



Published in final edited form as:

*J Ment Health*. 2020 April ; 29(2): 234–241. doi:10.1080/09638237.2020.1739251.

## Conversational topics of social media messages associated with state-level mental distress rates

Daniel A. Bowen<sup>a</sup>, Jing Wang<sup>a</sup>, Kristin Holland<sup>a</sup>, Brad Bartholow<sup>a</sup>, Steven A. Sumner<sup>b</sup>

<sup>a</sup>Division of Violence Prevention, National Center for Injury Prevention and Control, U.S. Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA;

<sup>b</sup>Office of Strategy and Innovation, National Center for Injury Prevention and Control, U.S. Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA

### Abstract

**Background:** Upstream public health indicators of poor mental health in the United States (U.S.) are currently measured by national telephone-based surveys; however, results are delayed by 1–2 years, limiting real-time assessment of trends.

**Aim:** The aim of this study was to evaluate associations between conversational topics on Twitter from 2018 to 2019 and mental distress rates from 2017 to 2018 for the 50 U.S. states and capital.

**Method:** We used a novel lexicon, Empath, to examine conversational topics from aggregate social media messages from Twitter that correlated most strongly with official U.S. state-level rates of mental distress from the Behavioral Risk Factor Surveillance System.

**Results:** The ten lexical categories most positively correlated with rates of frequent mental distress at the state-level included categories about death, illness, or injury. Lexical categories most inversely correlated with mental distress included categories that serve as proxies for economic prosperity and industry. Using the prevalence of the 10 most positively and 10 most negatively correlated lexical categories to predict state-level rates of mental distress via a linear regression model on an independent sample of data yielded estimates that were moderately similar to actual rates (mean difference = 0.52%; Pearson correlation = 0.45,  $p < 0.001$ ).

**Conclusion:** This work informs efforts to use social media to measure population-level trends in mental health.

### Keywords

Depression; mental health; mental distress; social media; Twitter

---

CONTACT Steven A. Sumner, hvo5@cdc.gov, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, 4770 Buford Highway, NE, Mailstop F-63, Atlanta, GA 30341, USA.

This work was authored as part of the Contributor's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 USC. 105, no copyright protection is available for such works under US Law.

Disclosure statement

The authors declare no competing interests.

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

Population-level indicators of mental health in the United States (U.S.) have exhibited worsening trends in recent years; rates of suicide have increased more than 30% from 1999 to 2016 in half of states and rates of Emergency Department (ED) visits for non-fatal self-harm have increased 42% from 2001 to 2016 nationally (Stone et al., 2018). However, examining only suicide and self-harm present an incomplete picture as many more Americans are at risk of suicide; for example, over 10 million individuals had serious thoughts of suicide in the past year (Substance Abuse and Mental Health Services Administration, 2018).

Unfortunately, timely indicators of mental health at the population-level that are upstream or antecedent of suicide and self-harm are not routinely measured. The National Survey on Drug Use and Health, administered by the Substance Abuse and Mental Health Services Administration (SAMHSA) tracks self-reported suicide ideation and major depressive episodes while the Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) tracks self-reported mental distress (Centers for Disease Control and Prevention, 2018; Substance Abuse and Mental Health Services Administration, 2018). However, survey data from both of these surveillance systems experiences up to a two year delay in publication, limiting utility for timely monitoring of national and state-level behavioral health trends, particularly in response to public health emergencies or rapidly emerging risks.

### Use of Social Media for Mental Health Trend Monitoring

Consequently, there is considerable interest in the use of novel, emerging data sources and methodologies to provide real-time insights into trends in the public health burden of illnesses (Ayers et al., 2014; Brownstein et al., 2009; Sinnenberg et al., 2017). While the testing and development of such methodologies have been studied moreso for public health surveillance of infectious diseases and a limited number of chronic health conditions (McGough et al., 2017; Nagar et al., 2014; Nguyen et al., 2017; Park & Conway, 2018; Paul et al., 2014), some researchers have begun to examine the utility of social media data to estimate population-level mental health metrics or to study risk factors (De Choudhury et al., 2013; Jashinsky et al., 2014; Ueda et al., 2017).

Though social media data is available in real-time and at a large scale, questions about the ideal methodologies to utilize such data for epidemiologic trend monitoring exist. To date, the available research on the use of social media data for assessing national or U.S. state-level mental health indicators has explored two different methodologies. Jashinsky et al. (2014) used a researcher-curated list of terms related to suicide to filter the Twitter stream. The approach yielded a correlation coefficient of 0.53 when comparing U.S. state-level estimates of suicide derived from Twitter data with values from the National Vital Statistics System. Additionally, De Choudhury et al. (2013) have explored the utility of Twitter data in estimating population levels of depression. Using a machine learning algorithm trained on messages from depressed individuals, their research identified a 0.51 correlation between actual U.S. state-level depression rates from BRFSS and those predicted using Twitter data.

