This is the accepted version of a publication at the Australian and New Zealand Journal of Psychiatry: https://doi.org/10.1177/00048674221126649

Association of 7 million+ tweets featuring suicide-related content with daily calls to the Suicide Prevention Lifeline and with suicides, United States, 2016-2018

Thomas Niederkrotenthaler, MD, PhD,[1,2*] Ulrich S. Tran, DSc,[2,3] Hubert Baginski,[4,5] Mark Sinyor, MSc, MD, FRCPC,[6,7] Markus J. Strauss,[1] Steven A. Sumner, MD, MSc,[8] Martin Voracek, DSc, PhD,[2,3] Benedikt Till, DSc,[1,2] Sean Murphy, PhD,[9] Frances Gonzalez,[9] Madelyn Gould, PhD,[10] David Garcia, PhD,[4,5,11] John Draper, PhD,[9] Hannah Metzler, PhD,[1,4,5,11]

1. Medical University of Vienna, Center for Public Health, Department of Social and Preventive Medicine, Unit Suicide Research & Mental Health Promotion, Vienna, Austria
2. Wiener Werkstaette for Suicide Research, Vienna, Austria
3. School of Psychology, Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Vienna, Austria
4. Complexity Science Hub Vienna, Austria
5. Institute of Information Systems Engineering, Vienna University of Technology,
6. Vienna, Austria
7. Department of Psychiatry, Sunnybrook Health Sciences Centre, Toronto, Canada
8. Department of Psychiatry, University of Toronto, Canada
9. Centers for Disease Control and Prevention (CDC), National Center for Injury Prevention and Control, Atlanta, United States of America
10. Vibrant Emotional Health, National Suicide Prevention Lifeline, New York, United States of America
11. Departments of Psychiatry and Epidemiology, Columbia University Irving Medical Center, New York State Psychiatric Institute, New York, United States of America
12. Institute of Interactive Systems and Data Science, Faculty of Computer Science and Biomedical Engineering, Graz University of Technology, Graz, Austria

*Corresponding Author: Dr Thomas Niederkrotenthaler, Department of Social and Preventive Medicine, Center for Public Health, Medical University of Vienna, Kinderspitalgasse 15, A-1090 Vienna, Austria. E-mail: thomas.niederkrotenthaler@meduniwien.ac.at . Phone: +43-1-4016034611, Fax: +43-1-40160-934882.

Abstract

Objective

To assess associations of various content areas of Twitter posts with help-seeking from the US National Suicide Prevention Lifeline (Lifeline) and with suicides.

Methods

We retrieved 7,150,610 suicide-related tweets geolocated to the United States and posted between January 1, 2016, and December 31, 2018. Using a specially devised machine-learning approach, we categorized posts into content about prevention; suicide awareness; personal suicidal ideation without coping; personal coping and recovery; suicide cases; and other. We then applied seasonal autoregressive integrated moving-average (SARIMA) analyses to assess associations of tweet categories with daily calls to the US National Suicide Prevention Lifeline (Lifeline) and suicides on the same day. We hypothesized that coping-related and prevention-related tweets are associated with greater help-seeking and potentially fewer suicides.

Results

The percentage of posts per category was 15.4% (SD: 7.6%) for awareness, 13.8% (SD: 9.4%) for prevention, 12.3% (SD: 9.1%) for suicide cases, 2.4% (SD: 2.1%) for suicidal ideation without coping, and 0.8% (SD: 1.7%) for coping posts. Tweets about prevention were positively associated with Lifeline calls (B=1.94, SE=0.73, p=0.008), and negatively associated with suicides (B=-0.11, SE=0.05, p=.038). Total number of tweets were negatively associated with calls (B=-0.01, SE=0.0003, p=0.007) and positively associated with suicide, (B= 6.4x10-5, SE: 2.6x10-5, p=.015).

Conclusions

This is the first large-scale study to suggest that daily volume of specific suicide-prevention-related social media content on Twitter corresponds to higher daily levels of help-seeking behavior and lower daily number of suicide deaths.

Preregistration: As Predicted, #66922, May 26, 2021.

Keywords: Twitter, suicide, help-seeking, suicide prevention, interrupted time series, Papageno effect, social media, media effects, United States

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Introduction

It is increasingly well established that the specific content of media exposures can influence suicide rates differently. The overarching content focus of a post or report (e.g., whether it is focused on telling a story of death or one of prevention) may have a key role on its influence on subsequent suicides (Pirkis et al., 2006; Niederkrotenthaler et al., 2010; Sinyor et al., 2018; 2021; Ueda et al., 2017). For example, news reports featuring suicide deaths or suicide methods (i.e., reports likely focused on narratives of death or how to die), have frequently been found to be associated with subsequent increases in suicides (i.e., a Werther effect; Niederkrotenthaler et al., 2010). In contrast, texts that feature individuals coping with suicidal ideation without making a suicide attempt or dying by suicide, has been associated with a decrease in suicide subsequently (i.e., a Papageno effect; Niederkrotenthaler et al., Sinyor et al., 2021).

Most studies in the field focus on news reports, but some have focused on entertainment media,[6] and social media including tweets (Ueda et al., 2017; Sinyor et al., 2021; Won et al., 2013). With regard to Twitter, a previous study specifically assessed social media contents in suicide-related tweets by high influencers (i.e., Twitter users with a large number of followers) geolocated to Toronto, Canada, over a 1-year period (July 2015 to June 2016), and tested associations with suicides. Similar to findings from news media studies, tweets about a suicide by a celebrity and suicide death were associated with increases in suicides, whereas tweets on suicidal thoughts and messages of hope were associated with decreases in suicides (Sinyor et al., 2021).

