Exploring experiences of COVID-19-positive individuals from social media posts

Jia-Wen Guo PhD, RN | Shawna M. Sisler MS, NP | Ching-Yu Wang MS, RN | Andrea S. Wallace PhD, RN, FAAN

Abstract

Aims: This study aimed to explore the experience of individuals who claimed to be COVID-19 positive via their Twitter feeds.

Background: Public social media data are valuable to understanding people’s experiences of public health phenomena. To improve care to those with COVID-19, this study explored themes from Twitter feeds, generated by individuals who self-identified as COVID-19 positive.

Design: This study utilized a descriptive design for text analysis for social media data.

Methods: This study analysed social media text retrieved by tweets of individuals in the United States who self-reported being COVID-19 positive and posted on Twitter in English between April 2, 2020, and April 24, 2020. In extracting embedded topics from tweets, we applied topic modelling approach based on latent Dirichlet allocation and visualized the results via LDAvis, a related web-based interactive visualization tool.

Results: Three themes were mined from 721 eligible tweets: (i) recognizing the seriousness of the condition in COVID-19 pandemic; (ii) having symptoms of being COVID-19 positive; and (iii) sharing the journey of being COVID-19 positive.

Conclusion: Leveraging the knowledge and context of study themes, we present experiences that may better reflect patient needs while experiencing COVID-19. The findings offer more descriptive support for public health nursing and other translational public health efforts during a global pandemic.

KEYWORDS
coronavirus, patient experience, public health, resilience, text analysis

Summary statement
What is already known about this topic?

- Social media data can be used to predict potential outbreak areas, which has been beneficial to informing decision makers.
- Public opinions about the pandemic can be extracted from social media data, which helps understand patients and the public’s experiences and potential needs.
- In the early stage of the novel COVID-19 pandemic in the United States, little was known about the experiences and needs of those with COVID-19.
What this paper adds?

- Individuals who claimed to be COVID-19 positive used the social media platform (e.g., Twitter) as a broader communication strategy to their social networks during the social distancing phase at the early stage of the COVID-19 pandemic.
- Key messages delivered through the social media included warning their family, friends and network take the pandemic seriously, expressing their symptom profiles related to COVID-19 and sharing their journey of being a COVID-19 patient.
- Although latent Dirichlet allocation may be underutilized in nursing research, valuable information regarding mental health issues can be discovered and amplified from the social media data using this methodology.

The implications of this paper:

- Substantial evidence encourages the public health community to engage in more detailed mental health screening of patients with COVID-19.
- By leveraging the knowledge and context of the social media threads (e.g., patient perspective), there is ample opportunity for more focal and descriptive public health interventions to mitigate the stress and uncertainty of a pandemic.

1 | INTRODUCTION

The wide-scale interaction and massive real-time data of social media users provide unprecedented opportunities for public health research and practice. Within the current COVID-19 crisis, robust data suggest that social media data were helpful for not only disseminating information and sharing research knowledge in a timely and rapid manner to the public but also for supporting the surveillance of diseases or symptoms (Cuello-Garcia et al., 2020; González-Padilla & Tortolero-Blanco, 2020; Guo et al., 2020). As such, researchers have been able to utilize social media search indices for COVID-19 symptoms to predict potential outbreak areas, which has been beneficial to informing decision makers and thus limiting the spread of COVID-19 (Qin et al., 2020). During the stricter COVID-19 social distancing period (since April 2020), social media played a critical role in maintaining interaction and communication with their family, friends and others. Social media data from this time period offer a valuable perspective to better understand public perceptions and opinions towards COVID-19 and a window into individual health profiles and symptom reporting.

