Using tweets to understand how COVID-19 related health beliefs are affected in the age of social media

Hanyin Wang, Yikuan Li, Meghan Hutch, Andrew Naidech, Yuan Luo

Submitted to: Journal of Medical Internet Research
on: December 06, 2020

Table of Contents

Using tweets to understand how COVID-19 related health beliefs are affected in the age of social media

Hanyin Wang[1] BMed; Yikuan Li[1] MSc, BSc; Meghan Hutch[1] BSc; Andrew Naidech[2] MD, MPH; Yuan Luo[1] PhD

[1]Department of Preventive Medicine Northwestern University Chicago US
[2]Department of Neurology Northwestern Universtity Chicago US

Corresponding Author:
Yuan Luo PhD
Department of Preventive Medicine
Northwestern University
750 N Lakeshore Dr.
Chicago
US

Abstract

Background: The emergence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 or COVID-19) has given rise to a global pandemic affecting 215 countries and over 40 million people as of October 2020. Meanwhile, we are also experiencing an infodemic induced by the overabundance of information, some accurate and some not, spreading rapidly across social media platforms. Social media has arguably shifted the information acquisition and dissemination of a considerable large population of internet users towards higher interactivities.

Objective: This study aimed to investigate the COVID-19 related health beliefs on one of the mainstream social media platforms, Twitter, as well as the potential impacting factors associated with the fluctuations in health beliefs on social media.

Methods: We used COVID-19-related posts from the mainstream social media platform Twitter to monitor health beliefs. 92,687,660 tweets corresponding to 8,967,986 unique users from January 6 to June 21, 2020 were retrieved. To quantify the health beliefs, we employed the health belief model (HBM) with four core constructs, namely perceived susceptibility, perceived severity, perceived benefits, and perceived barriers. We utilized natural language processing (NLP) and machine learning techniques to automate the process of judging the conformity of each tweet with each of the four HBM constructs. 5,000 tweets were manually annotated for training such machine learning architecture.

Results: The machine learning classifiers yielded AUCs over 0.86 for the classification of all the four HBM constructs. Our analyses revealed a basic reproduction number R_0 of 7.62 for trends in the number of Twitter users posting health belief-related contents over the study period. The fluctuations in the number of health belief-related tweets could reflect dynamics in cases and death statistics, systematic interventions, and public events. Specifically, we observed scientific events, such as scientific publications, and non-scientific events, such as speeches of politicians, are comparable in their abilities to influence health beliefs trends on social media through a Kruskal-Wallis test (P-value = .78 and .92 for perceived benefits and perceived barriers, respectively).

Conclusions: As an analogy of the classic epidemiology model where an infection is considered to be spreading in a population with an R_0>1, the number of users tweeting about COVID-19 health beliefs is amplifying in an epidemic manner and could partially intensify the infodemic. It is "unhealthy" that both scientific and non-scientific events constitute no disparity in impacting the health belief trends on Twitter, since non-scientific events, such as politicians' speeches, might not be endorsed by solid evidence and could be misleading sometimes.

(JMIR Preprints 06/12/2020:26302)
DOI: https://doi.org/10.2196/preprints.26302

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
✔ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
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

Using tweets to understand how COVID-19 related health beliefs are affected in the age of social media

Abstract

Background: The emergence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 or COVID-19) has given rise to a global pandemic affecting 215 countries and over 40 million people as of October 2020. Meanwhile, we are also experiencing an infodemic induced by the overabundance of information, some accurate and some not, spreading rapidly across social media platforms. Social media has arguably shifted the information acquisition and dissemination of a considerable large population of internet users towards higher interactivities.

Objective: This study aimed to investigate the COVID-19 related health beliefs on one of the mainstream social media platforms, Twitter, as well as the potential impacting factors associated with the fluctuations in health beliefs on social media.

Methods: We used COVID-19-related posts from the mainstream social media platform Twitter to monitor health beliefs. 92,687,660 tweets corresponding to 8,967,986 unique users from January 6 to June 21, 2020 were retrieved. To quantify the health beliefs, we employed the health belief model (HBM) with four core constructs: perceived susceptibility, perceived severity, perceived benefits, and perceived barriers. We utilized natural language processing (NLP) and machine learning techniques to automate the process of judging the conformity of each tweet with each of the four HBM constructs. 5,000 tweets were manually annotated for training the machine learning architecture.

Results: The machine learning classifiers yielded AUCs over 0.86 for the classification of all the four HBM constructs. Our analyses revealed a basic reproduction number R_0 of 7.62 for trends in the number of Twitter users posting health belief-related contents over the study period. The fluctuations in the number of health belief-related tweets could reflect dynamics in cases and death statistics, systematic interventions, and public events. Specifically, we observed scientific events, such as scientific publications, and non-scientific events, such as speeches of politicians, are comparable in their abilities to influence health beliefs trends on social media through a Kruskal-Wallis test (*P-value* = .78 and .92 for perceived benefits and perceived barriers, respectively).

