BMJ Open

Studying expressions of loneliness in individuals using twitter: an observational study

Sharath Chandra Guntuku,[1,2,3] Rachelle Schneider,[2,3] Arthur Pelullo,[1,2,3] Jami Young,[3,4] Vivien Wong,[2,3] Lyle Ungar,[1,5] Daniel Polsky,[3,6] Kevin G Volpp,[3,6] Raina Merchant[2,3]

¹Computer and Information Science, University of Pennsylvania, Philadelphia, Pennsylvania, United States
²Center for Digital Health, Penn Medicine, Philadelphia, PA, United States
³Perelmen School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
⁴Children's Hospital of Philadelphia, Philadelphia, PA, United States
⁵Positive Psychology Center, University of Pennsylvania, Philadelphia, PA, United States
⁶The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

Correspondence to
Dr Sharath Chandra Guntuku;
sharathg@sas.upenn.edu

ABSTRACT

Objectives Loneliness is a major public health problem and an estimated 17% of adults aged 18–70 in the USA reported being lonely. We sought to characterise the (online) lives of people who mention the words 'lonely' or 'alone' in their Twitter timeline and correlate their posts with predictors of mental health.

Setting and design From approximately 400 million tweets collected from Twitter in Pennsylvania, USA, between 2012 and 2016, we identified users whose Twitter posts contained the words 'lonely' or 'alone' and compared them to a control group matched by age, gender and period of posting. Using natural-language processing, we characterised the topics and diurnal patterns of users' posts, their association with linguistic markers of mental health and if language can predict manifestations of loneliness. The statistical analysis, data synthesis and model creation were conducted in 2018–2019.

Primary outcome measures We evaluated counts of language features in the users with posts including the words lonely or alone compared with the control group. These language features were measured by (a) open-vocabulary topics, (b) Linguistic Inquiry Word Count (LIWC) lexicon, (c) linguistic markers of anger, depression and anxiety, and (d) temporal patterns and number of drug words. Using machine learning, we also evaluated if expressions of loneliness can be predicted in users' timelines, measured by area under curve (AUC).

Results Twitter timelines of users (n=6202) with posts including the words lonely or alone were found to include themes about difficult interpersonal relationships, psychosomatic symptoms, substance use, wanting change, unhealthy eating and having troubles with sleep. Their posts were also associated with linguistic markers of anger, depression and anxiety. A random forest model predicted expressions of loneliness online with an AUC of 0.86.

Conclusions Users' Twitter timelines with the words lonely or alone often include psychosocial features and can potentially have associations with how individuals express and experience loneliness. This can inform low-resource online assessment for high-risk individuals experiencing loneliness and interventions focused on addressing morbidities in this condition.

INTRODUCTION

Loneliness is a major public health epidemic and an estimated 17% of adults aged 18–70

Strengths and limitations of this study

► Novel focus on timelines of social media users to study mentions of loneliness and correlation with predictors of mental health.
► The study sample consists of social media, specifically Twitter users and is not representative of the general population.
► Though we manually annotated a subset of posts mentioning loneliness, some may have been metaphorical or non-sequiturs.

in the USA are reported being lonely.[1] Loneliness is defined as the discrepancy between a person's desired and actual social relationships. Loneliness is also one of the primary underlying causes and correlates for chronic mental health conditions and physician visits in some populations.[1–6] It has also been linked with a 30% increased risk of heart disease, stroke, dementia, depression and anxiety.[1 2 7–9]

Reducing morbidity from loneliness requires identifying who experiences it. Traditionally, this has occurred through surveys but unfortunately, this is not common and not scalable to screen large populations.[10] Rather than relying on the traditional screening approaches, social media platforms, like Facebook, Twitter and Instagram, are being investigated to shed light on an individual's health and well-being.[11] With people increasingly using social media platforms to inform others about their mental states, solicit social support, as well as to keep records of their daily activities, preferences and interests,[12 13] social media has emerged as a potentially relevant tool to passively measure health states and behaviours of people.[14 15] For example, individuals who are stressed and depressed use more first-person singular pronouns suggesting higher self-focus and communities with heart disease discuss hate more

frequently.[12][13] Social media posts have also been used to predict first documented diagnosis of depression using posts 6 months prior yielding an area under curve (AUC) of 0.72.[16]

Although the use of social media is increasingly common, less is known about how individuals use the platform to explicitly share feelings of loneliness. In this study, we sought to characterise the Twitter timelines of individuals' whose posts include the words lonely or alone. Based on the language of such Twitter users, we analysed the correlations between posting about loneliness and users' mental health and psycholinguistic attributes (eg, anger, anxiety, and depression).