As the first country hit by COVID-19 in December 2019, researchers from China quickly analysed their COVID-19 data posted via Weibo, a popular Twitter-like social media site in the country (Han et al., 2020; Li et al., 2020). Public opinions were extracted in social media messages posted by individual users and by organizations (e.g., government agencies gathered during the peak of China’s COVID-19 pandemic; Han et al., 2020; Li et al., 2020). Themes that emerged from these studies included a strong sense of public uncertainty, fear and worry, knowledge and attitudes about COVID-19, symptom reports, health-protective behaviours and risk behaviours (Han et al., 2020; Li et al., 2020). Although such context is important, the collected data lack information shared by those claiming to have tested positive for COVID-19. Thus, information discussed or shared by those diagnosed with COVID-19 may prove to be useful in developing interventions to improve care during the pandemic. Public health nursing stands in an important and unique position to bring the voices of our patients and community forward to better inform care practices.

As clinicians and researchers, we anticipate a myriad of public health problems as a result of the COVID-19 pandemic (Edmonds et al., 2020). By understanding the experiences and needs of the general public, public health nursing practice is well positioned to address these challenges and build more patient-centred solutions in response. Given strict social distancing and quarantine requirements during the COVID-19 pandemic, social media data may be one of few ways to feasibly understand experiences specific to those with COVID-19, especially during the earliest stage of the pandemic in the United States where public anxiety and the ‘unknowns’ of the virus were rampant. These initial responses to the novel coronavirus also help track public sentiment, concerns and symptom expression at a larger public scale. As such, the objective of this study was to explore themes gathered from social media posts generated by those who claimed they were COVID-19 positive as a way to explore people’s needs in coping with the pandemic.

2 | METHODS

2.1 | Aims

This study aimed to analyse Twitter tweets posted by users who claimed they were COVID-19 positive as an effort to gain insight and
context about how coping and information processing occurred in the earliest state of the pandemic within the United States.

2.2 | Design

This study utilized a descriptive design for text analysis for social media data according to steps suggested by Stiegitz and Dang-Xuan (2013) and Stiegitz et al. (2018). Using R program with the ‘rtweet’ package via the Twitter Application Programming Interface to retrieve tweets, we limited our search to English language posts in the United States made from April 2, 2020, to April 24, 2020 (23 days) in order to capture context and characteristics of the first round of peak social distancing in the United States.

2.3 | Sample and data collection

Our sample was built from search criteria that included synonyms of COVID-19 (e.g., corona, coronavirus, covid, covid-19 and ncov-19) and phrases that indicated a positive COVID-19 diagnosis (e.g., ‘I tested positive’, ‘I tested pos’, ‘am positive for’, ‘I’m positive with’, ‘I’m positive for’, ‘am positive with’, ‘I was positive for’ and ‘I was positive with’). Retweets, which are tweets disseminated by another user, were excluded from the sample. To ensure retrieved tweets mentioned individual experiences of being COVID-19 positive, we manually reviewed each collected tweet to include or exclude entries of users claiming as COVID-19 positive.

To conduct our text analysis, we initially cleaned tweets via four steps: (i) removing non-English characters (e.g., punctuation, numbers, hyperlinks and symbols); (ii) converting uppercase letters into lowercase characters; (iii) eliminating stopwords that are words with little or no meaning such as articles (e.g., a and an) and prepositions (e.g., on and at); and (iv) normalizing words by stemming, a process of reducing different word forms (e.g., feel feels feeling) to common roots (e.g., feel feel feel) for enhancing text mining process.

2.4 | Ethical considerations

This study does not involve any human participants; it utilizes a public, deidentified dataset. Therefore, the study is not considered human subjects research, as defined at 45 CFR 46.102 and does not require institutional review board (IRB) review. As with any public data, however, there is an ongoing need to evaluate contemporary questions involving possible ethics and human dignity concerns via privacy policies or informed consent. In order to protect the sampled population, we extracted the data from the general Twitter community (not from established subgroups in the Twitterverse) and proposed a descriptive study that was minimally invasive. All data were stored on a password-protected computer and used the Health Insurance Portability and Accountability Act (HIPPA) compliant storage.

