

Original Paper

COVID-19 Health Beliefs Regarding Mask Wearing and Vaccinations on Twitter: Deep Learning Approach

Si Yang Ke^{1*}, BS; E Shannon Neeley-Tass^{1*}, PhD; Michael Barnes^{2*}, PhD; Carl L Hanson^{2*}, PhD; Christophe Giraud-Carrier^{3*}, PhD; Quinn Snell^{3*}, PhD

¹Department of Statistics, Brigham Young University, Provo, UT, United States

²Department of Public Health, Brigham Young University, Provo, UT, United States

³Computer Science Department, Brigham Young University, Provo, UT, United States

* all authors contributed equally

Corresponding Author:

Quinn Snell, PhD

Computer Science Department

Brigham Young University

3361 TMCB

Provo, UT, 84602

United States

Phone: 1 801 422 5098

Email: snell@cs.byu.edu

Abstract

Background: Amid the global COVID-19 pandemic, a worldwide infodemic also emerged with large amounts of COVID-19–related information and misinformation spreading through social media channels. Various organizations, including the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC), and other prominent individuals issued high-profile advice on preventing the further spread of COVID-19.

Objective: The purpose of this study is to leverage machine learning and Twitter data from the pandemic period to explore health beliefs regarding mask wearing and vaccines and the influence of high-profile cues to action.

Methods: A total of 646,885,238 COVID-19–related English tweets were filtered, creating a mask-wearing data set and a vaccine data set. Researchers manually categorized a training sample of 3500 tweets for each data set according to their relevance to Health Belief Model (HBM) constructs and used coded tweets to train machine learning models for classifying each tweet in the data sets.

Results: In total, 5 models were trained for both the mask-related and vaccine-related data sets using the XLNet transformer model, with each model achieving at least 81% classification accuracy. Health beliefs regarding perceived benefits and barriers were most pronounced for both mask wearing and immunization; however, the strength of those beliefs appeared to vary in response to high-profile cues to action.

Conclusions: During both the COVID-19 pandemic and the infodemic, health beliefs related to perceived benefits and barriers observed through Twitter using a big data machine learning approach varied over time and in response to high-profile cues to action from prominent organizations and individuals.

(*JMIR Infodemiology* 2022;2(2):e37861) doi: [10.2196/37861](https://doi.org/10.2196/37861)

KEYWORDS

COVID-19; Health Belief Model; deep learning; mask; vaccination; machine learning; vaccine data set; Twitter; content analysis; infodemic; infodemiology; misinformation; health belief

Introduction

On January 30, 2020, the World Health Organization (WHO) declared the Chinese outbreak of SARS-CoV-2 (ie, COVID-19)

to be a public health emergency of international concern [1]. The following day, the United States Department of Health and Human Services (HHS) secretary declared a US public health emergency to respond to COVID-19 [1]. The president of the United States signed a “Proclamation on Suspension of Entry

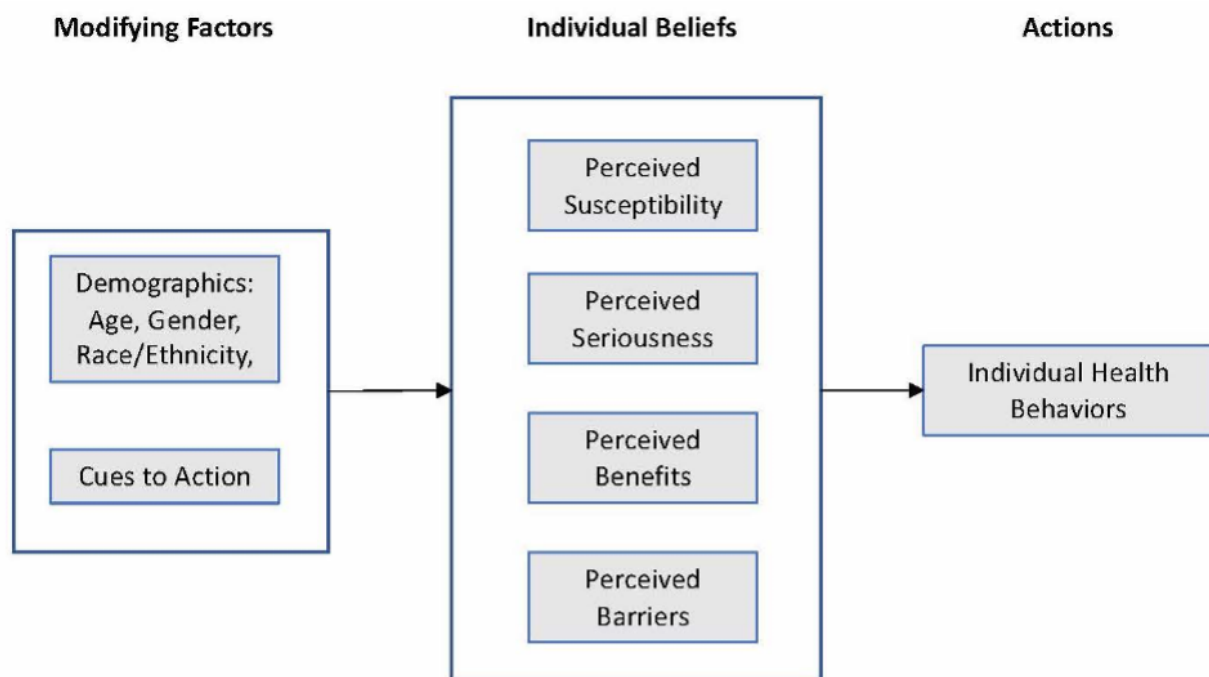
as Immigrants and Nonimmigrants of Persons Who Pose a Risk of Transmitting 2019 Novel Coronavirus,” limiting entry into the United States of persons who traveled to mainland China. Subsequently, on March 11, 2020, WHO declared COVID-19 to be a global pandemic.

With the emergence of a global pandemic came another concern, the emergence of a worldwide infodemic. As it pertained to COVID-19, WHO described the infodemic as an overabundance of information and misinformation related to the COVID-19 pandemic that led to mistrust of health authorities and hampered public health efforts [2]. With the growth in social media use, information about COVID-19 spread quickly, necessitating infodemic management or the need to manage false and misleading information in such a way that would reduce the impact on health behaviors [2]. Greater attention is being paid to sources of COVID-19 information, especially low-credibility sources responsible for spreading COVID-19 misinformation through social media channels [3]. As such, several researchers have begun to address methods for fighting the COVID-19 infodemic and acknowledge the influential role of social media [4-7]. The continuous monitoring and analysis of social media information (infodemiology) has been heralded as a critical tool for understanding the influence of social media and combating misinformation [5]. Although social media data are not specifically designed for public health purposes, they are a valuable and accessible resource for public health surveillance purposes [8]. For example, topic modeling of Twitter posts has been used in understanding topics and sentiments related to COVID-19 in general [9-11], face masks [12,13], and vaccine discussions [14]. Although simply monitoring social media information can provide valuable insight into COVID-19 information/misinformation, understanding the influence on health beliefs and behaviors during an infodemic can assist

public health in better managing information during health emergencies, such as COVID-19, through risk communication [14]. In addition, less is known about the influence of higher-credibility sources of information, such as prevention guidelines coming from WHO and the Centers for Disease Control and Prevention (CDC).

The Health Belief Model (HBM) was developed to explain how beliefs impact health decisions [15]. The theory posits that people engage in health-related behaviors based on (1) their perception of the health condition (eg, COVID-19), (2) their perception of the advantages and disadvantages of the health behavior (eg, mask wearing or receiving a vaccine), and (3) cues to action or stimuli that encourage them to participate in the behavior (eg, health organization recommendations). This theory consists of 5 main elements: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, and cues to action (Figure 1). The model has been successfully used to assess health beliefs on social media regarding physical distancing during the COVID-19 pandemic [16], Zika virus [17], and the human papillomavirus vaccine [18]. Although traditional polling methods require a substantial number of resources and have limitations in assessing public health beliefs (eg, difficulty reaching a large-scale population in large geographic areas and tracking changes in real time), social media has provided millions of people, worldwide, a chance to voluntarily and continuously express their thoughts and opinions on issues that they deem important [18]. Although 1 study used machine learning of Twitter posts to monitor health beliefs regarding COVID-19, health care treatments, and the influence of various external cues to action [19], no identified study has used this methodology to explore the HBM regarding important COVID-19-related behavioral outcomes—mask wearing and vaccinations.

Figure 1. Health Belief Model (HBM).



Research has demonstrated that health organizations, physicians, and the media during COVID-19 represent important HBM cues to action [20]. WHO and the CDC have both issued COVID-19 prevention recommendations; however, recommendations have evolved over time. For example, in the first official advice document regarding the need for and usage of masks, WHO stated that a medical mask is not required for healthy individuals, as no evidence was available on its usefulness to protect nonsick persons [21]. Later, in the updated mask guidelines, the previous advice was modified, and the general public was encouraged to wear masks [22]. Similarly, the CDC initially asserted that the wearing of face masks was an unnecessary public health tool, but a short while later, it issued new guidelines advising people to wear face coverings in public settings where social distancing was difficult [23]. Understanding the individual beliefs regarding COVID-19-preventive behaviors in response to various cues to action from these high-credibility sources is crucial toward helping manage an infodemic. The fallout from these types of COVID-19 shifts in prevention guidelines have created controversy among many sources.