Conclusions: As an analogy of the classic epidemiology model where an infection is considered to be spreading in a population with an $R_0 > 1$, the number of users tweeting about COVID-19 health beliefs is amplifying in an epidemic manner and could partially intensify the infodemic. It is "unhealthy" that both scientific and non-scientific events constitute no disparity in impacting the health belief trends on Twitter since non-scientific events, such as politicians' speeches, might not be endorsed by substantial evidence and could be misleading sometimes.

Keywords: COVID-19; social media; health belief; Twitter; infodemic; machine learning; natural language processing.

Introduction

Beginning in December 2019, the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) rapidly evolved into a global pandemic [1-3]. As of the time this paper is written, over 40 million cases and 1 million deaths from 215 countries or regions have been confirmed [4]. However, spreading faster than the virus is information. Sylvie Briand, director of the Infectious Hazards Management at World Health Organization (WHO) 's Health Emergencies Programme, pointed out that "We know that every outbreak will be accompanied by a kind of tsunami of information, but also within this information you always have misinformation, rumors, etc." [5]. WHO used the term "Infodemic" to describe the overabundance of information occurring during the COVID-19 pandemic. Though the term "Infodemic" was first coined in 2002 [6], the concerns over infodemics have become dramatic recently with the amplification effect from social media. WHO held the first Infodemiology Conference in June 2020 as the phenomenon has escalated to a level that requires a coordinated response [7]. Even though we cannot avoid an infodemic, we can still manage it. Previous studies and commentaries proposed several perspectives to detect and fight the COVID-19 infodemic [5,8-10]. However, one of the critical points absent from these studies is an investigation of health beliefs. Understanding how the general public's health beliefs are expressed and altered can facilitate our management of both the pandemic and infodemic. In conjunction, it is also essential to evaluate any concurrent or ongoing interventions.

The health belief model (HBM) quantifies health beliefs [11-13]. The HBM was developed to investigate people's beliefs about health problems. It consists of the following four core constructs that can be tailored for given hypotheses: 1) perceived susceptibility, 2) perceived severity, 3) perceived benefits, and 4) perceived barriers. The HBM has been widely used to investigate individual opinions towards diseases and interventional approaches, such as HIV risk behaviors [14], HPV vaccines [15], and the gender difference in food choices [16]. In those cases, HBM was employed to evaluate the people's beliefs towards the given health problem and their perceived benefits or barriers of action, for which each of the core constructs of the HBM is assessed based on the corresponding definitions. During pandemics, researchers have employed HBM to investigate the health beliefs towards public interventional policies, such as stay-at-home orders [17], to analyze public health communication on Instagram during the Zika outbreak [18], to examine public perceptions of physical distancing [19], and to guide the community pharmacists in their communication with patients [20]. However, because these are survey-based or merely commentary studies, results are limited to the analyzed population and, therefore, may be biased. In this study, we expanded and diversified our study population by using crowdsourcing data from one of the mainstream social media platforms, Twitter, in order to investigate the health beliefs of the general public towards COVID-19 and its potential treatments.

In addition to quantifying health beliefs, we aimed to identify factors influencing fluctuations in public opinions. For instance, the pandemic dynamics, i.e., the number of cases and deaths due to COVID-19, constitute one of the leading factors influencing attitudes towards the pandemic. Additionally, interventional government policies may also impact the opinions of the general public. Furthermore, it is reasonable to believe that health belief-related posts can also be self-regulated as a consequence of their nature to induce or sooth panic for readers. Potential treatments trigger massive discussions as well, such as the debate over the appropriate use of the anti-malarial drug Hydroxychloroquine (HCQ) or Chloroquine (CQ), advocated by the American President as a "game-changer," then subsequently discarded. Public attitudes regarding potential treatments may be altered by public events such as the news or politicians' speeches. Furthermore, rapidly emerging scientific publications can also influence the point of view of the general public. In this paper, we aim to identify factors that impact health beliefs on social media, which may serve as a probe for identifying better strategies to manage both the pandemic and the infodemic.

Contributions of the current study include,

- An evaluation of utilizing a mainstream social media platform, Twitter, to facilitate comprehensive understandings of health beliefs towards COVID-19 and potential treatments.
- A publicly available data set annotated by multiple professionals for studying the health beliefs related to COVID-19 and potential treatments.
- Identification and comparison of factors that influence the health beliefs towards COVID-19 and potential treatments, Hydroxychloroquine or Chloroquine in particular.
- An extendable framework for monitoring the general public's health beliefs during a pandemic and infodemic, which could be feasibly transferred to facilitate the management of future infodemic outbreaks, such as when COVID-19 vaccines become available to the public.

Methods

The entire workflow of data extraction, filtering, and classification is illustrated in Figure 1.