2.5 | Data analysis

To extract embedded topical themes in collected tweets, we used a topic modelling approach based on latent Dirichlet allocation (LDA), a machine learning approach to analyse narrative texts. Topic modelling is a statistical method for discovering patterns and themes from the collected narrative data. LDA is ‘a generative model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar’ (Blei et al., 2003). According to LDA, each tweet is composed of a combination of hidden topics, and each topic is made up a set of words. Because words are generated by topics, meaningful words with high frequency in a topic represent the theme of the topic. Therefore, the LDA allows us to discover latent topics from the tweets and determine co-occurring words in each topic (Blei et al., 2003).

There are two critical steps of using topic modelling with LDA: determining the preferred number of latent topics and then naming each topic. To estimate the number of topics per sample, we estimated the number of topics from 2 to 20 by using the minimization approach, ‘CaoJuan2009’ (Cao et al., 2009); the minimum value of ‘CaoJuan2009’ suggests the optimal number of topics for a model (i.e., returns a number) from adaptively selecting an optimal LDA model based on its density (Cao et al., 2009). After the topic number was determined, we used LDA to generate the representative word list for each topic and the value of the probability distribution of topics for each tweet.

To interpret the theme for each latent topic, we reviewed the most frequent and relevant words in each topic, checking for the context and use of those words within the tweet. To facilitate this process, we used LDAvis (Blei et al., 2003), a visualization data analysis method. The output of LDAvis generates an interactive web-based presentation with two panels (see Figure 1 as an example). On the LDAvis left panel, an intertopic distance map presents how the various topics distance from each other and the proportion to which each topic relates to the overall topic of the tweets. The frequency is presented via the size of each circle (i.e., the circle size is proportional to the number of words in that given topic); then, the program plots the topics on a multidimensional scale to gauge how similar each topic is to the others. The right panel in the LDAvis dashboard lists the Top-30 Most Relevant Terms for Topics, presented by (i) the most frequent terms in the descending order in each topic and (ii) the most representative terms in each topic. In order to identify a representative term, we selected a term with a majority of the proportion of the red bar (the frequency of the term in its particular topic) in the blue bar (the frequency of the term in the overall texts) (Figure 1). The LDA output is dynamic, allowing for adjustments in the bar chart to explore relationships within the data corpus, paying specific attention to relevance and saliency within and between topics. All statistical analyses were conducted by R (Version 3.6.3) with packages (e.g., rtweet, tm, topicmodels, LDAvis and ldatuning). Ethics approval was not required for this study as it uses publicly available data exclusively.
A total of 1148 tweets posted from April 2, 2020, to April 24, 2020, were retrieved for initial analysis in this study sample. As Twitter offers no age, gender and other personal information, the demographic data of the sample were not able to describe in this study. The majority of excluded tweets were related to content that did not indicate any of the following: individual user’s experience as COVID-19 positive, tweets related to news or regulations for COVID-19, users assuming being COVID-19 positive in the future or users sharing stories about others who were COVID-19 positive. Of the remaining, 721 tweets were eligible for inclusion in this study; these tweets originated from 709 unique users.

From the topic modelling analysis, results of the ‘CaoJuan2009’ modelling suggested the optimal number of three total topic groups for this sample. According to intertopic distance map generated by LDAvis for a three-topic model, the three topics remain distinct from each other because they do not overlap (Figure 1). Overall, the proportion of each of the represented topics is 46.2% for Topic 1 (Figure 1), 31.0% for Topic 2 (Figure 2) and 22.8% for Topic 3 (Figure 3). The most representative terms in Topic 1 were ‘stay’, ‘home’, ‘social’, ‘distancing’, ‘save’, ‘life’, ‘safe’, ‘live’, ‘serious’, ‘fine’, ‘family’ and ‘friend’. In Topic 2, these terms included ‘symptom’, ‘fever’, ‘cough’, ‘short’, ‘breath’, ‘chest’, ‘pain’, ‘ache’, ‘smell’, ‘taste’ and ‘headache’. The most representative terms in Topic 3 were ‘health’, ‘scare’, ‘quarantine’, ‘finally’, ‘wait’, ‘result’, ‘donate’, ‘antibodies’ and ‘blood’.