We hypothesise that language usage patterns would both confirm the existing understanding of loneliness and give new insights into the daily lives of those who express being lonely. As loneliness can impact health outcomes, determining ways to track prevalence and manifestations of loneliness online would be useful for developing approaches for identifying and offering support for these individuals. While prioritising the privacy of individuals, specifically with the number of health insights that can be gleaned from social media, this research presents the opportunity of digital platforms to not only provide markers of health but also potentially serve as platforms that can be used for developing interventions.[17][18]

METHODS
This was a retrospective analysis of publicly available data on users posting about loneliness on Twitter. This study was exempt by the University of Pennsylvania Institutional Review Board.

Twitter data
Twitter is a popular social media platform which allows users to send and receive short 140-character messages, or 'tweets' (at the time of this study; the character limit was later increased to 280). First, from the Twitter Streaming API, we collected tweets from the 1% sample using a bounding box of location coordinates around Pennsylvania, USA. To increase the sample size of tweets from the state, all unique user IDs were recorded, and the Twitter API was used to extract timelines (each user's prior 3200 tweets) filtered by timestamps ranging from 2012 to 2016.

Patient and public involvement
Patients and the public were not involved in the development of the research questions and outcome measures.

Study sample
We identified users who posted the word 'alone' or 'lonely' at least once in their timeline in Pennsylvania, USA(25 966 users). As social media includes colloquial, metaphorical and light-hearted language (eg, 'If I see Justin Bieber, I will have a heart attack'), we sought to identify the proportion of tweets in which lonely seemed to refer to the public health meaning rather than other

uses of the term (eg, metaphor and joke).[19] Two coauthors independently coded a random set of 100 tweets from individuals who used the words lonely/alone at least five times in their timeline and identified them as presumed to be associated with the feeling of loneliness or other (Cohen's κ=0.70, and 76% of users' tweets indicate presumably feeling lonely). A few examples are as follows: '*I'm feelin real depressed, confused, & lonely*', '*Im always the only up around this time, feeling a lil lonely*' and '*I'm so Lonely in life :-(I just wish I can have love again it feels so go to be in love with someone whom loves you*'. A total of 6202 users posted messages with 'alone' or 'lonely' at least five times.

Control group
We then identified a control group of users by matching each user in the above dataset to another user by age, gender and period of activity (dates of first and last posting on Twitter). We obtained the age and gender estimates by using lexica developed previously.[20] Then, we selected users with a minimum of 500 words across all their posts to have sufficient language for linguistic analyses.[11] We excluded non-English tweets, re-tweets and tweets containing 'alone' and/or 'lonely' that were used to identify users in all analyses. Hereafter, we indicate users who had more than five posts with the words 'lonely' or 'alone' as 'users with posts including the words lonely or alone' and the matched set of users who had no such posts as 'control' group.'

Deriving language features to characterise individuals expressing loneliness
We used four sets of language features: (a) dictionary-based psycholinguistic features,[21] (b) open-vocabulary topics,[22] (c) mental well-being attributes such as anxiety, depression by applying previously developed statistical models,[23][24] (d) temporal patterns and number of drug words as past research has shown an association between loneliness and substance use.[25][26] These language features have been shown to be predictive of several health outcomes, such as depression, schizophrenia, attention deficit hyperactivity disorder (ADHD) and general well-being.[16][25][27]

Open vocabulary: From each post, we extracted the relative frequency of single words and phrases (consisting of two or three consecutive words). Then, all words used by <1% of users were removed from the analysis so as to remove uncommonly used words (outliers). Additionally, all tweets used to identify our study group were removed prior to further analysis. Topics consist of clusters of co-occurring words created using Latent Dirichlet Allocation (LDA).[28] The LDA generative model assumes that tweets contain a combination of topics and that topics are a distribution of words. As the words in a tweet are known, topics, which are latent variables, can be estimated through Gibbs sampling.[29] We use the Mallet implementation of the LDA algorithm, adjusting one parameter (alpha=5) to favour fewer topics per tweet.[30] All other

parameters were kept at their default. An example of such a model is the following sets of words ('Tuesday', 'Monday', 'Wednesday', …) which clusters together days of the week by exploiting their similar distributional properties across tweets. In our study, 200 topics were generated using tweets across all users in the dataset including the words lonely or alone and control users.