All of the previous analyses assessing the associations of putatively harmful and protective media content have used human coding and/or keyword searches to classify media content. This approach means that either only a very limited number of media items can be characterized based on the human resources necessary for reliable coding, or the specific meaning of keywords within the context of a post is not accounted for. Further, previous approaches have typically only investigated associations with one relevant outcome, suicide. Knowledge about other relevant outcomes, particularly help-seeking, is very limited. Calling a crisis line, for example, is a rarely used indicator of help-seeking in spite of the fact that crisis lines such as the National Suicide Prevention Lifeline in the United States are often the most commonly referenced help service in media portrayals of suicide. A huge majority of primary presenting problems of callers represent risk factors of suicidal thoughts and behaviours. As for the US Lifeline, data suggest that more than 25% of callers present with the problem suicide or suicidal thoughts specifically (see Supplemental Text S3 in Niederkrotenthaler, Tran, Gould, et al., 2021, National Suicide Prevention Lifeline, 2018).

Machine-learning approaches can help in the effort to study large numbers of social media posts, while still distinguishing between particular content and narrative elements. In a recent study, which forms the basis of the present work, we used natural language processing and machine-learning methods to automatically label large quantities of social media data according to characteristics considered important for media effects research on suicide (Metzler et al., 2021). This scheme differentiates postings based on the type of post into six main content categories that might be expected to differentially influence suicidal behaviour in some individuals exposed to them: (or (1) general posts intended to spread awareness of the problem of suicide (without any focus on prevention); or (2) general posts to spread prevention-related information; (3) personal postings about coping with suicidal crises (i.e., tweets putatively related to the Papageno effect); (4) personal postings of suicidal ideation and attempts in the absence of coping; (5) reporting of suicide cases (i.e., tweets putatively related to the Werther effect); and (6) irrelevant tweets (i.e., those not clearly fitting within any of the other five categories). The deep-learning approach achieved high intercoder-reliability with a coding accuracy that was comparable to human coding (see methods section) (Metzler et al., 2021; Burnap et al., 2017).

In the present study, we conducted a time-series analysis of all tweets geolocated to the United States from January 1, 2016, to December 31, 2018. We assessed correlations of the various tweet categories with daily calls to the US National Suicide Prevention Lifeline (Lifeline) and daily suicides across the US on the same day. Due to the very short lifetime of attention to tweets, which lasts for only a few hours (Bray, 2012) we focused on associations on the same day, consistent with a recent study (Niederkrotenthaler, Tran, Gould et al., 2021). We hypothesized that coping-related and prevention-related tweets would be associated with stronger help-seeking in terms of calls to the Lifeline and potentially lower numbers of suicides, whereas posting on suicide cases would be associated with more suicides. Further, we explored the respective associations for the total volume of suicide-related tweets.

Methods

Datasets

Lifeline calls and suicide data. Daily calls routed to the Lifeline, the primary US national resource for suicidal crises, were obtained from the Lifeline, for the time period of January 1, 2016, to December 31, 2018 (1,096 days / observations). Daily US suicide death counts for the same time period were obtained from the Centers for Disease Control and Prevention’s (CDC) National Vital Statistics System. Suicide deaths were defined by International Classification of Diseases, 10th Revision (ICD-10) underlying cause of death codes X60–X84, Y87.0, and U03.

Twitter data. Full daily volumes of original English language tweets from users in the US that contained any of a comprehensive list of suicide-related keywords (see Metzler et al., 2021 and *Supplementary Appendix, Supplemental Text S1*) were obtained using a commercial analytics platform (https://brandwatch.com).

Tweet categories

The present analysis considered six different tweet categories for this analysis. We selected and defined these categories based on the following considerations. First, due to the current interest in the effect of prevention messaging (Till et al., 2021), we wanted to include a prevention category. There are ongoing discussions that prevention messages including a message on how to prevent suicide might differ in terms of impact from pure awareness messages aiming to create awareness of the problem of suicide. Because suicide awareness messages typically highlight the prevalence of suicide and risk factors like mental disorders without mentioning solutions to the issue, attempts to spread awareness of suicide may normalize suicidal behaviour (Niederkrotenthaler et al., 2014). In the present work, we therefore differentiated between prevention tweets and tweets focusing on suicide awareness. Second, based on ongoing research on the effects of messaging on personal coping (the Papageno effect), we wanted to create a category for coping with suicidal thoughts. As many tweets appeared to address personal suicidal ideation but without any coping aspect, we differentiated these tweets into two categories: coping and suicidal ideation / behaviour without coping. Third and finally, due to research showing that messaging on suicide cases can trigger additional suicides (the Werther effect), we established a category for suicide cases. All other tweets were put into the sixth category, other / irrelevant. A detailed description of the creation of the annotation system is provided in Metzler et al., 2021. Note that these categories appeared to tackle content areas of major interest in current research and prevention efforts, but the list of categories is not necessarily exhaustive.

A machine-learning model based on BERT (bidirectional encoder representations from transformers (Devlin et al., 2019, see below) was trained to distinguish between the six categories of tweets which we defined as follows. For a detailed coding book with definitions and examples please see Metzler et al. (2021).