Topic 1 presents the ideas of the seriousness of COVID-19 and the importance of staying home and social distancing; moreover, terms of ‘family’ and ‘friend’ were particularly salient in the sample. Topic 2 houses many keywords related to COVID-19 symptoms experiences (e.g., fever and loss of smell and taste). Topic 3 revolved around individual experiences of being COVID-19 positive, such as the feelings of uncertainty surrounding COVID-19 diagnosis and their relationship with the course of the virus (e.g., being quarantine, fear, waiting or getting test results, the experience of being sick and wanting to contribute to research). After reviewing the tweets classified in each topic (Table 1), we named these three topics as ‘recognizing the seriousness of the condition in COVID-19 pandemic’ for Topic 1, ‘having symptoms of being COVID-19 positive’ for Topic 2 and ‘sharing the journey of being COVID-19 positive’ for Topic 3.

The three extracted topical themes present a myriad of patient experiences being COVID positive and can be used to reflect patients’ needs, which in turn can be leveraged to better support public health actions during this global pandemic. At the early stage of the pandemic, very little was known about COVID-19. Most reported COVID-19 patients were only asked about flu-like symptoms, and the virus was assumed to be dangerous to only elderly and vulnerable
**FIGURE 2** Intertopic distance map generated by LDAvis for Topic 2, ‘having symptoms of being COVID-19 positive’ (e.g., size of red circle presents the proportion of Topic 2 information in the overall texts; each red bar presents a term frequency in Topic 2)

**FIGURE 3** Intertopic distance map generated by LDAvis for Topic 3, ‘sharing the journey of being COVID-19 positive’ (e.g., size of red circle presents the proportion of Topic 3 information in the overall texts; each red bar presents a term frequency in Topic 3)
groups, even though we now know that children and young adults are also at risk for COVID-19 (Glasper, 2020). As such, this study’s sample has a heavy emphasis on countering early misunderstandings about the severity of the condition by communicating its seriousness and by disseminating information to their friends and family about infection control actions such as social distancing, sheltering in place and hand hygiene. It also offers an original methodology to harness social media data and application in translational research. Nurses and providers stand as healthcare’s front line in navigating patient-reported symptoms. Finding a medium to utilize these dynamic findings may help flag emerging issues or concerns for infectious disease tracking and research.

The study’s findings showed that a significant number of individuals used social media platforms as a broader communication strategy to warn their family, friends and network to take the pandemic seriously, expressing novel profiles of symptoms and experiences that mainstream media and research were just starting to explore. Practical information about symptoms offered by this sample largely aligns with symptoms reported through public health agencies (Centers for Disease Control and Prevention, 2020) and likely indicates the success of education at testing sites and of public health outreach and resources. The fact that Twitter users openly and spontaneously tweet about their specific symptoms provide further opportunity to explore disease spread and emerging symptoms on a population level and to identify other COVID-19-related symptoms, which have not yet been listed by public health agencies (Guo et al., 2020).

In communicating the overall experience of being COVID-19 positive, the wide range of experiences from screening, diagnosis, to recovery offers important context to the struggles and needs of patients and families affected by the pandemic. Symptoms of mental distress related to waiting for screening results or being diagnosed as COVID-19 positive were presented, implying there is a need for assessing the mental health status of patients with COVID-19. Even in general populations of those not experiencing COVID symptoms, there are still substantial mental health ramifications; it is estimated that half of the general public are experiencing depression and/or anxiety during the pandemic, one third of whom are experiencing severe symptoms (Wang et al., 2020). It is clear that a targeted mental health strategy is needed for those with a COVID-19 diagnosis (Bo et al., 2021; Holmes et al., 2020; Hutchings et al., 2020), but we also need to recognize the need for mental health interventions in the general population.