Dictionary based: The Linguistic Inquiry Word Count (LIWC) dictionary is a language-specific, many-to-many mapping of tokens (including words and word stems) and psychologically validate categories. Each category (a curated list of words) is found to be correlated with and also predictive of several psychological traits and outcomes.[13] For each user, we measure the proportion of word tokens that fall into a given LIWC category.

Mental well-being attributes: We used automatic text-regression methods to assign to each user scores on the depression, anxiety and anger facets for users.[23 24] This model was trained on a sample of over 28 749 users who had taken the International Personality Item Pool Neuroticism-Extraversion-Openness Personality Inventory Revised (IPIP NEO-PI-R) survey that contains the depression, anxiety and anger facets of the neuroticism factor.[31 32] The machine learning model trained on words and phrases from Facebook posts to predict survey measures of depression, anger and anxiety resulted in a performance of $r=0.32$, which is consistent with other reports of mental health states identified via social media.[13] The model was trained using status updates of users from another study[23] and has been shown to generalise to Twitter users.[24]

Use of drug words: We also extracted the frequency of most common drug words as used on social media for every user in our analysis.[26]

Temporal patterns: We determined the frequency of posts across different hours of the day by users in both users with posts including the words lonely or alone and control groups to understand the diurnal patterns in posting.

Identifying differentially expressed language features in users with posts including the words lonely or alone

We isolated the patterns in Twitter timelines of users who post the words alone or lonely using linguistic attributes and mental health attributes. We used logistic regression to distinguish the different features associated with users with posts containing the words lonely and alone and control groups and measure the effect size using Cohen's *d*. The models were set up to predict the group of users with posts including the words lonely or alone against the control group (i.e, group was the dependent variable). For identifying themes from topics, researchers looked at 20 messages each with the highest topic prevalence. We used Benjamini-Hochberg p-correction and $p<0.001$ for indicating meaningful correlations . We also tested that the results hold if the frequency of posting is used as an additional variable on which to match the users with lonely expressions and the control group. The

statistical analysis, data synthesis and model creation were conducted in 2018–2019.

Predicting the likelihood of posting about loneliness online

We then looked at the feasibility of predicting if a user is likely to mention expressions of loneliness based on their social media language. Automated analysis of social media is accomplished by building predictive models, which use linguistic features that have been extracted from social media posts. For this analysis, we used LIWC and topics as features. Features are then treated as independent variables in a machine learning algorithm (Random Forests) to predict the dependent variable of an outcome of interest (eg, users' expressing that they are lonely or not). For cross-validation, the predictive model was trained, using Random Forests, on the training set and then evaluated on a test set to avoid overfitting. The prediction performances are reported as area under the receiver operating curves (AUC) on an out-of-sample five-fold cross-validation setting.

RESULTS

Of the 408 296 620 tweets posted by users geo-located in Pennsylvania, USA, 25 966 users with 46 160 774 posts in their timelines, had at least 1 post with the words 'lonely' or 'alone', and 6202 users with 17 995 084 posts in their timelines, had >5 such posts. Users with posts including the words lonely or alone had 1.9 times more posts in the study time period as the control (table 1). The median estimated age of this cohort was 21 years, and 69% women.

Open-vocabulary approach: figure 1 illustrates the words and phrases most prominently associated with the group of users with posts including the words lonely or alone and the control group. Analysing differences in individual words and phrases used across both groups, we observed (figure 1A) that users with the words lonely or alone in their Twitter timeline referred to themselves ('myself' ($d=0.18$), 'I' ($d=0.16$)) significantly more than the control group. They also posted about relationship issues ('want_somebody' ($d=0.08$), 'no_one_to' ($d=0.1$), needs and feelings ('I_just_wanna ($d=0.12$), 'in_my_feelings' ($d=0.1$), 'I_need' ($d=0.12$), 'I_cant' ($d=0.1$)) and included more expletives. Users in the control group (figure 1B) engaged in a lot more conversations as indicated by '<user>' ($d=-0.2$) (anonymised '@' mentions in users tweets as '<user>') compared with users with posts