## Assessing Conversational Properties

One challenge to effectively using social media data for assessing public health trends is the need to extract messages of interest by determining the words, terms, and topics of interest in an automated fashion, balancing computational efficiency and accuracy with interpretability of results. As noted above, some authors have used expert-curated lists of terms to achieve this goal, while others have applied machine learning algorithms to recognize relevant patterns (De Choudhury et al., 2013; Jashinsky et al., 2014). Identifying social media messages that are useful for better understanding population-level mental health trends is particularly challenging given the multitude of expressions and considerable diversity in language which is used to express feelings of poor mental health.

In this research, we aimed to further explore and inform methodological strategies for the use of social media data for the monitoring of population level mental health trends by examining a novel lexicon generated from a machine learning procedure, as discussed further in the methodology section (Fast et al., 2016). A lexicon is a list of vocabulary related to specific items, topics, or themes and offers the benefit of allowing public health practitioners and researchers to straight-forwardly assess which words and topics are being most often discussed.

Our research questions and aims in this exploratory analysis are to (1) assess the utility of the novel Empath lexicon for real-time measurement of correlates of mental health at the U.S. state-level and (2) explore which conversational topics on social media most strongly correlate with traditional measures of poor mental health.

For the purposes of comparing results from the Empath lexicon to traditional data, we chose rates of past 30-day mental distress as our main outcome measure. Mental distress is the only mental health incidence measure assessed nationally with each wave of CDC's BRFSS survey and represents a construct that is upstream of suicide, self-harm, and interpersonal violence and is under-explored in previous research.

## Materials and Methods

### Population and Data Sources

This exploratory study analyzed a sample of publicly-available Twitter messages from the U.S. over two years: 2018 and 2019. Data from 2018 were used first to identify topic categories likely associated with U.S. state-level mental distress rates (as described in further detail below). Data from 2018 were collected over an approximate five-month period (sampled from 23 February 2018 to 8 April 2018; and 21 September 2018 to 31 December 2018). All messages available from the public 1% random sample of the Twitter feed were assessed, a total of 425,430,143 tweets. Retweets of original messages were excluded and tweets were geocoded to the U.S. state-level using Carmen, an open-source Python library developed by researchers at Johns Hopkins University and commonly used for such research (Chary et al., 2017; Dredze et al., 2013), yielding 29,477,365 messages that could be resolved to the U.S. state-level. In the second phase of the analysis, we validated the predictive ability of the categories identified in the first phase to make estimates of U.S. state-level mental distress rates. For this validation work, we used a new sample of public

Twitter messages which included all messages available from the public 1% sample of the Twitter feed from January 1<sup>st</sup>, 2019, through March 31<sup>st</sup>, 2019. This data was geocoded to the state-level using the above process, resulting in a total of 17,695,309 messages.

## Outcome

Our population-level mental health outcome of interest in the U.S. was frequent mental distress, which is calculated for each of the 50 U.S. states and the capital (Washington D.C.), yielding 51 geographic entities for analysis. Poor mental health is measured through CDC's Behavioral Risk Factor Surveillance Survey (BRFSS), a telephone-based, probability sample of U.S. adults which is conducted each year (Centers for Disease Control and Prevention, 2018). Specifically, the BRFSS asks participants, "Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?" (Centers for Disease Control and Prevention, 2016). We classified frequent mental distress as having experienced poor mental health for 23 days in the past 30 days (e.g., approximately 75% or more days of the month) as this cutoff has been used in prior research and shown to most strongly correlate with physical health measures (Boyer et al., 2018). Thus, this cutoff represents a significantly severe disease state. We used 2017 BRFSS data to identify potential lexical categories of interest and used 2018 BRFSS data as a second independent sample to validate the ability of these lexical categories to infer U.S. state-level mental distress rates. Although the BRFSS data in our study lagged 1-year behind the Twitter data used, this is common in research with novel data sources due to the delay in reporting of traditional data sources (Nguyen et al., 2017).

## Predictors

To explore conversational and topic predictors of frequent mental distress we utilized Empath, a tool recently developed by researchers at Stanford University that automatically generates lexical/topic categories and related words using word embeddings (Fast et al., 2016). In summary, using a large volume of text data (nearly 2 billion words) and a neural network, the authors developed complex numerical representations of words, known as word embeddings. The approach to generate the word embeddings employed an algorithm recently published by researchers at Google, known as Word2Vec (Mikolov et al., 2013). In brief, each word was represented by a string of 150 unique numbers which defined the position (or coordinates) of the word in a 150-dimensional coordinate space. These embeddings allow for the accurate detection of related words by measuring the distance between words using calculations from linear algebra, in particular, measuring the cosine of the angle between any two words. In this approach, more semantically similar words will have a smaller angle between their coordinate positions. The authors of Empath used this approach to create and validate lexicons for 194 categories. Each category began with a small number of seed terms and then the constructed word embeddings were used to identify all words that fell within a prespecified proximity to those words (as measured by the cosine of the angle between words). Subsequently, all identified words for each of the 194 lexical categories underwent manual human review to validate the final list of words for each category. Lexical categories developed by Empath were validated against an older,

manually-developed lexicon known as Linguistic Inquiry and Word Count (LIWC) (Fast et al., 2016).

