February 1-13, 2022:1-6; New Delhi, India. [doi: [10.1109/delcon54057.2022.9753295](https://doi.org/10.1109/delcon54057.2022.9753295)]
17. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. *J Med Internet Res*. Jul 10, 2017;19(7):e243. [[FREE Full text](#)] [doi: [10.2196/jmir.7276](https://doi.org/10.2196/jmir.7276)] [Medline: [28694239](https://pubmed.ncbi.nlm.nih.gov/28694239/)]
18. Ruiz V, Shi L, Quan W, Ryan N, Biernesser C, Brent D. CLPsych2019 shared task: predicting suicide risk level from reddit posts on multiple forums. In: *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*. New York, NY. Association for Computational Linguistics; 2019. Presented at: The Sixth Workshop on Computational Linguistics and Clinical Psychology; June 6, 2019:162-166; Minneapolis, MN. URL: <https://aclanthology.org/W19-3020> [doi: [10.18653/v1/w19-3020](https://doi.org/10.18653/v1/w19-3020)]
19. Shan Y. A promising form of suicide intervention model: distant supervision in social media the application of artificial intelligence in suicide intervention. In: *Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences*. 2021. Presented at: The 2nd International Symposium on Artificial Intelligence for Medicine Sciences; October 29-31, 2021; Beijing, China. [doi: [10.1145/3500931.3500941](https://doi.org/10.1145/3500931.3500941)]
20. Tan Z, Liu X, Liu X, Cheng Q, Zhu T. Designing microblog direct messages to engage social media users with suicide ideation: interview and survey study on Weibo. *J Med Internet Res*. Dec 12, 2017;19(12):e381. [[FREE Full text](#)] [doi: [10.2196/jmir.8729](https://doi.org/10.2196/jmir.8729)] [Medline: [29233805](https://pubmed.ncbi.nlm.nih.gov/29233805/)]
21. Varathan K, Talib N. Suicide detection system based on Twitter. 2014. Presented at: 2014 Science and Information Conference; August 27-29, 2014; London, UK. [doi: [10.1109/sai.2014.6918275](https://doi.org/10.1109/sai.2014.6918275)]
22. Goel A, Gupta L. Social media in the times of COVID-19. *J Clin Rheumatol*. Sep 2020;26(6):220-223. [[FREE Full text](#)] [doi: [10.1097/RHU.0000000000001508](https://doi.org/10.1097/RHU.0000000000001508)] [Medline: [32852927](https://pubmed.ncbi.nlm.nih.gov/32852927/)]
23. Hussain W. Role of social media in COVID-19 pandemic. *Int J Front Sci*. 2020;4(2):59-60. [[FREE Full text](#)] [doi: [10.37978/tjfs.v4i2.144](https://doi.org/10.37978/tjfs.v4i2.144)]
24. r/SuicideWatch. Reddit. URL: <https://www.reddit.com/r/SuicideWatch/> [accessed 2023-03-26]
25. Kusuma K, Larsen M, Quiroz JC, Gillies M, Burnett A, Qian J, et al. The performance of machine learning models in predicting suicidal ideation, attempts, and deaths: a meta-analysis and systematic review. *J Psychiatr Res*. Nov 2022;155:579-588. [doi: [10.1016/j.jpsychires.2022.09.050](https://doi.org/10.1016/j.jpsychires.2022.09.050)] [Medline: [36206602](https://pubmed.ncbi.nlm.nih.gov/36206602/)]
26. Sawhney R, Joshi H, Gandhi S, Shah R. Towards ordinal suicide ideation detection on social media. In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 2021. Presented at: ACM International Conference on Web Search and Data Mining; March 8-12, 2021; New York, NY. [doi: [10.1145/3437963.3441805](https://doi.org/10.1145/3437963.3441805)]
27. Zirikly A, Resnik P, Uzuner O, Hollingshead K. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In: *Proceedings of The Sixth Workshop on Computational Linguistics and Clinical Psychology*. New York, NY. Association for Computational Linguistics; 2019. Presented at: The Sixth Workshop on Computational Linguistics and Clinical Psychology; June 6, 2019:24-33; Minneapolis, MN. URL: <https://aclanthology.org/W19-3003.pdf> [doi: [10.18653/v1/w19-3003](https://doi.org/10.18653/v1/w19-3003)]
28. Li Q, Xiong Q, Ji S, Wen J, Gao M, Yu Y, et al. Using fine-tuned conditional probabilities for data transformation of nominal attributes. *Pattern Recognition Letters*. Dec 2019;128:107-114. [doi: [10.1016/j.patrec.2019.08.024](https://doi.org/10.1016/j.patrec.2019.08.024)]
29. Zdravevski E, Lameski P, Kulakov A, Kalajdziski S. Transformation of nominal features into numeric in supervised multi-class problems based on the weight of evidence parameter. 2015. Presented at: 2015 Federated Conference on Computer Science and Information Systems (FedCSIS); September 13-16, 2015; Lodz, Poland. [doi: [10.15439/2015f90](https://doi.org/10.15439/2015f90)]
30. Kumar M, Dredze M, Coppersmith G, De CM. Detecting changes in suicide content manifested in social media following celebrity suicides. In: *HT '15: Proceedings of the 26th ACM Conference on Hypertext & Social Media*. New York, NY. Association for Computing Machinery; 2015. Presented at: HT '15: 26th ACM Conference on Hypertext and Social Media; September 1-4, 2015:85-94; Guzelyurt, Northern Cyprus. [doi: [10.1145/2700171.2791026](https://doi.org/10.1145/2700171.2791026)]
31. Monselise M, Yang C. "I'm always in so much pain and no one will understand"-Detecting patterns in suicidal ideation on Reddit. In: *WWW '22: Companion Proceedings of the Web Conference 2022*. New York, NY. Association for Computing

