# Texas Public Agencies' Tweets and Public Engagement during the COVID-19 Pandemic: Natural Language Processing Approach

Lu Tang, Wenlin Liu, Benjamin Thomas, Hong Thoai Nga Tran, Wenxue Zou, Xueying Zhang, Degui Zhi

Submitted to: JMIR Public Health and Surveillance
on: December 22, 2020

# Table of Contents

**Original Manuscript** 5
**Supplementary Files** 27
Figures 28
Figure 1 29

# Texas Public Agencies' Tweets and Public Engagement during the COVID-19 Pandemic: Natural Language Processing Approach

Lu Tang[1] PhD; Wenlin Liu[2] PhD; Benjamin Thomas[3]; Hong Thoai Nga Tran[4] MSc; Wenxue Zou[1] MA; Xueying Zhang[5] PhD; Degui Zhi[4] PhD

[1]Department of Communication Texas A&M University College Station US
[2]Jack J. Valenti School of Communication University of Houston Houston US
[3]Department of Computer Science Rice University Houston US
[4]School of Biomedical Informatics University of Texas Health Science Center at Houston Houston US
[5]Department of Journalism and Mass Communication North Carolina A&T State University Greensboro US

**Corresponding Author:**
Lu Tang PhD
Department of Communication
Texas A&M University
Texas A&M University
College Station
US

## *Abstract*

**Background:** The ongoing COVID-19 pandemic is characterized by different morbidity and mortality rates across different states, big cities and rural areas, and diverse neighborhoods within the same cities. The absence of a national strategy in battling the pandemic also leaves state and local governments responsible for creating their own response strategies and policies.

**Objective:** This study examines the content of the tweets sent by public health agencies in Texas about COVID-19 and how such content predicts the level of public engagement.

**Methods:** All COVID-19 related tweets (n=7269) posted by Texas public agencies were downloaded. These tweets were classified in terms of each tweet's functions (whether the tweet provides information, promotes action, or builds community), preventative measures mentioned, and health beliefs discussed using natural language processing. Hierarchical linear regressions were run to explore how tweet content predicted public engagement.

**Results:** Information was the most prominent function, followed by action and community. Susceptibility, severity, and benefits were the most frequently covered health beliefs. Tweets serving the action function was most likely to be retweeted, while tweets performing the action and community functions were more likely to be liked. Tweets communicating susceptibility and severity information led to more public engagement.

**Conclusions:** Public health agencies should continue to use Twitter to disseminate information, promote action, and build communities. They need to improve social media message strategies regarding the benefit of disease prevention behaviors.

(JMIR Preprints 22/12/2020:26720)
DOI: https://doi.org/10.2196/preprints.26720

### Preprint Settings

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

# Original Manuscript

# Texas Public Agencies' Tweets and Public Engagement during the COVID-19 Pandemic: Natural Language Processing Approach


## Abstract

**Background:** The ongoing COVID-19 pandemic is characterized by different morbidity and mortality rates across different states, cities, rural areas, and diverse neighborhoods. The absence of a national strategy in battling the pandemic also leaves state and local governments responsible for creating their own response strategies and policies.

**Objective:** This study examines the content of the tweets sent by public health agencies in Texas about COVID-19 and how content characteristics predict the level of public engagement.

**Methods**: All COVID-19 related tweets (*n*=7269) posted by Texas public agencies during the first six months of 2020 were classified in terms of each tweet's functions (whether the tweet provides information, promotes action, or builds community), preventative measures mentioned, and health beliefs discussed using natural language processing. Hierarchical linear regressions were run to explore how tweet content predicted public engagement.

**Results**: Information was the most prominent function, followed by action and community. Susceptibility, severity, and benefits were the most frequently covered health beliefs. Tweets serving the information or action functions were more likely to be retweeted, while tweets performing the action and community functions were more likely to be liked. Tweets communicating susceptibility information led to most public engagement in terms of both retweeting and liking.

**Conclusions**: Public health agencies should continue to use Twitter to disseminate information, promote action, and build communities. They need to improve social media message strategies regarding the benefit of disease prevention behaviors and audiences' self-efficacy.

**Texas Public Agencies' Tweets and Public Engagement during the COVID-19 Pandemic: Natural Language Processing Approach**

## Introduction

Coronavirus disease 2019 (COVID-19) is a new infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a new and potentially deadly coronavirus. The ongoing COVID-19 pandemic is characterized by different morbidity and mortality rates across different states, cities, rural areas, and diverse neighborhoods. The absence of a national strategy in battling the pandemic also leaves state and local governments responsible for creating their own response strategies and policies [1]. Meanwhile, misinformation and disinformation circulate on social media with unprecedented volume and velocity, which in turn affect the public trust in and response to governmental restrictions and corrective actions [2,3]. Thus, it is crucial to examine how state and local health departments communicate with their stakeholders on social media.

