



Online disclosure of suicide method: What can online posts tell us about suicidal planning?

Brooke A. Ammerman^{a,*} , Kenneth McClure^a , Keyne C. Law^b , Caitlin M. O'Loughlin^a ,
Ross Jacobucci^a

^a University of Notre Dame, Department of Psychology, Notre Dame, IN, USA

^b Seattle Pacific University, Department of Psychology, Seattle, WA, USA

ABSTRACT

Rates of self-disclosure of, and subsequently formal help-seeking for, suicidal thoughts and behaviors (STBs) remains low. However, online discussions of STBs have become increasingly prevalent, creating potential opportunities to identify those at potentially elevated suicide risk (i.e., engaging in suicidal planning). This study used linguistic analysis to examine online STB discussions to identify posts that reference a suicide method, as well as the lethality and timeframe of the identified suicide method. A corpus of 3974 posts from the online forum Reddit SuicideWatch were extracted for data analysis. Approximately one-fifth of analyzed posts mentioned a suicide method; over 35% were classified as severe risk with regard to lethality and almost 70% discussed a suicide method yet to occur. Analyses demonstrated that three categories of words – Substances, Physical, and Past Focus – were associated with increased odds of the presence of a suicide method in a post, whereas five different word categories were associated with a decrease in odds of method presence; most notably was the use of Social words. Together, results support the utility of this approach in detecting discussions of a suicide method. However, they may be less fruitful in distinguishing planning severity among posts including method discussions.

Suicide is a global health concern (World Health Organization, 2021) and represents a leading cause of death across demographics in the U.S. (Center for Disease Control and Prevention, 2021). Furthermore, over 5% of U.S. adults have thought about suicide in the past year, and suicide rates have continued to rise over the prior decade (Substance Abuse and Mental Health Services Administration, 2024). Despite the prevalence, only one-third of those struggling with suicidal thoughts and behaviors (STBs) engaging with mental health care (Hom et al., 2015). Self-disclosure of STBs is a key help-seeking behavior (Sheehan et al., 2019), facilitating both treatment engagement and accurate risk assessment across the STB risk continuum (i.e., presence suicidal thoughts or planning, or imminent risk for suicide attempt). However, STBs are a stigmatizing experience (Ammerman et al., 2022; Bond et al., 2022), which often results in low rates of self-disclosure and subsequent treatment engagement (Fulginiti and Frey, 2019; Monteith et al., 2020). Consequently, it is critical to examine environments, such as online forums and social media, that may promote STB self-disclosure and increase risk detection. Over 70% of Americans use social media to connect with others and share personal information (Pew Research Center, 2021). Given the balance between social connection and identity concealment (e.g., Andalibi et al., 2017; Moors and Webber, 2013), social media communities have also become primary avenues for

disclosing mental health problems (Berry et al., 2017; Naslund et al., 2016) and discussing STBs (O'Dea et al., 2017; O'Dea et al., 2017).

Several benefits of online community engagement surrounding mental health have been identified. Beyond facilitating self-disclosures, online discussions of stigmatizing topics (e.g., mental health) help to decrease stigma and improve social wellness, as well as facilitate greater feelings of social support and connectedness (Brown et al., 2021). As online discussions of mental health topics, and specifically STBs, have become increasingly prevalent, there is an opportunity to learn about these experiences in vivo. This could ultimately inform our ability to identify those at high suicide risk. Yet, few studies have focused on examining these discussions: among all published studies of mental health-focused social media discussions, only 17% of the studies reported findings specific to STBs (Wongkoblapp et al., 2017). Moreover, much of the work in this area has been descriptive in nature (e.g., examining phrases that commonly occur in posts to an STB-focused online community page; Jiang et al., 2020) or aimed to identify the presence of suicidality in posts (i.e., reports of suicidal thinking; Guan et al., 2015; Shang et al., 2018), rather than risk severity.

Linguistic features of written communication may be informative in determining suicide risk severity (e.g., expressions of passive suicidal ideation vs. imminent suicidal behavior), offering a framework for

* Corresponding author. 390 Corbett Family Hall Notre Dame, IN, 46656, USA.
E-mail address: baammerman@uisc.edu (B.A. Ammerman).

studying online STB discussions. For example, among Twitter posts mentioning suicide ideation, word count, number of personal pronouns used, and frequent references to death differentiated between posts that were deemed “safe to ignore” and those that were “strongly concerning” (O’Dea et al., 2017). This approach has also been applied in two case studies of online postings leading up to adolescents’ death by suicide, finding changes in emotional expression leading up to the individual’s deaths (Gunn and Lester, 2012). Together this work highlights the potential utility of linguistic analysis for detecting STBs in online discussions.