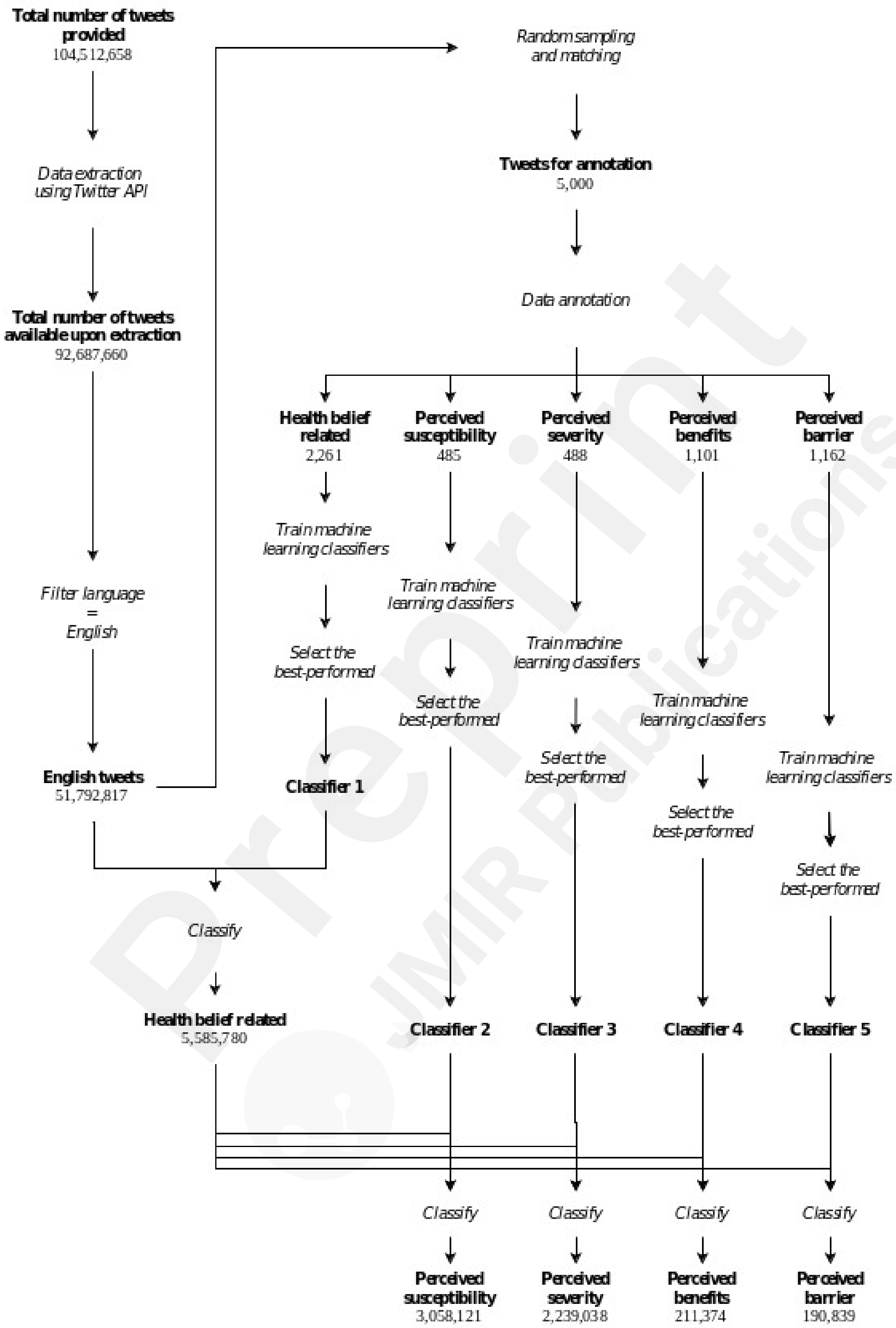


Figure 1.

Data

We used version 15.0 of the COVID-19 Twitter chatter dataset constructed by the Panacea

Lab [21], which collected all COVID-19 related tweets between January 6, 2020 and June 21, 2020. The provided dataset only contains the identifiers of corresponding tweets, so we used the application programming interface (API) provided by Twitter to extract the full content of each tweet. The used Social Media Mining Toolkit (SMMT) tool provided by the Panacea Lab hydrate the dataset [22]. There is no limitation regarding the days prior to the extraction. A language filter ("lang" attribute in the tweet object = "en") was then applied to identify tweets written in the English language.

Data annotation for constructing the health belief model

We employed the health belief model (HBM) to quantify health beliefs. It consists of the following four core constructs that can be tailored for given hypotheses: 1) perceived susceptibility, 2) perceived severity, 3) perceived benefits, and 4) perceived barriers. The HBM was developed to investigate people's beliefs about health problems and has been widely used to investigate individual opinions towards diseases and interventional approaches, such as HIV risk behaviors [14], HPV vaccines [15], and the gender difference in food choices [16]. To construct the health belief model (HBM), labels (positive or negative) were assigned for each tweet based on whether the contents of the given tweet fell into the definition of the four HBM core constructs. Specifically, for perceived benefits and barriers, we focused on Hydroxychloroquine (HCQ) or Chloroquine (CQ), the anti-malarial drug advocated by the American President as a "game-changer," then subsequently discarded. Tweets were also labeled positive or negative for HBM related, meaning they could be mapped to at least one of the four aforementioned constructs. Thus, each tweet could potentially have up to five labels. The annotation process was performed by three senior Ph.D. students in biomedical informatics (authors HW, YL (Li), and MH). All annotators classified the first 500 tweets individually, then reconciled different opinions and built final annotation rules. The definitions for each construct of the HBM are described in Table 1 with examples. Based on the rule, HW and YL annotated the rest of 5,000 tweets independently and evaluated the agreement by Cohen's Kappa score [23]. Finally, MH resolved the divergent annotations between HW and YL with further consideration. We made the dataset with the 5,000 annotated tweets available for researchers (Supplemental Dataset). To protect the privacy of Twitter users and per the policy of Twitter, we did not include any tweet content in the dataset. Instead, the unique identifiers for each tweet (tweet ID) were provided.