Awareness: Tweets spreading awareness about the problem of suicide, often focusing on high suicide rates or associations with bullying, racism, depression, or specific risk groups e.g., veterans etc. These tweets do not hint at any solution and are often reporting research findings or suicide statistics. Mere expressions of a personal intent to help are coded as awareness, but any hint at something that can be practically done to prevent suicide counts as prevention rather than awareness.

Prevention: Tweets that focus on spreading information about a solution or an attempt to prevent suicide, including prevention at an individual (e.g., suggestion not to leave people alone in suicidal crisis situations) or public health-level (e.g., safety nets on bridges). Vaguely hinting at a solution or a way of dealing with the problem is sufficient to qualify for prevention. Announcements of prevention events and broad recommendations for actions are included in this category: donations, prayers with a focus on a solution for suicide, being there for someone, telling people that they matter, taking a course about suicide prevention, warning signs to watch out for, provision of the Lifeline number as a resource for suicide prevention.

Coping: Personal stories about an individual’s experience with suicidal thoughts or a suicide attempt, with a sense of hope, recovery, coping, or mentioning an alternative to suicide. Descriptions of suicidal experiences that clearly lie in the past qualify for coping as well, because they suggest the person coped with the crisis even if this is not directly stated. The post does not have to have a clearly positive connotation but cannot be purely negative. This category does not include news items (see below, category *other/irrelevant*). Any items referring to a suicide death or attempt are not included in this category. Previous research suggests coping messages may have a Papageno effect (Niederkrotenthaler et al., 2010).

Suicidal ideation & attempts without coping: Personal stories about an individual's negative experience with suicidal thoughts, related suffering (e.g., depression), suicidal communication/announcements or suicide attempts from the perspective of an affected individual, in 1st or 3rd person perspective. News items about an individual experiencing suicidal ideation were not included here (see below, category *other/irrelevant*). If the item included any content on a case of suicide death, this is coded under *suicide cases*.

Suicide cases: Stories about an individual suicide, or a timely or geographical suicide cluster, including news items. This category takes priority over other categories with regards to suicide deaths. This means that any tweet including a suicide case (in addition to any other content) would be coded here. Previous research suggests such messages may have a Werther effect (Niederkrotenthaler, Braun, Pirkis, et al., 2020).

Other/Irrelevant: This category subsumes posts not clearly related to any of the above categories. First, this includes posting types which were part of our original annotation scheme[8] but had a very low quantity. A sufficient number of training tweets is generally necessary for adequate model training (Metzler et al., 2021), and particularly, posts about news reports and lives saved were too infrequent for model training. Second, this includes postings that were not part of our categories of interest, e.g. murder-suicides. Finally, this category includes off-topic postings, e.g., postings that are (suspected) jokes, irony, sarcasm, flippant remarks, or unclear in terms of authenticity, use suicide as a metaphor or about suicide bombings.

Summary of machine-learning analyses

Please see Metzler et al. for a detailed explanation of the machine-learning basis of this work (Metzler et al., 2021). A summary is also provided in the *Supplementary Appendix (Supplemental Text S3)*. We labelled 3,200 English language tweets using an annotation scheme specifically developed for social media data (Metzler et al., 2021). Classification performances were comparable to the state-of-the-art on similar tasks (e.g., Burnap et al., 2017). The model's agreement with human annotations was comparable to the agreement between two trained human raters. For the five categories of interest (which were analysed using time-series models), human interrater-reliability (Cohen's kappa) was 0.85 (percent agreement: 88%) between human raters, and 0.81 (85%))/ 0.80 (84%), respectively, between the model and each of the two human raters. These values were from an independent sample of Twitter posts that were not used during the development of the coding scheme.

Time-series analysis

Time-series models of daily counts of calls to the Lifeline and of daily suicide deaths were fitted to the time series from January 1, 2016 to December 31, 2018. For the selection of models, we used the SPSS IBM version 26 Expert Modeler function, to choose the model with the lowest Bayesian information criterion value, highest stationary R^2 value (the variance accounted for by the fitted time-series model), and a not significant Box-Ljung Q statistic (indicating whether residuals could be assumed to be white noise). Seasonal components were included as necessary. For additional considerations on the statistical model, see Supplementary Appendix (*Supplemental Text S4*). For call data, we controlled the analysis for several outliers which, based on discussions with the Lifeline, likely reflect technical glitches in registrations of calls, similar to previous research (Niederkrotenthaler, Tran, Gould, et al., 2021)). The models also included explanatory variables for the proportions of each of the six specific tweet categories and the overall number of suicide-related tweets on a given day. We used proportions of tweets (i.e., the proportion of a specific category across the total number of tweets on any given day) of each category. Tools to retrieve large numbers of tweets generally draw a random sample of tweets if the number exceeds 10,000 tweets on a given day, and thus produce an accurate estimate of the proportion of tweets per category (but do not provide the total counts per category). The use of proportions is the state-of-the art approach to any content analysis of large amounts of tweets (e.g., Jashinsky et al., 2014; Fahey, Matsubayashi & Ueda, 2018). Because proportions of categories do not allow to make assessments of the overall number of tweets, we also included the total number of suicide-related tweets on the respective day in the model.

Models for each tweet category (adjusted only for the total number of tweets on the specific day), as well as fully adjusted models, including additionally all other content categories, were calculated. We modelled the transfer function to assess any immediate impacts of the proportion of tweets in a specific category with calls and with suicides on the same day. This was based on research indicating that social media traffic about a specific media event is closely related to calls and potentially to suicide occurrence on the same day (Niederkrotenthaler, Tran, Gould, et al., 2021). This, together with evaluations indicating that the overall impact period amounts to hours rather than days (Bray, 2012), resulted in the assumption that any effects would occur soon after the posting. Of note, any large-scale media events that endured over longer time spans would typically create social media traffic for several days or weeks and would therefore remain represented in the analysis via their daily related tweet volume.