1. Current study

The overarching goal of the current study was to determine whether linguistic analysis is useful in detecting severity of STBs in online STB discussions. While much of the existing work has solely examined the presence vs. absence of suicidal ideation (Guan et al., 2015), STBs span a continuum, ranging from passive and active suicidal ideation to suicidal planning and preparatory behaviors, ultimately progressing to suicide attempts; each step on the continuum represents increased risk of attempting suicide. For example, while less than one-third of those who think about suicide will go on to attempt suicide, 56% of those who plan for suicide will make an attempt (Kessler et al., 1999; Nock et al., 2008). Better detection of STBs, such as suicidal planning, may serve to prevent suicide attempts and even death by suicide. Moreover, the detection and prevention of STBs, at any point along the continuum, is important in its own right: experiencing STBs is linked to later adverse outcomes, including persistent mental (Mortier et al., 2018) and physical (Herba et al., 2007) health problems, relationship dysfunction (Goldman-Mellor et al., 2014), and unemployment (Ettner et al., 1997). Thus, the specific aim of the current study was to utilize linguistic analysis to: (a) identify online STB discussion posts that reference a suicide method, a key component of a suicide plan (e.g., Linthicum and Ribeiro, 2022); and (b) among those posts that include a suicide method, determine the severity of suicidal planning, as defined by the lethality of the identified suicide method and the timeframe, or immediacy, of post information. Given the limited research in this area, particularly with regard to suicidal planning, no specific hypotheses were made.

2. Method

2.1. Data extraction and coding procedures

Posts found on the social media platform Reddit, which is a forum where content is socially curated and promoted by site members, were utilized for the current study given its popularity (Oladipo, 2024), growing number of users (i.e., over 500 million) (Statistica, 2024), and number of forums (or subreddits) dedicated to mental health discussions (Reddit, n.d.), including a forum specific to the discussion of STBs: r/SuicideWatch. A corpus of publicly available posts from r/SuicideWatch were extracted for this study: 3974 posts that were in English and legible (i.e., did not consist solely of symbols, included coherent phrases) were included in the study. In line with study aims, a coding manual was developed (see below) and trained research assistants coded each post; approximately 30% of posts were coded by two raters. For posts where there was rater disagreement on presence of method, a third party weighed in. To assess interrater reliability of coding, we examined absolute agreement using two-way random intraclass correlations (ICCs). This research was classified as non-human subjects research by the authors’ institutional review board; to support the ethical use of this data, data was stored in a secure location, available only to members of the research team on a need-to-know basis, and no specific posts from the corpus are included in this publication (Ford et al., 2021).

2.2. Data coding

Suicide Method. All posts were coded for the presence of a suicide method. For the purpose of this study, the presence of any behavior or action associated with the intent to attempt suicide, regardless of the timeframe of the behavior (e.g., “I know I could just jump off a bridge”; “I just took some pills”). Information pertaining to the plan could be found in the title or the body of the post. Posts were coded based on the presence vs. absence of a suicide method (ICC = 0.84).

Suicide Method Lethality. Among posts where a suicide method was discussed, content was also coded for the lethality of the suicide method. Based on text included in the post title or body, the suicide method lethality was coded in accordance with the Lifetime Suicide Attempts Self-Injury Interview (Linehan et al., 2006), which entails a 1 (Very Low) to 6 (Severe) rating of lethality, wherein higher numbers represent greater lethality. See Supplemental Table 1 for detailed description (ICC = 0.83).

Suicide Method Timing. Among posts where a suicide method was discussed, content was also coded for the imminence, or timing, of the discussed suicide method. Based on text included in the post title or body, the suicide method timing was coded as follows: 1 = discussed a suicide method that occurred more than 24 h ago (e.g., “When I was younger ...”; 2 = discussed a suicide method that occurred within the past 24 h (e.g., “Last night I took a handful of pills.”); 3 = discussed a suicide method that has yet to happen (e.g., “I know that my grandpa has a gun”); or 4 = discussed any action, timing unknown, associated with a suicide method (e.g., “I was holding a knife to my arm”). Timing of the suicide method was interpreted such that a higher code was associated with higher suicide risk (i.e., a post coded as 1 (i.e., discussed a suicide method that occurred more than 24 h ago) was interpreted as lower risk than a post coded as a 3 (i.e., discussed a suicide method that has yet to happen)) (ICC = 0.78). During data analysis, categories 1 and 2 were collapsed into a single category of discussion of prior use of a suicide method.