Table 1. Health Belief Model constructs, definitions, and examples

Constructs	Definition of the Constructs	Examples
Perceived Susceptibility	The assessment of the risk of getting COVID-19 infection.	*"Across the UK, 194,990 people had tested positive for coronavirus as of 9am on Tuesday, up from 190,584 at the same point on Monday. Find out how many cases there are in your area."*
Perceived Severity	The assessment of whether COVID-19 is a sufficient health concern.	*"US Recorded 1,297 Coronavirus Deaths in Past 24 Hours."*

Perceived Benefits	The benefits of HCQ/CQ[a] in prevention or treatment of COVID-19; Positive statements or reports about HCQ/CQ.	*"Dr. Zelenko In NY has now treated 699 Coronavirus patients with 100% success using Hydroxychloroquine."*
Perceived Barriers	The side effects of HCQ/CQ; The unaffordable cost of HCQ/CQ; The inaccessibility of HCQ/CQ; Negative statements or reports about HCQ/CQ.	*"Family of New York woman blames hydroxychloroquine combo for fatal heart attack."*
HBM related	Can be mapped to at least one of the above.	

[a] HCQ: Hydroxychloroquine; CQ: Chloroquine

Machine learning classifiers

We trained machine learning classifiers on the annotated data, evaluated the performance, and automatically classified the 50 million tweets. The entire annotated data set was split into training and testing set with the proportion 8:2. Feature selection was applied by ignoring terms with document frequency less than 0.01 or greater than 0.99 were ignored. Terms with only letters are considered (ignored when there are numbers or special characters). Before vectorization, we removed all the URLs, unified all the contractions, punctuation marks, and white spaces, and lowercased all terms in the corpus. The free-text tweets were vectorized using both Bag-of-Words and term frequency-inversed document frequency (tf-idf) algorithms. A list of English stop words provided by Natural Language Toolkit (NLTK) [24] was used to rule out unrelated words. Five-fold cross-validation was performed to select the best-suited classifier for the task. The machine learning classifiers that we experimented on include Ridge classifier, perceptron, passive-aggressive classifier, k-nearest neighbors classifier, random forest, support vector machine with linear kernel and l1 or l2 penalty, support vector machine with rbf or poly or sigmoid kernel, stochastic gradient descent classifier with l1 or l2 or elastic net penalty, multinomial naïve Bayesian classifier, Bernoulli naïve Bayesian classifier, and logistic regression. Performance across classifiers is evaluated using the area under the receiver operating characteristic curve (AUC), and the classifier that yielded the highest AUC was chosen. In total, five classifiers were built to construct the final HBM. First, a classifier was trained to classify whether a tweet was HBM-related or not. Then, we collected all the tweets that were identified as HBM-related for the following task. Lastly, we built four classifiers to label each core construct of the HBM separately.

The entire pipeline was built with Python V 3.6.8. Bag-of-Words, machine learning classifiers, and model evaluations were implemented with the scikit learn V 0.22.1 package [25].

The overall trend of health beliefs in tweets

To quantify whether the information spread constitutes an infodemic, we applied one of the classic measurements in the epidemiological models, the basic reproduction number (R_0). We employed the SIR model [26], for which the detailed calculation can be found in the

Supplementary Information. In our case of infodemic, we consider the users who tweeted about COVID-19 as the susceptible population, among which "being infected" means a user tweeting about health beliefs defined in our HBM scope, "recovering" would then indicate a user stopping tweeting about health beliefs. Thus, "contacting with infected individuals" could be considered as reading health belief-related tweets posted by other users.

The trend of health belief towards the disease

There are two core constructs in the HBM that focus specifically on the disease of interest: perceived susceptibility and perceived severity. We visualized these two constructs together with the dynamics of the pandemic in *Figure 2*. We observed a similar pattern in COVID-19 case dynamics and the number of tweets regarding perceived susceptibility, as well as in the dynamics of COVID-19 deaths and the number of tweets indicating perceived severity. For the first pair, we observed an earlier increase in the number of perceived susceptibility tweets prior to a surge in COVID-19 cases, while for the second pair, there is a delay in the increase of COVID-19 deaths compared with the number of perceived severity tweets. To investigate how many days the trend dynamics of health belief discussions proceeded or postponed the actual case or deaths increase, we calculated Spearman's correlation coefficient under various time lags, i.e., where for a one-day lag, the correlation between the number of tweets and COVID-19 situation was calculated by moving the COVID-19 trend one day forward. Moreover, we conducted a change point analysis using the dynamic programming algorithm [27] to detect the significant turning point of the trends.