Table 1 Descriptive statistics for users in the dataset

Descriptive statistics of the dataset		
	Users with posts including the words lonely or alone (n=6202)	Control group (n=6202)
Median age (y)	21±3	21±3
No. females	4400	4400
No. males	1802	1802

Figure 1 Words/Phrases more likely to be posted by Twitter users with (A) posts including the words lonely or alone compared with (B) the control group. Word size indicates the strength of the correlation and word colour indicates relative word frequency (p<0.001, Bonferroni p-corrected).

including the words lonely or alone. The control group also posted more about games ('season' (d=-0.09),'coach' (d=-0.07), 'team' (d=-0.1)) and positivity ('!' (d=-0.13), 'awesome' (d=-0.09), ':)' (d=-0.08)).

Table 2 shows the effect sizes between most prominent topic distributions generated from LDA and the users with mentions of loneliness. Themes which occur more frequently in Twitter timelines of users with posts including the words lonely or alone were about interpersonal relationships (d=0.28) (and associated issues (d=0.22)), self-reflection (d=0.21) (accompanied with wondering about the future (d=0.12)), drug/alcohol use (d=0.29) (considering them to be the 'only friend'), insomnia (d=0.27), uncontrolled emotions (d=0.28) (accompanied by confusion (d=0.11)) and psychosomatic symptoms (d=0.29).

Dictionary-based: association of LIWC categories of users with posts including the words lonely or alone are shown in table 3. Individuals who had the words lonely or alone in their Twitter timeline used increased self-references (first-person pronouns, d=0.18), words indicating cognitive processes (including certainty, d=0.15, discrepancies,

d=0.14, differentiation, d=0.13 and tentativeness, d=0.13), and negative emotions (swearing, d=0.11).

Mental well-being: Users with posts including the words lonely or alone were more likely to have posts associated with anger (d=0.95), depression (d=0.81) and anxiety (d=0.75) when compared with the control group.

Use of drug words: We also identified the distribution of words pertaining to drugs in the Twitter timeline of users with posts including the words lonely or alone, and these were more likely to reference a blunt (d=0.16), smoke (d=0.13) and heroin (d=0.1), and included prescribed medications for treatment, recreational drug use and recreational drugs.

Temporal patterns: Users with posts including the words lonely or alone were found to post more during the night (d=0.1), shown in figure 2. We also see themes associated with nighttime posting and having difficulty sleeping (d=0.27) in the open-vocabulary analysis (table 2).

Predictive analysis: Table 4 shows that the random forest model using topics as input features predicted mentions of loneliness in users with an AUC of 0.854 (F1 score=0.778) and LIWC features resulted in AUC of 0.859

Table 2 Highly correlated topics with mentions of loneliness. Effect size is measured using Cohen's D. Only significant topics after Benjamini-Hochberg p-correction and use p<0.001 are shown.

Topic theme	Highly correlated words in topic	Effect size (Cohen's d)
Interpersonal relationships	Relationships, matter, perfect hurt, feelings, trust, forget	0.28 0.22
Self-reflection	Times, changed, lost, I've	0.21
Drug/Alcohol use	Smoke, weed, blunt, drugs, drunk	0.29
Psychosomatic symptoms	Bad, stomach, hurt, head, sick	0.29
Insomnia	Sleep, awake, tired, bed	0.27
Emotional dysregulation	People, f***ing, hate, stupid	0.28
Food/Hunger	Food, breakfast, eat, pizza, hungry	0.26

Table 3 Association of LIWC categories, mental health attributes and drug words with mentions of loneliness. *Effect size is measured using Cohen's D. Only significant topics after Benjamini-Hochberg p-correction and use p<0.001 are shown.

Category	Cohen's *d**
Pronouns	
First-person pronouns	0.18
Cognitive processes	
Certainty	0.15
Discrepancies	0.15
Differentiation	0.14
Tentativeness	0.13
Negative emotions	
Swearing	0.11
Mental well-being	
Depression	0.81
Anger	0.95
Anxiety	0.75
Drug words	
Blunt	0.16
Smoke	0.13
Heroin	0.1

LIWC, Linguistic Inquiry Word Count.

(F1 score=0.777). A combination of LIWC and topics resulted in the best AUC of 0.863 (F1 score=0.782).