The study was preregistered (As Predicted, #66922, May 26, 2021; https://aspredicted.org/m5sf8.pdf).

Results

There were 18,236,130 tweets geolocated to the US and posted between January 1, 2016, and December 31, 2018, as identified through the search query. On average, there were 16,639 (SD: 16,782; minimum: 6,049, maximum: 270,431) suicide-related tweets per day.

In total 7,150,610 tweets were retrieved for categorization by the machine-learning model (see methods section; this includes daily random samples from all tweets for tweet volumes larger than 10,000 tweets on a given day day). Tweets about awareness made up about 15.4% (SD: 7.6%, min: 2.1%, max: 72.9%); prevention tweets 13.8% (SD: 9.4%, min: 0.6%, max: 80.1%); coping with suicidal ideation 0.8% (SD: 1.7%, min: 0%, max: 30.1%); suicidal ideation and attempt without coping tweets 2.4% (SD: 2.1%, min: 0%, max: 51.1%); and suicide cases made up about 12.3% (SD: 9.1%, min: 0.7%, max: 79.4%); ; . Tweets not relevant to this categorization made up 55.4% on average (SD: 12.6%, minimum: 4.4%, maximum: 93.4%). Descriptive statistics of daily percentages per category are provided in Table 1.

The time series for total tweets, the six categories assessed, and for Lifeline calls and suicides are shown in Figure 1. A more fine-grained figure of tweet categories over time is provided in Metzler et al, 2021. Estimates from the regression models are shown in Table 2.

Lifeline calls

The preregistered analysis with a fully adjusted model including all categories of relevant tweets and controlling for the total amount of postings revealed that prevention-focused tweets (B=1.94, SE=0.73, p=0.008) were positively associated with calls to the Lifeline, as were tweets with coping (Papageno) content (B=9.40, SE=3.53, p=0.008). Tweets about suicidal ideation or attempts without coping were negatively associated with calls (B=-5.79, SE=2.52, p=0.02). Additionally, exploratory analyses showed that the total number of tweets was negatively associated with calls (B=-0.01, SE=0.0003, p=0.005).

Table 1. Descriptive statistics of tweet categories, United States, 2016-2018 (1,096 days).

Predicted category	Mean	Model recall	Adjusted mean[1]	SD[3]	Min	Max
Awareness	15.37	0.70	21.96	7.56	2.09	72.85
Prevention	13.76	0.89	15.46	9.39	0.58	80.1
Coping	0.83	0.69	1.2	1.73	0.06	30.07
Suicidal ideation/ attempts without coping	2.39	0.45	5.31	2.06	0.16	51.05
Suicide cases	12.31	0.77	15.99	9.1	0.73	79.41
Other/ irrelevant[2]	55.35	**---**	40.08	12.59	4.39	93.38

[1]The mean proportion per category on any given day in the observation period divided by the model's recall (i.e., the proportion of true cases the model detects in each category).

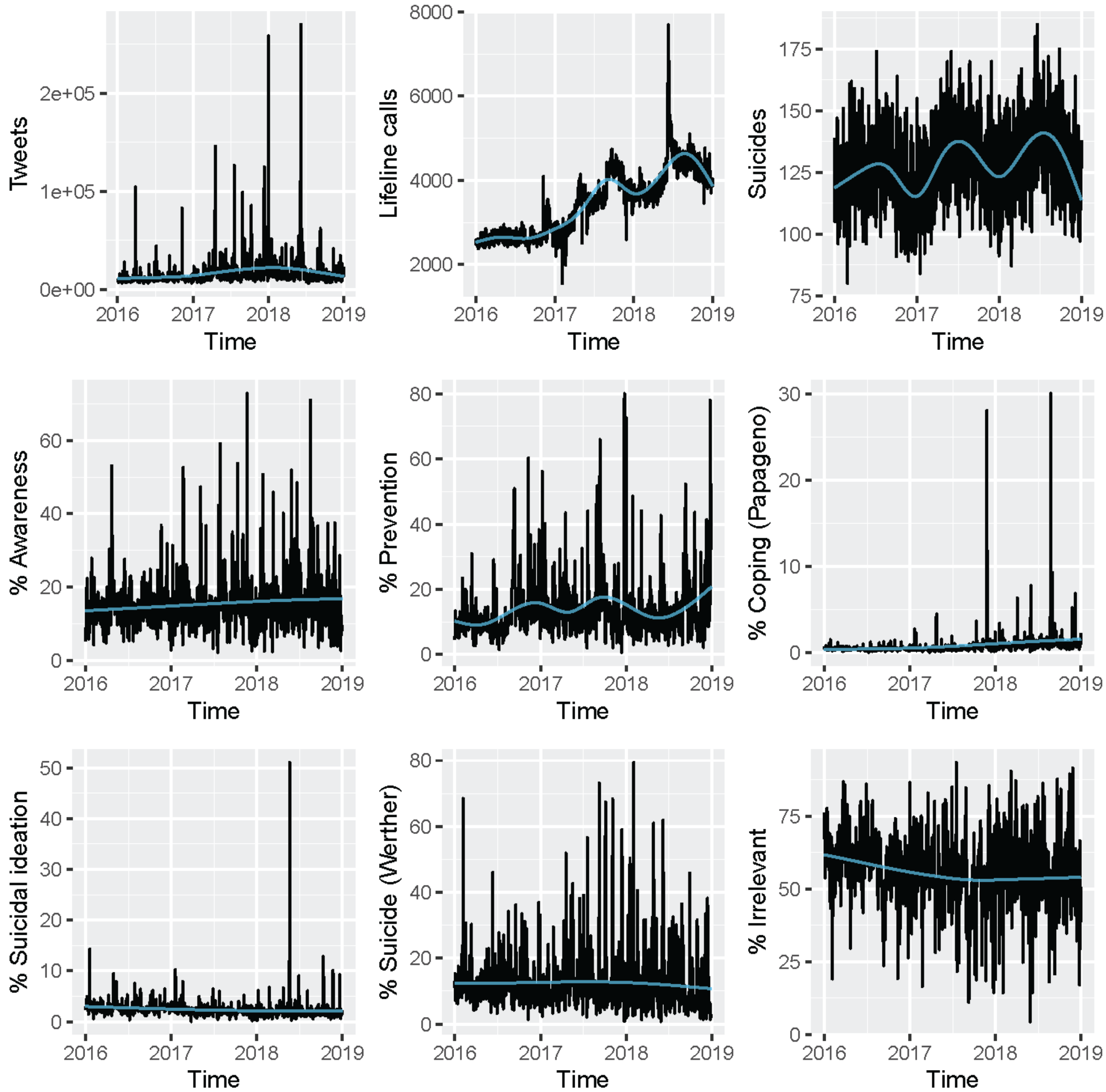
[2]The percentage for this category was calculated by subtracting all other adjusted mean percentages from 100.