Generally, US guidelines emerging from national, state, and local public health organizations received prominent attention. Lessons have been learned from global and national guidelines, including the following: (1) Travel restriction delays allowed citizens traveling from high-risk areas to pass freely through airports without screening; (2) quarantine delays in high-risk areas allowed potentially infected individuals to spread the

infection; (3) public misinformation allowed racism, incorrect public precautions, and unprecedented fear surrounding COVID-19, allowing rumors, speculation, and misinformation to spread; and (4) emergency announcement regarding the outbreak severity was delayed and not widely broadcast for a month when WHO declared the public health emergency of international concern [24]. Additionally, the WHO guidelines came under political scrutiny by the US president, who blamed WHO for delays and dysfunctions to investigate early cases of COVID-19 and suspended WHO funding [25].

The purpose of this study is to investigate health beliefs and cues to action for mask wearing and vaccination using machine learning of COVID-19-related Twitter posts. External cues to action from prominent pandemic declarations (eg, WHO and the CDC) regarding mask wearing and vaccines and prominent examples of displayed preventive behaviors (eg, presidential mask wearing) were explored for possible influence on health beliefs, as explained by HBM constructs. Although these prominent events could not be studied for a cause-effect relationship, given the surveillance approach of this study, observing the prominent events along with the ongoing Twitter posts may help provide clues to their potential effect on cues to action. This unique approach is an important way to begin exploring how infodemics may influence cues to action. Tables 1 and 2 show the HBM constructs for masks and vaccines, respectively. Findings from this study revealed that cues to action are associated with increased conversations around the perceived health beliefs about mask wearing and vaccinations.

Table 1. HBM^a constructs related to COVID-19 and face coverings.

Construct	Definition
Perceived susceptibility	Assessment of the likelihood or risk of contracting COVID-19; increased likelihood of contracting the disease (eg, increased/decreased prevalence, high/low number of COVID-19 cases)
Perceived severity	Assessment of the perceived seriousness and consequences of contracting COVID-19 (eg, hospitalization, death, mortality, disability)
Perceived benefits	Comments mentioning the benefits of masks or face coverings to reduce the transmission of COVID-19 or the removal of barriers (eg, promotion of mask or face coverings)
Perceived barriers	Comments mentioning the difficulties, challenges, and negative effects of masks and face coverings or the perceived ineffectiveness of masks and face coverings (eg, negative reports of masks or face coverings)

^aHBM: Health Belief Model.

Table 2. HBM^a constructs related to COVID-19 and vaccines.

Construct	Definition
Perceived susceptibility	Assessment of one's likelihood or risk of contracting COVID-19 if not vaccinated; references increased/decreased prevalence, high/low number of cases, and high/low risk/chance/probability
Perceived severity	Assessment of the seriousness of COVID-19 and the major consequences that contracting COVID-19 would have on one's life, such as hospitalization, death, mortality, or disability
Perceived benefits	Assessment of the benefits of COVID-19 vaccines or being vaccinated against COVID-19; the removal of barriers (see the Perceived Barriers section for more information); positive opinion
Perceived barriers	Assessment of the barriers to COVID-19 vaccination, including difficulties, challenges, conspiracies, negative effects, dangers, and perceived ineffectiveness; the removal of benefits; negative opinion

^aHBM: Health Belief Model.

Methods

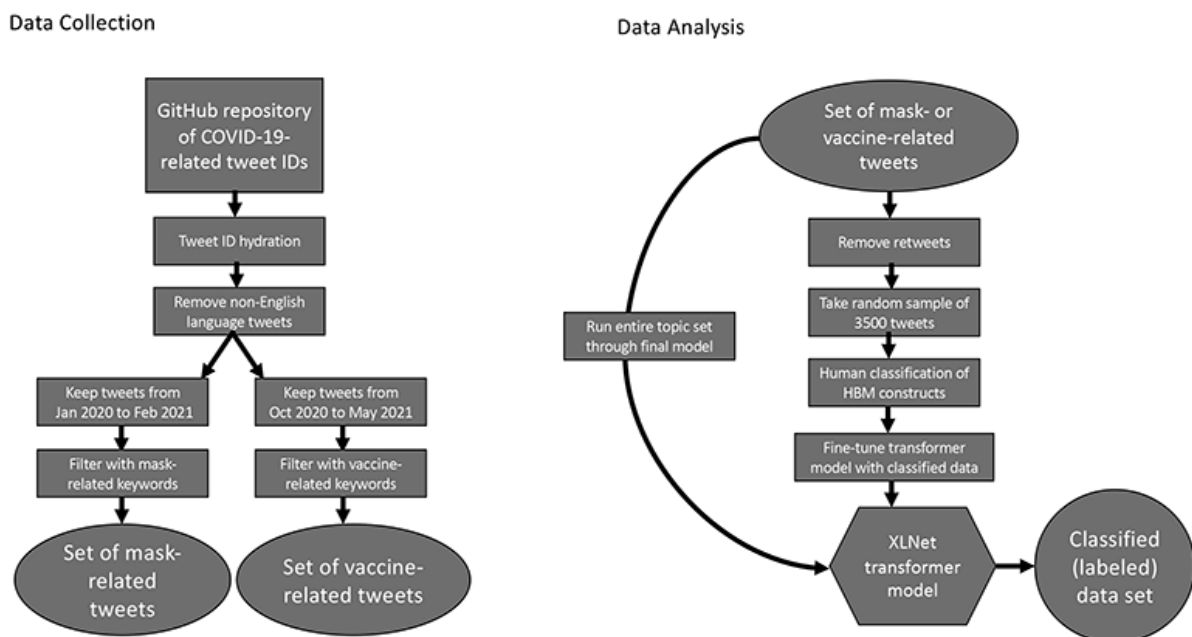
Data Collection

For this paper, a large, publicly available data set of COVID-19-related tweets was used [26,27]. Since Twitter's terms of service only allow the tweet IDs to be publicly available, the authors hydrated the tweet IDs with their own Twitter developer accounts.

The diagram on the left in Figure 2 outlines the data collection process. Since the transformer models used are pretrained and can handle rather raw information, only minor preprocessing of the data was necessary. Non-English tweets were excluded,

and all text was converted to lowercase. Tweets were then filtered by date, and an iterative process was used to filter the tweets by keywords. Research students came up with initial lists of keywords for both mask-related and vaccine-related tweets. The lists were reviewed by the extended research group and modified, as appropriate. For example, the first list included the keyword "face," which picked up a lot of tweets talking about "Facebook." Hence, the keyword "face" was changed to "face " (with a space after the "e"). The keyword lists may not be perfect and may cause the inclusion of tweets outside of the scope of face masks or vaccines related to COVID-19. However, they do produce a more relevant population than the entire corpus, and together with the hand labeling, it was felt that the model would be able to identify HBM-related tweets.

Figure 2. Data collection and analysis procedures. HBM: Health Belief Model.



The set of mask-related tweets was created by filtering tweets from January 2020 through January 2021 with the following keywords: "mask," "face cover," "facemask," "cloth cover," "cover your face," "face covering," "maskup," and "face." Likewise, the set of vaccine-related tweets was created by filtering tweets from October 2020 through November 2021 with the following keywords: "vaccine," "antiva," "anti-va," "vax," "shot," "inoculat," "needle," "booster," "pfizer," "biontech," "immun," "mrna," "trials," "moderna," "novacax," "astrazeneca," "johnson," "sanofi," and "glaxosmithkline." As expected, vaccine-related Twitter conversation came later than mask-related conversation, which is reflected in the different start dates of the data collection. As to the end dates, their selection was simply a matter of choice based on the authors' hypotheses. Upon examination of the mask-related data and resultant graphs, the hypothesized correlations were clearly exhibited, and it was felt that additional data would not significantly change these results (the HBM is all about beliefs affecting behavior), so data collection for the mask-related tweets was interrupted (similarly for the vaccine-related tweets, but for the relevant, later date range).

The final data set statistics are as follows: (1) 1.8 TB on disk, (2) 646,885,238 total tweets, (3) 59,724,507 mask-related tweets, and (4) 113,542,400 vaccine-related tweets.

Data Analysis

Once the topic-related tweets sets were created, they were classified according to the HBM constructs. The classification process was performed separately for each topic-related set, but the process was the same. A random sample of 3500 tweets was selected for manual labeling. No retweets were included in these sets to avoid biasing the models, since retweets would cause repetition of content. Hence, the labeled data consisted of unique tweets, and only these were used in the subsequent model-building phase. Three independent reviewers manually classified the sample according to their relevance to the HBM constructs. Every tweet in the sample was classified with either a positive or a negative label for each of the 4 HBM constructs (ie, perceived susceptibility, perceived severity, perceived benefits, and perceived barriers). Next, a tweet that was labeled positive for at least 1 of the 4 constructs was also classified as HBM related. Tables 1 and 2 show the criteria for a positive label for each of the 4 HBM constructs for the mask- and vaccine-related tweets, respectively. These criteria were defined

using constructs from similar work. Compiled together, each construct is assumed to impact the likelihood for persons taking (or not taking) action. The reviewers classified the tweets independently but compared classifications together after labeling the first 100 tweets and again after the first 500 tweets to resolve conflicts. During these calibration meetings, the reviewers came together and examined samples of tweets that differed in classifications and came to consensus based on the criteria defined in [Tables 1](#) and [2](#). The final label for each tweet was the label that received the majority vote (ie, at least 2 votes out of 3 from the reviewers).

Note that each tweet could be labeled as belonging to more than 1 of the 4 HBM constructs as 1 part of a tweet could fall into 1 HBM category and another part of the tweet could fall into a different HBM category. As an example, consider the following tweet from the set of mask-related tweets:

Buying reusable masks is generally a waste of time and hurts healthcare workers (hello) who need to use them in everyday care during the flu season. Is the corona virus scary? The weirdo incubation time and severity/rapid spread is honestly wild.