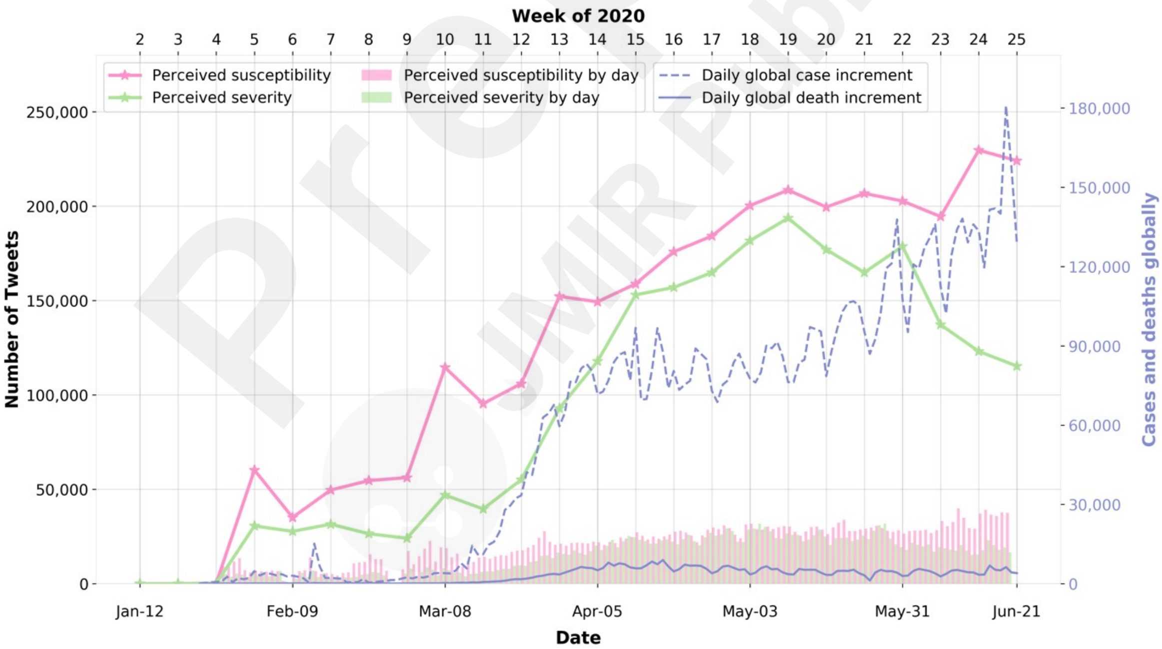


Figure 2.

The effect of interventions

To evaluate the impact of interventions on the infodemic and the pandemic, we further investigated the lockdown in the United States (US). Because we are analyzing only tweets written in the English language, and there were systematic official lockdowns issued in the

US, we chose to study the effect of US-based interventions on health beliefs. The location information is not available for each tweet. We analyzed 136,641 tweets with the *place* (a variable in the tweet object: when the present, indicates that the tweet is associated with a place) available and subsequently identified 54,164 tweets corresponding to the US. We investigated the effect of interventions by visualizing the trends along with the timeline of lockdowns in the US.

News in the top topics

To understand major topics in the tweets related to health beliefs, we extracted the top 10 phrases from tweets each week. We considered uni-grams and bi-grams in this case. The frequency of each phrase is not only calculated as the count; instead, we used the term frequency-inversed document frequency (tf-idf) score to find the highlight topics of each week. A higher tf-idf score is obtained if a given word or phrase frequently appears in one document but only appears in a small number of documents.

The influence of scientific and non-scientific events

To evaluate the difference between the impact of science and non-scientific events on health beliefs, we conducted a Kruskal-Wallis test. Kruskal-Wallis was chosen since there is no reasonable assumed distribution for the influence of the two types of events, and the two groups being compared have different sample sizes. We collected events associated with HCQ / CQ on the internet during the study period with no exclusion criteria. All the events are classified as scientific events if they are based on scientific evidence or endorsed by authorities, while all remaining events are treated as non-scientific events. Full references for each event can be found in Supplementary Table 7. To quantify the influences, for each event, we calculated the sum of the number of tweets that expressed perceived benefits or perceived barriers regarding HCQ / CQ on the day of the event and the day after.

Results

Data and machine learning classifiers

The dataset contains identifiers for 104,512,658 unique tweets, of which 92,687,660 were still available upon extraction. After applying the language filter, our final set for analysis consisted of 51,792,817 English tweets. Cohen's Kappa score for inter-rater reliability of data annotation was 0.94 for identifying whether a given tweet was health belief model (HBM)-related and around 0.9 for the annotation of all four individual HBM constructs (Table 2). Random forest was found to be the best-performing classifier for the HBM-related classification and three of the HBM constructs (perceived susceptibility, perceived benefits, and perceived barriers), while the passive-aggressive classifier was found to be the most suitable choice for classifying whether a tweet indicates perceived severity. The area under the receiver operating characteristic curve (AUC) for HBM-related and the two disease-related constructs were all above 0.9, while AUCs for the two treatment-related constructs were around 0.86 (Table 2).

