Health Data Science

A SCIENCE PARTNER JOURNAL

## Research Article

# The COVID-19 Pandemic and Mental Health Concerns on Twitter in the United States

**Senqi Zhang**,[1] **Li Sun**,[1] **Daiwei Zhang**,[1] **Pin Li**,[1] **Yue Liu**,[1] **Ajay Anand**,[1] **Zidian Xie**,[2] **and Dongmei Li**[2]

[1]*Goergen Institute for Data Science, University of Rochester, Rochester, New York, USA*
[2]*Department of Clinical & Translational Research, University of Rochester Medical Center, Rochester, NY, USA*

Correspondence should be addressed to Zidian Xie; zidian_xie@urmc.rochester.edu
and Dongmei Li; dongmei_li@urmc.rochester.edu

*Background.* During the COVID-19 pandemic, mental health concerns (such as fear and loneliness) have been actively discussed on social media. We aim to examine mental health discussions on Twitter during the COVID-19 pandemic in the US and infer the demographic composition of Twitter users who had mental health concerns. *Methods.* COVID-19-related tweets from March 5th, 2020, to January 31st, 2021, were collected through Twitter streaming API using keywords (i.e., "corona," "covid19," and "covid"). By further filtering using keywords (i.e., "depress," "failure," and "hopeless"), we extracted mental health-related tweets from the US. Topic modeling using the Latent Dirichlet Allocation model was conducted to monitor users' discussions surrounding mental health concerns. Deep learning algorithms were performed to infer the demographic composition of Twitter users who had mental health concerns during the pandemic. *Results.* We observed a positive correlation between mental health concerns on Twitter and the COVID-19 pandemic in the US. Topic modeling showed that "stay-at-home," "death poll," and "politics and policy" were the most popular topics in COVID-19 mental health tweets. Among Twitter users who had mental health concerns during the pandemic, Males, White, and 30-49 age group people were more likely to express mental health concerns. In addition, Twitter users from the east and west coast had more mental health concerns. *Conclusions.* The COVID-19 pandemic has a significant impact on mental health concerns on Twitter in the US. Certain groups of people (such as Males and White) were more likely to have mental health concerns during the COVID-19 pandemic.

## 1. Introduction

Coronavirus disease 2019, known as COVID-19, was first reported to be detected in China in December 2019. On March 13th, 2020, the declaration of a national emergency in the US marked the full outbreak of the pandemic. By July 5th, 2021, there were 30 million confirmed COVID-19 cases and 0.6 million related deaths in the US [1]. During this COVID-19 pandemic, the US people endured living in isolation and communicating in distance, and the country suffered from huge economic losses. The vaccine is a huge step towards the revitalization of lives. The US is striving to promote COVID-19 vaccines, and vaccinations are happening at an astounding rate. By June 30th, 2021, 3 billion vaccine doses have been administered worldwide [1]. Although a study pointed out that the pandemic will not end immediately even with the

prevalence of vaccines [2], people's physical health condition has greatly improved as vaccines offer protection to at least the degree of preventing severe diseases [3].

During the COVID-19 pandemic, there is another pressing issue—mental health conditions. In the US, 51.5 million adults have mental health issues according to the 2019 National Survey on Drug Use and Health data [4]. The cases of mental health illness are expected to drastically increase during the pandemic because of the restrictions, isolations, and sufferings. Many studies have discussed the impacts and consequences of the COVID-19 pandemic on mental health [5–7]. Studies found that COVID-19 sequelae include depression, anxiety, psychiatric disorder, and other mental health conditions [8, 9].

Previous studies have shown that social media is an ideal data source for studying mental health issues and monitoring sentiments towards COVID-19 vaccines [10, 11]. Other

studies have taken mental health-related Twitter data into practical usage. For example, one study used Twitter data to build a model that can detect heightened interest in mental health topics [12]. Another study applied geographic information system (GIS) analysis on Twitter users who expressed depression [13]. During the COVID-19 pandemic, many studies emerged focusing on mental health topics using social media data. One study focused on tracking "loneliness" on one-month Twitter data [14]. Another study monitored the shift of mental-health-related topics and revealed the responsiveness of Twitter [15]. All those studies contributed greatly to raising awareness against mental health issues.

In this study, we tried to understand the mental health concerns during the COVID-19 pandemic in the US using Twitter data. Furthermore, we aimed to examine which demographic groups were most likely to have mental health concerns during the pandemic.

## 2. Methods

### 2.1. Data Collection and Preprocessing.
We used the Twitter streaming API to collect COVID-19-related Twitter posts (tweets) between March 5th, 2020, and January 31st, 2021, using COVID-19-related keywords, except from May 18th, 2020, to May 19th, 2020, and from August 24th, 2020, to September 14th, 2020, due to technical issues. The COVID-19-related keywords include abbreviations and aliases ("corona," "covid19," "covid," "coronavirus," and "NCOV") [16]. The dataset was filtered with health-related keywords from seven health-related categories including mental health, cardiovascular, respiratory, neurological, psychological, digestive, and others [17, 18]. There were 8,108,004 health-related tweets in the dataset. After removing duplicates, 8,044,576 tweets remained. To avoid the potential impact of promotion tweets, we filtered out the tweets that contained promotion-related keywords ("promo code," "free shipping," "percent off," "% off," "use the code," "check us out," "check it out," "% discount," and "percent discount"). In this process, 4195 promotion-related tweets were removed. In our study, we focused solely on the mental health category. Therefore, mental health-related keywords were used to derive a mental health subset ("depression," "depressed," "depress," "failure," "hopeless," "nervous," "restless," "tired," "worthless," "unrested," "fatigue," "irritable," "stress," "dysthymia," "anxiety," "adhd," "loneliness," "lonely," "alone," "boredom," "boring," "fear," "worry," "anger," "confusion," "insomnia," and "distress") [18]. In this subset, there were 5,088,049 mental health-related tweets (Appendix Figure 1).