This tweet was labeled for both perceived barriers and perceived susceptibility because the user first advocated not buying reusable masks (perceived barriers) and then proceeded to comment on how fast the virus spreads (perceived susceptibility).

Once labeled, each set of 3500 tweets (mask related and vaccine related) could be used for model-building purposes. A random stratified 2450/1050 (70%/30%) split was applied to create the training and test sets. Each model was trained exclusively on the training data ($n=2450$, 70%). The only hyperparameter considered was the dropout rate, which did not significantly change the results. The test data ($n=1050$, 30%) were then used to assess the quality of the models, and the results presented here are based on those data alone. Following their construction, the models were applied to label all tweets, as a real system would indeed be expected to label both original tweets as well as all retweets. The diagram on the right of [Figure 2](#) illustrates the steps taken to process the data from the GitHub repository through the creation of 2 topic-related tweet sets and finally through the classification process.

State-of-the-art bidirectional transformer models [28] were used, combined with custom classification layers. Three different pretrained transformer models were considered and fine-tuned: bidirectional encoder representations from transformers (BERT) [29], a distilled version of bidirectional encoder representations from transformers (DistilBERT; more memory efficient), and XLNet [30]. Three additional simpler models were also included for comparison, namely logistic regression, RepresentationNet (a vanilla version of BERT with a custom classification network), and a bidirectional gated recurrent unit (BiGRU) network [31]. All models were implemented in Python with the *pytorch* library. Each model's predictive ability was evaluated by the area under the receiver operating characteristic (AUROC) curve, accuracy, precision, recall, and F_1 score. In addition, the final model size (pytorch binary format) and a measure of

performance in the form of the number of tweets the model can classify per second were also computed.

After classifying all the tweets using the transformer model, the tweets were separated into calendar weeks and counted. HBM-positive label percentages were computed by taking raw counts of HBM-positive labels divided by the total raw count of COVID-19-related tweets filtered for mask- or vaccine-related keywords, respectively. Potential linear relationships between HBM label percentages by week and COVID-19-related statistics from the corresponding weeks, such as US confirmed case counts, US COVID-19 death counts, and US COVID-19 vaccine doses administered, were also investigated using scatter plots, Spearman correlation matrices, simple linear regression models, and regression with added quadratic terms.

Ethical Considerations

Ethical approval was not needed, since the study only analyzed publicly available data from existing data sets, and results do not contain any identifiable information and are represented in aggregate only.

Results

Manual Labeling and Interrater Reliability

For the mask-related data set, interrater reliability coefficients, measured by Gwet AC1, were 0.784, 0.762, 0.957, and 0.938 for perceived benefits, perceived barriers, perceived severity, and perceived susceptibility, respectively. Interrater reliability coefficients for the vaccine-related data set, measured by Gwet AC1, were 0.825, 0.814, 0.937, and 0.915, respectively, for the 4 HBM constructs in the same order as above. These interrater reliability coefficients are all interpreted as substantial agreement to almost perfect agreement according to the Landis-Koch benchmarking scale [32]. Compared to other interrater reliability calculations, Gwet AC1 is more stable than traditional κ coefficient calculations [33].

For the mask-related data set, 2135 (61%) tweets were manually labeled as related to the HBM model. For the vaccine-related data set, 1330 (38%) tweets were labeled as related to the HBM model.

Machine Learning Model

[Table 3](#) reports on the AUROC, accuracy, precision, recall, and F_1 score for each model. It also includes the model size, as well as the number of tweets the model can classify per second. All 3 pretrained transformer models outperformed the simpler models. Among the transformer models, the XLNet transformer model (with a custom dense 3-layer classification network and a dropout rate set to 0.25) was clearly the best model for this task.

To provide a finer-grained analysis of performance, [Figure 3](#) displays the complete AUROC curve for each of the models. The graph confirms the superiority of the transformer models. Furthermore, it shows that the XLNet transformer generally dominates, and at worse is on par with, the other transformer models.

Table 3. Model evaluation metrics.

Model	AUROC ^a	Accuracy	Precision	Recall	F_1 score	Size (MB)	Evaluations/second
XLNet	0.878	0.824	0.775	0.761	0.768	467	234.8
BERT ^b	0.850	0.786	0.685	0.760	0.721	440	321.1
DistilBERT ^c	0.858	0.795	0.760	0.692	0.725	261	563.5
RepresentationNet	0.717	0.735	0.644	0.648	0.646	484	316.0
BiGRU ^d	0.737	0.752	0.665	0.678	0.672	27	473.9
Logistic regression	0.566	0.651	0.584	0.229	0.329	0.0016	15,925.3

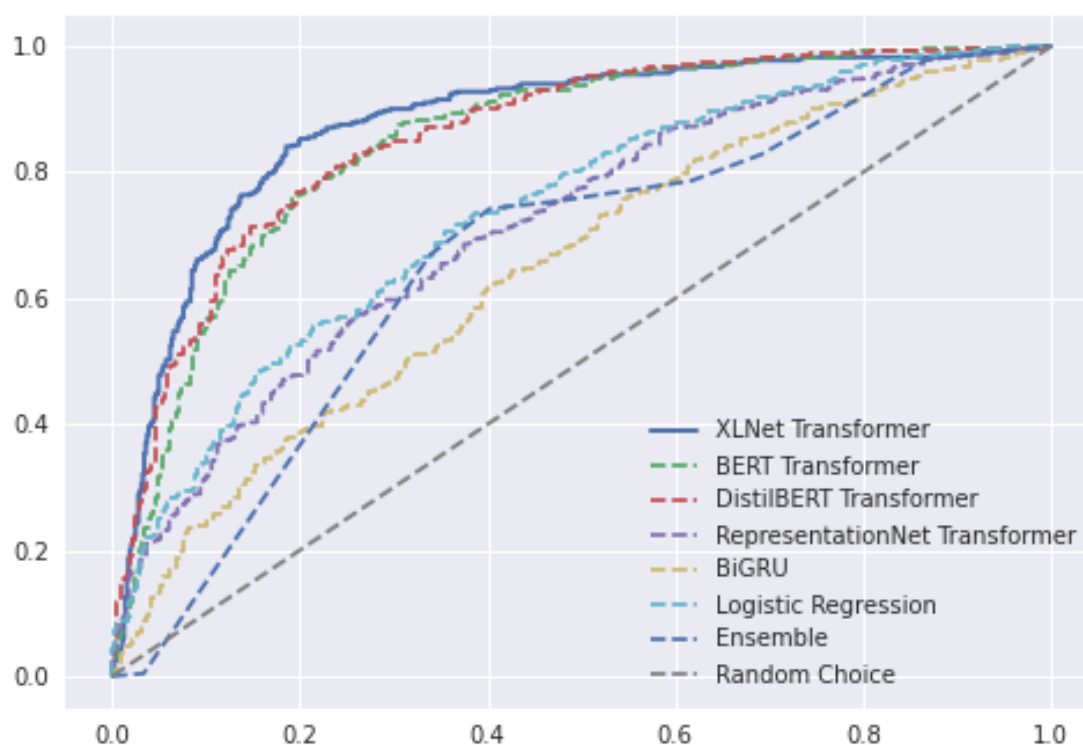
^aAUROC: area under the receiver operating characteristic.

^bBERT: bidirectional encoder representations from transformers.

^cDistilBERT: distilled version of bidirectional encoder representations from transformers.

^dBiGRU: bidirectional gated recurrent unit.

Figure 3. AUROC curves. AUROC: area under the receiver operating characteristic; BERT: bidirectional encoder representations from transformers; BiGRU: bidirectional gated recurrent unit; DistilBERT: distilled version of bidirectional encoder representations from transformers.



It is worth noting the strength of transformer models, especially pretrained ones. Indeed, it is quite remarkable that XLNet achieved over 82% accuracy with only 3500 training data points. By contrast, logistic regression achieved only about 65% accuracy. It is likely that a larger training corpus would improve the performance of logistic regression, but transformer models can do rather well with relatively small corpora. It is also of note that although XLNet is better across predictive evaluation metrics, its memory footprint is significantly larger, and its speed of execution is slower. In cases of low memory, the DistilBERT model may be a viable alternative at the slight cost of some performance loss in classification. Here, the XLNet model was preferred because it fit well within available memory constraints.

Using the XLNet transformer model, 5 models were trained to classify the tweets for HBM relatedness and for the 4 HBM constructs (in the order of perceived benefits, perceived barriers, perceived severity, and perceived susceptibility). The 5 models each achieved over 81% classification accuracy (81%, 97%, 96%, 85%, 82%, respectively) on test data for the mask-related tweets and over 79% classification accuracy (82%, 81%, 86%, 79%, 85%, respectively) on test data for the vaccine-related tweets. Again, the pretrained transformers proved effective at embedding the tweets, making it simple to train the custom classification networks.

The HBM-positive label percentages are plotted by week in Figure 4 for mask-related tweets and Figures 5 and 6 for vaccine-related tweets. In Figures 4-6, the HBM-positive label percentages were computed by dividing raw counts of HBM

labels by the total raw counts of COVID-19-related tweets filtered for mask- or vaccine-related keywords, respectively. Due to the nature of these calculations, the analysis and comparisons synthesized from Figures 4-6 focused on the

direction of change in individual HBM labels rather than on the magnitude of change. Moreover, the magnitude of individual HBM label percentages were not compared across HBM label categories.

Figure 4. US COVID-19 case counts and each HBM scale across time for mask-related tweets. There was a statistically significant correlation between case counts and perceived benefits. Important events corresponding to the time intervals are listed in Table 4. HBM: Health Belief Model.