After classification, 5,585,780 tweets were HBM-related, among which 3,058,121 (54.75%)

tweets expressed perceived susceptibility of COVID-19, 2,239,038 (40.08%) tweets expressed perceived severity of COVID-19, 211,374 (0.04%) tweets expressed perceived benefits of Hydroxychloroquine (HCQ) or Chloroquine (CQ), and 190,839 (0.03%) tweets expressed perceived barriers towards HCQ or CQ. To further ensure the validity of the classification, we performed additional spot checks on the final results; examples can be found in Supplementary Information.

Table 2. Performance of machine learning classifiers

Constructs	*Kappa*[a]	*Classifier*[b]	*AUC*[c]	*Accuracy*	*Precision*[d]	*Recall*[e]	*F1*[f]
Perceived Susceptibility	0.92	Random Forest	0.97	0.94	0.92	0.85	0.88
Perceived Severity	0.88	Passive-Aggressive	0.92	0.90	0.88	0.77	0.81
Perceived Benefits	0.92	Random Forest	0.87	0.79	0.78	0.78	0.78
Perceived Barriers	0.92	Random Forest	0.86	0.77	0.77	0.77	0.77
HBM related	0.94	Random Forest	0.90	0.84	0.84	0.84	0.84

[a] Cohen's kappa coefficient for inter-rater reliability of annotation
[b] The Machine Learning classifier selected by the best performance
[c] AUC: Area Under the Receiver Operating Characteristic Curve
[d] Macro averaged precision
[e] Macro averaged recall
[f] Macro averaged F1 score

The overall trend of health beliefs in tweets

The visualization of the overall trend of health beliefs is shown in Figure 3 with the number of tweets that fell into each core construct of the health belief model (HBM). Each construct was displayed in a different color chronologically, starting from the third week of 2020, and stacked together to show the total number of HBM-related tweets. A dramatic increase can be observed from January to June, which indicates an increasing amount of discussions regarding personal health beliefs. The R_0 was 7.62 for the users who tweeted regarding health beliefs in our data.

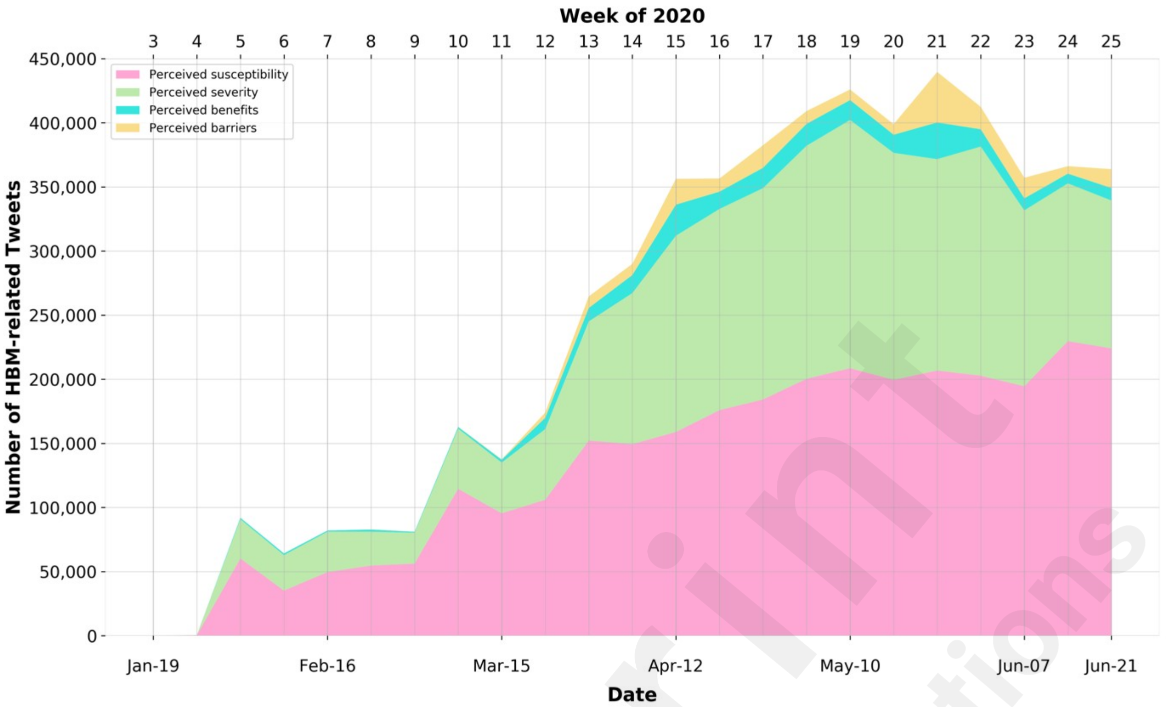


Figure 3.