2.3. Data analysis

Text from the body of the posts were utilized in analysis. Posts with fewer than 10 words were removed from analysis (N = 140). One post without a presence of method rating was also removed resulting in a final dataset of N = 3834 (736 with method present; 19.2%). Method timing and lethality analyses only included individuals with a method present, resulting in N = 727 and N = 695, respectively.

Text responses were tokenized using the tidytext package (Silge and Robinson, 2016) and automated spelling correction was conducted with the hunspell package (Ooms, 2020) in R (R Core Team, 2021).

The Linguistic Inquiry and Word Count (LIWC; Boyd et al., 2022) software was used to analyze text. The LIWC-22 dictionary consists of over 12,000 words organized in many subdictionaries designed to assess psychosocial contexts (for a full overview of LIWC-22, please see Boyd et al., 2022). As an example, LIWC-22 includes four dictionaries – Drives, Cognition, Affect, and Social Processes, that are categorized as Psychological Processes; these four dictionaries are made up of 32 subdictionaries, which, in total, include 8639 words. LIWC output in the current study can be understood as the percentage of words in each Reddit post associated with the corresponding subdictionary. Regularized Bayesian regression using horseshoe priors (Carvalho et al., 2009) in brms (Bürkner, 2017) was used to identify a sparse set of text variables predictive of suicide method, method timing, and method lethality. This approach was selected given the imbalance in the categorical outcome. Further, regularization methods reduce the potential for overfitting by penalizing small effects towards zero (e.g., Hastie, 2020; McNeish, 2015). The horseshoe priors used in the Bayesian regression models emphasize sparse solutions and function similar to traditional regularization strategies (Carvalho et al., 2009).

Logistic regression using 6 chains each containing 2500 samples (500

burnin) was used to examine predictors of suicide method presence; $\text{adapt_delta} = 0.99$ was specified to prevent divergence given the horseshoe prior. Cumulative probit regressions using 6 chains each containing 3000 samples (1500 burnin) were conducted for the ordinal timing and lethality outcomes; $\text{adapt_delta} = 0.99$ was specified. Multinomial regression with a logit link using 6 chains of 2500 samples (1000 burnin) and $\text{adapt_delta} = 0.95$ was used to examine method timing; posts discussing past suicidal behaviors (coded 1 or 2)¹ were combined into a single reference group and contrasted with future methods (coded 3) and methods with unknown timing (coded 4). Analyses for method lethality and timing only included posts with a method present and scores on the outcome variables. All chains were evaluated for convergence using \hat{R} and effective sample size (ESS; Vehtari et al., 2021). \hat{R} close to 1.00 and ESS larger than 400 are desired for inference (Gelman et al., 2015).

3. Results

3.1. Preliminary analysis

Of the post corpus, 736 posts (19.2%) were identified as discussing a suicide method. Of posts with a suicide method present, on average, posts were rated as 4.30 on lethality (SD = 1.57); 6.2% of posts were identified as very low risk, 4.8% as low risk, 28.6% as moderate risk, 11.4% as high risk, 13.8% as very high risk, and 35.2% as severe risk. With regard to suicide method timing, 16.7% mentioned a method more than 24 h ago, 7.5% mentioned a method in the past 24 h, 69.3% mentioned a method that had yet to happen; and 6.5% mentioned action associated with the method.

3.2. Primary analysis

Regression analyses demonstrated chain convergence for all parameters across analyses (Rhat = 1.00, ESS >1701). Odds ratios (OR) and associated credible intervals (CI) for non-zero effects are provided below. See Table 1 for full results and Supplemental Table 2 for LIWC-22

Table 1
Regularized regression results.

<i>Method Presence (N = 3834)^a</i>				
LIWC Predictor	Est (err)	95% CI	OR	OR 95% CI
Substance	0.73 (0.12)	[0.51, 0.97]	2.08	[1.67, 2.64]
Physical	0.14 (0.02)	[0.09, 0.19]	1.15	[1.10, 1.20]
Past Focus	0.05 (0.02)	[0.01, 0.09]	1.05	[1.01, 1.10]
Social	-0.08 (0.03)	[-0.14, -0.03]	0.92	[0.87, 0.97]
Adjectives	-0.06 (0.02)	[-0.10, -0.01]	0.94	[0.90, 0.99]
Present Focus	-0.05 (0.02)	[-0.10, -0.00]	0.95	[0.91, 1.00]
Allure	-0.04 (0.02)	[-0.07, -0.01]	0.96	[0.93, 0.99]
Clout	-0.02 (0.01)	[-0.03, -0.00]	0.98	[0.97, 0.99]
<i>Method Lethality (N = 727)^b</i>				
LIWC Predictor	Est (err)	95% CI	OR	OR 95% CI
Articles	0.08 (0.03)	[0.00, 0.13]	1.08	[1.00, 1.14]
<i>Method Timing (N = 695)^c</i>				
LIWC Predictor	Est (err)	95% CI	OR	OR 95% CI
Past Focus (FM)	-0.34 (0.04)	[-0.42, -0.25]	0.71	[0.66, 0.78]

Note: Only the results for predictors with non-null effects after regularization are reported for each set of analyses.