[3]SD = standard deviation

Suicides

The preregistered fully adjusted analysis showed that prevention tweets were negatively associated with suicides (B=-0.11, SE=0.05, p=.038). The other preregistered analyses of associations with suicide cases (e.g., positive association with tweets about suicide cases), and also the hypothesized negative association with personal coping narratives yielded nonsignificant results. An exploratory analysis revealed that the total number of tweets was positively associated with suicides (B= $6.4x10^{-5}$, SE=$2.6x10^{-5}$, p=.015).

Figure 1. Number of tweets, tweets for each content category, Suicide Prevention Lifeline calls and suicides from 1 January 2016 to 31 December 2018, United States.



The light blue trend line was made with ggplot2 in R (version 3.3.5) via generalized additive model (GAM) smoothing with default parameters (Wickham, 2016).

Table 2.
A. Fitted SARIMA Models and Excess Number of Calls to the Suicide Prevention Lifeline by Twitter Content Category, 2016-2018 (1,096 days), United States

Content category	Raw association (only controlled for total tweets)		Adjusted for all other categories and total tweets	
	Estimate (*SE*)	*p*	Estimate (*SE*)	*p*
Awareness	0.69 (0.73)	.35	0.96 (0.76)	.20
Prevention	**1.95 (0.71)**	**.007**	**1.94 (0.73)**	**.008**
Coping	**9.34 (3.51)**	**.008**	**9.40 (3.53)**	**.008**
Suicidal ideation/ attempt without coping	-4.74 (2.54)	.062	**-5.79 (2.52)**	**.022**
Suicide case	-0.58 (0.60)	.33	0.06 (0.64)	.93
Total tweets	**-0.01 (0.0003)**	**.022**	**-0.01 (0.0003)**	**.007**

SARIMA = Seasonal Autoregressive Integrated Moving Average; SE = Standard ErrorThe time series (Jan 1, 2016 to December 31, 2018) data were checked for additive outliers (i.e., outliers affecting only one observation) and innovative outliers (i.e., outliers affecting several consecutive observations) and level shifts, which were integrated when necessary. There were 20 outliers in total, with some of them related to specific events, others to possible technical glitches in call registration. A SARIMA(0,1,5)(1,0,1) model, stationary R^2 = .60, Box-Ljung-Q = 16.28, df = 13, p =.20, was fitted to the data. Note that this model required not five, but only three moving average terms (at lags 1, 3, and 5). We subsequently added explanatory models to the basic model. We modelled outliers most likely related to technical glitches in call registration as discussed with the Lifeline, see Supplemental Appendix (Supplemental Text S2). In total, nine outliers were modelled in the final model, SARIMA(0,1,5)(1,0,1), stationary R^2 = .43, Box-Ljung-Q = 18.11, df = 13, p =.15.

B. Fitted SARIMA Models and Excess Suicides by Twitter Content Category, 2016-2018 (1,096 days), United States

Content category	Raw association (only controlled for total tweets)		Adjusted for all other categories and total tweets	
	Estimate (*SE*)	*p*	Estimate (*SE*)	*p*
Awareness	-0.002 (0.06)	.98	-0.025 (0.06)	.67
Prevention	-0.08 (0.05)	.092	**-0.11 (0.05)**	**.038**
Coping	-0.14 (0.26)	.60	-0.18 (.25)	.47
Suicidal ideation / attempt without coping	0.052 (0.20)	.80	-0.011 (0.20)	.96
Suicide case	-0.045 (0.05)	.33	-0.078 (0.05)	.11
Total tweets	4.9×10^{-5} (2.6×10^{-5})	.059	**6.4×10^{-5} (2.6×10^{-5})**	**.015**

SARIMA = Seasonal Autoregressive Integrated Moving Average; SE = Standard Error. The time series (Jan 1, 2016 to December 31, 2018) data were checked for additive and innovative outliers and level shifts. There were no outlier. A SARIMA(2,0,1)(1,0,1) model, stationary R^2 = .39, Box-Ljung-Q = 19.32, df = 13, p = .11, was fitted to the data. The final multivariate model was a SARIMA(2,0,1)(1,0,1) model, stationary R^2 = .40, Box-Ljung-Q = 15.72, df = 13, p = .27. Suicide deaths were defined by International Classification of Diseases, 10th Revision (ICD-10) underlying cause of death codes X60–X84, Y87.0, and U03.