To identify tweets from the United States, we further applied a geological filtering process to derive a US subset based on a US keyword list [19]. The US keyword list contained full names and the abbreviation of the country, states, and some major cities in each state. The filtering process was applied to the place of the tweets. Since most of the Twitter users preferred not to share the location for a single tweet [20], we continued the filtering on the "location" feature of users if "place" is empty. After this process, we derived our US mental health dataset, which contains a total of 1,270,218 tweets.

We used the "user_name" feature to identify the distinct users in the US mental health dataset. We examined the number of Twitter users posted their first mental health-related tweet and the number of mental health-related tweets they have posted during the study period. There were 591,022 distinct users in the dataset.

To study the relationship between the number of mental health-related tweets and daily COVID-19 cases in the US, we downloaded US daily case data from COVID Tracking Project (https://covidtracking.com/data/download) on March 19, 2021. We performed time series analysis to determine the association of COVID-19 cases and tweets related to mental health in a log scale using the statistical analysis software SAS v9.4 (SAS Institute Inc., Cary, NC).

### 2.2. Topic Modeling.
The Latent Dirichlet Allocation (LDA) model was applied to extract the most frequent topics that people discussed relating to mental health during the pandemic. LDA is an unsupervised generative probabilistic model which, typically given the number of topics, allocates each word in a document to a specific topic and calculates a weight for each word representing the probability of appearance in each topic [14]. First, we removed all punctuation, converted all texts to lowercase, and tokenized every sentence. Then, the Natural Language ToolKit package was applied to remove stop-words (e.g., the, is, and a) [21]. Next, the Gensim package was applied to convert frequent bigrams and trigrams into a single term [22]. In this way, those phrases would be considered one element in the modeling process. Lastly, we lemmatized all texts by converting all tenses to present tense and keeping only nouns, adjectives, verbs, and adverbs using spaCy [23]. We determined the optimal number of topics based on the coherence score, which measured the relative distances between keywords in each topic and the intertopic distance map generated by LDAvis that visualized the overlap between topics [24]. We selected the topic number that had a relatively high coherence score.

### 2.3. Demographic Inference of Twitter Users.
We utilized a facial detection API provided by Face++ and a race/ethnicity prediction package called Ethnicolr to extract the demographic information of the users. Face++ is an AI open platform that applies deep learning to predict gender and age from an image [25]. Ethnicolr is a collection of several machine learning-based race and ethnicity classifiers trained on different data sets [26, 27].

First, we sorted the distinct user dataset based on the number of tweets per user posted in a descending order. Since the average number of tweets per Twitter user posted is 2.15, we focused on 101,492 users who posted at least three mental health-related tweets. After downloading the profile image using the "profile_image_url" feature, we utilized the API to identify the number of faces, age, and gender in the image. Age and gender would be collected if the image only contains one face. As reported in a previous study [16], Face++ has an accuracy of 93% in predicting gender. Age is much harder to be accurately determined, and the accuracy is 41%. To accommodate the situation, we choose to classify

users into age groups to achieve better accuracy. We grouped the users into five age groups (<18, 18-29, 30-49, 50-64, and ≥65) based on the criteria of the Pew Research Center [28]. We eliminated the 18-age group due to the small sample size (only 49 Twitter users).

We utilized the "census_ln" function in Ethnicolr, which trained on US census data in 2010 to predict race and ethnicity. It has an average accuracy of 79% on four races [17]. After removing all emoji and special characters, the algorithm takes a list of clean usernames with valid age and gender information as input to predict the probabilities of non-Hispanic Whites, non-Hispanic Blacks, Asians, and Hispanics for each name. We accepted the category with the highest probability as the prediction result. We obtained 11,330 Twitter users with valid age, gender, and race information.

For the comparison purpose, we have obtained the demographics (including age, gender, and race/ethnicity) of the US general population in 2019 from the United States Census Bureau (http://www.census.gov). Since we could only estimate Black, White, Asian, and Hispanic for Twitter users, we recalculated the relative proportion of these four race/ethnicities in the US general population. To compare the gender and race/ethnicity composition between different age groups, we performed 2-proportion $Z$-tests at a significance level of 0.05 using statistical analysis software R version 4.1.2 (R Core Team, 2017).