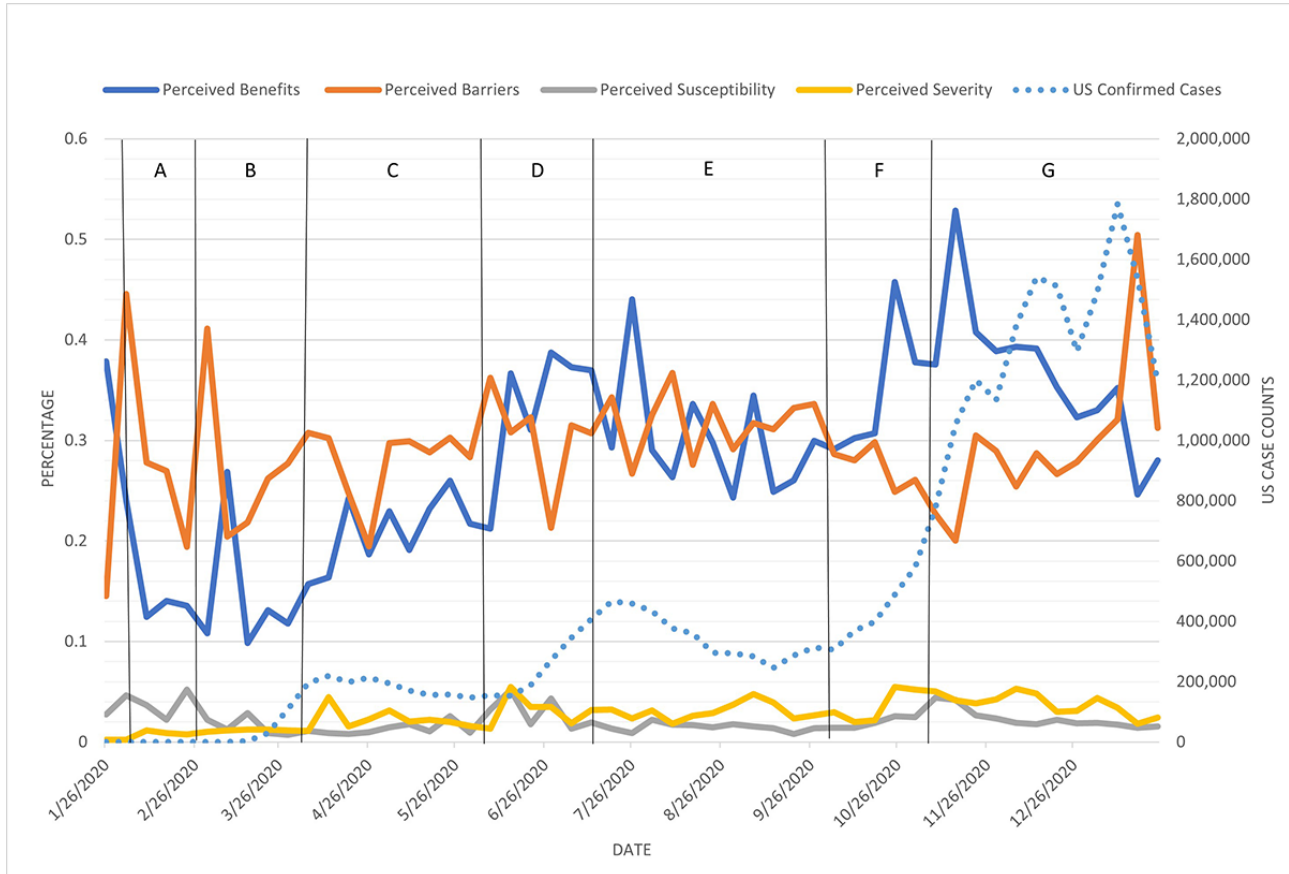


Figure 5. US COVID-19 case counts and each HBM scale across time for vaccine-related tweets. There was a statistically significant correlation between case counts and perceived barriers. Important events corresponding to the time intervals are listed in Table 5. HBM: Health Belief Model.

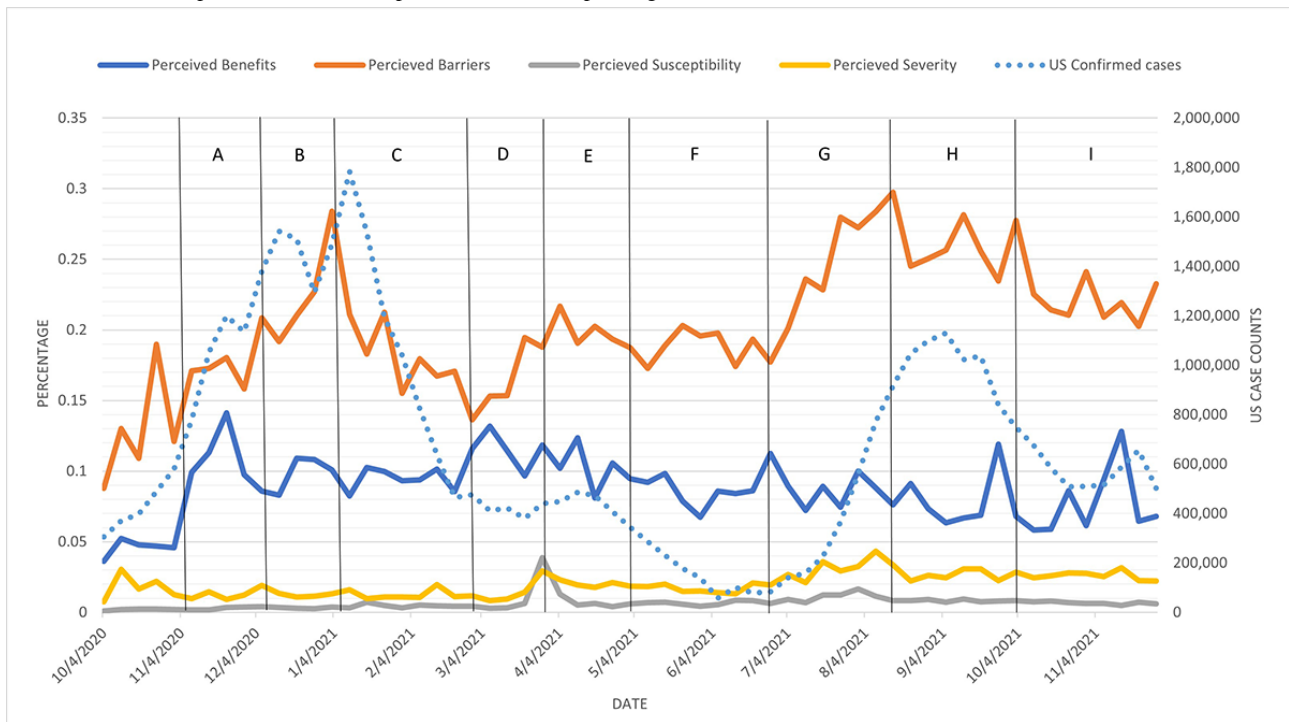


Figure 6. US COVID-19 vaccine doses administered and each HBM scale across time for vaccine-related tweets. There was a statistically significant negative correlation between vaccination counts and perceived barriers. Important events corresponding to the time intervals are listed in Table 5. HBM: Health Belief Model.

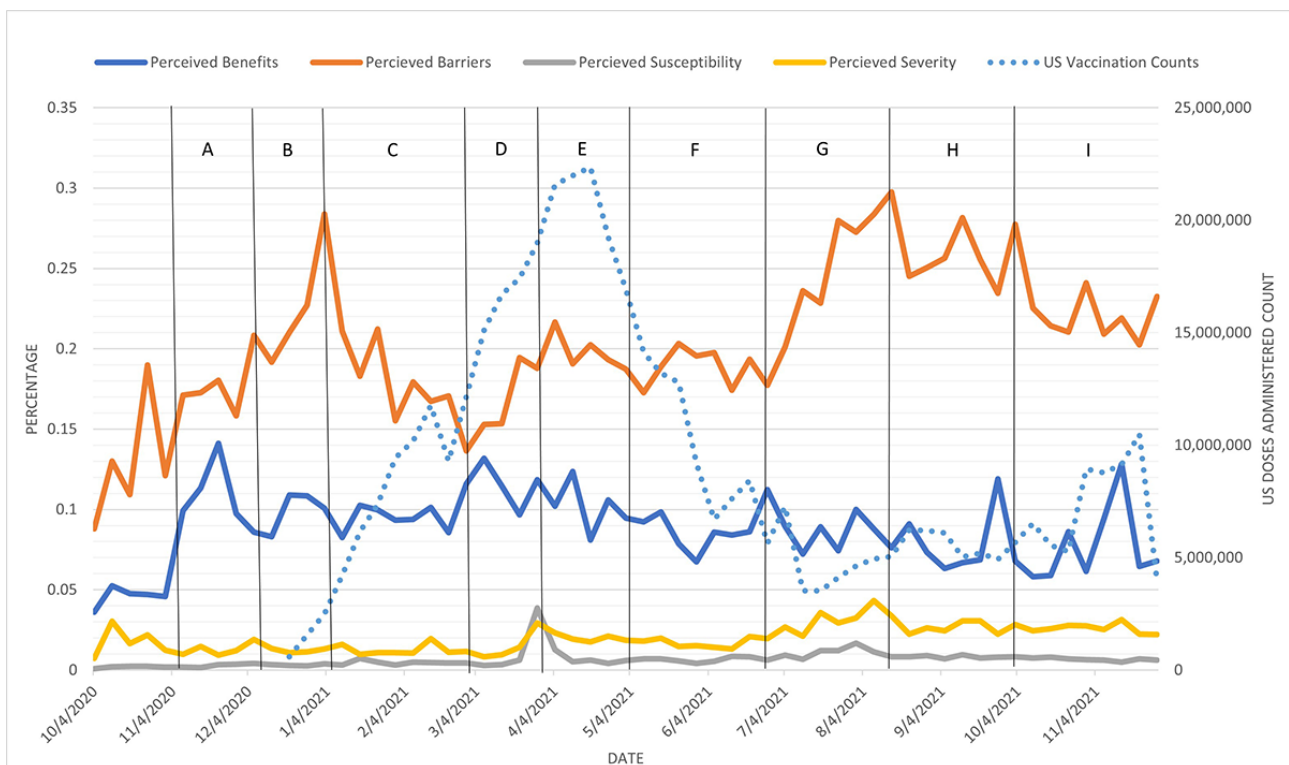


Table 4. Mask-related events and cues to action.