Public health agencies have been actively using platforms such as Twitter and Facebook to communicate with their stakeholders during public health crises. The accumulating literature on organizational social media use has identified three primary functions: information, action, and community [4]. The information function refers to the organizational use of social media to provide the public with emergency and risk information [5]. It includes a wide range of activities such as making emergency updates, advisories, and warnings; providing scientific explanations and public education; and clarifying misinformation about an unfolding epidemic [6]. The action function means organizations can mobilize followers to adopt or avoid certain behaviors through the use of social media [4], such as attending events, making monetary donations, volunteering, and adopting other recommended behaviors. In the context of health and risk communication, action-oriented messages may specify how individuals can protect themselves when an imminent threat arises, and this function is directly related to the overarching goal of public health agencies to mitigate risk behaviors during an epidemic [7]. Finally, the community function revolves around building relationships with community members, providing social, emotional support, and communicating about collective identities. Providing emotional support and boosting community morale can enhance public trust and cooperative behaviors [8], both of which are essential for effective risk mitigation. While health agencies are generally advised to leverage multiple social media functions as outlined above, a large body of earlier work suggests that most public agencies' social media content was one-way information dissemination [9]. Thus, we propose the first research question about the functions of public agencies' tweets during the COVID-19 pandemic:

*RQ1: To what extent do public health agencies' Twitter messages fulfill the functions of information, action, and community during the COVID-19 pandemic?*

According to the Health Belief Model (HBM), whether a person adopts a recommended health behavior is influenced by her desire to avoid an illness and her belief that the recommended behavior can help prevent the illness [10]. Two beliefs affect one's desire to avoid an illness: perceived severity and perceived susceptibility. When a person thinks that the illness is serious (perceived severity) and she has a high chance of getting it (perceived susceptibility), she will be more alarmed and want to avoid the illness. Meanwhile, an individual's preventative behavior is also influenced by her beliefs about (1) whether the recommended behavior can indeed provide health benefits, such as preventing the illness (perceived benefits), (2) obstacles associated with adopting the recommended behavior, such as cost and time (perceived barriers), and (3) one's belief about her ability to engage in the behavior (self-efficacy). A meta-analysis study of decades of research using the HBM indicates that perceived benefits and perceived barriers are the strongest predictors of behavioral change [11].

The original HBM was a psychological model created to predict an individual's health behaviors. It has recently been used to guide the design of health messages to effectively promote health behaviors and the evaluate the presence or absence of elements in media content that might contribute to these health beliefs [12]. Understanding the extent to which public health agencies' tweets addressed different health beliefs could offer insights into how these tweets might inform the public about the threats of COVID-19 and encourage proper preventative measures. Hence, we asked the next set of RQs.

*RQ2a: What are the preventative behaviors recommended?*

*RQ2b: To what extent do public health agencies communicate severity, susceptibility, benefits, barriers, and self-efficacy in their Twitter messages about COVID-19?*

In addition to behavioral outcomes, public engagement is another indicator of the effectiveness of public agencies' crisis communication efforts. Public engagement refers to the various forms of communicative interaction between the public and government agencies, such as the public sharing of or replying to governmental agencies' communication [13]. Public engagement has several benefits. First, greater public engagement with public health agencies' social media content typically indicates a higher level of exposure, attention, and information absorption of the content communicated (e.g., advisories, warnings, or other educational material), which is essential in helping the public to form accurate risk perception and encourage risk-reduction behaviors [14]. Second, public engagement could be an indicator or precursor of institutional health-related trust, which further leads to better health adherence and other positive behavioral changes [15]. Finally, public engagement thus could help identify, clarify, and correct misinformation to assist more effective health promotion [16]. Although public engagement is generally associated with positive outcomes, it should be noted that scholars distinguish between positive and negative engagement, suggesting the latter may lead to "denial, rejection, avoidance and negative word-of-mouth" of an organization [17]. In the context of crisis, for example, it has been found that certain types of engagement may breed misinformation and undermine the authority of crisis management agencies [18].

We adopt Johnson and Taylor's conceptualization of public engagement at the individual level as the psychological and behavioral involvement and participation with public health agencies' messages [19]. In social media-mediated crisis communication, such individual-level engagement is manifested in two forms, the public's re-sharing behavior on social media [19,20], and the behavior of "liking" or showing endorsement of public agencies' social media messages. The first form of engagement, sharing public agencies' social media content with one's own social networks, is viewed as an important outcome of effective health risk communication. Individuals' social media sharing behavior as a key mechanism that enables the amplification of public health agencies' messages [21]. By sharing these messages via functions such retweets, the public not only relays relevant health contents to their immediate communities, but the collective sharing behaviors can generate normative influence leading to intended behavioral change [22]. Meanwhile, endorsing public health agencies' messages through Twitter's "favorite" or Facebook's "like" function has been examined as a distinct form of public engagement from re-sharing [23,24]. Specifically, the endorsement behavior has been conceptualized as a kind of affective engagement, which indicates the audience's feeling of support or symbolic alignment with the organization on a specific issue [25]. Although endorsement does not fully equate to the psychological acceptance of the message, research suggests that the positive assessment is significantly associated with health message acceptance, especially when such endorsement is made by celebrities [26]. We thus ask:

*RQ3: How do the features of tweets predict public engagement in terms of the numbers of favorites and retweets during the COVID-19 pandemic?*

# Method

## Sampling and Data Collection

The current study focused on public agencies in the state of Texas. Texas was chosen first because the state was one of the disease epicenters following Governor Abbott's state re-opening measures in April 2020. At the time of data collection (mid July 2020), Texas was facing the second peak of COVID cases with the highest seven-day average of daily new cases (*n*=15,038) [27]. In addition, with Texas being the second largest and second most populous state in the U.S., its public agencies may face a particularly challenging task of reaching diverse population and coordinating among peer agencies. Since this study examined the public tweets of governmental agencies, the study was exempt from Human Subjects Ethics Review.