### Statistical Analysis

We explored whether lexical categories generated by Empath would identify conversational topics and language associated with U.S. state-level mental distress rates. We first preprocessed all tweets in our dataset using standard natural language processing procedures, including converting text to lowercase, removing @-replies, removing hyperlinks, and removing emojis (emojis, while potentially valuable in some sentiment scoring algorithms, are not scored with Empath). The Empath algorithm, implemented in Python 3.6.3, then scores each tweet by counting the number of words in the tweet belonging to a given lexical category. For example, if a given tweet contained two words in the “suffering” lexical category, the tweet would receive a score of two for that category. Consequently, each tweet has a score for each of the 194 lexical categories that Empath contains. We then calculated the mean number of words in each lexical category for a given U.S. state by taking the average Empath generated score for each category among all tweets occurring in a state. Thus, resultant values for each state represent the average number of words for Twitter messages in a given state that fall into a certain lexical category.

To investigate topic categories, we used 2018 Twitter data to identify the 10 lexical categories with the highest positive Pearson correlation with actual U.S. state-level mental distress rates in 2017, as well as the 10 lexical categories with the strongest inverse correlation with 2017 state-level mental distress rates. As not all indicators of mental distress can be easily captured using a single conversational or topic construct, we then combined these 20 categories as predictor variables using a linear regression model with U.S. state-level mental distress rates as the outcome. Next, to evaluate the predictive ability of these categories on a new sample of data, we applied stored parameter estimates from this linear regression to new 2019 Twitter data to predict U.S. state-level rates of mental distress. These predicted rates were then compared to actual rates from the most recent 2018 BRFSS data. We calculated the mean error of such predicted mental distress rates and the Pearson correlation of these predictions with actual results from BRFSS survey data. Statistical analysis was conducted in R version 3.4.3. Our examination of publicly available media was granted IRB exemption.

### Results

The percentage of the population in each state experiencing frequent mental distress as calculated from BRFSS data is presented in Figure 1. In 2017, state-level percentages range from 4.5% of the population in Minnesota to 10.8% in West Virginia.

Overall, 29,477,365 Twitter messages from the 2018 sample could be resolved to the U.S. state-level. The three states with the highest number of messages were California, Texas, and New York, with 4,597,432, 2,703,536, and 2,630,215 messages, respectively. The three states with the lowest volume of messages were Wyoming, North Dakota, and Vermont, with 18,900, 30,789, and 33,673 messages, respectively. 41 states had greater than 100,000 messages. Table 1 shows the 10 lexical categories with the highest positive Pearson

correlation as well as the 10 lexical categories with the strongest inverse relationship with state mental distress levels and their associated correlation coefficients. Lexical topics positively associated with state-level mental distress levels include the categories of worship, masculine, death, weakness, divine, religion, sleep, swearing terms, injury, and envy. Categories negatively associated with state mental distress levels include the categories of negotiate, air travel, tourism, technology, payment, programming, independence, banking, reading, and gain. Pearson correlation coefficients for positively associated lexical categories range from 0.270 to 0.508. Pearson correlation coefficients for negatively associated lexical categories range from  $-0.504$  to  $-0.654$ . A linear regression model fitting these 20 lexical categories as independent variables to the U.S. state-level mental distress rates as the dependent variable yielded an adjusted  $R^2$  of 0.71 with an  $F$ -statistic of 7.2 on 20 and 30 degrees of freedom.

To more fully elucidate the concept of lexicons as generated by Empath and utilized in this research, Table 2 displays the lexical categories identified in this research and shows the number of words per topic category along with example words from each category.

For the most positively correlated category of “worship,” Mississippi had the most words per Tweet falling into the “worship” lexical category, with nearly 1.5 words (1.44) per 100 messages being related to worship. This was almost twice as high as the amount of worship-related words in the state with the lowest amount of such content, Vermont (0.840 worship related words per 100 messages).

Figure 2 displays the results of the validation of the lexical categories on an independent set of data. This is accomplished by using 2019 Twitter data to predict 2018 mental distress rates using stored model parameter estimates on the lexical categories from the linear regression as described above. Figure 2 displays a scatter plot of predicted levels of state mental distress levels from the linear regression with actual mental distress levels from BRFSS data. Each of the 50 states and the District of Columbia is labeled, allowing for assessment of deviation of predicted levels from actual levels. The mean difference between estimated and actual percentage of frequent mental distress was less than a percentage point (0.52%) and the Pearson correlation between predicted and actual mental distress levels was 0.45 ( $p < 0.001$ ).

## Discussion

This exploratory research aimed to assess whether use of aggregate, publicly available social media data and a novel lexicon could potentially provide more real-time insight into population-level measures of mental distress as well as inform conversational topics associated with population-level mental distress. We found that the lexical categories most strongly associated with actual state-levels of mental distress were plausible and theoretically consistent with mental health problems. Application of these categories in predictive modeling to validate their usefulness in estimating state-level mental distress rates suggested moderate utility, comparable with previous studies.

This research adds to the emerging body of literature which attempts to employ social media analysis for monitoring of population level trends of health indicators and behaviors. Although there has been limited research in the area of estimating state-level mental health indicators using social media, several researchers have explored methodologies for detecting messages related to poor mental health from a broad stream of general messages. These methodologies often include the use of a keyword or lexicon-based strategy alone or in combination with a machine learning algorithm. Procedures to generate the keyword list or train the machine learning algorithm have employed individuals as diverse as members of the lay public, patients, researchers, or medical clinicians (Burnap et al., 2015; De Choudhury et al., 2013; Jashinsky et al., 2014; O’Dea et al., 2015).