Discussion

Main findings

This is the first large-scale analysis of specific suicide-related content on Twitter and associations with help-seeking behaviours and suicides. There was evidence for a positive association between the daily volume of prevention and coping postings with calls to the Lifeline. Further, as hypothesized, the volume of daily prevention tweets was associated with decreases in suicides. Yet, there were no significant associations between tweet content about suicide cases or coping stories and suicides. Controlling for the different tweet categories, the total number of tweets was weakly associated with fewer calls and more suicides.

Overall, the present patterns support ideas about the relevance of the specific content of social media posts and their potential effects on suicide (Sinyor et al., 2021). Media stories featuring individual coping have been found to be associated with decreases in suicidal ideation in randomized controlled trials (Niederkrotenthaler, Till, Kirchner, et al., 2022), whereas media representations of suicide deaths appear to increase both suicidal ideation (Till et al., 2015) and suicide (Niederkrotenthaler, Stack, Till, et al., 2019). Although the present findings did not support the hypothesis of a positive association between tweets about suicide deaths and actual suicides, the overall number of tweets was weakly associated with an increase in suicides. Possible explanations for this finding include that during times of media reports on specific suicides, different types of tweets are released, including awareness, prevention, and other/irrelevant types of tweets. The overall number of tweets might better reflect widespread reporting of suicides and public reactions to these cases than the category reflecting the proportion of specific tweets about suicide cases. This is corroborated by the observation that the category of suicide cases made up only a relatively small proportion of all tweets. Further, findings from media studies suggest that not all news reports of suicides are associated with subsequent increases of suicides, but rather only suicides which receive strong media attention, most importantly suicides by celebrities (Niederkrotenthaler, Braun, Pirkis, et al., 2020).

The finding that the overall quantity of suicide-related media stories is associated with increases in suicides is well-known and, in fact, one of the most robust findings in this research field (Pirkis et al., 2006; Niederkrotenthaler et al., 2009). The volume of social media postings appears particularly high during times of harmful media events, such as celebrity suicides, which garner considerably larger, and more enduring, media attention compared to prevention events (see e.g., Niederkrotenthaler, Tran, Gould et al., 2021; Niederkrotenthaler, Till & Garcia, 2019; International Association for Suicide Prevention, 2020).

Prevention postings, different from other categories, frequently mentioned the Lifeline as a resource for suicide prevention (the Lifeline was mentioned in 24.6% of tweets in this category, as also indicated in the word clouds for each category) (Metzler et al., 2021). Any effect on Lifeline calls is both plausible and relevant from a suicide prevention perspective. Previous analyses have found that changes in volumes of Lifeline calls often occur in response to media events, but these events were most often about suicides of celebrities, which were found to be associated not only with increases in calls but also with increases in suicides (Ramchand et al., 2019). The present analysis, however, suggests that, days with larger proportions of prevention-related postings are associated with a decrease in suicides and an increase in calls to the Lifeline, which may indicate an increase in help-seeking behaviour. This might mean that a change in suicide-related conversation toward a large proportion of prevention tweets (with many of them including help resources such as helpline numbers) might influence both in the desired direction (i.e., may well increase calls and decrease suicides).

With regard to the role of postings on personal coping stories, we hypothesized positive effects, based on previous studies, which showed decreased suicidal ideation in audiences exposed to content about hope and recovery (Niederkrotenthaler, Till, Kirchner, et al., 2022). Contrary to this assumption, we did not identify any associations for coping postings with suicide, but there was a strong positive association with help-seeking from the Lifeline. In contrast, posts that were about suicidal ideation or attempts but without any indication of coping were associated with fewer calls to the Lifeline. We speculate that posts about hope and recovery might trigger help-seeking in some parts of the audience even in the absence of the Lifeline number in these posts (only 0.1% of posts in this category included a reference to the Lifeline). A possible pathway might be that individuals feel encouraged to take action against suicidal thinking when reading these posts, whereas posts that feature current suicidal ideation and behaviour, mainly suicide announcements, might discourage others rather than encourage them in taking action. Overall, coping tweets made up only a tiny minority of tweets, which may have limited our ability to detect an impact on suicides. The small proportion (2.4%) but positive association with Lifeline calls highlights that these narratives are underrepresented, but potentially useful from a mental health promotion perspective. Based on the present findings, exposing a population to frequent, low-volume prevention and coping stories or themed statements might provide the dual impact of increasing helpline calls and reducing suicides. Efforts to change the conversation on social media, in order to increase the proportion of prevention and coping tweets compared to other categories (rather than increasing the overall quantity of conversations), appear warranted.