Dates	Cues to action
Time frame A	
January 21, 2020	The CDC ^a confirms the first COVID-19 case.
January 31, 2020	WHO ^b declares a global emergency.
Time frame B	
February 27, 2020	WHO issues interim guidance on masks.
March 8, 2020	This National Institutes of Health's (NIH) Dr Fauci says face masks are not fully effective.
March 11, 2020	WHO declares a pandemic.
March 13, 2020	President Trump declares an emergency.
Time frame C	
April 3, 2020	The CDC recommends wearing face masks in public.
April 10, 2020	The first state (New Jersey) mandates face masks in public.
April 16, 2020	The White House begins a formal discussion to open the economy.
June 3, 2020	The surgeon general (HHS ^c) asks Americans to stop buying face masks.
Time frame D	
June 5, 2020	WHO recommends masks for areas with community transmission.
June 17, 2020	WHO halts hydroxychloroquine production.
July 6, 2020	WHO is asked by scientists to revise guidelines to acknowledge airborne transmission.
July 9, 2020	WHO declares that COVID-19 is airborne-transmissible.
Time frame E	
July 12, 2020	President Trump is seen wearing a mask in public for the first time.
August 13, 2020	Presidential candidate Joe Biden calls for a 3-month mask mandate.
August 17, 2020	The United States declares COVID-19 as the third-leading cause of death.
August 28, 2020	The first US reinfection case is found.
Time frame F	
September 16, 2020	President Trump releases a vaccine distribution plan.
September 21, 2020	The CDC withdraws guidance saying COVID-19 is airborne-transmissible.
October 2, 2020	President Trump and the First Lady are diagnosed with COVID-19 and the president is hospitalized.
October 15, 2020	WHO declares conclusive evidence that hydroxychloroquine is ineffective.
Time frame G	
November 4, 2020	The United States reports an unprecedented 100,000 cases in 1 day.
December 14, 2020	WHO reports the first of the COVID-19 variants (in the United Kingdom).

^aCDC: Centers for Disease Control and Prevention.^bWHO: World Health Organization.^cHHS: Department of Health and Human Services.

Table 5. Vaccine-related events and cues to action.

Dates	Cues to action
Time frame A	
November 4, 2020	Joe Biden is elected the 46th President of the United States over Donald Trump.
November 9, 2020	Pfizer publishes vaccine results.
November 16, 2020	Moderna reveals vaccine efficacy results.
Time frame B	
December 4, 2020	President Biden asks Americans to commit to 100 days of wearing masks, his first act as president-elect.
December 11, 2020	The FDA ^a approves the emergency use of the Pfizer vaccine.
December 18, 2020	The FDA approves the emergency use of the Moderna vaccine.
December 21, 2020	President Biden get the first dose of the vaccine.
December 31, 2020	The United States falls short of the goal to give 20 million vaccinations by year-end (2.8 million).
Time frame C	
January 4, 2021	The White House says more individuals will be vaccinated using reserve supplies.
January 6, 2021	The HHS ^b provides US \$22 billion to fund testing and vaccine distribution.
January 7, 2021	The CDC ^c says COVID-19 vaccine benefits outweigh allergic reaction risks for Pfizer or Moderna.
January 12, 2021	The CDC and the HHS update vaccine allocation to release all available doses.
February 4, 2021	The FDA begins considering the J&J ^d vaccine.
February 26, 2021	The Kaiser Family Foundation (KFF) poll shows vaccine acceptance increases among Americans.
February 27, 2021	The FDA approves the J&J vaccine with emergency use authorization.
March 1, 2021	The former president and First Lady urge followers to get vaccinated.
Time frame D	
March 4, 2021	The COVID-19 UK variant does not affect vaccine efficacy.
March 8, 2021	The CDC releases guidance on safe activities for fully vaccinated.
March 11, 2021	The Public Broadcasting Service (PBS) poll says nearly half of Republican men will not get COVID-19 vaccines.
March 11, 2021	President Biden pushes for expanded vaccine eligibility to all adults aged 18 years and older by May 1.
March 15, 2021	The White House unveils an expansive public relations vaccine confidence campaign.
Time frame E	
April 2, 2021	The CDC expands travel guidelines for those fully vaccinated.
April 6, 2021	The COVID-19 variant is detected in all 50 states.
April 13, 2021	The CDC and the FDA recommend pausing the J&J vaccine.
April 23, 2021	The CDC lifts the pause on the J&J vaccine.
April 27, 2021	The CDC eases mask restrictions for fully vaccinated individuals.
Time frame F	
May 4, 2021	The FDA prepares to authorize the Pfizer vaccine in adolescents.
May 4, 2021	President Biden announces a new goal of having 70%, or 160 million, American adults with at least 1 dose of a COVID-19 vaccine by July 4.
May 10, 2021	The Pfizer/BioNTech vaccine is approved for adolescents.
May 11, 2021	WHO ^e declares the Delta variant a variant of concern.
May 24, 2021	Heart problems are investigated in vaccinated Teens.
June 1, 2021	Employers can require a COVID-19 vaccine.
June 3, 2021	The Biden administration announces a National Month of Action with the goal of immunizing at least 70% Americans by July 4.

Dates	Cues to action
June 10, 2021	President Biden is set to announce the purchase of 500 million doses of Pfizer's COVID-19 vaccine to be donated to the rest of the world.
June 25, 2021	<i>Associated Press</i> reports: Almost all COVID-19 deaths recorded in the United States are among those who are not vaccinated.
Time frame G	
July 9, 2021	Pfizer says it will pursue booster shots.
July 12, 2021	The FDA warns that the COVID-19 J&J vaccine can lead to increased risk of Guillain-Barré syndrome.
July 13, 2021	HHS officials say fully vaccinated individuals do not need a COVID-19 vaccine booster shot.
July 21, 2021	New research published in the <i>New England Journal of Medicine</i> shows the Pfizer vaccine is not as effective against the Delta variant.
July 29, 2021	President Biden calls for federal worker vaccines.
August 3, 2021	Around 70% of Americans are vaccinated.
Time frame H	
August 12, 2021	The CDC recommends pregnant people get vaccinated.
August 13, 2021	A booster shot is endorsed for those immunocompromised.
August 18, 2021	US health officials announce a plan for booster shots for the general public.
August 23, 2021	The Pfizer/BioNTech vaccine gains full FDA approval.
August 24, 2021	Several large organizations issue vaccine mandates for workers.
September 9, 2021	President Biden announces all companies with over 100 employees must mandate COVID-19 vaccinations.
September 10, 2021	Los Angeles schools mandate vaccines.
September 17, 2021	The FDA committee votes against boosters for the general public.
September 20, 2021	Pfizer says its vaccine is safe and effective in children.
September 22, 2021	The FDA authorizes the Pfizer booster.
Time frame I	
October 15, 2021	The United States opens to vaccinated travelers.
October 25, 2021	The Moderna dose is effective in children.
October 29, 2021	The FDA authorizes the Pfizer vaccine for children.
November 1, 2021	The FDA investigates Moderna vaccine adverse effects.
November 5, 2021	At least 2 groups file lawsuits against implementation of President Biden's vaccine mandate.
November 13, 2021	The appeals court affirms hold on the employer vaccine mandate.
November 19, 2021	The FDA approves a vaccine booster for all adults.

^aFDA: Food and Drug Administration.

^bHHS: Department of Health and Human Services.

^cCDC: Centers for Disease Control and Prevention.

^dJ&J: Johnson & Johnson.

^eWHO: World Health Organization.

Mask-Related Tweet Results

Figure 4 displays HBM label percentages. The U.S. confirmed cases by week were added to the secondary y-axis and letters within vertical lines represent a selection of events that could potentially help interpret the fluctuations in the figure from the U.S. government and related sources from the American Journal of Managed Care. Case count data came from the New York Times COVID-19 data GitHub repository [34].

Perceived benefits and perceived barriers frequently corresponded to cues for action, such as official government or public health guidance or policies (Figure 4). Perceived benefits for mask wearing trended with the rise and fall of the US COVID-19 case counts for the majority of the 2020 calendar year until mid-November and early January when vaccines were available. This observation was supported by a statistically significant ($P < .001$) Spearman correlation of 0.686. The lowest levels of perceived benefits occurred between February and mid-March, where barriers to mask wearing overshadowed benefits. Benefits first emerged above barriers after the

pandemic was declared by WHO and when WHO recommended mask wearing for health care workers and sick individuals, but mask wearing was not sustained. Mask wearing emerged again more steadily when WHO declared COVID-19 was airborne-transmissible and when the CDC began recommending masks in early April. Benefits emerged most quickly after WHO changed its stance on mask wearing. The highest levels of perceived barriers were from February to mid-March. According to the top 5 retweets categorized for perceived barriers from February 24 to March 1, the “stop buying masks, save it for the health care workers” idea was the content of 4 of the top 5 most retweeted retweets. The top retweets correspond well with the initial stance of WHO regarding mask wearing for the general public [21] and the elevated perceived barriers to the action shown in Figure 4. Although WHO did not change its official stance on public mask wearing until June 5, the CDC first recommended face masks on April 3, and the first state (New Jersey) began a mask mandate on April 10. Perceived benefits gradually trended upward and peaked to correspond with each of the peaks in US case counts in late July and again in October and November and exceeded perceived barriers during much of that time. Perceived benefits trended with case counts until November around the US presidential election and in anticipation of the US Thanksgiving holiday. The lowest perceived barriers and the most significant divergence from benefits occurred as the COVID-19 case counts mounted toward their highest level in late November 2020. Table 4 lists a few of the major events and cues to action during the time frame in Figure 4.