We took the following steps to select the sample tweets for analysis. First, we identified all active Twitter accounts of public health departments and the Office of Emergency Management Organizations (OEMs) at city, county, and state levels in Texas. To identify public health departments, we started with the list of health department directories through the CDC and the U.S. Department of Health and Human Services (HHS). Additionally, a list of local-level health agencies was obtained from the National Association of County and City Health Officials (NACCHO). This step identified a total of 26 Texas public health departments that actively tweeted during the studied period. We also used a list of Texas city and county names to search on Twitter and identified an additional 56 OEMs' official Twitter accounts, generating a total of 82 organizations. Second, we created a list of 25 COVID-19 related keywords (covid, corona, koronavirus, ncov, sars, pandemic, epidemic, quarantine, outbreak, handwash, wuhan, panic, chinese virus, lock down, sheltering in place, shelter in place, flatten the curve, safer at home, stay home, face covering, wear mask, get tested, quarantine, ppe, and n95). All tweets containing at least one of these keywords from the above 82 organizations published between January 1 and June 30, 2020 were downloaded using Twitter's developer API (*n* = 15,382).

## Measurement

A codebook was developed to guide the coding of the training dataset. It included the following variables: functions, types of actions recommended, and HBM variables. Each tweet was coded in terms of the presence or absence of COVID-19 related contents. Those tweets containing COVID-19 related content were further coded.

First, each tweet was coded in terms of the functions it served: (1) information: the tweet shared information about COVID-19, such as the symptoms, risks of the disease, prevention information, current infection rate/cases, testing information; or the tweet describes actions that agencies are taking to contain the disease spread, (2) action: the tweet urged readers to adopt a certain health behavior, and (3) community: the tweet built community by asking readers to interact with each other and with the sender, providing emotional support, and boosting morale (adapted from [25]). Each tweet was evaluated in terms of whether it contained any of these three types of information.

Second, each tweet was coded in terms of whether it included one or more of the following actions: (1) handwashing, (2) social distancing, (3) mask/face covering, (4) staying at home/shelter in place, (5) getting tested, (6) learning more information, and (7) other behaviors.

Finally, HBM variables were coded, including severity (any reference to the magnitude and seriousness of COVID-19), susceptibility (the likelihood that a person, a group, or the public in general will contract COVID-19), benefits (benefits of recommended behaviors, its effectiveness in prevention or treatment, or in containing the pandemic on the societal level), barriers (the difficulties associated with adopting or implementing the behaviors recommended), and self-efficacy (highlighting one's ability to engage in recommended behavior) (adapted from [12]).

## Development of Training Dataset

Several rounds of training sessions were held to assist the two coders to understand each item in the codebook. Afterwards, around 20% (*n*=3000) of the tweets were used for the development of a training data set. Two coders coded 150 tweets randomly selected from the remaining 80% of the tweet. They reached satisfactory intercoder reliability (Cohen's kappa ranges between .56 and .96, with a mean of .83). Two items (barriers, self-efficacy) were dropped from the codebook because they were nearly completely absent. Afterward, each coder independently coded half of the training dataset.

## Computer-assisted Classification Based on Natural Language Processing

Data cleaning was conducted first following the steps laid out in Du et al [28]. BERT, a natural language processing program developed by Google, was trained to automatically classify tweets [29]. The pretrained ‘BERT-large’ model from the Huggingface was used [30]. We divided the initial manually coded datasets (*n*=3000) into a training dataset (80%) and a testing dataset (20%). Of our training set, some labels had a relatively low frequency (<250 occurrences), which resulted in them being mostly ignored in the model’s training process. To train such low-frequency categories, we doubled all instances of tweets with minority labels, giving them a stronger signal in the model. Model was trained for three epochs, using AdamW optimizer with a learning rate of 2e-5.

Precision, recall, and overall F1 score (harmonic mean of precision and recall) were calculated for each variable. We also calculated the micro-averaging F1 score and macro-averaging F1 score to evaluate their performance on each classification task. We summed up all the individual true positives, false positives, and false negatives for the micro-averaged score. For the macro-averaged score, we took the average of the F1 score of different categories. Overall, our model achieved good results (see Table 1). Afterward, we used the program to classify all the tweets in the sample automatically.