## 3. Results

*3.1. Longitudinal Trend of Tweets Related to Mental Health in the US.* To understand how the COVID-19 pandemic might affect mental health in the United States over time, we performed a temporal analysis on the number of tweets mentioning mental health in the US. As shown in Figure 1, the number of mental health-related tweets fluctuated over time with three major peaks, including from late April to early May in 2020, middle June to late July in 2020, and late October to early November in 2020. On October $6^{th}$, 2020, there were 7,033 mental health-related tweets which was the highest number. To better understand the correlation between the number of mental health-related tweets and the severity of the COVID-19 pandemic, the number of daily COVID-19 cases in the US is shown in Figure 1. Time series analysis showed that the correlation between the number of mental health tweets and the number of COVID-19 cases is 0.1196 with $P$ value = 0.0005, which indicates that there is a mild positive correlation between the number of mental health-related tweets and the number of COVID-19 cases.

To examine which mental health keywords were the most mentioned on Twitter during the pandemic, we counted the appearances of each mental health keyword. After combining words that have the same meaning (for example, "depress" includes "depress," "depression," and "depressed"), we divided the number of tweets for each keyword by the total number of tweets to calculate the proportion of each symptom. "Fear" was the most frequently mentioned symptom along with COVID-19, followed by "alone," "failure," and "depress" (Appendix Figure 2).

*3.2. Major Topics Discussed in COVID-19 Tweets Mentioning Mental Health.* To understand what might contribute to these mental health concerns mentioned in COVID-19 tweets, we performed topic modeling. As shown in Table 1, the first common topic is "stay-home and loneliness," which has the highest percentage (24.70%) in all mental health-related tweets. The second topic is "death toll," which is 16.70% of all tweets. The remaining topics have similar percentages, including "Politics and policy" (13.20%), "Personal symptom" (12.30%), "Covid and vaccine" (11.50%), "Help and relief" (10.90%), and "Government responses" (10.70%).

*3.3. Twitter Users Who Had Mental Health Concerns Related to COVID-19.* To analyze how the COVID-19 pandemic affects the public on mental health over time, we examined the Twitter users who posted their first mental health-related tweet and calculated the number of unique Twitter users who posted their first tweet each day (Figure 2). Even though the number of Twitter users who posted their first mental health tweet varied over time, the overall number of new users who had mental health concerns was decreasing. On March $6^{th}$, 2020, there were 4,451 Twitter users who posted their first tweet related to COVID-19 and mental health, while there were no more than 1,000 new Twitter users in each day in January 2021. In addition, we examined the number of Twitter users who mentioned mental health each day during the pandemic (Appendix Figure 3), which showed a similar trend as the number of mental health-related tweets (Figure 1).

*3.4. Geographic Distribution of Twitter Users Who Had Mental Health Concerns in the US.* While we have shown that the number of Twitter users who had mental health concerns in the US was large, it is important to examine whether there were some geographic differences in these Twitter users. To address this, we calculated the proportion of distinct Twitter users who mentioned mental health in each state, which was normalized by the state population. As shown in Figure 3, the states with a high proportion of Twitter users who mentioned mental health were centered to the east and west coast, such as Washington (359 Twitter users per 100,000 people), New York (244 Twitter users per 100,000 people), and Maine (471 Twitter users per 100,000 people). In addition, we examined the average number of mental health-related tweets per Twitter user in different US states, which showed that the states in the middle west (such as South Dakota) had a higher average number of mental health tweets per user (Appendix Figure 4).

*3.5. Demographic Characteristics of Twitter Users Who Had Mental Health Concerns in the US.* To better understand the demographic composition of Twitter users, especially those who posted several mental health-related posts during the pandemic, we estimated their demographic information using deep learning algorithms including Face++ API and Ethnicolr. As shown in Figure 4(a), 58.41% of Twitter users who were concerned about mental health were males while 41.59% were females. In contrast, 49.22% of US general population were female and 50.74% were male (Figure 4(b)).
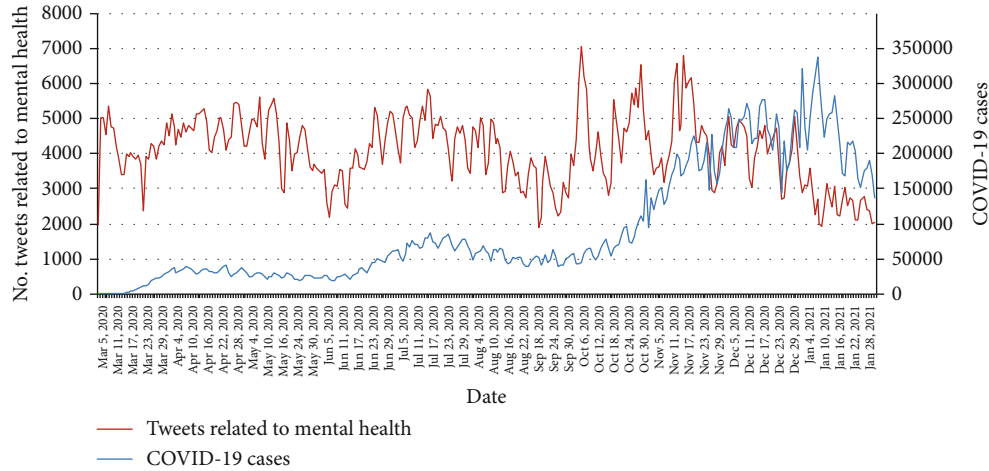
Figure 1: Number of COVID-19 tweets mentioning mental health and daily COVID-19 cases in the US.