DISCUSSION

From a widely used social network, Twitter, we characterised what and when individuals post about loneliness, association of such individuals with mental health and if manifestations of loneliness can be predicted in individuals using their social media language. Our hypothesis was that the language of users with posts including the words lonely or alone would be significantly different from matched controls, that this language would reveal differences in characteristics such as mental health attributes between both groups, and that the language usage patterns would both confirm existing understanding of loneliness and give new insights into the daily lives of those who post about loneliness. Towards this goal, we took an inductive approach of computationally analysing the large volumes of Twitter data with the aim of better understanding the varying manifestations of loneliness. This paper has three main findings. First, we identified themes and contexts associated with users posting about loneliness on Twitter. Second, we observed that users posting about loneliness used language associated with linguistic models for anger, depression and anxiety. Third, posts about loneliness were more likely to occur in the evening or night.

Themes associated with people mentioning loneliness on Twitter are consistent with prior literature about substance use, emotional dysregulation and troubles with relationships. For example, in one study, a high positive correlation was found between alcoholism and groups of lonely people, and lonely people were also found to express negative feelings towards relationships.[33] This expression of negativity related to relationships is likely related to a hypervigilance to social threat, associated with loneliness.[34] Lonely individuals were also reported to focus on overcoming past events as well as showing feelings of helplessness.[34]

Association of users with posts including the words lonely or alone with linguistic estimates of anger, depression and anxiety corroborate prior research, showing that loneliness and social isolation influence psychological functioning, specifically the ability to self-regulate emotion.[2 3 35] Anxiety, anger and negative mood were reported as higher in lonely young adults.[36] Tweets by users with posts including the words lonely or alone were more self-focused compared with the control group. Prior researchers have found that 'first-person singular pronouns are a modest linguistic marker of depression'.[37] Also, previous research has shown that loneliness has been associated with greater self-disclosure in Facebook posts.[38] This presents the potential for early identification and assessment to intervene on loneliness as well as mental health conditions for this group.

Trends in temporal variation in posting may reflect that sleep deprivation can contribute to social withdrawal and loneliness.[39] This finding corroborates prior research associating loneliness with diminished sleep quality.[35] A better understanding of the diurnal patterns of posting could inform the timing of interventions designed to address loneliness, as well as provide insight for other researchers to test the inter-relationships between loneliness and the motivations for using social media during nighttime.

Loneliness is known to be one of the primary underlying causes and correlates for chronic mental health conditions.[40] As loneliness is becoming increasingly recognised as a public health epidemic, several entities have taken action to address it. For example, the United Kingdom appointed a Minister for Loneliness responsible for addressing loneliness within communities.[41] CareMore, a health plan and delivery system providing care for enrollees in Medicare Advantage and Medicaid in seven states across the United States, launched the 'Togetherness Program' in a clinical setting to address loneliness in elderly patients[42] CareMore's intervention led to an increased participation in their exercise program by 56.6% and decreased utilization in emergency room and hospital admissions by 3.3% and 20.8% respectively per thousand compared with the 'intent to treat population'.[43] Additionally, social network interventions targeting loneliness have been found to be effective in reducing social isolation among individuals with severe mental health conditions but these interventions are not
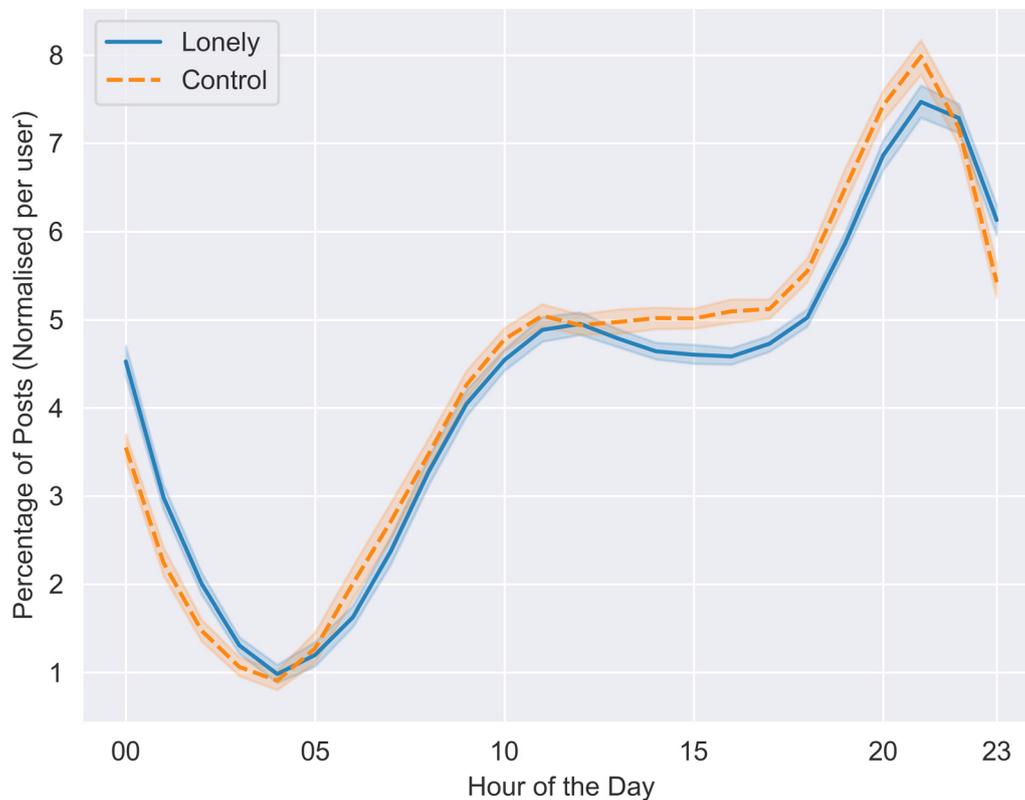
Figure 2 Temporal variation showing diurnal patterns of post frequency of both the users with posts including the words lonely or alone and control group. The solid line indicates the percentage of posts at different hours of the day by the group of users with at least five posts containing the word 'lonely' or 'alone' and the dotted line indicates users who do not have any posts about loneliness. The x-axis represents the hour of the day and the y-axis indicates the percentage of posts normalized per user for each group.