**Table 1**. BERT Classification Performance of tweets published by Texas Public Health Agencies about the COVID-19 pandemic between January 1, and June 30, 2020.

| | Precision | Recall | F1-score |
|---|---|---|---|
| COVID or not | .93 | .93 | .93 |
| Information | .85 | .92 | .88 |
| Action | .68 | .83 | .75 |
| Community | .58 | .58 | .58 |
| Handwashing | .75 | 1.00 | .53 |
| Social distancing | .80 | .80 | .80 |
| Mask/Face covering | .85 | .96 | .90 |
| Staying at home/Shelter in place | .74 | .78 | .76 |
| Getting tested | .69 | .90 | .78 |
| Learning more information | .77 | .92 | .84 |
| Other behaviors | .27 | .54 | .36 |
| Severity | .69 | .92 | .79 |
| Susceptibility | .84 | .86 | .85 |
| Benefit | .42 | .70 | .52 |
| Micro average | .78 | .88 | .83 |
| Macro average | .70 | .83 | .76 |
| Weighted average | .80 | .88 | .84 |
| Sample average | .46 | .50 | .47 |

## Data Analysis

Hierarchical linear regressions, or stepwise linear regressions, were used to answer RQ3 (how various tweet features predicted the numbers of favorites and retweets). This method enabled the assessment of separate effects from different blocks of variables. Because both variables measuring engagement were highly skewed, we adopted the standard practice to log-transform these metrics before they were entered into regression models. In the two regression models, the independent variables consisted of the following three blocks: 1) the *information*, *action*, and *community* message type; 2) the dichotomous thematic categories including *social distancing*, *face-covering*, *shelter-in-place*, *testing*, *information seeking*, and *other* behaviors; and 3) the health belief variables including *severity*, *susceptibility*, and *benefit*. To control for the effect of account popularity (popular accounts were more likely to attract greater public engagement), we entered the log-transformed number of followers as a control variable in each model.

## Results

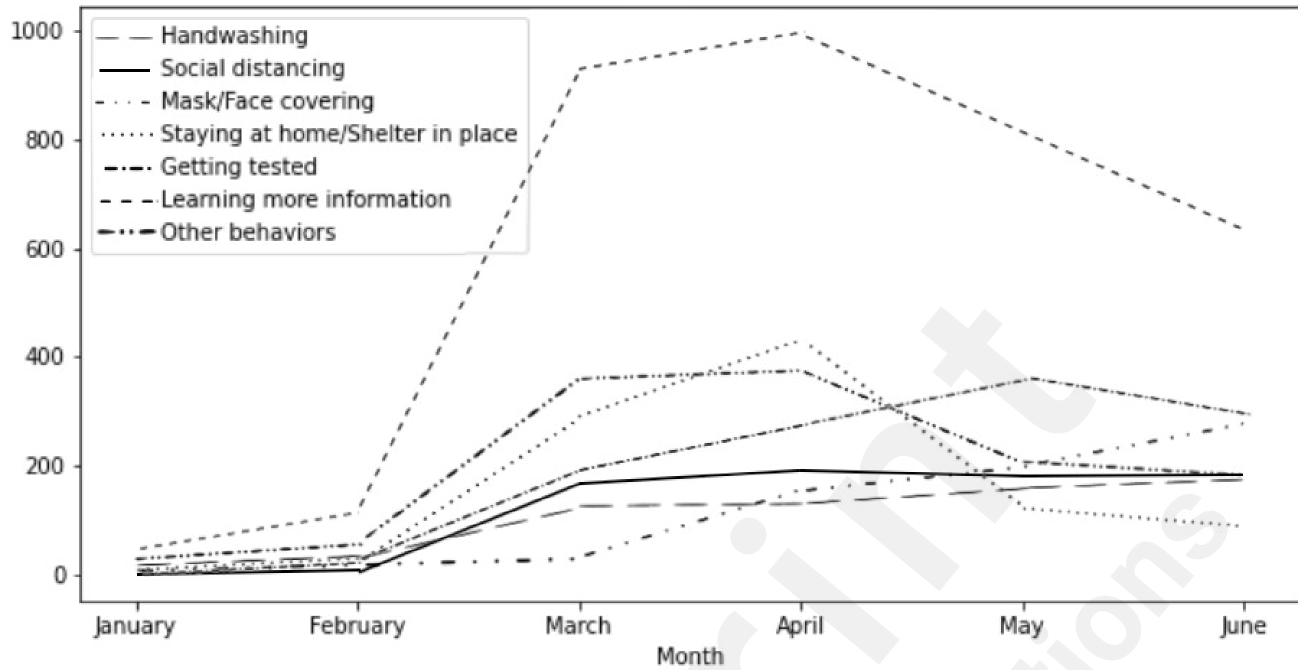
A total of 7,269 tweets were COVID-19 related. Among the 82 public health and EM agencies, only 61 tweeted about COVID-19 for an average of 119 times (*SD* = 203.09).

RQ1 asked the functions of tweets. Sharing information was the primary function of the tweets posted by public health agencies (94%, *n*=6835), followed by the action function (34.27%, *n*=2491). Community building was the least salient function as only 10.19% (*n*=741) of the tweets engaged community members and provided emotional support.