Strengths and limitations

Strengths of the present study include the very large number of tweets categorized with a machine-learning approach that yields reliable results. Study limitations include the observational design, which does not allow to infer causality. External events including National Suicide Prevention Week, World Mental Health Day, or specific suicides by celebrities might have influenced both help-seeking and / or suicide. Although we did not differentiate between these specific events, such events typically involve tweeting, and tweets produced during these periods were included in the analysis and therefore included in the analysis via tweets. Further, this study did not investigate any media effects at longer time scales. Due to the very short lifetime of attention to tweets, which lasts for only a few hours (Bray, 2012), it however appears unlikely that behavioural effects would be seen later on, when there were no immediate associations. In this context, it is also important to point out that any external events related to suicide or suicide prevention (e.g., National Suicide Prevention Week) would typically result in tweets over several days and therefore such events would be represented in the present analysis across the entire period of Twitter activity.

Further, we only examined a single social media platform. More than half of all tweets did not fit into our five categories of interest. It is possible that other groups of tweets that we considered irrelevant may have had an impact on help-seeking and/or suicide. Finally, the prevention category frequently included the Lifeline number. It remains unclear if the increase in calls was due to the Lifeline number in these posts or due to other prevention content. Word clouds for each category indicate that the content of tweets in this category was diverse and not limited to communication about the Lifeline specifically (Metzler et al., 2021).

Further, this study does not reveal any mechanisms behind the identified associations. Specifically, it remains unclear if Lifeline calls might have played any role in the association of prevention tweets with fewer suicides. Particularly experimental research designs are needed to better understand how various categories of suicide-related tweets impact audiences and their help-seeking as well as suicidal thoughts and behaviours. A further limitation is that we only retrieved English tweets for the present analysis. This was consistent with our Lifeline call data which included English calls but did not include calls to the Spanish Lifeline Network.

Further, help-seeking as assessed with calls to the Lifeline represent one indicator of help-seeking, but there are many others, e.g. hospital admissions and emergency department visits, which have not been considered for this analysis. Social isolation and suicidal ideation however, appears to be high among frequent callers of crisis lines (Pirkis et al., 2015), indicating that the threshold of calling a telephone crisis center is lower as compared to help services that require on-site visits. Finally, the involvement of authors employed by the funder of this study might be perceived as a study limitation. Specifically, some of the study authors have been employed by the funder of the current study (see section on conflicts of interest). These authors, however, were only involved in the acquisition of outcome data and revision of drafts, with no involvement in concept and design; provision of machine learning data; statistical analyses or drafting of the work. The authors have not had any financial or other benefit from the research or findings.

Conclusion

The present analysis offers potentially important implications for suicide prevention which is a preventable outcome (Stone et al., 2017). The study suggests that a steady flow of prevention tweets (and possibly coping tweets) that contributes to an overall change of the conversation in terms of an increase in the proportions of these categories within the overall suicide discourse might potentially be safer and more beneficial than efforts that increase the sheer quantity of tweets, especially, if such efforts are agnostic to tweet content. Further research is needed to investigate associations of prevention tweets with help-seeking and suicide in different countries, and effects of such tweets in various audiences.

Ethics Statement

This analysis of secondary, deidentified trend data was exempt from Institutional Review Board review.

Acknowledgements

We thank Vibrant Emotional Health for their generous grant support for this work.

Conflict of Interests

TN and BT were partially funded by a grant from Vibrant Emotional Health related to this project. Authors SM, FG, and JD are employees of Vibrant Emotional Health and provided data on calls to the Lifeline, as well as contributed to revising this report and approved its submission. There was no other involvement of the funding source.

Funding

This work was supported by the project Vibrant Emotional Health grant “Suicide Prevention media campaign Oregon” to TN, and Grant No. DG is recipient of a grant (VRG16-005) from the Vienna Science and Technology Fund to DG. MS reports a grant from the American Foundation for Suicide Prevention not directly related to this work (grant No SRG-0-153-19).

Authors and contributions

Study concept and design (TN, HM); provision of machine learning data (TN, HB, HM, DG), acquisition of outcome data (TN, SAS, SM, FG, JD), statistical analyses (TN, UST), interpretation of data (TN, UST, MS, MJS, SAS, MV, BT, MG, DG, HM); drafting the work (TN, HM); visualization (MJS), revising it critically for important intellectual content (all authors); final approval of the version to be published (all authors); agreement to be accountable for all aspects of the work (all authors).

References

Bray P (2012). *When is my Tweet's prime of life? (A brief statistical interlude.)* Published November 12, 2012. Accessed May 16, 2022. https://moz.com/blog/when-is-my-tweets-prime-of-life

Burnap, P, Colombo G, Amery R, Hodorog A, Scourfield J (2017). Multi-class machine classification of suicide-related communication on Twitter. *Online Social Networks and Media*: 2, 32–44. https://doi.org/10.1016/j.osnem.2017.08.001

Devlin J, Chang MW, Lee K, Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv* 1810.04805 [Cs]. http://arxiv.org/abs/1810.04805

Fahey RA, Matsubayashi T, Ueda,M (2018). Tracking the Werther Effect on social media: Emotional responses to prominent suicide deaths on twitter and subsequent increases in suicide. *Social Science & Medicine* 219: 19–29.

International Association for Suicide Prevention (2020). World Suicide Prevention Day 2020. Impact report. Available at: https://www.iasp.info/wspd2020/

Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, Argyle T (2014). Tracking Suicide Risk Factors Through Twitter in the US. *Crisis* 35(1), 51–59.

Metzler H, Baginski H, Niederkrotenthaler T, Garcia D (2022). Detecting Potentially Harmful and Protective Suicide-related Content on Twitter: A Machine Learning Approach. *Journal of Medical Internet Research*, 24(8), e34705. https://doi.org/10.2196/34705.