^a Bernoulli link.

^b cumulative probit link.

^c multinomial link with past methods as reference category; CI = credible interval, OR = odds ratio; FM = future suicide method group; Rhat = 1.00 and ESS >1701 for all analyses.

¹ Results did not differ when examining categories 1 and 2 separately; combined results are presented for ease of interpretation.

dictionary overview.

3.3. Method presence

Only 8 text variables exhibited nonzero effects after regularization (see Fig. 1). Increases in words from the Substances (OR = 2.07, 95% CI [1.67, 2.63]), Physical (OR = 1.15 95% CI [1.10, 1.20]), and Past Focus (OR = 1.05 95% CI [1.01, 1.10]) dictionaries were associated with corresponding increases in the odds of a present suicide method in the post. Conversely, words in the Social (OR = 0.92 95% CI [0.87, 0.97]), Adjectives (OR = 0.94, 95% CI [0.90, 0.99]), Present Focus (OR = 0.95, 95% CI [0.91, 1.00]), and Allure (OR = 0.96 95% CI [0.93, 0.99]) dictionaries exhibited a small decrease in odds of method presence. Clout (OR = 0.98 95% CI [0.97, 0.99]; i.e., language of leadership or status; Boyd et al., 2022) also showed a small negative relationship with suicide method.

3.4. Method lethality

Among respondents with a method present, only an increase in the frequency of Articles (OR = 1.08, 95% CI [1.00, 1.14]) was associated with increased lethality.

3.5. Method timing

Among respondents with a method present, Past Focus words were associated with decreased odds for future method (OR = 0.71; 95% CI [0.66, 0.78]). No text variables distinguished between discussions of past methods and posts with unknown timing. Similarly, no variables distinguished between future methods and posts with unknown timing.

4. Discussion

This current study used linguistic analysis to identify the presence of suicide method, and the severity of that method (i.e., lethality and timing) from postings on a popular social media forum for STB discussions, Reddit SuicideWatch. Results support the utility of this approach in detecting discussions of a suicide method, however, they may be less fruitful in distinguishing planning severity among posts including method discussions.

Approximately one-fifth of analyzed posts mentioned a suicide method. Analyses demonstrated that three categories of words – Substances, Physical, and Past Focus – were associated with increased odds of the presence of a suicide method in the post. These findings are in line with prior work that highlight a relationship between both acute (Kaplan et al., 2013) and chronic (Schneider, 2009) substance use and suicide risk, as well as medical or physical conditions (Druss and Pincus, 2000). It is possible, however, that these word categories may also be capturing the discussion of specific suicide methods, such as overdosing with substance or prescription medication, which is a common suicide attempt method especially among young people (Farah et al., 2023) and have overlap with the linguistic dictionaries (i.e., medic*). It was also found that suicide method presence was associated with increased use of Past Focus words, such as “was”, “had”, “were”, and “been”. It’s possible this association can be explained by the timing of the discussed suicide method (i.e., events that occurred in the past); however, the majority of posts were future oriented. Rather, this finding may reflect the strong association between a ruminative cognitive style and suicide risk (Rogers and Joiner, 2017), suggesting that individuals whose posts included a suicide method were more likely to also be focusing or reflecting on past negative events.

Alternatively, five different word categories were associated with a decrease in odds of method presence. Most notably was the use of Social words (i.e., “we”, “he”, “she”). Those at elevated suicide risk have been found to be particularly sensitive to interpersonal stressors (Herres et al., 2016), with such events found to preceding suicidal crises (Bagge et al.,