Perceived severity and perceived susceptibility possessed a much lower percentage of HBM influence compared to perceived benefits and perceived barriers. The overall rate of perceived severity tended to be slightly higher than that of perceived susceptibility except at the beginning of the pandemic.

Vaccine-Related Tweet Results

Figure 5 displays the HBM label percentages (raw count of HBM labels divided by the total raw count of COVID-19–related tweets filtered for vaccines). US confirmed cases by week were added to the secondary y axis, and letters within vertical lines represent a selection of events that could potentially help interpret the fluctuations in the figures from the US government and related sources from the *American Journal of Managed Care* [35,36]. In addition, US confirmed case count data were collected from the *New York Times* COVID-19 data GitHub repository [34]. Figure 6 looks at how US vaccine doses administered trended with the HBM label percentages. US vaccine data were collected from the Bloomberg Covid-19 Vaccine Tracker Open Data GitHub repository [37].

Perceived barriers and US case counts had a statistically significant ($P=.03$), positive Spearman correlation of 0.28. This means that as US case counts increased, perceived barriers also increased. Perceived barriers and US vaccination counts had a statistically significant ($P<.001$), negative Spearman correlation of -0.67 . This means that there was an inverse relationship between US vaccination counts and perceived barriers. The displayed findings do not reflect magnitude but simply identify the relative volume of tweets for each of the HBM constructs.

Perceived barriers for vaccination consistently remained higher than perceived benefits, but they did not appear to trend inversely as expected. Like the face mask findings, perceived susceptibility and perceived severity accounted for a noticeably lower percentage than perceived benefits and perceived barriers. Unlike the face mask findings, perceived barriers had no crossovers with perceived benefits (Figures 5 and 6). Analysis of US case counts tended to trend with perceived barriers throughout fall 2020 through winter 2021. The percentage of perceived barriers was steadily higher than US case counts in 2021. The initial peak of perceived barriers to vaccination occurred at the beginning of the availability and distribution of vaccines. Still, it then steadily declined until early March 2021, when the CDC released guidance that safe activities were available for fully vaccinated individuals (March 8) and when President Biden pushed for expanded vaccination eligibility for all adults aged 18 years and older (March 11). The highest level of perceived barriers to vaccine acceptance peaked in late summer and early fall 2021 around the time that schools resumed operations. This rapid incline seemed to correspond with the greatest decline in US vaccination counts (Figure 6). The second greatest rise in barriers occurred early during the vaccination process from late December 2020 to early January 2021. Perceived benefits did not rise until late October to early November 2020, when vaccine efficacy was established, and then trended and held relatively steady in 2021.

Vaccination distribution rates continued to climb steadily from mid-March until mid-April 2021, but perceived barriers also continued to climb up and down. Overall vaccination counts dropped steadily in mid-April, near the time when the CDC and the Food and Drug Administration (FDA) recommended pausing the Johnson & Johnson (J&J) vaccine (April 13) and continued despite those restrictions being lifted 10 days later (April 23). The vaccination drop continued to plummet even though the FDA formally expanded the availability of vaccines for adolescents aged 12–15 years in early May (May 4). At that time, barriers again climbed upward through June 1. An inverse correlation between vaccine doses administered and perceived barriers to obtaining vaccination was observed. That is, as doses increased, perceived barriers decreased. During this inverse correlation period, WHO declared the COVID-19 Delta variant a concern (May 11), heart problems were noted in vaccinated teenagers (May 24), the government allowed employers to require the COVID-19 vaccine (June 1), and research in the *New England Journal of Medicine* indicated the Pfizer vaccine was not as effective (July 21). From June through September 2021, the conversation on perceived barriers for vaccines continued to climb, while the case counts also climbed, and doses administered declined. Table 5 lists some of the major events and cues to action during the time frame in Figures 5 and 6.

Discussion

Principal Findings

With COVID-19 also came the emergence of a worldwide infodemic, or the overabundance of information and misinformation about COVID-19. A central informational theme

for COVID-19 in 2020 in public health revolved around controlling its spread using social distancing and face mask wearing and included ramping up vaccines that could be quickly distributed. A continuing theme in 2021 for COVID-19 involved efforts to end the pandemic through vaccine distribution and continuing some level of preventative measures (eg, stay-at-home orders, mask mandates, and capacity limits in certain businesses or other settings). However, social media played an influential role in creating an infodemic [21]. These social media influences are closely related to the political and personal reactions to COVID-19 from early in the US pandemic and appear to have continued throughout [38]. Although some social media research during the infodemic pointed to the influence of low-credibility content regarding COVID-19 [3], less has been done to understand the influence of higher-credibility content largely from prominent sources on COVID-19 health beliefs. This study used social media to investigate the COVID-19-related Twitter posts in the United States to understand health beliefs related to mask wearing and vaccinations in the midst of an infodemic. We also explored external cues to action from prominent pandemic declarations (eg, WHO and the CDC) regarding mask wearing and vaccines and notable examples of displaying preventive behaviors (eg, presidential mask wearing) and their possible association with health beliefs, as explained by the HBM constructs. Understanding the influence of social media information on COVID-19 health beliefs and preventative behaviors is important for information management during emergencies.

First, health beliefs relating to the perceived benefits of and perceived barriers to mask wearing appeared to be most influenced by external cues to action. Furthermore, they tended to inversely weakly mirror each other over time. That is, as more perceived benefits of mask wearing were discussed, fewer perceived barriers were discussed. These findings are consistent with other HBM research, particularly among student pharmacists [39]. Although previous research on COVID-19 mask-wearing beliefs pointed to the influence of perceived severity [40], in this study, the perceived susceptibility to and perceived severity of COVID-19 were much less prominent. This remained fairly consistent over time despite numerous high-profile cues to actions, such as major announcements, COVID-19 case counts, or COVID-19 deaths. This contrasts with early assumptions about what health beliefs influence face mask wearing in context with the HBM [41]. Regardless, the perceptions of disease susceptibility and severity seemed muted in this study, presumably because the worldwide pandemic may have appeared obviously relevant for most people in most places. Another possible explanation is that the immediacy of an ongoing and rapidly changing pandemic tended not to discuss the severity and seriousness of getting the disease but seemed to emphasize the importance of behaviors (ie, face masks and vaccines) and taking action.

Perceived benefits of and perceived barriers to mask wearing varied over time with sometimes dramatic swings that appeared to align with specific major cues to action. The most important or consistent cue to action was the rise and fall of total US case counts from March through December 2020 with perceived benefits. This observation is reinforced by the fact that the

perceived benefits of mask wearing occurred before formal mask wearing was recommended by WHO (June 5). The benefits of mask wearing continued despite the CDC and other sources asking that masks not be hoarded so they could be available for health care workers and other individuals who were sick.

The second-most important cues to action included how the timing of certain messaging cues may have prompted temporary perceptions of the benefits of mask wearing. For example, perceived benefits peaked around the same time that WHO recommended masks in June 2020, WHO asked scientists to revise guidelines to acknowledge airborne transmission in mid-July, and President Trump was seen wearing a mask for the first time in mid-July [35]. By contrast, other kinds of messaging cues to action may have prompted a rise in perceived barriers to mask wearing when contradictory messages from WHO and the CDC urged people to not use masks early in the pandemic or messages were released about vaccines' immediate availability at the end of 2020. Thus, the greatest cues to action for mask wearing identified from these data were primarily the rise in US case counts followed by episodic messaging that promoted the health belief that mask wearing was important. Such observed cues to action do not reflect a cause-and-effect relationship, but findings from this study are verified by what others have recently suggested [42].

Second, health belief findings regarding vaccinations from this study demonstrated several important implications for cues to action and the 4 HBM constructs. Similar to mask wearing, health beliefs relating to perceived benefits of and perceived barriers to vaccination appeared to be influenced by specific cues to action and often inversely mirrored each other over time, while not being as pronounced as mask-wearing perceptions. Twitter conversations regarding perceived vaccine barriers generally increased up until January 1 before declining and flattening. These trends tended to precede and mirror US confirmed COVID-19 case counts and death counts, suggesting the potential influence of several high-profile announcements (cues to action) during those peak times. Although previous HBM survey research in Malaysia indicated perceived barriers, such as vaccine efficacy, safety, affordability, and side effects [19], findings from this study appear to be linked more with vaccine access. Perceived barriers among COVID-19 vaccine Twitter conversations began to increase sharply when emergency use began (December 11), only peaking after the White House announced the use of reserve supplies (January 4) and the government committed funds for vaccine distribution (January 6).