- Machinery; 2022. Presented at: WWW '22: The ACM Web Conference; April 25-29, 2022:686-691; Virtual Event/Lyon, France. URL: <https://dl.acm.org/doi/10.1145/3487553.3524700> [doi: [10.1145/3487553.3524700](https://doi.org/10.1145/3487553.3524700)]
32. Lee K, Agrawal A, Choudhary A. Real-time disease surveillance using Twitter data: demonstration on flu and cancer. New York, NY. Association for Computing Machinery; 2013. Presented at: KDD' 13: The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 11-14, 2013:1474-1477; Chicago, IL. [doi: [10.1145/2487575.2487709](https://doi.org/10.1145/2487575.2487709)]
 33. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: longitudinal analysis of Twitter data. *J Med Internet Res*. Dec 14, 2020;22(12):e21418. [FREE Full text] [doi: [10.2196/21418](https://doi.org/10.2196/21418)] [Medline: [33284783](https://pubmed.ncbi.nlm.nih.gov/33284783/)]
 34. Dutta R, Gkotsis G, Velupillai S, Bakolis I, Stewart R. Temporal and diurnal variation in social media posts to a suicide support forum. *BMC Psychiatry*. May 19, 2021;21(1):259. [FREE Full text] [doi: [10.1186/s12888-021-03268-1](https://doi.org/10.1186/s12888-021-03268-1)] [Medline: [34011346](https://pubmed.ncbi.nlm.nih.gov/34011346/)]
 35. Coppersmith G, Ngo K, Leary R, Wood A. Exploratory analysis of social media prior to a suicide attempt. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016. Presented at: The Third Workshop On Computational Linguistics and Clinical Psychology; June 16, 2016:106-117; San Diego, CA. URL: <https://aclanthology.org/W16-0311> [doi: [10.18653/v1/w16-0311](https://doi.org/10.18653/v1/w16-0311)]
 36. Feldhege J, Wolf M, Moessner M, Bauer S. Psycholinguistic changes in the communication of adolescent users in a suicidal ideation online community during the COVID-19 pandemic. *Eur Child Adolesc Psychiatry*. Jun 2023;32(6):975-985. [FREE Full text] [doi: [10.1007/s00787-022-02067-7](https://doi.org/10.1007/s00787-022-02067-7)] [Medline: [36018514](https://pubmed.ncbi.nlm.nih.gov/36018514/)]
 37. Biester L, Matton K, Rajendran J, Provost EM, Mihalcea R. Understanding the impact of COVID-19 on online mental health forums. *ACM Trans Manage Inf Syst*. Sep 08, 2021;12(4):1-28. [doi: [10.1145/3458770](https://doi.org/10.1145/3458770)]
 38. Zhang S, Liu M, Li Y, Chung JE. Teens' social media engagement during the COVID-19 pandemic: a time series examination of posting and emotion on Reddit. *Int J Environ Res Public Health*. Sep 25, 2021;18(19):10079. [FREE Full text] [doi: [10.3390/ijerph181910079](https://doi.org/10.3390/ijerph181910079)] [Medline: [34639381](https://pubmed.ncbi.nlm.nih.gov/34639381/)]
 39. Shing H, Nair S, Zirikly A, Friedenber M, Daumé IH, Resnikxpert, Crowdsourced, et al. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. New York, NY. Association for Computational Linguistics; 2018. Presented at: The Fifth Workshop on Computational Linguistics and Clinical Psychology; June 5, 2018:25-36; New Orleans, LA. URL: <https://aclanthology.org/W18-0603.pdf> [doi: [10.18653/v1/w18-0603](https://doi.org/10.18653/v1/w18-0603)]
 40. Rudd MD, Mandrusiak M, Joiner TE. The case against no-suicide contracts: the commitment to treatment statement as a practice alternative. *J Clin Psychol*. Feb 2006;62(2):243-251. [doi: [10.1002/jclp.20227](https://doi.org/10.1002/jclp.20227)] [Medline: [16342293](https://pubmed.ncbi.nlm.nih.gov/16342293/)]