Table 1: Topics discussed in Tweets mentioning COVID-19 and mental health.

| Topic | % Token | Keywords | Examples |
|---|---|---|---|
| Stay-home and loneliness | 24.70% | covid, get, alone, tired, go, people, worry, mask, know, home | Joe, I am old… been home alone since mid-Feb. I have auto-immune issues. I have been sick 3 times this year after going to the market. I was very careful. I fear Covid and worry how long it will be for someone like me to get a vaccine. There are many like me. |
| Death toll | 16.70% | fear, covid, people, death, die, alone, virus, many, number, live | Yea but fearing it causes more damage than actual virus. Virus deaths is lie created to cause that fear. We cannot live in fear always. I have been around a lot with covid and not once have caught it. Covid is not the big bad killer that dems and media want u to believe. |
| Politics and policy | 13.20% | covid, fear, biden, trump, country, vaccine, great, american, bill | I do not fear Covid specifically at all, though I should since I'm high-risk. But yes, I do have that generalized anxiety. I think it's a combo of Covid, murder hormets, Trump, Biden, and riots in the streets. |
| Personal symptom | 12.30% | covid, fatigue, test, vaccine, get, depression, patient, hospital, symptom | Lost a job, could not attend my grandfather's funeral, may never get to see my other grandfather again as he has been given little time to live, have had to cope with severe depression in total isolation, actually had vivid, developed chronic fatigue and urticaria as a result… |
| Covid and vaccine | 11.50% | vaccine, covid, new, worry, alone, case, coronavirus, pandemic, day, worker | Vaccines offer protection against COVID-19, but experts worry side effects may scare people off |
| Help and relief | 10.90% | covid, stress, anxiety, pandemic, help, health, relief, mental, time | THIS VIDEO FROM @PrinceEa is going to help give relief from covid anxiety because you are being proactive and keeping in the best possible immune health! |
| Government responses | 10.70% | failure, covid, trump, american, response, coronavirus, dead, pandemic, lie, death | Literally what rules? This is Failure of Leadership. Slow to close and quick to reopen, like trump wanted. Require Masks. Get Money to Citizens for Rent, Food, Utilities. Close Bars + Restaurants. Ban Large Gatherings. This does not take a rocket scientist. just a leader |

The age 30-49 group has the highest percentage (40.81%) in Twitter users who had mental health concerns, followed by age 50-64 (30.10%), age 18-29 (15.38%), age 65, and above (13.28%) (Figure 4(a)). In 2019 US general population (Figure 4(b)), age group 30-49 was also the largest one (33.10%), followed by age 50-64 (24.68%) and age 18-29 (21.05%). For race/ethnicity, among Twitter users who had mental health concerns, White was the most (85.28%), followed by Asian (7.06%), Hispanic (5.25%), and Black (2.42%) (Figure 4(a)). White was also the most (62.34%) in

the US general population (Figure 4(b)), but less than that in Twitter users who had mental health concerns.