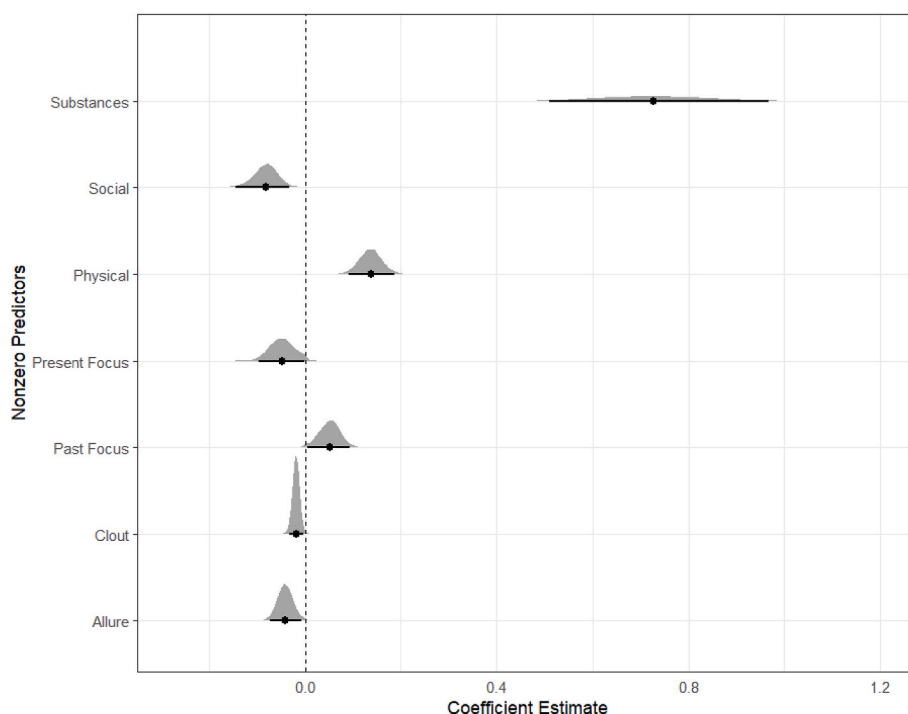


Fig. 1. Posterior Distributions of Logistic Regression Coefficients: Method Presence

Note: Horizontal lines below each posterior represent the 95% Credible Interval; Rhat = 1.00 and ESS >3798 for all predictors.

2013); however, potentially in line with current findings, social connection and social support are key protective factors for suicide risk (i.e., Kleiman and Liu, 2013). Consequently, it may be that greater inclusion of Social words in a post is indicative of greater connection, or reasons for living, and thus, reduces risk of suicide method presence. It is possible that the words falling within the dictionaries of Allure and Clout are also representations of positive social processes, and thus garner a similar protective effect.

Posts that utilized more Present Focus words and Adjectives were also less likely to have a suicide method present. Although speculative, it is possible that these individuals were driven to post of the forum as a means of “venting” about a current problem or recent events, a common coping strategy among those at risk for suicide (Konkan et al., 2014); while such events may have contributed to suicide risk (i.e., elevations in suicidal thinking), the act of expressing their feelings (i.e., expressive writing) may serve as a coping strategy (Travagin et al., 2015), preventing escalation from suicidal thinking to the discussion of a method. For example, prior research has found that “off my chest” is a common phrase used among users on the SuicideWatch forum (Monselise and Yang, 2022), highlighting the potential benefits of such online communities in management suicide risk.

On average, discussed suicide methods were rated high to very high in lethality based on the LSASI ratings of lethality, with over one-third of posts being identified as severe risk. Similarly, the majority of methods discussed had yet to happen, suggesting the individual had a plan for how they may attempt suicide in the, potentially near, future, rather than discussing past suicidal planning or behavior. Together, findings may indicate that a subset of Reddit SuicideWatch users are at high current risk for suicide as their posts are detailing a highly lethal plan for future suicidal behavior. With regard to linguistic analysis, an increase in the use Articles (i.e., “an”, “the”, “a lot”) were associated with greater lethality of the discussed method. This is somewhat in line with prior analyses of online posts finding that those including suicide-related phrases (i.e., suicidal, not worth living) used more quantifiers, prepositions, and adverbs (O’Dea et al., 2017). Our results also demonstrated that greater use of Past Focus words were associated with a reduced

likelihood of discussing a future method. Although initially appearing to be in contrast with earlier results (i.e., Past Focus word use increasing risk of suicide method presence), it is possible that this findings is reflective of the inverse relationship between the language required to discuss a future method (i.e., “will”, “going to”, “may”) and Past Focus words (i.e., “had”, “were”, “been”).

It is important to note limitations associated with the current study. The quality of results for any linguistic analysis depends on the quality of the lexicon used. In this study, we compared words found in online posts to words in the LIWC dictionaries. Although it is comprehensive, LIWC is designed to assess psychosocial contexts and not explicitly designed to be used for suicide-related content, and thus doesn’t contain keywords that may be relevant to suicide risk. Although a Chinese suicide dictionary has been developed for studying and identifying suicide risk in Chinese social media platforms (Lv et al., 2015), to the best of our knowledge, there is currently no standard English suicide dictionary is readily available or has been empirically supported for use in linguistic analyses. Instead, most studies use general-purpose sentiment lexicons to study suicide risk and suicide-related content, which are limited in their ability to capture and categorize words as they relate to suicide (Bittar et al., 2021). Therefore, an essential next step in using linguistic analysis in suicide research would be to compile suicide-related words and phrases that have been identified by past studies (i.e. O’Dea et al., 2017) and convene a panel of individuals with lived experiences of suicide, clinical experts, and research experts to systematically develop an official, standardized lexicon for future research on suicide. The posts analyzed in the current study also offer only a snapshot of the individual’s state at a particular point in time. We were, therefore, unable to capture the greater affective or cognitive context surrounding the post (i.e., thoughts not expressed in writing), examine the individual’s experience prior to or after posting, or link post content to STB-related outcomes.