 41. Bryan CJ, Butner JE, May AM, Rugo KF, Harris J, Oakey DN, et al. Nonlinear change processes and the emergence of suicidal behavior: a conceptual model based on the fluid vulnerability theory of suicide. *New Ideas Psychol*. Apr 2020;57:100758. [FREE Full text] [doi: [10.1016/j.newideapsych.2019.100758](https://doi.org/10.1016/j.newideapsych.2019.100758)] [Medline: [32123464](https://pubmed.ncbi.nlm.nih.gov/32123464/)]
 42. Nguena Nguéack HL, Pagé MG, Katz J, Choinière M, Vanasse A, Dorais M, et al. Trajectory modelling techniques useful to epidemiological research: a comparative narrative review of approaches. *Clin Epidemiol*. 2020;12:1205-1222. [FREE Full text] [doi: [10.2147/CLEP.S265287](https://doi.org/10.2147/CLEP.S265287)] [Medline: [33154677](https://pubmed.ncbi.nlm.nih.gov/33154677/)]
 43. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Annu Rev Psychol*. 2003;54:547-577. [doi: [10.1146/annurev.psych.54.101601.145041](https://doi.org/10.1146/annurev.psych.54.101601.145041)] [Medline: [12185209](https://pubmed.ncbi.nlm.nih.gov/12185209/)]
 44. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. *Biomed Inform Insights*. 2018;10:1178222618792860. [FREE Full text] [doi: [10.1177/1178222618792860](https://doi.org/10.1177/1178222618792860)] [Medline: [30158822](https://pubmed.ncbi.nlm.nih.gov/30158822/)]
 45. Coppersmith G, Leary R, Whyne E, Wood T. Quantifying suicidal ideation via language usage on social media. In: Joint Statistics Meetings Proceedings, Statistical Computing Section. 2015. Presented at: Joint Statistics Meetings, Statistical Computing Section; August 8-15, 2015:8-15; Seattle, WA. URL: <https://api.semanticscholar.org/CorpusID:54026822>
 46. Adil Jaafar E, Abdul-Salam Jasim H. A corpus-based stylistic analysis of online suicide notes retrieved from Reddit. *Cogent Arts & Humanities*. Mar 06, 2022;9(1):2047434. [doi: [10.1080/23311983.2022.2047434](https://doi.org/10.1080/23311983.2022.2047434)]
 47. Jaafar EA, Jaafar EA. Studies on linguistic stylistic analysis of suicide notes and suicidal thought posts. *IJRSSH*. Mar 31, 2022;12(01):100-124. [doi: [10.37648/ijrssh.v12i01.006](https://doi.org/10.37648/ijrssh.v12i01.006)]
 48. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. *Proc SIGCHI Conf Hum Factor Comput Syst*. May 2016;2016:2098-2110. [FREE Full text] [doi: [10.1145/2858036.2858207](https://doi.org/10.1145/2858036.2858207)] [Medline: [29082385](https://pubmed.ncbi.nlm.nih.gov/29082385/)]
 49. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related Twitter posts. *Crisis*. Sep 2017;38(5):319-329. [doi: [10.1027/0227-5910/a000443](https://doi.org/10.1027/0227-5910/a000443)] [Medline: [28228065](https://pubmed.ncbi.nlm.nih.gov/28228065/)]
 50. Reddit. Reddit API Access. Reddit. 2023. URL: <https://www.reddit.com/wiki/api/> [accessed 2023-03-26]
 51. Liu X, Huang J, Yu NX, Li Q, Zhu T. Mediation effect of suicide-related social media use behaviors on the association between suicidal ideation and suicide attempt: cross-sectional questionnaire study. *J Med Internet Res*. Apr 28, 2020;22(4):e14940. [FREE Full text] [doi: [10.2196/14940](https://doi.org/10.2196/14940)] [Medline: [32343249](https://pubmed.ncbi.nlm.nih.gov/32343249/)]

52. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC 2015. Austin, TX. University of Texas; 2015. URL: <http://hdl.handle.net/2152/31333>
53. Chen J, Hsieh G, Mahmud J, Nichols J. Understanding individuals' personal values from social media word use. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing;. New York, NY. Association for Computing Machinery; 2014. Presented at: CSCW'14: Computer Supported Cooperative Work; February 15-19, 2014:405-414; Baltimore, MD. [doi: [10.1145/2531602.2531608](https://doi.org/10.1145/2531602.2531608)]
54. Jones BL, Nagin DS. A note on a Stata plugin for estimating group-based trajectory models. *Sociological Methods & Research*. Sep 30, 2013;42(4):608-613. [doi: [10.1177/0049124113503141](https://doi.org/10.1177/0049124113503141)]
55. Kim SY, Cho S. Developmental trajectories of tobacco use and risk factors from adolescence to emerging young adulthood: a population-based panel study. *BMC Public Health*. Aug 29, 2022;22(1):1636. [FREE Full text] [doi: [10.1186/s12889-022-14070-3](https://doi.org/10.1186/s12889-022-14070-3)] [Medline: [36038859](https://pubmed.ncbi.nlm.nih.gov/36038859/)]
56. Baumgartner SE, Leydesdorff L. Group - based trajectory modeling (GBTM) of citations in scholarly literature: dynamic qualities of “transient” and “sticky knowledge claims”. *Asso for Info Science & Tech*. Nov 20, 2013;65(4):797-811. [doi: [10.1002/asi.23009](https://doi.org/10.1002/asi.23009)]
57. Frankfurt S, Frazier P, Syed M, Jung KR. Using group-based trajectory and growth mixture modeling to identify classes of change trajectories. *The Counseling Psychologist*. Aug 05, 2016;44(5):622-660. [doi: [10.1177/0011000016658097](https://doi.org/10.1177/0011000016658097)]
58. Nagin D. *Group-Based Modeling of Development*. Harvard, MA. Harvard University Press; 2005.
59. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. Dec 1974;19(6):716-723. [doi: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705)]
60. Schwarz G. Estimating the dimension of a model. *Ann Statist*. Mar 1, 1978;6(2):461-464. [doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)]
61. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods*. Jun 2012;17(2):228-243. [FREE Full text] [doi: [10.1037/a0027127](https://doi.org/10.1037/a0027127)] [Medline: [22309957](https://pubmed.ncbi.nlm.nih.gov/22309957/)]
62. Cavanaugh JE, Neath AA. The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Stats*. Mar 14, 2019;11(3):e1460. [doi: [10.1002/wics.1460](https://doi.org/10.1002/wics.1460)]
63. Neath AA, Cavanaugh JE. The Bayesian information criterion: background, derivation, and applications. *WIREs Computational Stats*. Dec 14, 2011;4(2):199-203. [doi: [10.1002/wics.199](https://doi.org/10.1002/wics.199)]
64. Mohammed E, Naugler C, Far B. Emerging business intelligence framework for a clinical laboratory through big data analytics. In: Tran QN, Arabnia H, Mohammed EA, editors. *Emerging Trends in Computational Biology, Bioinformatics and System Biology (1st edition)*. New York, NY. Elsevier/Morgan Kaufmann; 2015:577-602.