By examining the gender distribution in each age group (Appendix Figure 5), we showed that at the young age group (age 18-29), the proportion of females is significantly higher than the proportion of males ($P < 0.0001$). However, in the middle and old age groups (age 30-49, age 50-64, and age 65+), the proportion of males is significantly higher than that of females ($P = 0.032$, $P = 0.0048$, and $P = 0.013$, respectively). Furthermore, we compared the proportion of
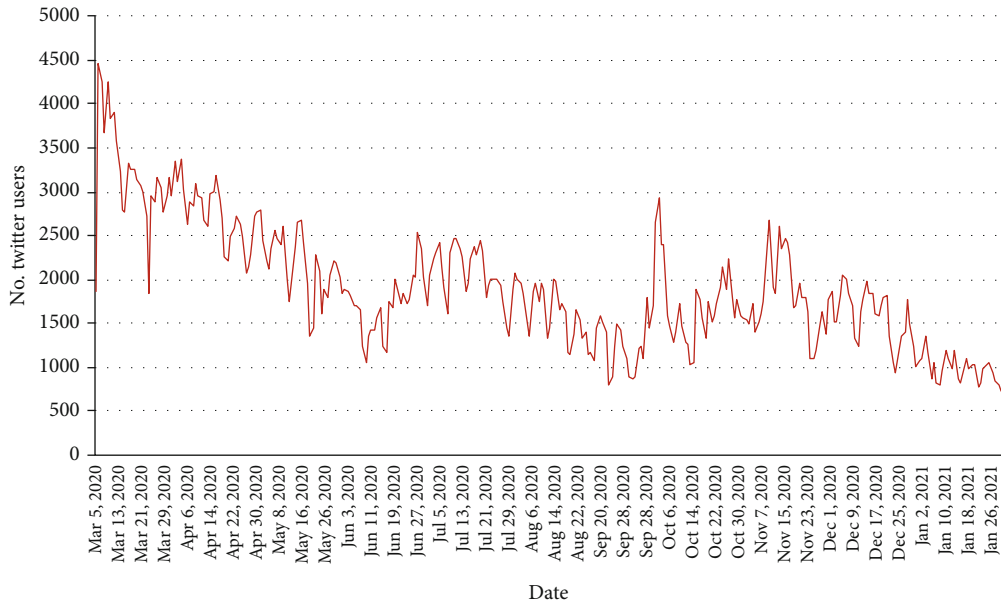
FIGURE 2: The number of US Twitter users having their first tweets about mental health over time.
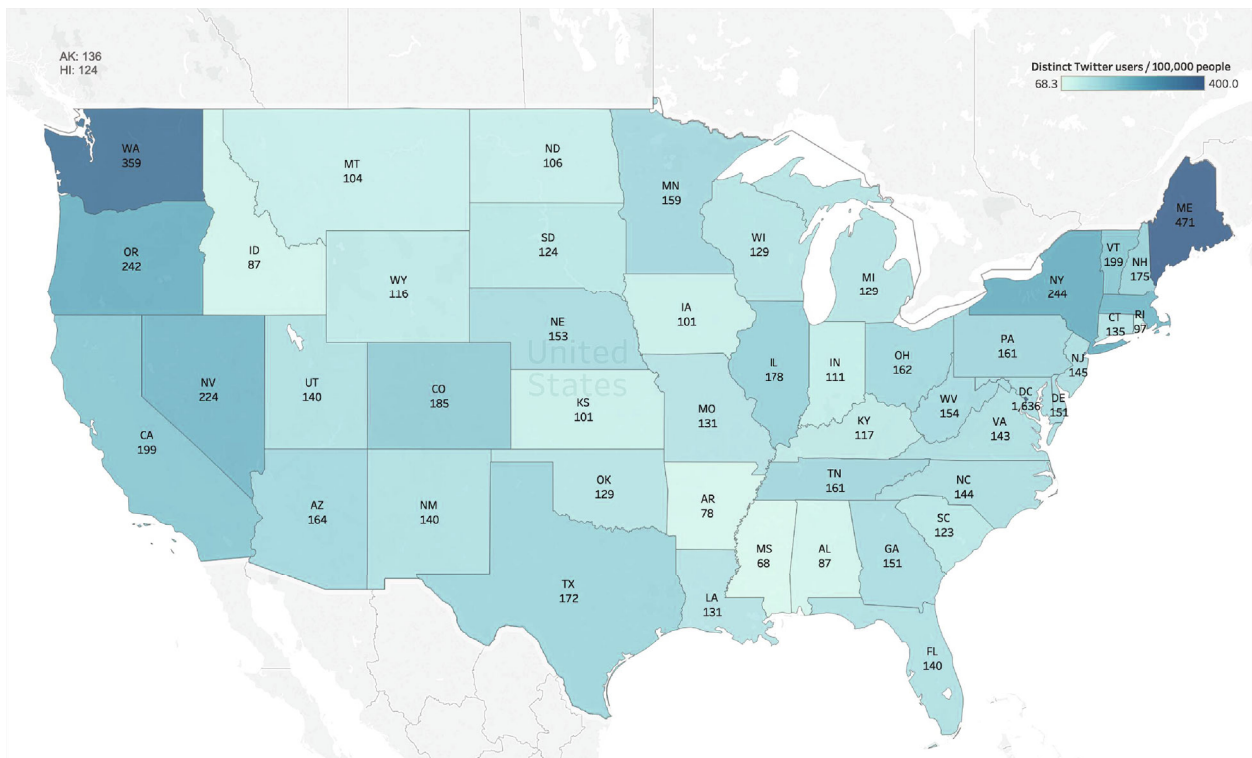


FIGURE 3: The proportion of Twitter users who had mental health concerns in different US states normalized by the population.

each age group in each race/ethnicity (Appendix Figure 6). By comparison, the proportion of Twitter users aged 30-49 with mental health concerns in the Asian group is significantly higher than that in the Black group ($P = 0.0014$) and White Twitter users ($P < 0.0001$). The proportion of Twitter users aged 18-29 with mental health concerns in the Asian group is significantly higher than