The current study contributes to this body of work by being the first paper to assess the utility of Stanford University’s Empath lexicon to population-level mental health measurement. Empath uses a novel procedure to develop their lexicon, based on word embeddings generated from an unsupervised machine learning procedure (Fast et al., 2016; Mikolov et al., 2013). We found that categories produced through this procedure help to identify groups of words associated with actual mental distress rates.

Indeed, the leading positively associated categories all can be grounded in our current understanding of mental health. Some categories, such as injury, weakness, envy, and death directly express adverse experiences and emotions associated with poor mental health and are related to constructs used in clinical mental health screening instruments (Kessler et al., 2003). Categories such as worship, divine, and religion may be positively correlated with mental distress for multiple reasons; for example, some words in this lexicon, such as “pray” capture a need or request for help. It has been previously shown in the literature that religiosity often increases in times of adversity (Ironson et al., 2006; Kessler et al., 2006). Similarly, use of swearing terms may directly capture a response to adversity present in one’s life and also has been posited as a psychological response to reduce stress or pain (Vingerhoets et al., 2013). The positive correlation observed between masculine terms and mental distress is more challenging; however, constructs including harmful masculine roles and norms, such as stoicism and difficulties in seeking help, have been found to be associated with depression (Seidler et al., 2016). Other categories, such as “sleep”, may help to capture messages reflecting the diurnal sleep imbalance commonly associated with mental health problems (Gotlib et al., 1995).

Interestingly, the lexical categories most inversely associated with mental distress rates are also theoretically plausible. Categories such as gain, reading, banking, independence, programming, payment, technology, air travel, tourism, and negotiate capture messages reflective of commercial, economic, and intellectual activity. As job loss and unemployment have been linked with poor mental health, the social media based markers of economic activity that we detected may serve as useful proxy measurements for community-level mental wellbeing associated with greater economic health (Classen and Dunn, 2012; Faria et al., 2020).

Using these lexical categories to predict U.S. state-level mental distress rates yields a moderate correlation with actual mental distress rates, findings that are comparable with the

two prior studies that used Twitter data to assess a mental health outcome (De Choudhury et al., 2013; Jashinsky et al., 2014). While lexicon or keyword-based approaches are seemingly more basic than natural language processing approaches based on supervised machine learning algorithms, lexical approaches can help provide interpretability to results and can also inform the development of machine learning models. The wide variety of constructs present in the Empath lexicon may also help to capture a wide variety of emotions, situations, and topics associated with poor mental health, helping to improve its predictive ability. However, it is important to remember that any correlations detected do not represent causal associations.

Some important limitations of this research should be noted. First, this study focuses on the outcome of frequent mental distress. This outcome was selected as it is the only mental health measure captured by CDC's BRFSS that assesses recent incidence of poor mental health. Furthermore, mental distress is "upstream" of previous outcomes that have been assessed, such as suicide fatalities and hospitalizations; such outcomes represent the most severe events and our goal was to detect disease states preceding suicide and suicide attempts to advance early prevention work. Nonetheless, mental distress represents a broad construct and examining the applicability of lexicons to more narrow mental health and suicide-related constructs is needed in future work. This however presents a more challenging task as accurately estimating suicide fatalities depends not just on a knowledge of underlying mental distress but access to lethal means, barriers to treatment or emergency care, and other factors (Arnaez et al., 2019; Miller et al., 2012). It may also be possible to model these contributors through novel data sources, but additional work is needed. Secondly, while we detected important correlations between certain lexical categories and actual mental distress rates, further research is needed to assess whether these relationships are consistent over time and to perform additional prospective validation of such an approach. It should also be noted that although the time points were offset one year between the outcome data and the social media data, this commonly occurs in comparable studies (Nguyen et al., 2017) because official data typically lags social media data sources. Third, this study assessed only one social media data source, Twitter. It is possible that use of additional social media data sources may further improve the accuracy of results. Similarly, a large body of literature exists on classic determinants of mental health, such as socioeconomic, environmental, and community-level factors; future research may consider using information on these factors from traditional data sources in combination with social media data as combining data from multiple sources may enhance the stability and accuracy of predictions (Lu et al., 2018). Fourth, it is important to acknowledge the limitations of social media data. Users of social media are not representative of the complete population and efforts to better weight social media data may yield improved results (Wood-Doughty et al., 2017). Nonetheless, the strong associations detected in this and other research offers evidence that such limitations do not preclude the utility of novel data sources. Finally, although our study examined a large and high-quality lexicon, it is useful to note that other lexicons exist, such as the LIWC; assessment of additional lexicons would be an additional meaningful methodological contribution. Nevertheless, the EMPATH lexicon includes more categories than the LIWC lexicon, which may aid in identifying more narrow constructs associated with mental health, and the EMPATH lexicon is based on quantitatively analyzing



large volumes of actual human language usage patterns through the word embedding procedures, which may further help in constructing useful lexical categories that are more challenging for human reviewers to assemble. It is also useful to note that predictive performance from any data source is also influenced by the type of modeling approach employed. Our work utilizes a generalized linear model for interpretability; however, testing of additional machine learning models may boost accuracy of predictions made from the underlying lexical data. Similarly, although we selected the 10 most positively and 10 most inversely associated categories to utilize in this initial exploratory study because this represented a straightforward and feasible number of categories to inspect and to employ in regression analysis; future prospective work with more years of data could validate the ideal number of categories to use.