The trend of health belief towards the disease

The left panel of *Figure 4* displays the strongest correlation (0.92) between perceived susceptibility-related tweets and the global case increment when imposing a three-day lag, i.e., moving the trend of COVID-19 cases three days forwards so that the number of perceived susceptibility tweets on January 13, 2020 will be aligned with the number of COVID-19 cases on January 16, 2020. The patterns detected by the change points analysis depicted by colors in *Figure 4* also show similarities within the pair. For the second pair (perceived severity and COVID-19 deaths trend in the right panel of *Figure 4*), the strongest correlation was found at -6 days (0.87), which indicates that changes in the perceived severity are lagging the actual death dynamics by six days, i.e., the strongest correlation was found when moving the deaths trend six days backward. The change point analysis unraveled similar patterns between the trends of perceived severity and COVID-19 deaths.

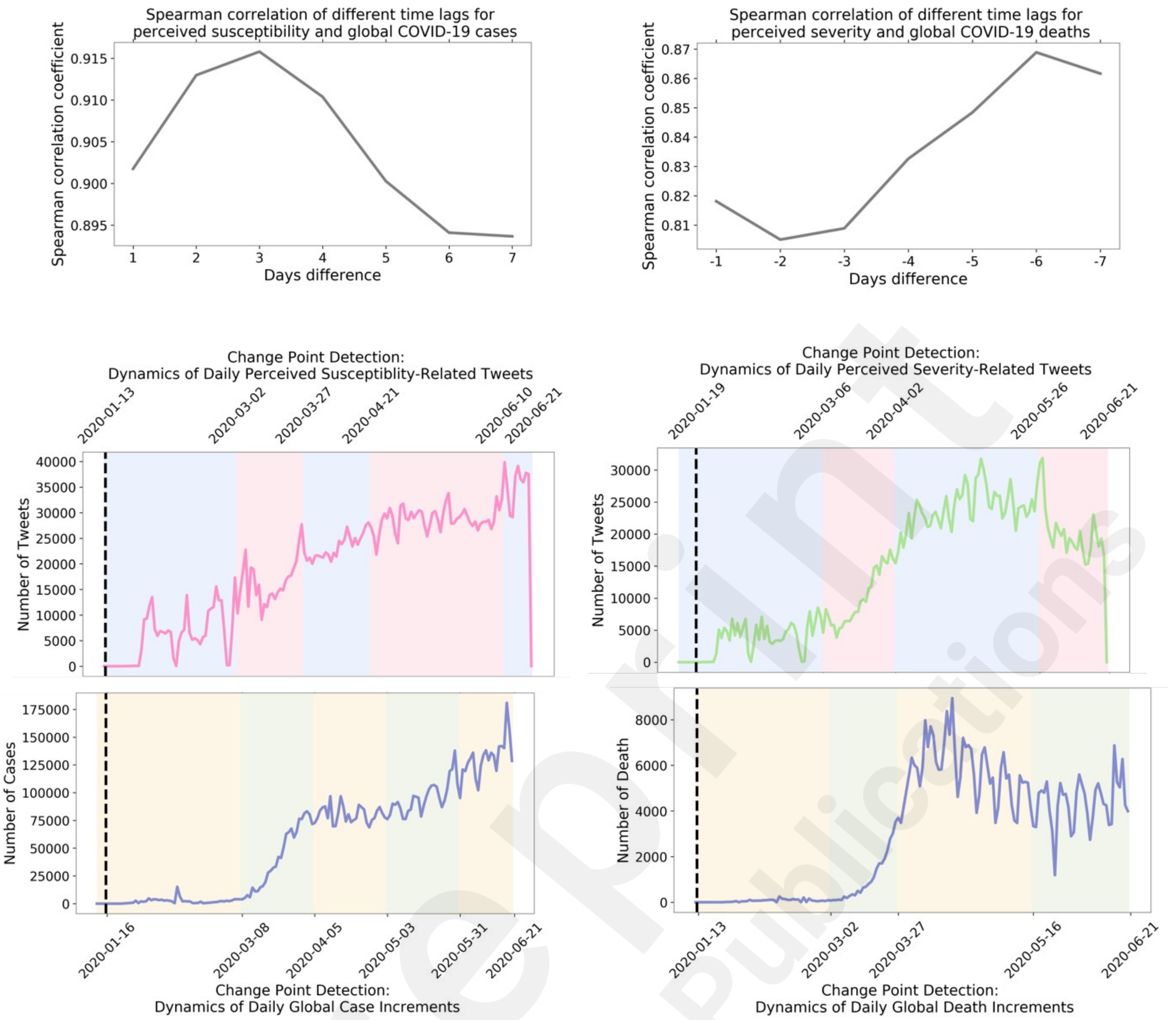


Figure 4.

The effect of interventions

We visualized the trends of perceived susceptibility and perceived severity along with the daily cases and death dynamics in the US in Figure 5. Meanwhile, lockdown information for each state is also listed by the timeline. Decisions made by Republican or Democratic governors are colored red and blue. South Dakota (Republican), North Dakota (Republican), Iowa (Republican), Nebraska (Republican), and Arkansas (Republican) did not announce official lockdowns and are not included in this figure. The official documents of lockdown and reopen decisions for each state are listed in Supplementary Table 6.