RQ2 asked the types of actions promoted and the health beliefs communicated. Among the behaviors recommended by agencies, learning more information was the most recommended action (*n*=3402, 46.8%), followed by getting tested (*n*=1076, 14.8%), staying at home/shelter in place (*n*=911, 12.53%), social distancing (*n*=700, 9.63%), face-covering (*n*=651, 8.96%), and handwashing (*n*=616, 8.47%). Figure 1 shows the number of tweets mentioning different health behaviors in public health agencies' tweets overtime. Handwashing was initially the most frequently recommended behavior, and its importance was continuously emphasized. Staying at home or sheltering in place saw the sharpest increase until April and dropped precipitously afterward. The mention of getting tested increased between February and May, but decreased in May and June. The number of tweets mentioning social distancing had plateaued since March. The discussion of wearing face coverings was minimal in the first three months but had been consistently increasing since March. In terms of HBM variables, severity (*n*=1389, 19.11%), susceptibility (*n*=2057, 28.3%), and benefits (*n*=1238, 17.03%) were the three concepts frequently mentioned in public health agencies' tweets.

**Figure 1**. Longitudinal Changes in the Number of Tweets Promoting Different Health Behaviors by

Public Health Agencies in Texas (January 1 to June 30, 2020)




RQ3 examined the relationship between the content of tweets and public engagement. Overall, the public engagement with the tweets sent by public health agencies was relatively low, with each tweet fetching an average of 13.05 retweets (*SD*=43.16) and 19 favorites/likes (*SD*=59.97). Table 2-3 presents the two hierarchical regression models predicting the two public engagement variables.

In terms of promoting public sharing or retweeting behaviors, tweeting fulfilling the functions of information and action were more likely to be retweeted. Tweets that contained face-covering, shelter-in-place, getting tested, and Covid-19 information seeking were also more likely to be retweeted, whereas those containing handwashing information were significantly less likely to be retweeted. Finally, severity and susceptibility significantly promoted retweeting tendencies.

In predicting the number of favorites received, the results showed slightly different patterns. Tweets that were primarily about action and community, rather than information, were more likely to receive favorites from the public. Meanwhile, content that included social distancing, shelter-in-place, and getting tested were more likely to be favorited, whereas referring to hand washing behaviors consistently reduced the chance of getting public favorites. Consistent with the other engagement indicator, severity and susceptibility health beliefs also significantly predicted the chance of getting public favorites.

**Table 2.** Hierarchical OLS regression of predictors on number of retweets (based on Texas Public Health Agencies' COVID-19 Related tweets between January 1 and June 30, 2020, *n*=7,269)

| | Number of Retweets | | | | | |
|---|---|---|---|---|---|---|
| | Model 1 | | Model 2 | | Model 3 | |
| Variables | β (SE) | *P* value | β (SE) | *P* value | β (SE) | *P* value |
| *Control* | | | | | | |
| Followers | .43 (.01) | < .001 | .43 (.01) | < .001 | .50 (.01) | < .001 |
| *Functions* | | | | | | |
| Information | .10 (.03) | < .001 | .09 (.02) | < .001 | .04 (.03) | < .001 |
| Action | .16 (.01) | < .001 | .09 (.01) | < .001 | .12 (.02) | < .001 |
| Community | .01 (.02) | .35 | .02 (.01) | .07 | .04 (.02) | .001 |
| *Types of actions proposed* | | | | | | |
| Handwashing | | | -.01 (.03) | .36 | -.05 (.03) | < .001 |
| Social distancing | | | .02 (.03) | .13 | .02 (.03) | .16 |
| Mask/face covering | | | .01 (.03) | .64 | .03 (.03) | .01 |
| Stay at home | | | .05 (.02) | < .001 | .04 (.02) | .001 |
| Get tested | | | .03 (.02) | .004 | .11 (.02) | < .001 |
| Learn more information | | | .05 (.01) | < .001 | .06 (.01) | < .001 |
| Other behaviors | | | -.04 (.02) | .01 | .03 (.02) | .08 |
| *HBM variables* | | | | | | |
| Severity | | | | | .10 (.01) | < .001 |
| Susceptibility | | | | | .20 (.01) | < .001 |
| Benefits | | | | | -.002 (.01) | .91 |
| Model of *f* values | 494.52 | < .001 | 186.74 | < .001 | 207.11 | < .001 |
| Change of $R^2$ | .21 | | .01 | | .07 | |
| Total $R^2$ | .21 | | .22 | | .29 | |

Note: β is standardized coefficients, and SE represents standard errors.