On an individual level, social media data such as those used in this study may help clinical public health nurses prepare patients for their illness experience and empower patients towards resiliency and recovery. Looking towards nursing theory, COVID-19 tweets can be broadly categorized into protecting, compensating and challenging coping concepts (Morse et al., 2021). Protecting concepts are internal or external resources that improve the adverse event (e.g., COVID-19 positive); compensatory concepts are the actions taken to counteract the adverse event; challenging concepts are the personal efforts that bring an individual to overcome the adverse event. In this study,

### Table 1: Three topical themes and their tweet examples

| Topic                                                        | Tweet examples                                                                                                                                                                                                 |
|---------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Recognizing the seriousness of the condition in COVID-19 pandemic | ‘Hey people, stay home. I tested positive for the covid but part of my job is to take care of sick people and unless you do the same or are considered an essential worker you should stay home to protect your family, friends and family members! It’s all LOVE ...’<br>‘Hey friends and family, I tested positive for COVID. Spirits are high. My family and I are doing great. This is just my way of telling everyone to stay safe and get tested if you can. This isn’t about you it’s about US. Much love ...’<br>‘I tested positive for COVID-19 if you work in a hospital or any health care facility I suggest you get tested and practice social distancing. That can save somebody you know life. I been practicing and; still caught it this shit not a game ...’<br>‘The study’s findings showed that a significant number of individuals used social media platforms as a broader communication strategy to warn their family, friends and network to take the pandemic seriously, expressing novel profiles of symptoms and experiences that mainstream media and research were just starting to explore. Practical information about symptoms offered by this sample largely aligns with symptoms reported through public health agencies (Centers for Disease Control and Prevention, 2020) and likely indicates the success of education at testing sites and of public health outreach and resources. The fact that Twitter users openly and spontaneously tweet about their specific symptoms provide further opportunity to explore disease spread and emerging symptoms on a population level and to identify other COVID-19-related symptoms, which have not yet been listed by public health agencies (Guo et al., 2020).’ |
| Having symptoms of being COVID-19 positive                     | ‘I tested positive for COVID-19 on March 23rd. I had fever, cough, chest pain, shortness of breath, neck pain, body aches. I think what helped me was that I took steam showers and drank warm liquids and room temperature water. I’m kicking its ass and beating it!’<br>‘Six days ago I tested positive for Covid-19. Today was my first day of having a solid meal, being able to taste and smell. My main symptoms have been fatigue, stuffy nose, lose of taste and smell. Everyday is a little better ...’<br>‘So, I developed cough, headache, and muscle aches the other night. Plus I was feeling really terrible yesterday. Since I take care of only positive COVID-19 patients, I decided to get tested today. My results came back, I am positive for COVID-19 ...’ |
| Sharing the journey of being COVID-19 positive                 | ‘Good news for me and my family. Last week I tested positive for Covid antibodies, confirming I cleared the virus after weeks without symptoms. I also signed up for a clinical trial to donate my blood plasma and expect to make the donation in the coming weeks ...’<br>‘Hi I’m 19 y.o and I tested positive with covid-19. I did the test few months back when this all started and got negative results, but then few weeks ago I started feeling like I had the really bad flu and then yesterday I got my results ...’<br>‘Last week I tested positive to COVID 19 and I’m so thankful that I’ve fully recovered from the virus. I will be returning to work tomorrow after 18 days off and I’d be lying if I said I wasn’t scared ...’ |
protection, hope, social support, and spirituality. COVID-19-positive patients can bear the fact of being diagnosed positive by utilizing these protection concepts. In terms of the compensatory concepts, the tweets suggested that patients will use humor, self-care, and mindfulness. Patients used these concepts to keep themselves stay on top of their condition and focus on recovery. Finally, the challenging concepts are knowledge, courage, patience, and meaning finding. These concepts give patients the power to endure the process of recovery and move towards equanimity (Morse et al., 2021). These findings might provide a preliminary understanding of being COVID-19 positive and, ultimately, help lead patients towards recovery.