65. Chen B, Gong W, Lai AY, Sit SM, Ho S, Wang M, et al. Patterns of perceived harms and benefits of the COVID-19 outbreak in Hong Kong adults: a latent profile analysis. *Int J Environ Res Public Health*. Apr 05, 2022;19(7):4352. [FREE Full text] [doi: [10.3390/ijerph19074352](https://doi.org/10.3390/ijerph19074352)] [Medline: [35410033](https://pubmed.ncbi.nlm.nih.gov/35410033/)]
66. McElduff F, Cortina-Borja M, Chan S, Wade A. When t-tests or Wilcoxon-Mann-Whitney tests won't do. *Adv Physiol Educ*. Sep 2010;34(3):128-133. [FREE Full text] [doi: [10.1152/advan.00017.2010](https://doi.org/10.1152/advan.00017.2010)] [Medline: [20826766](https://pubmed.ncbi.nlm.nih.gov/20826766/)]
67. StataCorp. *Stata LASSO reference manual release 17*. StataCorp. 2021. URL: <https://www.stata.com/manuals/lasso.pdf> [accessed 2023-03-26]
68. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med*. Aug 01, 2020;180(8):1081-1089. [FREE Full text] [doi: [10.1001/jamainternmed.2020.2033](https://doi.org/10.1001/jamainternmed.2020.2033)] [Medline: [32396163](https://pubmed.ncbi.nlm.nih.gov/32396163/)]
69. Bruckman A. Studying the amateur artist: a perspective on disguising data collected in human subjects research on the internet. *Ethics Inf Technol*. 2002;4:217-231. [FREE Full text] [doi: [10.1023/A:1021316409277](https://doi.org/10.1023/A:1021316409277)]
70. Statista. *Regional distribution of desktop traffic to Reddit*. Statista. 2022. URL: <https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/> [accessed 2023-03-26]
71. Centers for Disease Control and Prevention (CDC). *CDC Museum COVID-19 Timeline*. CDC. 2023. URL: <https://www.cdc.gov/museum/timeline/covid19.html> [accessed 2023-03-26]
72. Oh S, Syn SY. Motivations for sharing information and social support in social media: a comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr. *Asso for Info Science & Tech*. Apr 21, 2015;66(10):2045-2060. [doi: [10.1002/asi.23320](https://doi.org/10.1002/asi.23320)]
73. Park C, Shim M, Eo S, Lee S, Seo J, Moon H, et al. Empirical analysis of parallel corpora and in-depth analysis using LIWC. *Applied Sciences*. May 30, 2022;12(11):5545. [FREE Full text] [doi: [10.3390/app12115545](https://doi.org/10.3390/app12115545)]
74. Hawkins KA, Hames JL, Ribeiro JD, Silva C, Joiner TE, Cogle JR. An examination of the relationship between anger and suicide risk through the lens of the interpersonal theory of suicide. *J Psychiatr Res*. Mar 2014;50:59-65. [doi: [10.1016/j.jpsychires.2013.12.005](https://doi.org/10.1016/j.jpsychires.2013.12.005)] [Medline: [24388767](https://pubmed.ncbi.nlm.nih.gov/24388767/)]
75. Lightman E, McCarthy P, Dufty D, McNamara D. Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2007. Presented at: Annual Meeting of the Cognitive Science Society; August 1-4, 2007:1-4; Nashville, TN. URL: <https://escholarship.org/uc/item/0dh4553j>

76. Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A. Suicide note classification using natural language processing: a content analysis. *Biomed Inform Insights*. Aug 04, 2010;2010(3):19-28. [FREE Full text] [doi: [10.4137/bii.s4706](https://doi.org/10.4137/bii.s4706)] [Medline: [21643548](https://pubmed.ncbi.nlm.nih.gov/21643548/)]
77. Farrell DM. Suicide among gifted students. *Roeper Review*. Mar 1989;11(3):134-139. [doi: [10.1080/02783198909553189](https://doi.org/10.1080/02783198909553189)]
78. Hull-Blanks EE, Kerr BA, Robinson Kurpius SE. Risk factors of suicidal ideations and attempts in talented, at-risk girls. *Suicide Life Threat Behav*. 2004;34(3):267-276. [doi: [10.1521/suli.34.3.267.42782](https://doi.org/10.1521/suli.34.3.267.42782)] [Medline: [15385181](https://pubmed.ncbi.nlm.nih.gov/15385181/)]
79. Ware ZR. Adolescent Suicide Through the Lens of 13 Reasons Why: A Corpus Based Study. Doctor of Philosophy Dissertation. Oregon State University. 2022. URL: https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/41687r69m [accessed 2024-07-15]
80. Ornell F, Schuch JB, Sordi AO, Kessler FHP. "Pandemic fear" and COVID-19: mental health burden and strategies. *Braz J Psychiatry*. 2020;42(3):232-235. [FREE Full text] [doi: [10.1590/1516-4446-2020-0008](https://doi.org/10.1590/1516-4446-2020-0008)] [Medline: [32267343](https://pubmed.ncbi.nlm.nih.gov/32267343/)]
81. Mueller AS, Abrutyn S. Suicidal disclosures among friends: using social network data to understand suicide contagion. *J Health Soc Behav*. Mar 2015;56(1):131-148. [FREE Full text] [doi: [10.1177/0022146514568793](https://doi.org/10.1177/0022146514568793)] [Medline: [25722129](https://pubmed.ncbi.nlm.nih.gov/25722129/)]
82. Nan J, Salina N, Chong ST, Jiang H. Trajectory of suicidal ideation among medical students during the COVID-19 pandemic: the role of childhood trauma. *BMC Psychiatry*. Feb 06, 2023;23(1):90. [FREE Full text] [doi: [10.1186/s12888-023-04582-6](https://doi.org/10.1186/s12888-023-04582-6)] [Medline: [36747156](https://pubmed.ncbi.nlm.nih.gov/36747156/)]
83. Huang S, Wang D, Zhao J, Chen H, Ma Z, Pan Y, et al. Changes in suicidal ideation and related influential factors in college students during the COVID-19 lockdown in China. *Psychiatry Res*. Aug 2022;314:114653. [FREE Full text] [doi: [10.1016/j.psychres.2022.114653](https://doi.org/10.1016/j.psychres.2022.114653)] [Medline: [35671561](https://pubmed.ncbi.nlm.nih.gov/35671561/)]
84. Madsen T, van Spijker B, Karstoft K, Nordentoft M, Kerkhof AJ. Trajectories of suicidal ideation in people seeking web-based help for suicidality: secondary analysis of a Dutch randomized controlled trial. *J Med Internet Res*. Jun 30, 2016;18(6):e178. [FREE Full text] [doi: [10.2196/jmir.5904](https://doi.org/10.2196/jmir.5904)] [Medline: [27363482](https://pubmed.ncbi.nlm.nih.gov/27363482/)]

Abbreviations

- AIC:** Akaike information criterion
- APP:** average posterior probability
- BIC:** Bayesian information criterion
- CNORM:** censored normal
- GBTM:** group-based trajectory modeling
- LIWC:** Linguistic Inquiry and Word Count
- OR:** odds ratio
- ZIP:** zero-inflated Poisson

Edited by A Mavragani; submitted 11.05.23; peer-reviewed by E Jaafar, S Li; comments to author 15.03.24; revised version received 05.04.24; accepted 18.04.24; published 08.08.24

Please cite as:

Yan Y, Li J, Liu X, Li Q, Yu NX

Identifying Reddit Users at a High Risk of Suicide and Their Linguistic Features During the COVID-19 Pandemic: Growth-Based Trajectory Model

J Med Internet Res 2024;26:e48907

URL: <https://www.jmir.org/2024/1/e48907>

doi: [10.2196/48907](https://doi.org/10.2196/48907)

PMID: [39115925](https://pubmed.ncbi.nlm.nih.gov/39115925/)

©Yifei Yan, Jun Li, Xingyun Liu, Qing Li, Nancy Xiaonan Yu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 08.08.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.