**Table 3.** Hierarchical OLS regression of predictors on number of favorites (based on Texas Public Health Agencies' COVID-19 Related tweets between January 1 and June 30, 2020, *n*=7,269)

| | Number of Favorites | | | | | |
|---|---|---|---|---|---|---|
| | Model 1 | | Model 2 | | Model 3 | |
| Variables | β (SE) | *P* value | β (SE) | *P* value | β (SE) | *P* value |
| *Control* | | | | | | |
| Followers | .50 (.01) | < .001 | .50 (.01) | < .001 | .55 (.01) | < .001 |
| *Functions* | | | | | | |
| Information | .05 (.03) | < .001 | .05 (.03) | < .001 | .01 (.03) | .31 |
| Action | .09 (.01) | < .001 | .09 (.02) | < .001 | .11 (.02) | < .001 |
| Community | .08 (.02) | < .001 | .08 (.02) | < .001 | .09 (.02) | < .001 |
| *Types of actions proposed* | | | | | | |
| Handwashing | | | -.03 (.03) | .049 | -.06 (.03) | < .001 |
| Social distancing | | | .03 (.03) | .03 | .03 (.03) | .04 |
| Mask/face covering | | | -.001 (.03) | .91 | .03 (.03) | .12 |
| Stay at home | | | .04 (.02) | < .001 | .04 (.02) | .004 |
| Get tested | | | -.01 (.02) | .58 | .05 (.02) | < .001 |
| Learn more information | | | -.01 (.01) | .64 | .01 (.01) | .56 |
| Other behaviors | | | -.03 (.02) | .07 | .02 (.02) | .14 |
| *HBM variables* | | | | | | |
| Severity | | | | | .04 (.01) | .008 |
| Susceptibility | | | | | .19 (.01) | < .001 |
| Benefits | | | | | -.01 (.01) | .62 |
| Model of *f* values | 707.52 | < .001 | 269.29 | < .001 | 243.51 | < .001 |
| Change of $R^2$ | .28 | | .003 | | .04 | |
| Total $R^2$ | .28 | | .28 | | .32 | |

Note: β is standardized coefficients, and SE represents standard errors.

## Discussion

Governmental agencies are among the most trusted sources of COVID-19 related information [31]. Public health agencies shoulder the responsibility of promptly communicating locally relevant pandemic updates, prevention guidelines, and relevant policies to the public during the COVID-19 pandemic. This study integrates the HBM and social media functions to understand how public health agencies in Texas communicate to the public about the pandemic via Twitter and assesses the empirical relationships between various message features and social media engagement outcomes. It finds that public health agencies used Twitter mostly for information sharing, followed by promoting action and community building. Tweets serving the action function was most likely to be retweeted and liked. Susceptibility, severity, and benefits were the most frequently covered health beliefs. Tweets communicating susceptibility and severity information led to more public engagement in terms of both retweets and endorsement.

Information is the most prominent function of these tweets, followed by action and community. This is consistent with the findings of an earlier study examining the tweets of Canadian public health agencies [32]. Information is of paramount importance to the public, especially during the early stages of an infectious disease outbreak, which are characterized by a lack of information and a high level of uncertainty. In terms of public engagement, tweets serving the information function are more likely to be retweeted. An existing study based on a smaller sample size has also shown that science-based tweets about COVID-19 are more likely to be retweeted than tweets containing false information [33]. This means that useful information can get further disseminated through retweeting. Meanwhile, tweets promoting different preventive measures are most likely to be tweeted and liked, which shows that the Twitter users are spreading such recommendations through retweeting. Finally, retweets serving the functions of action and community are more likely to be liked. This means that readers tend to respond favorably to such tweets to show their support.

While the HBM has been traditionally used to study psychological predictors of individuals' adoption of preventative behaviors, it is used in the current study to examine the collective response to health messages in terms of public engagement. Susceptibility, severity, and benefits are the most frequently covered health beliefs, whereas information about barriers and self-efficacy is mostly absent. This means that communicating the risks of COVID-19 to the public is the priority for Texas public health agencies. Emphasizing the benefits is conducive to the adoption of preventative behaviors [11]. In addition, we find that tweets containing the susceptibility belief often lead to more public engagement in terms of the number of favorites and retweets, while the benefits of prevention methods do not increase public engagement. It appears that the public are more interested in finding out the risk of COVID-19 than learning about preventive behaviors during the early stage of a public health crisis, as laid out by the Crisis and Emergency Risk Communication Model [34]. While the currents study focused on how message characteristics affect public engagement, other research has shown that public health agencies' position in the network (e.g., whether the organization occupied a "star" position, which represents their network centrality) also improved the two-way communication between agencies and the public [35].

## Public Health Implication

The current findings identify several strategies that public agencies may adopt to more effectively communicate risk information during an unfolding pandemic. First, the fact that informative tweets are more likely to be retweeted suggests that public agencies should keep leveraging Twitter as an information dissemination tool to increase community outreach.

The sharing/retweeting function of social media can allow public health agencies to disseminate timely, credible, and easy-to-share information to a large scale, which directly and indirectly helps combat health misinformation [21]. Meanwhile, as action-oriented messages are more likely to be favored, public agencies may consider incorporating specific action items in their tweets. In other words, the public not only need factual information about the pandemic; specific guidance and concrete action-items can further boost public support of public agencies.

Second, while emphasizing susceptibility and severity of the disease increases public engagement, directly communicating about the benefit of preventive behaviors proves less effective in creating public engagement. Given the importance of educating the public about prevention behaviors for infectious diseases, public agencies need to be more creative in designing, framing, and implementing their social media messages regarding preventive behaviors. Furthermore, self-efficacy information is almost completely absent from the tweets of public health agencies. Communicating to the public that they are capable of performing the recommended behavior is essential in increasing the adoption of these behaviors.