In tweets from those recovered from COVID-19, there was a notable prosocial message—they wanted to make contributions to science or clinical trials and a strong desire to ‘give back and get involved’. In general, people who are infected by virus generate long-term antibodies against viral pathogens. However, the scientific community is still uncertain to what extent this applies to the novel COVID-19 coronavirus, as it requires substantial research and time to process biological samples from COVID-19-positive cases (Du et al., 2020; Kang et al., 2020). In our sample, the most common desire from these individuals who claimed to be COVID-19 positive expressed a willingness to make contributions to COVID-19 research via the donation of blood samples.

Armed with knowledge from this study, nurses can act as an amplifier to these efforts, whether it pertains to blood banking, mental health or related interventions. In the example above, those who recovered from the COVID-19 infection expressed desires to contribute to science by signing up for a clinical trial and educating the public for risk mitigation. It would also be helpful for social media platforms to assist organizations or researchers who are focusing on COVID-19 research to recruit participants as well as find local dissemination channels for clinical trial information to connect with those interested in being research subjects. By leveraging findings such as these, nursing science has the opportunity to extrapolate on this application, build more precise surveillance systems for anticipating patient care needs and generate more knowledge and context for the greater public health intervention.

4.1 Study limitations

There are limitations associated with using social media data that should be considered when interpreting the results of this study. First, we were unable to verify the status of those claiming to be COVID-19 positive in the social media platform, as the data collected were self-reported. Second, we did not account for what are likely important factors influencing social media users’ opinions, such as personal (e.g., geographical location and race/ethnicity) and clinical (e.g., hospitalized vs. recovering at home and duration of the illness) characteristics. Having such data would allow us to determine how their opinions may be influenced by their physical characteristics or their status of the disease. Finally, we analyzed only the initial social distancing wave in April 2020. With the pandemic’s trajectory since April 2020, there is certainly more opportunity to widen the timeline to include the entire timeline of COVID-19 in the United States.

For the scope of this study, we intentionally focused on the early communication themes and underlying social processes regarding COVID-19’s symptom presentation and patient experience. We believe that these limitations should not significantly affect the interpretations of our results, and the methodology remains sufficient to allow the study to present the experiences of being COVID-19 positive. Topic themes identified in the study add detail to the experience of COVID-19 for those with symptoms and highlight a novel, useful perspective for public health nursing to explore as the pandemic continues.

5 Conclusion

Aggregating self-reported patient experiences in the earliest days of COVID-19 via social media may offer both researchers and clinicians important insight into the struggles and symptoms of individuals reporting being COVID-19 positive. This study describes using the social media data as an approach to better track public sentiment and articulating surrounding the earliest sentiment and symptoms with the pandemic; this approach may help us align future intervention efforts with newer outbreaks. By utilizing the knowledge and context of the study themes, the findings offer more descriptive support for public health nursing and other translational public health efforts. As such, nursing practice and greater public health efforts will have a richer understanding of patient needs, as well as more alignment for interventions and programming. Future work may include investigating the tweets from those who have recovered from COVID-19 to understand and address their potential needs.

Funding Information

JG’s effort was supported by Grant Number K01NR016948 from the NIH National Institute of Nursing Research (PI: JG).

Conflict of Interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Authorship Statement

JG conceptualized and designed the study and collected the data and statistical analysis. JG and ASW analyzed and interpreted the data. JG, SMS, CW, and ASW drafted the manuscript. All the authors critically revised the manuscript for important intellectual content and approved the final version for submission.

ORCID

Jia-Wen Guo https://orcid.org/0000-0002-4698-4696
Shawna M. Sisler https://orcid.org/0000-0001-7076-6464
Ching-Yu Wang https://orcid.org/0000-0002-2984-6128
Andrea S. Wallace https://orcid.org/0000-0001-9694-2403