included in the treatment plans for individuals with a mental illness.[44 45]

Considering the advantage of large sample sizes and also the association between increased social media usage and individuals mentions of loneliness, it is promising to use natural language processing and machine learning to automatically identify if a person mentions the words alone or lonely on Twitter to inform interventions targeted at early identification and support for affected and at-risk individuals. To address loneliness will require being able to identify it passively, remotely and over time. Many people rarely visit a healthcare provider so would miss the opportunity for screening. Approaches for treatment will also need to harness the tools and technologies that are accessible and integrated with the things people use every day (eg, mobile phones). Future interventions would have to potentially rely on digital phenotyping of loneliness and using digital

platforms (eg, text messaging) to complement human-to-human interaction strategies to treat loneliness.

In this first study, our aim was to characterise loneliness mentions based on users' entire timelines. Future studies could perform a time-series analysis of the temporal variations associated with loneliness mentions. Other studies should replicate the findings in this study using more formal ground truth such as surveys and extend this work to investigate if Twitter can potentially map regional hotspots of loneliness to identify problematic loneliness for community public health monitoring.

LIMITATIONS AND ETHICS
The study sample consists of social media users and is not representative of the general population. An estimated 40% of US adults using Twitter are between the ages of 18 and

Table 4 Performance of different features at predicting mentions of loneliness, reported on an out-of-sample fivefold cross-validation setting

Feature	AUC	F1 score	Accuracy	Precision	Recall
Topics	0.854	0.778	0.778	0.780	0.778
LIWC	0.859	0.777	0.777	0.778	0.777
LIWC+topics	0.863	0.782	0.783	0.785	0.783

AUC, area under curve; LIWC, Linguistic Inquiry Word Count.

29, so our analysis is skewed towards younger individuals.[46] An automated machine learning tool could be a low-cost method to potentially detect posts about lonelinessthat may occur with other concerning signals from digital sensors (eg, changes in sleep, physical activity and time spent on specific apps etc). Though Twitter is far from perfect to be used as a diagnostic tool, these signals could trigger could then be referred to more formal screening methods or support resources.[47]

Considering we identified that 76% of users' tweets indicated presumably feeling lonely in the sample we hand-coded, posts mentioning the words alone or lonely may have been metaphorical or non-sequiturs. Researchers who coded the topics were attempting to identify these associations by looking at 20 messages each with the highest topic prevalence to identify themes, and this can be subjective. Also, considering the inclusion criteria based on the number of tweets mentioning the words alone or lonely, we are potentially selecting users with more posts than the average Twitter user. Further, the effects presented in this dataset may not be specific to loneliness considering the potential comorbidity with mental health conditions such as depression.