The complexity in how an individual can express or discuss STBs online makes both research and potential prevention efforts related to suicide-related content on social media challenging. Our study provides one piece of the puzzle to address this challenge. Foremost, results

highlight that a subset of Reddit users may be characterized as at increased risk for attempting suicide, based on the discussion of high lethality, future oriented suicide plans, suggesting an opportunity for prevention or intervention efforts. We also identified several linguistic patterns that seem consistent amongst posts discussing a suicide method, including the use of words included in the Substances and Physical dictionaries, as well as words of Past Focus. These findings provide insight into how individuals discuss suicide methods on Reddit. They may also help inform social-media focused suicide prevention efforts by identifying users discussing high-risk content, thus allowing for the promotion of relevant resources (e.g., crisis resources, chat-bots, etc. As responses to, or comment on, posts) or more direct intervention by site moderators.

CRedit authorship contribution statement

Brooke A. Ammerman: Conceptualization, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Keneth McClure:** Formal analysis, Investigation, Visualization, Writing – original draft. **Keyne C. Law:** Conceptualization, Writing – original draft, Writing – review & editing. **Caitlin M. O’Loughlin:** Writing – original draft, Writing – review & editing. **Ross Jacobucci:** Data curation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors have no conflicts of interest to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpsychires.2024.11.048>.