that in the White group ($P < 0.0001$). White has a significantly higher proportion of Twitter users aged 50-64 with mental health concerns than Asian ($P < 0.0001$), Black ($P = 0.026$), and Hispanic ($P = 0.0058$). The proportion of Twitter users aged 65+ with mental health concerns in the White group is significantly higher than that in the Asian group ($P < 0.0001$) and in the Hispanic group ($P = .00046$).
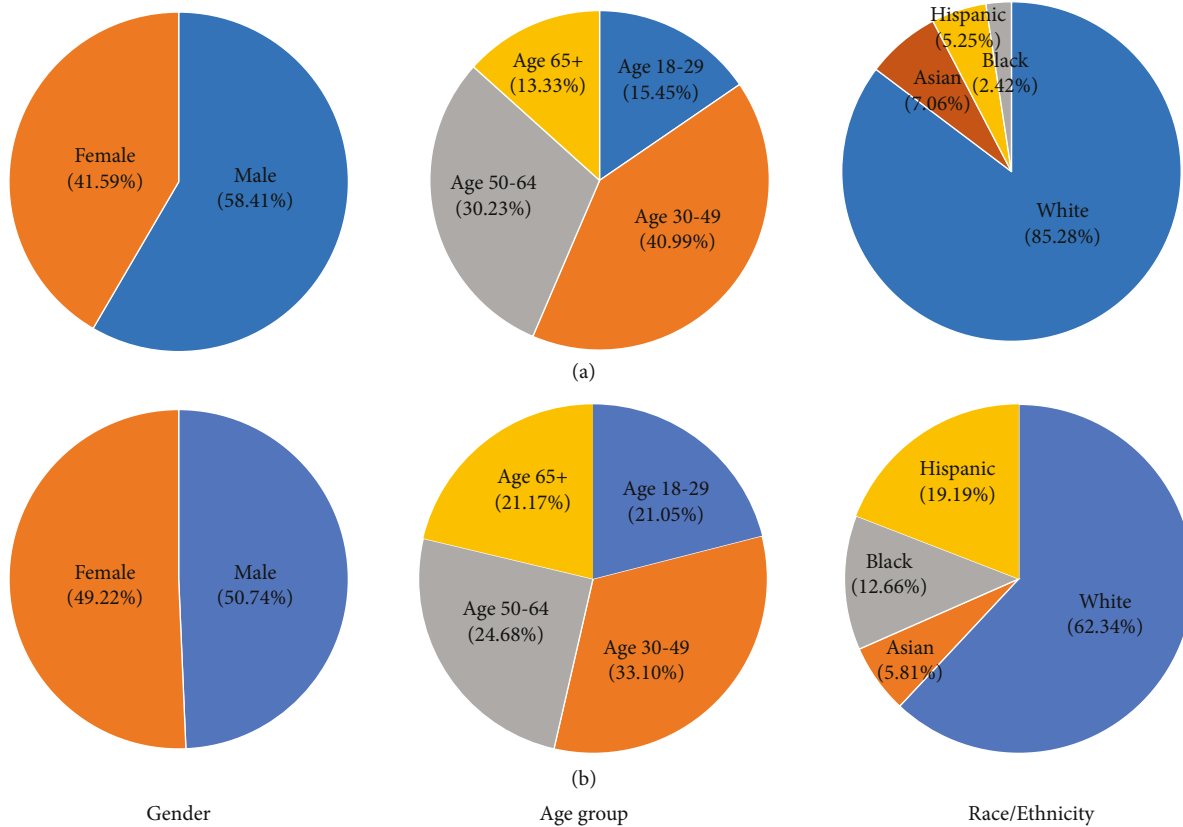
FIGURE 4: Demographic composition (including age, gender, and race/ethnicity). (a) Twitter users who had mental health concerns in the US. (b) US general population in 2019.

## 4. Discussion

*4.1. Principal Findings.* In this study, we observed a variation in mental health-related tweets over time and identified a moderate positive correlation between the number of mental health tweets and COVID-19 cases in the US, which suggests that the COVID-19 pandemic leads to more mental health concerns. We observed a downward trend of mental health-related tweets at the end of 2020, while the number of COVID-19 cases was still very high, which might be due to the success of COVID-19 vaccine development and the start of COVID-19 vaccination. Through keyword search and topic modeling, we identified the most pressing mental health concerns during the pandemic and related topics. The social distancing seemed to have severe effects on mental health concerns: people were forced to stay at home for a long time, and traveling became risky and prohibited, which naturally led to feelings like stress and loneliness. For fear and anxiety, we infer that they came from the uncertainty about how long the pandemic will last and the high COVID infection and death rate. During the pandemic, it seemed that the public was not satisfied with some government responses and related health precaution policies, which might lead to the high frequency of "failure."

In our demographic analysis, we estimated the demographic composition of Twitter users mentioning COVID-19 and mental health during the pandemic. Compared to the proportion of males in general US population (50.74%) and Twitter users (53.19%) [28], more male Twitter users (58.41%) had mental health concerns. Therefore, the males were more likely to express mental health concerns on Twitter. In the US, the proportion of people using Twitter decreases as the age increases (44.68% aged 18-29, 28.72% aged 30-49, 19.15% aged 50-64, 7.45% aged 65, or older) [28]. However, the majority of people posting mental health-related tweets during the pandemic were middle-aged and old-aged people (40.81% aged 30-49, 30.10% aged 50-64). The results indicate that middle-aged and old-aged people are more likely to express mental health concerns than young people on Twitter. Among all age groups, males were more likely to express mental health concerns except in the 18-29 age group. Compared with the distribution of the US general population and Twitter users provided by a previous study [25], the proportion of White in Twitter users who mentioned mental health issues (85.28%) is significantly higher than the proportion of White in the US general population (62.34%) and Twitter users (68%). However, the proportion of Asian (7.06%) and Black (2.42%) mentioning mental health issues is significantly lower than the proportion of general US Twitter users with Asian (18%) and Black (14%).

*4.2. Comparison with Prior Work.* Mental health is one of the major health issues during the COVID-19 pandemic. One study conducted two rounds of surveys to investigate

psychological impacts on people during the early phase of the COVID-19 pandemic in China [29]. Another study utilized an online survey and a gender-based approach to study the impact of the COVID-19 pandemic on mental health in Spain [30]. Both studies showed that the COVID-19 pandemic has a significant impact on mental health in public. Moreover, one previous study based on Twitter data showed that the volume of tweets on mental health was relatively constant before the COVID-19 and significantly increased during the COVID-19 pandemic compared to that before the pandemic [31]. In this study, we showed a positive correlation between the COVID-19 pandemic and mental health concerns on Twitter in the US.