## Methodological Implications

Methodologically, this study demonstrates the feasibility of using natural language processing to identify theoretical constructs such as social media functions and health beliefs. It shows that a relatively small training dataset can be used to create algorithms for the classification of a much larger corpus of Twitter data. The method established in this study can be easily applied to classifying COVID-19 related tweets by different types of organizations (e.g., hospitals, community organizations, media) and individuals (e.g., politicians and physicians) in the state of Texas and beyond.

## Limitations and Directions for Future Research

This study only examines public health agencies' tweets from a single state in the United State and our data only cover the first wave of the outbreak in the United States. According to the Crisis and Emergency Risk Communication Model, the public have different informational and emotional needs during different stages of the outbreak [34]. It is important to examine agencies' Twitter content during the later stages of the outbreak. Fortunately, our research method can be easily scaled up to study more Twitter content from different parts of the United States as well as longitudinally. Future work may examine how message features may vary across different stages of the pandemic, and how its public engagement outcome shifts over time. We only examined the texts but left out pictures and videos. Further study should examine how pictures or videos affect public engagement. Finally, in terms of the communication functions of governmental organizations, early study has suggested that such communication efforts are often fragmented in that there is a lack of mentioning, coordination, and mutual retweeting among different governmental organizations [36]. Future research could examine the coordination and inconsistency between the public health agencies on the local, state, national and international levels, as piloted in a recent study [37].

## Conclusion

This study examines the COVID-19 related tweet contents published by the public health agencies in Texas during the first six months of 2020. It finds while public health agencies mostly leverage Twitter to dissemination pandemic related information, they could use the Twitter platform to further promote preventative actions since the public respond to tweets promoting actions very well. Furthermore, the public are most likely to engage with tweets describing susceptibility as such information can help them understand the risk. However, there is a lack of information that convinces the public of the feasibility of the proposed preventative behaviors and increase their confidence. Overall, public health agencies would vastly expand their reach during public health crises by steadily building up their follower bases.

## References


1. Haffajee RL, Mello MM. Thinking globally, acting locally - The U.S. Response to Covid-19. *N Engl J Med*. 2020;382(22):e75. doi:10.1056/NEJMp2006740
2. Cheng Y, Chen ZF. Encountering misinformation online: Antecedents of trust and distrust and their impact on the Intensity of Facebook Use. *Online Infor Rev*. Published online December 4, 2020. https://doi.org/10.1108/OIR-04-2020-0130
3. Cheng Y, Luo Y. The presumed influence of digital misinformation in the U.S.: Examining publics' support for governmental restrictions versus corrective action in the COVID-19 Pandemic. *Online Infor Rev*. Published online December 2, 2020. https://doi.org/10.1108/OIR-08-2020-0386
4. Lovejoy K, Saxton GD. Information, community, and action: How nonprofit organizations use social media. *J Comput Mediat Commun*. 2012;17(3):337-353. doi:10.1111/j.1083-6101.2012.01576.x
5. Reynolds B, W Seeger M. Crisis and emergency risk communication as an integrative model. *J Health Commun*. 2005;10(1):43-55. doi:10.1080/10810730590904571
6. Lee NM, VanDyke MS. Set it and forget it: The one-way use of social media by government agencies communicating science. *Sci Commun*. 2015;37(4):533-541. doi:10.1177/1075547015588600
7. Jones T, Fowler MC, Hubbard D. Refining a tool to measure cues to action in encouraging health-promoting behavior--the CHAQ. *Am J Health Promot*. 2000;14(3):170-iii. doi:10.4278/0890-1171-14.3.170
8. Kock N, Mayfield M, Mayfield J, Sexton S, De La Garza LM. Empathetic leadership: How leader emotional support and understanding influences follower performance. *J Leader Organ Stud*. 2019;26(2):217-236. doi:10.1177/1548051818806290
9. Toppenberg-Pejcic D, Noyes J, Allen T, Alexander N, Vanderford M, Gamhewage G. Emergency

Risk Communication: Lessons Learned from a Rapid Review of Recent Gray Literature on Ebola, Zika, and Yellow Fever. *Health Commun*. 2019;34(4):437-455. doi:10.1080/10410236.2017.1405488

10. Becker MH, Rosenstock IM. Compliance with medical advice. In: Steptoe A, Matthews A, eds. *Health care and human behavior*. London: Academic Press; 1984:135-152.

11. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. *Health Commun*. 2010;25(8):661-669. doi:10.1080/10410236.2010.521906

12. Tang L, Park SE. Sun Exposure, Tanning Beds, and Herbs That Cure: An Examination of Skin Cancer on Pinterest. *Health Commun*. 2017;32(10):1192-1200. doi:10.1080/10410236.2016.1214223

13. Shahin S, Dai Z. Understanding public engagement with global aid agencies on Twitter: A technosocial framework. *Am Behav Sci*. 2019:63(12):1684-1707. doi:10.1177/0002764219835248

14. Valenzuela S, Halpern D, Katz JE, Miranda JP. The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism*. 2019;7(6):802-23. doi: 10.1080/21670811.2019.1623701

15. Hall MA, Dugan E, Zheng B, Mishra AK. Trust in physicians and medical institutions: what is it, can it be measured, and does it matter?. *Milbank Q*. 2001;79(4):613-v. doi:10.1111/1468-0009.00223

16. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. *Annu Rev Public Health*. 2020;41:433-451. doi:10.1146/annurev-publhealth-040119-094127

17. Witzling L, Shaw B, Amato MS. Incorporating information exposure into a theory of planned behavior model to enrich understanding of proenvironmental behavior. *Sci Commun*. 2015;37(5):551-574. doi:10.1177/1075547015593085

18. Lievonen M, Luoma-aho V, Bowden J. Negative engagement. In Johnson K, Taylor M, eds. *The Handbook of Communication Engagement.* Hoboken, NJ: John Wiley & Sons; 2018

19. Johnston KA, Taylor M. *The handbook of communication engagement*. Hoboken, NJ: John Wiley & Sons; 2018.

20. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on twitter. *PLoS One*. 2014;9(11):e112235. doi:10.1371/journal.pone.0112235

21. Vos S. *Using social networking sites during public health crises: Theorizing the diffusion of effective messages*. Unpublished Doctoral Dissertation. University of Kentucky; 2018.

22. Vaterlaus JM, Patten EV, Roche C, Young JA. #Gettinghealthy: The perceived influence of social media on young adult health behaviors. *Comput Human Behav*. 2015;45:151-157. doi:10.1016/j.chb.2014.12.013

23. Kim C, Yang SU, Like, comment, and share on Facebook: How each behavior differs from the other. *Public Relat Rev*. 2017;43(2):441-449. doi:10.1016/j.pubrev.2017.02.006

24. Liu W, Xu WW, Tsai, JYJ. Developing a multi-level organization-public dialogic communication framework to assess social media-mediated disaster communication and engagement outcomes. *Public Relat Rev*. 2020;46(4):101949. doi:10.1016/j.pubrev.2020.101949

25. Kang M. Understanding public engagement: Conceptualizing and measuring its influence on supportive behavioral intentions. *J Public Relat Res*. 2014;26(5):399-416. doi:10.1080/1062726X.2014.956107

26. Dekker K, Van Reijmersdal EA. Disclosing celebrity endorsement in a television program to mitigate persuasion: How disclosure type and celebrity credibility interact. *J Promot Mana*g. 2013;19(2):224-240. doi:10.1080/10496491.2013.769473

27. The New York Times. Texas Coronavirus map and case count. https://www.nytimes.com/interactive/2020/us/texas-coronavirus-cases.html

28. Du J, Tang L, Xiang Y, et al. Public Perception Analysis of Tweets During the 2015 Measles

Outbreak: Comparative Study Using Convolutional Neural Network Models. *J Med Internet Res*. 2018;20(7):e236. doi:10.2196/jmir.9413

29. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805v2*; 2018, October 11.

30. Huggingface Transformers. https://huggingface.co/transformers/index.html

31. Fridman I, Lucas N, Henke D, Zigler CK. Association Between Public Knowledge About COVID-19, Trust in Information Sources, and Adherence to Social Distancing: Cross-Sectional Survey. *JMIR Public Health and Surveillance*. 2020; 6(3): e22060. doi: 10.2196/22060

32. Slavik, CE, Buttle, C, Sturrock SL, Darlington, JC, Yiannakoulias N. Examining Tweet Content and Engagement of Canadian Public Health Agencies and Decision Makers During COVID-19: Mixed Methods Analysis. *J Med Internet Res*. 2021; 23(3): e24883 https://www.jmir.org/2021/3/e24883

33. Pulido CM, Villarejo-Carballido B, Redondo-Sama G, Gomez A. COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *Intl Sociol*.2020; 35(4): 377-392. https://doi.org/10.1177/0268580920914755

34. Centers for Disease Control and Prevention. Crisis & Emergency Risk Communication (CERC). https://emergency.cdc.gov/cerc/

35. Kim HM, Saffer, AJ, Liu W et al. How Public Health Agencies Break through COVID-19 Conversations: A Strategic Network Approach to Public Engagement. *Health Commu*. 2021; Published online first. https://doi.org/10.1080/10410236.2021.1886393

36. Zeemering ES. Functional Fragmentation in City Hall and Twitter Communication During the COVID-19 Pandemic: Evidence from Atlanta, San Francisco, and Washington, DC. *Gov Info Quarterly.* 2021; 38(1): 101539. https://doi.org/10.1016/j.giq.2020.101539

37. Wang Y, Hao H, Platt LS. Examining Risk and Crisis Communications of Government Agencies

and Stakeholders during Early-stages of COVID-19 on Twitter. *Computer in Human Behavior*. 2020; 114: 106568. https://doi.org/10.1016/j.chb.2020.106568

# Supplementary Files

# Figures

Longitudinal Changes in the Number of Tweets Promoting Different Health Behaviors by Public Health Agencies in Texas (January 1 to June 30, 2020).