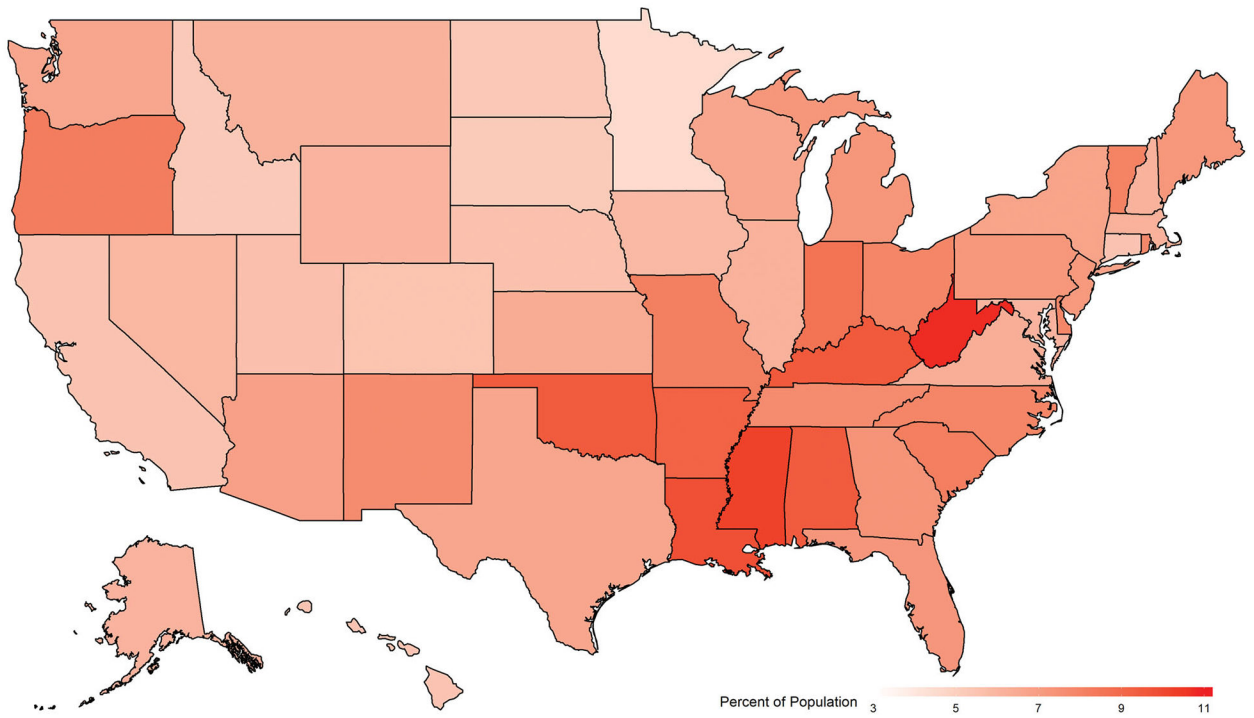
In spite of these limitations, this exploratory research makes several practical contributions. First, it is the first to explore the potential utility of the novel Empath lexicon, generated from word embeddings, for measurement of indicators of population level mental health. Second, the finding on which topical categories are most associated with mental distress rates provide insights to researchers and health professionals on constructs that may be useful in measuring proxies for mental health in novel ways. Use of innovative data sources and associated analytical methodologies requires considerable additional research to determine their role in public health, but nonetheless they may ultimately have a role in complementing traditional public health surveillance systems and help practitioners with early identification of epidemiologic trends to assist with timely mobilization of suicide prevention efforts.