Unique to vaccination Twitter conversations compared to this study's mask-wearing findings, discussions regarding perceived benefits were always below perceived barrier trends, especially later in the pandemic. Perceived barriers reflected difficulties, challenges, conspiracies, negative effects, dangers, and perceived ineffectiveness associated with vaccinations. One potential explanation is that vaccine-related conversations became more complex because of these factors as the pandemic emerged. This complexity likely involved factors outside our study, such as political distrust, contradictory messages, and others. The Twitter conversations seemed to focus on the

likelihood of the disease relative to face masks and the benefits of the behavior, but the vaccination barriers complicated the benefits of vaccination.

The number of vaccines administered peaked as perceived seriousness and perceived susceptibility discussions peaked (April 4). These findings are consistent with the HBM, which suggests behavior will change when the threat of disease increases one's perception of seriousness and susceptibility increases. Interestingly, although the volume of discussion for perceived benefits and perceived barriers related to vaccination nearly crossed on March 4, benefit discussions remained higher than barrier discussions. Altogether, however, it appears that various cues to action beginning March 1 influenced health beliefs and ultimately vaccine behavior. This helps demonstrate that understanding the individual beliefs regarding COVID-19-preventive behaviors in response to various cues to action from these high-credibility sources is crucial toward helping manage an infodemic. Moving forward, public health officials may better manage information and positively influence health beliefs and vaccine behaviors by using traditional risk communication approaches [43]. For example, focusing on building trust through announcing early findings, being transparent with what is known and unknown, respecting public concerns, and planning in advance may serve as important lessons learned from the COVID-19 pandemic, particularly in response to vaccinations [44].

Limitations

There are several limitations of this study. First, as noted in the study purpose justification, the results of the study are primarily exploratory in nature and should be interpreted in this context. Additional research is needed to further confirm the influence of high-profile cues to action and HBM constructs, ideally with some form of case-control or experimental design. Second, although publicly available Twitter data were used in this

analysis, these data were collected from a subset of tweets returned from the Twitter Application Programming Interface (API). Because there was no way of knowing the size of the subset in relation to the whole of tweets, or the sampling method used by Twitter to create the subset, there was a potential for bias. Third, because the COVID-19-related tweets were filtered for vaccines or masks, the perceived susceptibility and perceived severity constructs may have been minimized. It may also be described by the HBM construct definitions in [Tables 1 and 2](#), which emphasize COVID-19 itself. Regardless, filtering tweets for mask wearing and vaccines may have inadvertently diminished the susceptibility and seriousness of COVID-19 conversations with the HBM because most of these COVID-19 conversations tended to focus on the perceived benefits and perceived barriers. Fourth, future studies could use topic modeling techniques suitable for short-text documents, such as tweets, to better understand the topics surrounding masks, vaccines, and COVID-19 related to the HBM. In addition, future research might explore specifically which social media messages, communications channels, and voices are most influential on COVID-19-related prevention health beliefs. Fifth, although there are several explanations of the findings in this study, we did not attempt to establish causal relationships. Future longitudinal studies can explore questions of causation.

Conclusion

Throughout the pandemic, experts have provided aggressive recommendations for COVID-19 prevention [21,22,24,45]. During both the COVID-19 pandemic and the infodemic, this study used a machine learning approach to explore health beliefs related to mask-wearing and vaccination recommendations and the possible influence of high-profile cues to action. Findings suggest that although certain health beliefs on Twitter appear to respond to various high-profile cues to action, health belief trends differ between mask wearing and vaccination.

Acknowledgments

The authors would like to acknowledge the work of Kiersten DeCook, Sarah Callaway, McKaylee Smith, Zane Maguet, Si Yang Ke, Emily Freeman, and Anna Leman in hand-labeling the 2 training sets of 3500 tweets.

Authors' Contributions

All authors discussed and collaborated in the conception of the ideas presented in the paper. ESN-T gathered and filtered the data. QS developed the classification models and produced the resulting classification data. SYK developed the data-counting code and produced the graphs. MB and CLH developed the theory and analysis of the results. All authors collaborated on the writing and presentation of the research and contributed to the final manuscript.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Jernigan DB, CDC COVID-19 Response Team. Update: public health response to the coronavirus disease 2019 outbreak - United States, February 24, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Feb 28;69(8):216-219. [doi: [10.15585/mmwr.mm6908e1](https://doi.org/10.15585/mmwr.mm6908e1)] [Medline: [32106216](https://pubmed.ncbi.nlm.nih.gov/32106216/)]
2. World Health Organization. Infodemic. URL: https://www.who.int/health-topics/infodemic%23tab=tab_1 [accessed 2021-08-02]

3. Yang K, Pierri F, Hui P, Axelrod D, Torres-Lugo C, Bryden J, et al. The COVID-19 infodemic: Twitter versus Facebook. *Big Data Soc* 2021 May 05;8(1):205395172110138. [doi: [10.1177/20539517211013861](https://doi.org/10.1177/20539517211013861)]
4. Cinelli M, Quattrocioni W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. *Sci Rep* 2020 Oct 06;10(1):16598 [FREE Full text] [doi: [10.1038/s41598-020-73510-5](https://doi.org/10.1038/s41598-020-73510-5)] [Medline: [33024152](https://pubmed.ncbi.nlm.nih.gov/33024152/)]
5. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. *J Med Internet Res* 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: [10.2196/21820](https://doi.org/10.2196/21820)] [Medline: [32589589](https://pubmed.ncbi.nlm.nih.gov/32589589/)]
6. Zarocostas J. How to fight an infodemic. *Lancet* 2020 Feb;395(10225):676. [doi: [10.1016/s0140-6736\(20\)30461-x](https://doi.org/10.1016/s0140-6736(20)30461-x)]
7. Hua J, Shaw R. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: the case of China. *Int J Environ Res Public Health* 2020 Mar 30;17(7):2309 [FREE Full text] [doi: [10.3390/ijerph17072309](https://doi.org/10.3390/ijerph17072309)] [Medline: [32235433](https://pubmed.ncbi.nlm.nih.gov/32235433/)]
8. Fung IC, Tse ZTH, Fu K. The use of social media in public health surveillance. *Western Pac Surveill Response J* 2015;6(2):3-6 [FREE Full text] [doi: [10.5365/WPSAR.2015.6.1.019](https://doi.org/10.5365/WPSAR.2015.6.1.019)] [Medline: [26306208](https://pubmed.ncbi.nlm.nih.gov/26306208/)]
9. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. *J Med Internet Res* 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: [10.2196/20550](https://doi.org/10.2196/20550)] [Medline: [33119535](https://pubmed.ncbi.nlm.nih.gov/33119535/)]
10. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. *J Med Internet Res* 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: [10.2196/22624](https://doi.org/10.2196/22624)] [Medline: [33006937](https://pubmed.ncbi.nlm.nih.gov/33006937/)]
11. Guntuku SC, Bittenheim AM, Sherman G, Merchant RM. Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States. *Vaccine* 2021 Jul 05;39(30):4034-4038 [FREE Full text] [doi: [10.1016/j.vaccine.2021.06.014](https://doi.org/10.1016/j.vaccine.2021.06.014)] [Medline: [34140171](https://pubmed.ncbi.nlm.nih.gov/34140171/)]
12. He L, He C, Reynolds T, Bai Q, Huang Y, Li C, et al. Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic. *J Am Med Inform Assoc* 2021 Jul 14;28(7):1564-1573 [FREE Full text] [doi: [10.1093/jamia/ocab047](https://doi.org/10.1093/jamia/ocab047)] [Medline: [33690794](https://pubmed.ncbi.nlm.nih.gov/33690794/)]
13. Al-Ramahi M, Elnoshokaty A, El-Gayar O, Nasralah T, Wahbeh A. Public discourse against masks in the COVID-19 era: infodemiology study of Twitter data. *JMIR Public Health Surveill* 2021 Apr 05;7(4):e26780 [FREE Full text] [doi: [10.2196/26780](https://doi.org/10.2196/26780)] [Medline: [33720841](https://pubmed.ncbi.nlm.nih.gov/33720841/)]
14. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. *J Med Internet Res* 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: [10.2196/24435](https://doi.org/10.2196/24435)] [Medline: [34115608](https://pubmed.ncbi.nlm.nih.gov/34115608/)]
15. Rosenstock IM. Why people use health services. *Milbank Mem Fund Q* 1966 Jul;44(3):94. [doi: [10.2307/3348967](https://doi.org/10.2307/3348967)]
16. Sesagiri Raamkumar A, Tan SG, Wee HL. Use of health belief model-based deep learning classifiers for COVID-19 social media content to examine public perceptions of physical distancing: model development and case study. *JMIR Public Health Surveill* 2020 Jul 14;6(3):e20493 [FREE Full text] [doi: [10.2196/20493](https://doi.org/10.2196/20493)] [Medline: [32540840](https://pubmed.ncbi.nlm.nih.gov/32540840/)]
17. Guidry JPD, Carlyle KE, LaRose JG, Perrin P, Messner M, Ryan M. Using the Health Belief Model to analyze Instagram posts about Zika for public health communications. *Emerg Infect Dis* 2019 Jan;25(1):179-180 [FREE Full text] [doi: [10.3201/eid2501.180824](https://doi.org/10.3201/eid2501.180824)] [Medline: [30561302](https://pubmed.ncbi.nlm.nih.gov/30561302/)]
18. Du J, Cunningham RM, Xiang Y, Li F, Jia Y, Boom JA, et al. Leveraging deep learning to understand health beliefs about the human papillomavirus vaccine from social media. *NPJ Digit Med* 2019 Apr 15;2(1):27 [FREE Full text] [doi: [10.1038/s41746-019-0102-4](https://doi.org/10.1038/s41746-019-0102-4)] [Medline: [31304374](https://pubmed.ncbi.nlm.nih.gov/31304374/)]
19. Wang H, Li Y, Hutch M, Naidech A, Luo Y. Using tweets to understand how COVID-19-related health beliefs are affected in the age of social media: Twitter data analysis study. *J Med Internet Res* 2021 Feb 22;23(2):e26302 [FREE Full text] [doi: [10.2196/26302](https://doi.org/10.2196/26302)] [Medline: [33529155](https://pubmed.ncbi.nlm.nih.gov/33529155/)]
20. Al-Metwali BZ, Al-Jumaili AA, Al-Alag ZA, Sorofman B. Exploring the acceptance of COVID-19 vaccine among healthcare workers and general population using health belief model. *J Eval Clin Pract* 2021 Oct 07;27(5):1112-1122 [FREE Full text] [doi: [10.1111/jep.13581](https://doi.org/10.1111/jep.13581)] [Medline: [33960582](https://pubmed.ncbi.nlm.nih.gov/33960582/)]
21. World Health Organization. Advice on the Use of Masks in the Community, During Home Care and in Health Care Settings in the Context of the Novel Coronavirus (2019-nCoV) Outbreak: Interim Guidance, 29 January 2020. Technical Documents. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/330987/WHO-nCov-IPC_Masks-2020.1-eng.pdf?sequence=1&isAllowed=y [accessed 2021-08-02]
22. World Health Organization. Advice on the Use of Masks in the Context of COVID-19: Interim Guidance, 5 June 2020. Technical Documents. URL: https://apps.who.int/iris/bitstream/handle/10665/332293/WHO-2019-nCov-IPC_Masks-2020.4-eng.pdf?sequence=1&isAllowed=y [accessed 2021-08-02]
23. Berg M, Lin L. Prevalence and predictors of early COVID-19 behavioral intentions in the United States. *Transl Behav Med* 2020 Oct 08;10(4):843-849 [FREE Full text] [doi: [10.1093/tbm/ibaa085](https://doi.org/10.1093/tbm/ibaa085)] [Medline: [32893867](https://pubmed.ncbi.nlm.nih.gov/32893867/)]
24. Sohrai C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int J Surg* 2020 Apr 26;76:71-76 [FREE Full text] [doi: [10.1016/j.ijssu.2020.02.034](https://doi.org/10.1016/j.ijssu.2020.02.034)] [Medline: [32112977](https://pubmed.ncbi.nlm.nih.gov/32112977/)]
25. Kaiser Family Foundation. The US Government and the World Health Organization. URL: <https://www.kff.org/coronavirus-covid-19/fact-sheet/the-u-s-government-and-the-world-health-organization/> [accessed 2021-08-02]

26. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. *JMIR Public Health Surveill* 2020 May 29;6(2):e19273 [FREE Full text] [doi: [10.2196/19273](https://doi.org/10.2196/19273)] [Medline: [32427106](https://pubmed.ncbi.nlm.nih.gov/32427106/)]
27. Chen E. Covid-19-Tweet IDs. URL: <https://github.com/echen102/COVID-19-TweetIDs> [accessed 2021-12-11]
28. Vaswani A, Shazeer N, Parmar N. Attention is all you need. 2017 Presented at: NIPS'17: 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA URL: <https://dl.acm.org/doi/10.5555/3295222.3295349>
29. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv* 2018:4805. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
30. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV. XLNet: generalized autoregressive pretraining for language understanding. 2019 Presented at: NIPS'19: 33rd International Conference on Neural Information Processing Systems; December 8-14, 2019; Vancouver, Canada.
31. Cho K, van Merriënboer M, Bahdanau D, Bengio Y. On the properties of neural machine translation: encoder-decoder approaches. *arXiv* 2014 [FREE Full text] [doi: [10.3115/v1/w14-4012](https://doi.org/10.3115/v1/w14-4012)]
32. Gwet KL. Calculating Chance-Corrected Agreement Coefficients (CAC). URL: <https://cran.r-project.org/web/packages/irrCAC/vignettes/overview.html> [accessed 2022-09-21]
33. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol* 2013 Apr 29;13(1):61 [FREE Full text] [doi: [10.1186/1471-2288-13-61](https://doi.org/10.1186/1471-2288-13-61)] [Medline: [23627889](https://pubmed.ncbi.nlm.nih.gov/23627889/)]
34. nytimes / covid-19-data. URL: <https://github.com/nytimes/covid-19-data/blob/master/us.csv> [accessed 2022-01-26]
35. AJMC Staff. A Timeline of COVID-19 Vaccine Developments in 2020. URL: <https://www.ajmc.com/view/a-timeline-of-covid19-developments-in-2020> [accessed 2021-02-05]
36. AJMC Staff. A Timeline of COVID-19 Vaccine Developments in 2021. URL: <https://www.ajmc.com/view/a-timeline-of-covid-19-vaccine-developments-in-2021> [accessed 2022-02-02]
37. BloombergGraphics / covid-vaccine-tracker-data. URL: <https://github.com/BloombergGraphics/covid-vaccine-tracker-data/blob/master/data/historical-global-doses-administered.csv> [accessed 2022-02-02]
38. Christensen SR, Pilling EB, Eyring JB, Dickerson G, Sloan CD, Magnusson BM. Political and personal reactions to COVID-19 during initial weeks of social distancing in the United States. *PLoS One* 2020 Sep 24;15(9):e0239693 [FREE Full text] [doi: [10.1371/journal.pone.0239693](https://doi.org/10.1371/journal.pone.0239693)] [Medline: [32970761](https://pubmed.ncbi.nlm.nih.gov/32970761/)]
39. Li M, Lv G, Hsieh S, Shao R, Yuan J. Pandemic worry and preventive health behaviors during the COVID-19 outbreak. *Front Med (Lausanne)* 2021 Jun 17;8:700072 [FREE Full text] [doi: [10.3389/fmed.2021.700072](https://doi.org/10.3389/fmed.2021.700072)] [Medline: [34222295](https://pubmed.ncbi.nlm.nih.gov/34222295/)]
40. Kwan RYC, Lee PH, Cheung DSK, Lam SC. Face mask wearing behaviors, depressive symptoms, and health beliefs among older people during the COVID-19 pandemic. *Front Med (Lausanne)* 2021 Feb 5;8:590936 [FREE Full text] [doi: [10.3389/fmed.2021.590936](https://doi.org/10.3389/fmed.2021.590936)] [Medline: [33614680](https://pubmed.ncbi.nlm.nih.gov/33614680/)]
41. Sim S, Moey K, Tan N. The use of facemasks to prevent respiratory infection: a literature review in the context of the Health Belief Model. *Singapore Med J* 2014 Mar;55(3):160-167 [FREE Full text] [doi: [10.11622/smedj.2014037](https://doi.org/10.11622/smedj.2014037)] [Medline: [24664384](https://pubmed.ncbi.nlm.nih.gov/24664384/)]
42. Cherry TL, James AG, Murphy J. The impact of public health messaging and personal experience on the acceptance of mask wearing during the COVID-19 pandemic. *J Econ Behav Organ* 2021 Jul;187:415-430 [FREE Full text] [doi: [10.1016/j.jebo.2021.04.006](https://doi.org/10.1016/j.jebo.2021.04.006)] [Medline: [33994606](https://pubmed.ncbi.nlm.nih.gov/33994606/)]
43. Abraham T. Risk and outbreak communication: lessons from alternative paradigms. *Bull World Health Organ* 2009 Aug 1;87(8):604-607 [FREE Full text] [doi: [10.2471/blt.08.058149](https://doi.org/10.2471/blt.08.058149)] [Medline: [19705010](https://pubmed.ncbi.nlm.nih.gov/19705010/)]
44. Latkin CA, Dayton L, Miller JR, Yi G, Jaleel A, Nwosu CC, et al. Behavioral and attitudinal correlates of trusted sources of COVID-19 vaccine information in the US. *Behav Sci (Basel)* 2021 Apr 20;11(4):56 [FREE Full text] [doi: [10.3390/bs11040056](https://doi.org/10.3390/bs11040056)] [Medline: [33924118](https://pubmed.ncbi.nlm.nih.gov/33924118/)]
45. World Health Organization. COVID-19 Advice for the Public: Getting Vaccinated. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines/advice%20> [accessed 2021-08-02]

Abbreviations

- AUROC:** area under the receiver operating characteristic
- BERT:** bidirectional encoder representations from transformers
- BiGRU:** bidirectional gated recurrent unit
- CDC:** Centers for Disease Control and Prevention
- DistilBERT:** distilled version of bidirectional encoder representations from transformers
- FDA:** Food and Drug Administration
- HBM:** Health Belief Model
- HHS:** Department of Health and Human Services
- J&J:** Johnson & Johnson

WHO: World Health Organization

Edited by T Mackey; submitted 09.03.22; peer-reviewed by SF Tsao, M Memon; comments to author 01.07.22; revised version received 18.08.22; accepted 10.09.22; published 31.10.22

Please cite as:

*Ke SY, Neeley-Tass ES, Barnes M, Hanson CL, Giraud-Carrier C, Snell Q
COVID-19 Health Beliefs Regarding Mask Wearing and Vaccinations on Twitter: Deep Learning Approach
JMIR Infodemiology 2022;2(2):e37861*

URL: <https://infodemiology.jmir.org/2022/2/e37861>

doi: [10.2196/37861](https://doi.org/10.2196/37861)

PMID:

©Si Yang Ke, E Shannon Neeley-Tass, Michael Barnes, Carl L Hanson, Christophe Giraud-Carrier, Quinn Snell. Originally published in JMIR Infodemiology (<https://infodemiology.jmir.org>), 31.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Infodemiology, is properly cited. The complete bibliographic information, a link to the original publication on <https://infodemiology.jmir.org/>, as well as this copyright and license information must be included.