Social media data had been used to study mental health issues during the COVID-19 pandemic. One previous study applied machine learning models to track the level of stress, anxiety, and loneliness during 2019 and 2020 using Twitter data [32]. The results showed that all of the three mental health problems (stress, anxiety, and loneliness) increased in 2020. Another study developed a transformer-based model to monitor the depression trend using Twitter [33]. The results showed that there was a significant increase in depression signals when the topic is related to COVID-19. These results are consistent with our results that stress, anxiety, loneliness, and depression are the top mentioned mental health emotions during the pandemic. Our topic modeling results further showed that loneliness is related to the quarantine at home.

While it is important to examine the impact of the COVID-19 pandemic on mental health, it might be more important to understand who might be affected the most by the pandemic in terms of mental health. One study applied the M3 (multimodal, multilingual, and multiattribute) model to extract the age and gender information of Twitter users who posted COVID-19-related tweets from August 7 to 12, 2020, which showed that males and older people discussed more on COVID-19 and expressed more fear and depression emotion [34]. Another survey study showed that women and young people were more likely to have mental health issues and developed worse mental health outcomes during the pandemic [29]. In this study, we showed that there were more males, middle-aged people, and old-aged people discussing mental health-related topics on Twitter during the pandemic in the US. Besides gender and age, our study also estimated race/ethnicity information for Twitter users who tweeted about mental health during the pandemic, which provides a more comprehensive picture of the demographic portfolios of Twitter users having mental health concerns during the pandemic.

*4.3. Limitations.* In this study, the mental health concerns on Twitter during the COVID-19 pandemic do not necessarily mean that these Twitter users had a mental illness. The keywords that we used for mental health concerns are relatively limited, which might introduce some biases. Another limitation lies in our demographic analysis. In our study, only a small proportion of users shared a valid human face as their profile pictures and a valid name. Of the 101,481 Twitter users we used for inference, only 11,330 (11%) users have valid names and profile pictures. Even if it is a valid user, there is no guarantee that they are using photos and names of their own. In addition, due to the technical issue, we failed to collect the relevant Twitter data from May 18th, 2020, to May 19th, 2020, and from August 24th, 2020, to September 14th, 2020. Therefore, our data did not represent the whole population in the US. Due to lack of the distribution of Twitter users at the state level, we normalized the number of Twitter users who tweeted about mental health to the state population, which might introduce some biases.

## 5. Conclusions

During the COVID-19 pandemic, social media is one of the most popular platforms for the public to share their feelings. Our study successfully monitored the discussions surrounding mental health during the pandemic. As these topics revealed some causes of mental anxiety, they provided some directions for where efforts should be put to reassure confidence in the people. Our demographic analysis implicated that White and Males are more likely to have/express mental health concerns. Thus, more attention could be provided to them when mental health support becomes available. Furthermore, our study demonstrated the potential of social media data in studying mental health issues.

## Ethical Approval

The study has been reviewed and approved by the University of Rochester Office for Human Subject Protection (Study ID: STUDY00006570).

## Consent

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

ZX and DL conceived and designed the study. SZ, LS, DZ, PL, YL, and ZX analyzed the data. SZ and LS wrote the manuscript. AA, ZX, and DL assisted with the interpretation of analyses and edited the manuscript. Senqi Zhang and Li Sun contributed equally to this study.

## Acknowledgments

## Supplementary Materials

*Supplementary 1.* Appendix Figure 1: flow chart of data preprocessing.

*Supplementary 2.* Appendix Figure 2: mental health-related keywords that were mentioned the most in COVID-19-related tweets in the US.

*Supplementary 3.* Appendix Figure 3: number of Twitter users who had mental health concerns over time in the US.

*Supplementary 4.* Appendix Figure 4: average number of mental health-related tweets per user in different US states.

*Supplementary 5.* Appendix Figure 5: age composition in different gender groups for Twitter users who had mental health concerns in the US.

*Supplementary 6.* Appendix Figure 6: age composition in different race/ethnicity groups for Twitter users who had mental health concerns on Twitter in the US.

## References

[1] "Covid-19 Data in Motion," 2021, https://coronavirus.jhu.edu/.

[2] The Lancet Microbe, "COVID-19 vaccines: the pandemic will not end overnight," *The Lancet Microbe*, vol. 2, no. 1, p. e1, 2021.

[3] K. Katella, "Comparing the COVID-19 Vaccines: How Are they Different?," 2021, https://www.yalemedicine.org/news/covid-19-vaccine-comparison.

[4] "Mental Illness," 2021, https://www.nimh.nih.gov/health/statistics/mental-illness.

[5] B. Pfefferbaum and C. S. North, "Mental health and the Covid-19 pandemic," *The New England Journal of Medicine*, vol. 383, no. 6, pp. 510–512, 2020.

[6] K. Usher, J. Durkin, and N. Bhullar, "The COVID-19 pandemic and mental health impacts," *International Journal of Mental Health Nursing*, vol. 29, no. 3, pp. 315–318, 2020.

[7] A. Kumar and K. R. Nayar, "COVID 19 and its mental health consequences," *Journal of Mental Health*, vol. 30, no. 1, pp. 1-2, 2021.