## References

- Arnaez JM, Krendl AC, McCormick BP, Chen Z, & Chomistek AK (2019). The association of depression stigma with barriers to seeking mental health care: A cross-sectional analysis. *Journal of Mental Health*. Advance online publication. 10.1080/09638237.2019.1644494
- Ayers JW, Althouse BM, & Dredze M (2014). Could behavioral medicine lead the web data revolution? *JAMA*, 311(14), 1399–1400. 10.1001/jama.2014.1505 [PubMed: 24577162]
- Boyer WR, Indelicato NA, Richardson MR, Churilla JR, & Johnson TM (2018). Associations between mental distress and physical activity in US adults: A dose–response analysis BRFSS 2011. *Journal of Public Health*, 40(2), 289–294. 10.1093/pubmed/fox080 [PubMed: 28985352]
- Brownstein JS, Freifeld CC, & Madoff LC (2009). Digital disease detection—Harnessing the Web for public health surveillance. *New England Journal of Medicine*, 360(21), 2153–2157. 10.1056/NEJMp0900702 [PubMed: 19423867]
- Burnap P, Colombo W, & Scourfield J (2015, 9). Machine classification and analysis of suicide-related communication on twitter [Paper presentation] Proceedings of the 26th ACM Conference on Hypertext & Social Media, Guzelyurt, Northern Cyprus 10.1145/2700171.2791023
- Centers for Disease Control and Prevention (2016). Behavioral Risk Factor Surveillance System Questionnaire. Retrieved from [https://www.cdc.gov/brfss/questionnaires/pdf-ques/2016\\_BRFSS\\_Questionnaire\\_FINAL.pdf](https://www.cdc.gov/brfss/questionnaires/pdf-ques/2016_BRFSS_Questionnaire_FINAL.pdf)
- Centers for Disease Control and Prevention (2018). Behavioral Risk Factor Surveillance System. Retrieved 2018, from [https://www.cdc.gov/brfss/data\\_documentation/index.htm](https://www.cdc.gov/brfss/data_documentation/index.htm)
- Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, & Manini AF (2017). Epidemiology from Tweets: Estimating misuse of prescription opioids in the USA from social media. *Journal of Medical Toxicology*, 13(4), 278–286. 10.1007/s13181-017-0625-5 [PubMed: 28831738]

- Classen TJ, & Dunn RA (2012). The effect of job loss and unemployment duration on suicide risk in the United States: A new look using mass-layoffs and unemployment duration. *Health Economics*, 21(3), 338–350. 10.1002/hec.1719 [PubMed: 21322087]
- De Choudhury M, Counts S, & Horvitz E (2013, 5). Social media as a measurement tool of depression in populations [Paper presentation]. Proceedings of the 5th Annual ACM Web Science Conference, Paris, France. pp. 47–56. 10.1145/2464464.2464480
- Dredze M, Paul MJ, Bergsma S, & Tran H (2013). Carmen: A twitter geolocation system with applications to public health [Paper presentation]. AAAI workshop on expanding the boundaries of health informatics using AI (HIAI).
- Faria M, Santos MR, Sargento P, & Branco M (2020). The role of social support in suicidal ideation: A comparison of employed vs. unemployed people. *Journal of Mental Health*, 29(1), 52–59. 10.1080/09638237.2018.1487538 [PubMed: 30810419]
- Fast E, Chen B, & Bernstein MS (2016, 5). Empath: Understanding topic signals in large-scale text [Paper presentation]. Proceedings of the 2016 CHI Conference on Human Factors in Computing System, San Jose, CA, USA 10.1145/2858036.2858535
- Gotlib IH, Lewinsohn PM, & Seeley JR (1995). Symptoms versus a diagnosis of depression: differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology*, 63(1), 90–100. 10.1037/0022-006X.63.1.90 [PubMed: 7896995]
- Ironson G, Stuetzle R, & Fletcher MA (2006). An increase in religiousness/spirituality occurs after HIV diagnosis and predicts slower disease progression over 4 years in people with HIV. *Journal of General Internal Medicine*, 21(S5), S62–S68. 10.1111/j.1525-1497.2006.00648.x [PubMed: 17083503]
- 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. [PubMed: 24121153]
- Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, Howes MJ, Normand S-LT, Manderscheid RW, Walters EE, & Zaslavsky AM (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, 60(2), 184–189. 10.1001/archpsyc.60.2.184 [PubMed: 12578436]
- Kessler RC, Galea S, Jones RT, & Parker HA (2006). Mental illness and suicidality after Hurricane Katrina. *Bulletin of the World Health Organization*, 84(12), 930–939. 10.2471/BLT.06.033019 [PubMed: 17242828]
- Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, Hawkins J, Brownstein J, Conidi G, Gunn J, Gray J, Zink A, & Santillana M (2018). Accurate influenza monitoring and forecasting using novel internet data streams: A case study in the Boston Metropolis. *JMIR Public Health and Surveillance*, 4(1), e4 10.2196/publichealth.8950 [PubMed: 29317382]
- McGough SF, Brownstein JS, Hawkins JB, & Santillana M (2017). Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. *PLoS Neglected Tropical Diseases*, 11(1), e0005295 10.1371/journal.pntd.0005295 [PubMed: 28085877]
- Mikolov T, Chen K, Corrado G, & Dean J (2013). Efficient estimation of word representations in vector space. arXiv Preprint arXiv: 1301.3781.
- Miller M, Azrael D, & Barber C (2012). Suicide mortality in the United States: the importance of attending to method in understanding population-level disparities in the burden of suicide. *Annual Review of Public Health*, 33(1), 393–408. 10.1146/annurev-publhealth-031811-124636
- Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, & Brownstein JS (2014). A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *Journal of Medical Internet Research*, 16(10), e236 10.2196/jmir.3416 [PubMed: 25331122]
- Nguyen QC, McCullough M, Meng H-w., Paul D, Li D, Kath S, Loomis G, Nsoesie EO, Wen M, Smith KR, & Li F (2017). Geotagged US Tweets as predictors of county-level health outcomes, 2015–2016. *American Journal of Public Health*, 107(11), 1776–1782. 10.2105/AJPH.2017.303993 [PubMed: 28933925]
- O’Dea B, Wan S, Batterham PJ, Calear AL, Paris C, & Christensen H (2015). Detecting suicidality on Twitter. *Internet Interventions*, 2(2), 183–188. 10.1016/j.invent.2015.03.005