References

- Ammerman, B.A., Piccirillo, M.L., O’Loughlin, C.M., Carter, S.P., Matarazzo, B., May, A. M., 2022. The role of suicide stigma in self-disclosure among civilian and veteran populations. *Psychiatr. Res.* 309, 114408.
- Andalibi, N., Ozturk, P., Forte, A., 2017. Sensitive self-disclosures, responses, and social support on instagram: the case of # depression. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 1485–1500.
- Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G., Bucci, S., 2017. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. *J. Med. Internet Res.* 19 (4), e6173.
- Bagge, C.L., Glenn, C.R., Lee, H.J., 2013. Quantifying the impact of recent negative life events on suicide attempts. *J. Abnorm. Psychol.* 122 (2), 359.
- Bittar, A., Velupillai, S., Roberts, A., Dutta, R., 2021. Using general-purpose sentiment lexicons for suicide risk assessment in electronic health records: corpus-based analysis. *JMIR Med. Inf.* 9 (4), e22397.
- Bond, A.E., Bandel, S.L., Rodriguez, T.R., Anestis, J.C., Anestis, M.D., 2022. Mental health treatment seeking and history of suicidal thoughts among suicide decedents by mechanism, 2003–2018. *JAMA Netw. Open* 5 (3), e222101.
- Boyd, R.L., Ashokkumar, A., Seraj, S., Pennebaker, J.W., 2022. The Development and Psychometric Properties of LIWC-22. University of Texas at Austin, Austin, TX. <https://www.liwc.app>.
- Brown, G., Rathbone, A.L., Prescott, J., 2021. Social media use for supporting mental health (SMILE). *Ment. Health Rev.* 26 (3), 279–297.
- Bürkner, P., 2017. Brms: an R package for bayesian multilevel models using stan. *J. Stat. Software* 80 (1), 1–28. <https://doi.org/10.18637/jss.v080.i01>.
- Carvalho, C.M., Polson, N.G., Scott, J.G., 2009. Handling sparsity via the horseshoe. Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, pp. 73–80. In: <https://proceedings.mlr.press/v5/carvalho09a.html>.
- Center for Disease Control and Prevention, 2021. Data and statistics: Suicide. U.S. Department of Health and Human Services. Retrieved from: <https://www.cdc.gov/suicide/facts/data.html>.
- Druss, B., Pincus, H., 2000. Suicidal ideation and suicide attempts in general medical illnesses. *Arch. Intern. Med.* 160 (10), 1522–1526.
- Ettner, S.L., Frank, R.G., Kessler, R.C., 1997. The impact of psychiatric disorders on labor market outcomes. *ILR Rev.* 51 (1), 64–81.
- Farah, R., Rege, S.V., Cole, R.J., Holstege, C.P., 2023. Suspected suicide attempts bySelf-poisoning among persons aged 10–19 Years during the COVID-19 pandemic—United States, 2020–2022. *MMWR (Morb. Mortal. Wkly. Rep.)* 72 (16), 426.
- Ford, E., Shepherd, S., Jones, K., Hassan, L., 2021. Toward an ethical framework for the text mining of social media for health research: a systematic review. *Front. Digit. Health* 2, 592237.
- Fulginiti, A., Frey, L.M., 2019. Exploring suicide-related disclosure motivation and the impact on mechanisms linked to suicide. *Death Stud.* 43 (9), 562–569.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2015. *Bayesian Data Analysis*, third ed. Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>.
- Goldman-Mellor, S.J., Caspi, A., Harrington, H., Hogan, S., Nada-Raja, S., Poulton, R., Moffitt, T.E., 2014. Suicide attempt in young people: a signal for long-term health care and social needs. *JAMA Psychiatr.* 71 (2), 119–127.
- Guan, L., Hao, B., Cheng, Q., Yip, P.S., Zhu, T., 2015. Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. *JMIR Ment. Health* 2 (2), e17.
- Gunn, J.F., Lester, D., 2012. Twitter postings and suicide: an analysis of the postings of a fatal suicide in the 24 hours prior to death. *Suicidol.* 17 (3).
- Hastie, T., 2020. Ridge regularization: an essential concept in DataScience. *Technometrics* 62 (4), 426–433.
- Herba, C.M., Ferdinand, R.F., Verhulst, F.C., 2007. Long-term associations of childhood suicide ideation. *J. Am. Acad. Child Adolesc. Psychiatr.* 46 (11), 1473–1481.
- Herres, J., Ewing, E.S.K., Kobak, R., 2016. Emotional reactivity to negative adult and peer events and the maintenance of adolescent depressive symptoms: a daily diary design. *J. Abnorm. Child Psychol.* 44, 471–481.
- Hom, M.A., Stanley, I.H., Joiner Jr, T.E., 2015. Evaluating factors and interventions that influence help-seeking and mental health service utilization among suicidal individuals: a review of the literature. *Clin. Psychol. Rev.* 40, 28–39.
- Jiang, M., Ammerman, B.A., Zeng, Q., Jacobucci, R., Brodersen, A., 2020. Phrase-level pairwise topic modeling to uncover helpful peer responses to online suicidal crises. *Humanit. Soc. Sci. Commun.* 7 (1), 1–13.
- Kaplan, M.S., McFarland, B.H., Huguet, N., Conner, K., Caetano, R., Giesbrecht, N., Nolte, K.B., 2013. Acute alcohol intoxication and suicide: a gender-stratified analysis of the national violent death reporting system. *Inj. Prev.* 19 (1), 38–43.
- Kessler, R.C., Borges, G., Walters, E.E., 1999. Prevalence of and risk factors for lifetime suicide attempts in the national comorbidity survey. *Arch. Gen. Psychiatr.* 56, 617–626.
- Kleiman, E.M., Liu, R.T., 2013. Social support as a protective factor in suicide: findingsfrom two nationally representative samples. *J. Affect. Disord.* 150 (2), 540–545.
- Konkan, R., Erkuş, G.H., Güçlü, O., Şenormanci, Ö., Aydin, E., Ülgen, M.C., Sungur, M.Z., 2014. Coping strategies in patients who had suicide attempts. *Nöro Psikiyatri Arşivi* 51 (1), 46.
- Linehan, M.M., Comtois, K.A., Brown, M.Z., Heard, H.L., Wagner, A., 2006. Suicide attempt Self-Injury Interview (SASH): development, reliability, and validity of a scale to assess suicide attempts and intentional self-injury. *Psychol. Assess.* 18 (3), 303.
- Linthicum, K.P., Ribeiro, J.D., 2022. Suicide plan prevalence, recurrence, and longitudinal association with nonfatal suicide attempt. *Suicide and Life-Threatening Behav.* 52 (6), 1062–1073.
- Lv, M., Li, A., Liu, T., Zhu, T., 2015. Creating a Chinese suicide dictionary for identifying suicide risk on social media. *PeerJ* 3, e1455.
- McNeish, D.M., 2015. Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. *Multivariate Behav. Res.* 50 (5), 471–484.
- Monselise, M., Yang, C.C., 2022. I’m always in so much pain and no one will understand”-Detecting Patterns in Suicidal Ideation on Reddit. In: Companion Proceedings of the Web Conference 2022, pp. 686–691.
- Monteith, L.L., Smith, N.B., Holliday, R., Holliman, B.A.D., LoFaro, C.T., Mohatt, N.V., 2020. “We’re afraid to say suicide”: stigma as a barrier to implementing a community-based suicide prevention program for rural veterans. *J. Nerv. Ment. Dis.* 208 (5), 371–376.
- Moors, R., Webber, R., 2013. The dance of disclosure: online self-disclosure of sexual assault. *Qual. Soc. Work* 12 (6), 799–815.
- Mortier, P., Auerbach, R.P., Alonso, J., Bantjes, J., Benjet, C., Cuijpers, P., et al., 2018. Suicidal thoughts and behaviors among first-year college students: results from the WMH-ICS project. *J. Am. Acad. Child Adolesc. Psychiatr.* 57 (4), 263–273.
- Naslund, J.A., Aschbrenner, K.A., Marsch, L.A., Bartels, S.J., 2016. The future of mental health care: peer-to-peer support and social media. *Epidemiol. Psychiatr. Sci.* 25 (2), 113–122.
- Nock, M.K., Borges, G., Bromet, E.J., Alonso, J., Angermeyer, M., Beautrais, A., Williams, D., 2008. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br. J. Psychiatr.* 192, 98–105.
- O’Dea, B., Larsen, M.E., Batterham, P.J., Calear, A.L., Christensen, H., 2017. A linguistic analysis of suicide-related Twitter posts. *Crisis. J. Crisis Interv. Suicide Prev.* 38 (5), 319–329.
- Oladipo, T., 2024. *23 Top Social Media Sites for Your Brand in 2024*. Buffer. Inc. Retrieved from: <https://buffer.com/library/social-media-sites/>.
- Ooms, J., 2020. Hunspell: high-performance stemmer, tokenizer, and spell checker. Rpackage version 3.0.1. <https://CRAN.R-project.org/package=hunspell>.
- Pew Research Center, 2021. Social media facts sheet URL. <https://www.pewresearch.org/internet/fact-sheet/social-media/>.
- Reddit, n.d. Search results for “mental health”. [Online forum search page]. Retrieved from: <https://www.reddit.com/search/?q=mental+health&type=communities@cid=873adb7-d542-4cab-baaf-b42dba670435&ild=d2469303-3af8-4508-88c-5c89d039a2f70>.
- Rogers, M.L., Joiner, T.E., 2017. Rumination, suicidal ideation, and suicide attempts: a meta-analytic review. *Rev. Gen. Psychol.* 21 (2), 132–142.

- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Schneider, B., 2009. Substance use disorders and risk for completed suicide. *Arch. Suicide Res.* 13 (4), 303–316.
- Shang, J., Liu, J., Jiang, M., Ren, X., Voss, C.R., Han, J., 2018. Automated phrase mining from massive text corpora. *IEEE Trans. Knowl. Data Eng.* 30 (10), 1825–1837.
- Sheehan, L., Oexle, N., Armas, S.A., Wan, H.T., Bushman, M., Glover, L., Lewy, S.A., 2019. Benefits and risks of suicide disclosure. *Soc. Sci. Med.* 223, 16–23.
- Silge, J., Robinson, D., 2016. Tidytext: text mining and analysis using tidy data principles in R. *J. Open Source Softw.* 1 (3). <https://doi.org/10.21105/joss.00037>.
- Statista, 2024. Number of Reddit users worldwide 2018–2027. Retrieved from. <https://www.statista.com/forecasts/1143622/reddit-users-in-the-world>.
- Substance Abuse and Mental Health Services Administration, 2024, Jul 30. SAMSHA releases annual National Survey on Drug Use and Health. U.S. Department of Health and Human Services. Retrieved from. <https://www.samhsa.gov/data/sites/default/files/reports/rpt47095/National%20Report/National%20Report/2023-nsduh-annual-national.pdf>.
- Travagin, G., Margola, D., Revenson, T.A., 2015. How effective are expressive writing interventions for adolescents? A meta-analytic review. *Clin. Psychol. Rev.* 36, 42–55.
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.-C., 2021. Rank-normalization, folding, and localization: an improved R^2 for assessing convergence of MCMC (with discussion). *Bayesian Anal.* 16 (2), 667–718. <https://doi.org/10.1214/20-BA1221>.
- Wongkoblap, A., Vadillo, M.A., Curcin, V., 2017. Researching mental health disorders in the era of social media: systematic review. *J. Med. Internet Res.* 19 (6), e7215.
- World Health Organization, 2021. Suicide worldwide in 2019: global health estimates. Retrieved from. <https://apps.who.int/iris/bitstream/handle/10665/341728/9789240026643-eng.pdf?sequence=1>.