Social media use seeks to connect people but it has also been associated with increased perceived social isolation.[48] It is unclear if social media use causes perceived social isolation or if perceived social isolation causes social media use. Privacy of individuals is an ongoing concern, especially with social media users not fully realising the amount of health insights that can be gleaned by their online activity. As the notion that one does not have social connections and is alone might carry a social stigma and engender discrimination, data protection and ownership frameworks are needed to make sure the data are not used against the users' interest.[49] Further, transparency about which indicators are derived by whom for what purpose should be part of ethical and policy discourse. There are also open questions around the impact of misclassifications and how derived mental health indicators can be responsibly integrated into systems of care.[50]

CONCLUSIONS

In this study, we characterised mentions of loneliness in individuals on Twitter. Specifically, we identified specific contexts, themes and traits in the posts of individuals mentioning loneliness on Twitter and built a language-based predictive model for loneliness. As loneliness is a public health challenge, a better understanding of how loneliness is expressed online can inform passive assessment of loneliness and interventions targeted at addressing it in regards to the behaviour of lonely individuals who may be at risk of developing a severe mental health condition.

REFERENCES

1 Hyland P, Shevlin M, Cloitre M, et al. Quality not quantity: loneliness subtypes, psychological trauma, and mental health in the US adult population. Soc Psychiatry Psychiatr Epidemiol 2019;54:1089-1099.
2 Stravynski A, Boyer R. Loneliness in relation to suicide ideation and parasuicide: a population-wide study. Suicide Life-Threatening Behav 2001;31:32-40.
3 Blai B. Health consequences of loneliness: a review of the literature. J Am Coll Health 1989;37:162-7.
4 Heinrich LM, Gullone E. The clinical significance of loneliness: a literature review. Clin Psychol Rev 2006;26:695-718.
5 Richard A, Rohrmann S, Vandeleur CL, et al. Loneliness is adversely associated with physical and mental health and lifestyle factors: results from a Swiss national survey. PLoS One 2017;12:7.
6 Rico-Uribe LA, Caballero FF, Olaya B, et al. Loneliness, social networks, and health: a cross-sectional study in three countries. PLoS One 2016;11:e0145264.
7 Gerst-Emerson K, Jayawardhana J. Loneliness as a public health issue: the impact of loneliness on health care utilization among older adults. Am J Public Health 2015;105:1013-9.
8 de Jong-Gierveld J. Developing and testing a model of loneliness. J Pers Soc Psychol 1987;53:119-28.
9 Peplau LA DP. Loneliness: a Sourcebook of current theory, research, and therapy. New York: Wiley, 1982.
10 Draugalis JR, Coons SJ, Plaza CM. Best practices for survey research reports: a synopsis for authors and reviewers. Am J Pharm Educ 2008;72:11.
11 Jaidka K, Guntuku SC, Ungar LH. Facebook versus Twitter: Differences in Self-Disclosure and Trait Prediction. In: Twelfth int AAAI Conf web soc media, 2018.
12 Guntuku SC, Buffone A, Jaidka K. Understanding and measuring psychological stress using social media. Proceedings of the International AAAI Conference on Web and Social Media, 2019:214-25.
13 Guntuku SC, Yaden DB, Kern ML, et al. Detecting depression and mental illness on social media: an integrative review. Current Opinion in Behavioral Sciences 2017;18:43-9.
14 Kivran-Swaine F, Ting J, Naaman M. Understanding loneliness in social awareness streams: expressions and responses. ICWSM.
15 Kamvar S, Kamvar S JJH. We feel fine an Almanac of human emotion. New York: Scribner, 2009.
16 Eichstaedt JC, Smith RJ, Merchant RM, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018;115:11203-8.
17 Qntfy I. OurDataHelps. OurDataHelps. Available: https://ourdatahelps.org/
18 Coppersmith G, Ngo K, Leary R. Exploratory analysis of social media prior to a suicide attempt. Proceedings of the third workshop on computational Lingusitics and clinical psychology, 2016.
19 Sinnenberg L, DiSilvestro CL, Mancheno C, et al. Twitter as a potential data source for cardiovascular disease research. JAMA Cardiol 2016;1:1032-6.

20 Sap M, Park G, Eichstaedt J. Developing age and gender predictive Lexica over social media. *Proceedings of the 2014 conference on empirical methods in natural language processing*, 2014.