[8] S. Murata, T. Rezeppa, B. Thoma et al., "The psychiatric sequelae of the COVID-19 pandemic in adolescents, adults, and health care workers," *Depression and Anxiety*, vol. 38, no. 2, pp. 233–246, 2021.

[9] H. Estiri, Z. H. Strasser, G. A. Brat et al., "Evolving phenotypes of non-hospitalized patients that indicate long COVID," *BMC Medicine*, vol. 19, article 249, 2021.

[10] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in Twitter," in *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, Baltimore, Maryland USA, 2014.

[11] T. Hu, S. Wang, W. Luo et al., "Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective," *Journal of Medical Internet Research*, vol. 23, no. 9, article e30854, 2021.

[12] C. McClellan, M. M. Ali, R. Mutter, L. Kroutil, and J. Landwehr, "Using social media to monitor mental health discussions – evidence from Twitter," *Journal of the American Medical Informatics Association*, vol. 24, no. 3, pp. 496–502, 2017.

[13] W. Yang and L. Mu, "GIS analysis of depression among Twitter users," *Applied Geography*, vol. 60, pp. 217–223, 2015.

[14] J. X. Koh and T. M. Liew, "How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 Twitter feeds," *Journal of Psychiatric Research*, vol. 145, pp. 317–324, 2022.

[15] D. Valdez, M. Ten Thij, K. Bathina, L. A. Rutter, and J. Bollen, "Social media insights into US mental health during the COVID-19 pandemic: longitudinal analysis of twitter data," *Journal of Medical Internet Research*, vol. 22, no. 12, article e21418, 2020.

[16] Y. Gao, Z. Xie, and D. Li, "Electronic cigarette users' perspective on the COVID-19 pandemic: observational study using twitter data," *JMIR Public Health and Surveillance*, vol. 7, no. 1, article e24859, 2021.

[17] M. Hua, M. Alfi, and P. Talbot, "Health-related effects reported by electronic cigarette users in online forums," *Journal of Medical Internet Research*, vol. 15, no. 4, article e59, 2013.

[18] L. Chen, X. Lu, J. Yuan et al., "A social media study on the associations of flavored electronic cigarettes with health symptoms: observational study," *Journal of Medical Internet Research*, vol. 22, no. 6, article e17496, 2020.

[19] C. Zou, X. Wang, Z. Xie, and D. Li, "Public reactions towards the COVID-19 pandemic on twitter in the United Kingdom and the United States," 2020, medRxiv : the preprint server for health sciences.

[20] R. J. Gore, S. Diallo, and J. Padilla, "You are what you tweet: connecting the geographic variation in America's obesity rate to twitter content," *PLoS One*, vol. 10, no. 9, article e0133505, 2015.

[21] "NLTK 3.6.2 documentation," https://www.nltk.org.

[22] "Gensim," https://radimrehurek.com/gensim_3.8.3/auto_examples/index.html.

[23] "Industrial-Strength Natural Language Processing," https://spacy.io.

[24] C. Sievert and K. Shirley, "LDAvis: a method for visualizing and interpreting topics," in *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, Baltimore, Maryland USA, 2014.

[25] J. Messias, P. Vikatos, and F. Benevenuto, "White, man, and highly followed: gender and race inequalities in twitter," in *Proceedings of the International Conference on Web Intelligence*, Leipzig Germany, August 2017.

[26] "ethnicolr: Predict Race and Ethnicity From Name," https://ethnicolr.readthedocs.io/.

[27] G. Sood and S. Laohaprapanon, "Predicting race and ethnicity from the sequence of characters in a name," 2018, https://arxiv.org/abs/1805.02109.

[28] B. Auxier and M. Anderson, "Social Media Use in 2021," 2021, https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

[29] C. Wang, R. Pan, X. Wan et al., "A longitudinal study on the mental health of general population during the COVID-19 epidemic in China," *Brain, Behavior, and Immunity*, vol. 87, pp. 40–48, 2020.

[30] C. Jacques-Aviñó, T. López-Jiménez, L. Medina-Perucha et al., "Gender-based approach on the social impact and mental health in Spain during COVID-19 lockdown: a cross-sectional study," *BMJ Open*, vol. 10, no. 11, article e044617, 2020.

[31] O. El-Gayar, A. Wahbeh, T. Nasralah, A. El Noshokaty, and M. A. Al-Ramahi, "Mental Health and the COVID-19 Pandemic: Analysis of Twitter Discourse," in *Twenty-Seventh Americas Conference on Information Systems*, Montreal, 2021.

[32] S. C. Guntuku, G. Sherman, D. C. Stokes et al., "Tracking mental health and symptom mentions on twitter during COVID-19," *Journal of General Internal Medicine*, vol. 35, no. 9, pp. 2798–2800, 2020.

[33] Y. Zhang, H. Lyu, Y. Liu, X. Zhang, Y. Wang, and J. Luo, "Monitoring depression trend on Twitter during the COVID-19 pandemic," 2020, https://arxiv.org/abs/2007.00228.

[34] H. Lyu, L. Chen, Y. Wang, and J. Luo, "Sense and sensibility: characterizing social media users regarding the use of controversial terms for COVID-19," *IEEE Transactions on Big Data*, vol. 7, pp. 952–960, 2021.