National Suicide Prevention Lifeline (2018). Lifeline Backup Center Demographic Report 1/1/2017 - 12/31/2017 (internal report).

Niederkrotenthaler T, Reidenberg DJ, Till B, Gould MS (2014). Increasing Help-Seeking and Referrals for Individuals at Risk for Suicide by Decreasing Stigma The Role of Mass Media. Am J Prev Med; 47(3):S235–S243.

Niederkrotenthaler T, Braun M, Pirkis J, Till B, Stack S, Sinyor M, Tran US, Voracek M, Cheng Q, Arendt F, Scherr S, Yip PSF, Spittal MJ (2020). Association between suicide reporting and suicide: Systematic review and meta-analysis. BMJ 368:m575.

Niederkrotenthaler T, Kirchner S, Till B, Sinyor M, Tran US, Pirkis J, Spittal MJ (2021). Systematic review and meta-analyses of suicidal outcomes following fictional portrayals of suicide and suicide attempt in entertainment media. Eclinicalmedicine 36: 100922.

Niederkrotenthaler T, Stack S, Till B, Sinyor M, Pirkis J, Garcia D, Rockett IRH, Tran US (2019). Association of Increased Youth Suicides in the United States With the Release of 13 Reasons Why. *JAMA Psychiatry* 76(9):933-940.

Niederkrotenthaler T, Till B, Garcia D (2019). Celebrity suicide on Twitter: Activity, content and network analysis related to the death of Swedish DJ Tim Bergling alias Avicii. *Journal of Affective Disorders* 245:848-855. DOI: 10.1016/j.jad.2018.11.071.

Niederkrotenthaler T, Till B, Kirchner S, Sinyor M, Braun M, Pirkis J, Tran US, Voracek M, Arendt F, Ftanou M, Kovacs R, King K, Schlichthorst M, Stack S, Spittal MJ (2022). Effects of media stories of hope and recovery on suicidal ideation and help-seeking attitudes and intentions: Systematic review and individual participant data meta-analysis of randomised controlled trials. *Lancet Public Health* 7: e156–68.

Niederkrotenthaler T, Till B, Voracek M, Dervic K, Kapusta ND, Sonneck G (2009). Copycat effects after media reports on suicide: A population-based ecologic study. *Social Science & Medicine* 69:1085-1090.

Niederkrotenthaler T, Tran U, Gould M, Sinyor M, Sumner S, Strauss MJ, Voracek M, Till B, Murphy S, Gonzalez F, Spittal MJ, Draper J (2021). Association of Logic's Hip Hop Song 1-800-273-8255 with Lifeline Calls and Suicides in the United States: A Time-Series Analysis. *BMJ* 375:e067726.

Niederkrotenthaler, T, Voracek, M, Herberth, A, Till B, Strauss M, Etzersdorfer E, Eisenwort B, Sonneck G (2010). Role of media reports in completed and prevented suicide: Werther v. Papageno effects. British Journal of Psychiatry 197: 234–243.

Pirkis JE, Burgess PM, Francis C, Blood RW, Jolley DJ (2006). The relationship between media reporting of suicide and actual suicide in Australia*. Social Science & Medicine* 62: 2874–86.

Pirkis JE, Middleton A, Bassilios B, Harris M, Spittal MJ, Fedszyn I,1 Chondros P, Gunn J. Frequent Callers to Lifeline. Melbourne: University of Melbourne; 2015. Available at: https://www.lifeline.org.au/media/ynffmhu5/frequent-callers-to-lifeline-2015.pdf

Ramchand R, Cohen E, Draper J, Schoenbaum M, Reidenberg D, Colpe L, Reed J, Pearson J (2019). Increases in demand for crisis and other suicide prevention services after a celebrity suicide. *Psychiatric Serv*ices 70(8):728–731. doi:10.1176/appi.ps.201900007

Sinyor M, Schaffer A, Nishikawa Y, Redelmeier DA, Niederkrotenthaler T, Sareen J, Levitt AJ, Kiss A, Pirkis J (2018). The association between suicide deaths and putatively harmful and protective factors in media reports. *CMAJ* 190: E900-7. DOI: 10.1503/cmaj.170698.

Sinyor M, Williams M, Zaheer R, Loureiro R, Pirkis J, Heisel MJ, Schaffer A, Redelmeier DA, Cheung AH, Niederkrotenthaler T (2021). The association between Twitter content and suicide. Australian & New Zealand Journal of Psychiatry 1 –9 DOI: 0.1177/0004867420969805.

Stone DM, Holland KM, Bartholow B, Crosby AE, Davis S, Wilkins N (2017). *Preventing Suicide: A Technical Package of Policies, Programs, and Practices*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Till B, Strauss M, Sonneck G, Niederkrotenthaler T (2015). Determining the Effects of Films with Suicidal Content: A Laboratory Experiment. *British Journal of Psychiatry*. 207(1), 72–78. DOI: 10.1192/bjp.bp.114.152827.

Till B, Tran US, Niederkrotenthaler T (2021). The Impact of Educative News Articles about Suicide Prevention: A Randomized Controlled Trial. Health Communication Routledge; 6;36(14):2022–2029. PMID:32867541

Ueda M, Mori K, Matsubayashi T, Sawada Y (2017). Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on Twitter and subsequent increases in actual suicides. *Social Science & Medicine* 189: 158–166.

Wickham H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Won HH, Myung W, Song GY, et al (2013). Predicting National Suicide Numbers with Social Media Data. *PLOS ONE* 8(4), e61809. https://doi.org/10.1371/journal.pone.0061809