21 Pennebaker JW, Boyd RL, Jordan K, *et al*. *The development and psychometric properties of LIWC2015*, 2015.

22 Schwartz HA, Eichstaedt JC, Kern ML, *et al*. Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS One* 2013;8:e73791.

23 Schwartz HA, Eichstaedt J, Kern ML. Towards assessing changes in degree of depression through Facebook. *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2014:118–25.

24 Guntuku SC, Preotiuc-Pietro D, Eichstaedt JC. What Twitter profile and posted images reveal about depression and anxiety. *Proceedings of the International AAAI Conference on Web and Social Media*, 2019:236–46.

25 Guntuku SC, Ramsay JR, Merchant RM, *et al*. Language of ADHD in adults on social media. *J Atten Disord* 2019;23:1475–85.

26 Jones R. No Title. Drug Slang Dictionary - Words Starting With G. Available: http://www.noslang.com/drugs/dictionary.php

27 De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, New York, New York, USA, 2013.

28 Blei DM, AY N, Jordan MI, *et al*. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3:993–1022.

29 Gelfand AE, Smith AFM. Sampling-Based approaches to calculating marginal densities. *J Am Stat Assoc* 1990;85:398–409.

30 Graham S, Weingart S, Milligan I. *Getting started with topic modeling and Mallet*, 2012.

31 Costa Jr PT, McCrae RR. *The revised neo personality inventory (NEO-PI-R)*, 2008.

32 Bienvenu OJ, Samuels JF, Costa PT, *et al*. Anxiety and depressive disorders and the five-factor model of personality: a higher- and lower-order personality trait investigation in a community sample. *Depress Anxiety* 2004;20:92–7.

33 Booth R. Toward an understanding of loneliness. *Soc Work* 1983;28:116–9.

34 Qualter P, Vanhalst J, Harris R, *et al*. Loneliness across the life span. *Perspect Psychol Sci* 2015;10:250–64.

35 Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. *Ann Behav Med* 2010;40:218–27.

36 Cacioppo JT, Hawkley LC, Ernst JM, *et al*. Loneliness within a nomological net: an evolutionary perspective. *J Res Pers* 2006;40:1054–85.

37 Edwards T, Holtzman NSA. Meta-Analysis of correlations between depression and first person singular Pronoun use. *Journal of Research*.

38 Al-Saggaf Y, Nielsen S. Self-Disclosure on Facebook among female users and its relationship to feelings of loneliness. *Comput Human Behav* 2014;36:460–8.

39 Ben Simon E, Walker MP. Sleep loss causes social withdrawal and loneliness. *Nat Commun* 2018;9:3146.

40 Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago health, aging, and social relations study. *Psychol Aging* 2010;25:453–63.

41 Yeginsu C. Uk Appoints a Minister for loneliness, 2018. Available: https://www.nytimes.com/2018/01/17/world/europe/uk-britain-loneliness.html

42 Rubin R. Loneliness Might Be a Killer, but What's the Best Way to Protect Against It? *JAMA* 2017;318.

43 Jain S. CareMore health tackles the unmet challenges of the aging population. *Am Soc Aging* 2018;42:14–18.

44 Perese EF, Wolf M. Combating loneliness among persons with severe mental illness: social network interventions' characteristics, effectiveness, and applicability. *Issues Ment Health Nurs* 2005;26:591–609.

45 Webber M, Fendt-Newlin M. A review of social participation interventions for people with mental health problems. *Soc Psychiatry Psychiatr Epidemiol* 2017;52:369–80.

46 US. Twitter reach by age group, 2018. Available: https://www.statista.com/statistics/265647/share-of-us-internet-users-who-use-twitter-by-age-group/

47 Barnett I, Torous J, Ethics TJ. Ethics, transparency, and public health at the intersection of innovation and Facebook's suicide prevention efforts. *Ann Intern Med* 2019;170:565.

48 Primack BA, Shensa A, Sidani JE, *et al*. Social media use and perceived social isolation among young adults in the U.S. *Am J Prev Med* 2017;53:1–8.

49 McKee R. Ethical issues in using social media for health and health care research. *Health Policy* 2013;110:298–301.

50 Inkster B, Stillwell D, Kosinski M, *et al*. A decade into Facebook: where is psychiatry in the digital age? *Lancet Psychiatry* 2016;3:1087–90.