- Park A, & Conway M (2018). Leveraging discussions on reddit for disease surveillance. *Online Journal of Public Health Informatics*, 10(1), e52 10.5210/ojphi.v10i1.8373
- Paul MJ, Dredze M, & Broniatowski D (2014). Twitter improves influenza forecasting. *PLoS Currents*, 6 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117
- Seidler ZE, Dawes AJ, Rice SM, Oliffe JL, & Dhillon HM (2016). The role of masculinity in men's help-seeking for depression: A systematic review. *Clinical Psychology Review*, 49, 106–118. 10.1016/j.cpr.2016.09.002 [PubMed: 27664823]
- Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, & Merchant RM (2017). Twitter as a tool for health research: A systematic review. *American Journal of Public Health*, 107(1), e1–e8. 10.2105/AJPH.2016.303512
- Stone DM, Simon TR, Fowler KA, Kegler SR, Yuan K, Holland KM, & Report MW (2018). Vital signs: Trends in state suicide rates—United States, 1999–2016 and circumstances contributing to suicide—27 states, 2015. *Morbidity and Mortality Weekly Report*, 67(22), 617–624. 10.15585/mmwr.mm6722a1 [PubMed: 29879094]
- Substance Abuse and Mental Health Services Administration. (2018). Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health (HHS Publication No. SMA 18–5068, NSDUH Series H-53. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.
- Thissen D, Steinberg L, & Kuang D (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 27(1), 77–83. 10.3102/10769986027001077
- 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. 10.1016/j.socscimed.2017.06.032 [PubMed: 28705550]
- Vingerhoets AJ, Bylisma LM, & De Vlam C. J. P. t. (2013). Swearing: A biopsychosocial perspective. *Psihologijske Teme*, 22(2), 287–304.
- Wood-Doughty Z, Smith M, Broniatowski D, & Dredze M (2017, 8). How does Twitter user behavior vary across demographic groups? [Paper presentation] Proceedings of the Second Workshop on NLP and Computational Social Science, Vancouver, BC, Canada 10.18653/v1/W17-2912



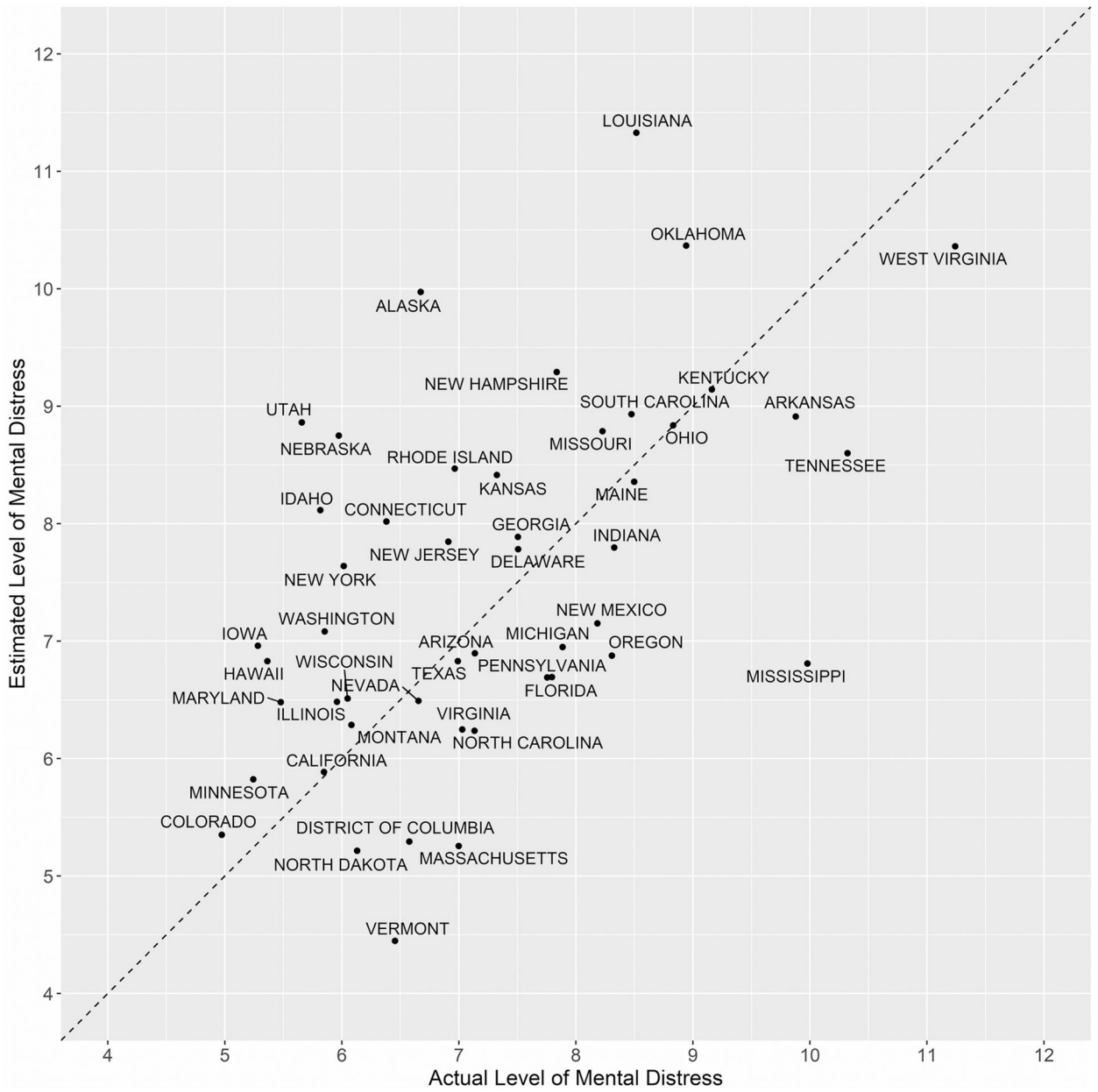
**Figure 1.** Percent of population reporting frequent mental distress, United States, 2017. Frequent mental distress defined as having experienced poor mental health for 23 days in the past 30 days, CDC’s Behavioral Risk Factor Surveillance Survey.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 2.** Comparison of predicted rates of frequent mental distress from Twitter data and actual rates, by state. Y-axis displays estimated percent of population experiencing frequent mental distress as calculated from 2019 Twitter data. X-axis displays actual percent of population experiencing frequent mental distress as calculated from CDC’s 2018 Behavioral Risk Factor Surveillance Survey. South Dakota, Alabama, and Wyoming are not shown on this plot as predicted values were outliers likely due to small numbers.

**Table 1.**

Lexical categories of Twitter messages with strongest positive and negative correlations with state-level mental distress rates.

Lexical category	Pearson correlation coefficient
Worship	0.508 *
Masculine	0.463 *
Death	0.394 *
Weakness	0.310
Divine	0.300
Religion	0.295
Sleep	0.294
Swearing terms	0.277
Injury	0.274
Envy	0.270
Gain	-0.504 *
Reading	-0.507 *
Banking	-0.519 *
Independence	-0.521 *
Programming	-0.534 *
Payment	-0.536 *
Technology	-0.579 *
Tourism	-0.584 *
Air travel	-0.585 *
Negotiate	-0.654 *

Results show the correlation between the mean number of words per social media message in a given state in a specific lexical category and actual state mental distress rates as calculated from CDC's Behavioral Risk Factor Surveillance Survey.

\*  $p < 0.05$  with Benjamini–Hochberg correction for multiple comparisons (Thissen et al., 2002).

Table 2.

Description of lexical categories with strongest positive and negative correlations with state-level mental distress rates.

Category	Number of words in category	Example words
Positive correlation		
Worship	101	Obligation, preach, church, redemption, forsake, priest, pray, savior, sacrificial, deliverance
Masculine	65	Guy, boys, intimidate, man, tattooed, rowdy, muscular, aggressive, dominate, overpowering
Death	83	Dying, deathbed, casket, condolence, perish, beloved, suffocation, cancer, misery, loss
Weakness	66	Weary, brittle, fragile, wobbly, defeated, damaged, puny, tremor, falter, powerless
Divine	83	Fate, bless, creator, angel, gospel, rejoice, mighty, pray, merciful, faith
Religion	124	Theology, govern, pilgrim, sermon, proverb, shrine, symbolic, ethical, immoral, church
Sleep	52	Slumber, bed, wakeup, snoring, exhaustion, sleepless, dream, tired, nap, slept
Swearing terms	22	*
Injury	108	Oozing, surgery, sprain, accident, swell, bruise, hurt, gash, infected, blood
Envy	45	Cynicism, jealous, hostility, discontent, deception, despised, frustration, disdain, dislike, resentment
Negative correlation		
Gain	78	Successful, reward, ownership, surpass, redeem, benefit, boost, prize, influential, promote
Reading	84	Creative, biography, thesis, compilation, poetry, fiction, script, history, literature, research
Banking	70	Fund, accounting, financial, investor, tax, payday, saving, rich, rent, revenue
Independence	51	Visionary, activist, adventurous, ambitious, ideal, politically, democracy, discipline, virtue, independent
Programming	60	Programmer, format, prototypes, internet, tech, computer, application, code, virtual, website
Payment	94	Customer, voucher, paid, bill, mortgage, monthly, loan, payroll, discount, clerk
Technology	118	Cellular, messaging, interface, data, firewall, android, technical, scientist, module, powered
Tourism	51	Theater, landmark, vacation, leisure, sightseeing, visit, tourist, hotel, embassy, travel
Air travel	71	Flight, plane, sightseeing, airport, gate, transportation, travel, departure, itinerary, luggage
Negotiate	80	Sell, profit, price, trading, partnership, wage, asset, loan, employ, supplier

The 20 categories with the highest positive and negative Pearson correlation coefficient with state mental distress rates are displayed for example purposes.

\* Example vocabulary for swearing terms not shown.