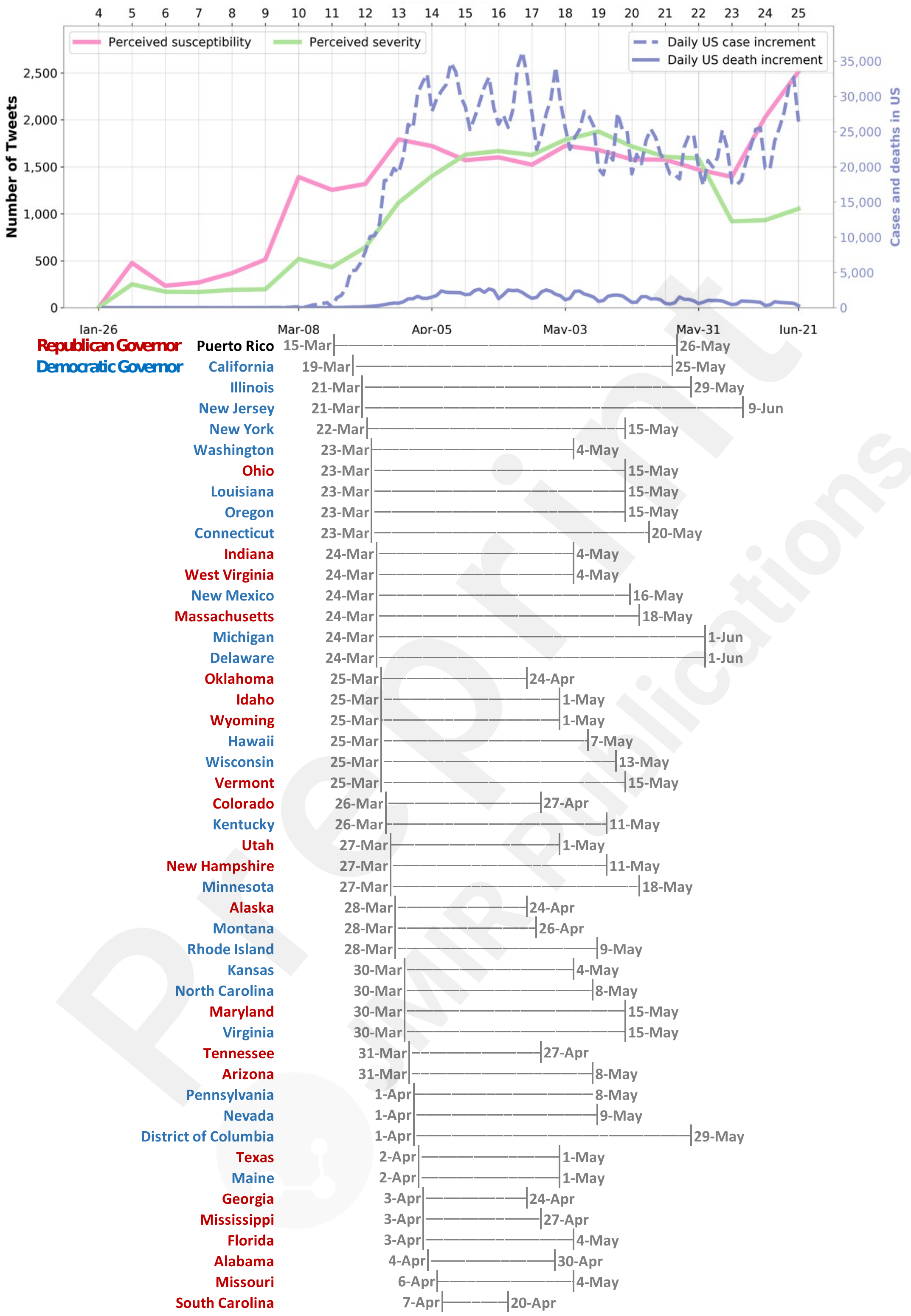


Figure 5.

News in the top topics

Top topics (all lowercased) according to the term frequency-reversed document frequency (tf-idf) scores are shown in Figure 6, where a darker shade of the cell indicates a higher tf-idf score. We highlighted the featured phrases that are closely related to the news during the corresponding time periods in purple.

Week	Data interval	1	2	3	4	5	6	7	8	9	10
2	Jan-06~Jan-12	**advisory chinese**	experts search	low experts	via tldr	viral nucleic	difficult confirm	search answers	notify health	weeks returning	outbreak ask
3	Jan-13~Jan-19	closed known	**market infection**	turn lethal	lethal likely	**spread seafood**	coronavirus rpts	**ncov japan**	report laboratory	major transport	dies novel
4	Jan-20~Jan-26	spreads multiple	**france confirms**	**confirmed chicago**	**texas student**	publishes early	warns grave	alarming consequences	consequenc es	points alarming	exposed lancet
5	Jan-27~Feb-02	**outside china**	**philippines**	spreads multiple	**germany**	**confirmed uk**	spreads regions	russia	**outbreak spreads**	rises outbreak	jumps outbreak
6	Feb-03~Feb-09	cruise	ship	**cruise ship**	li	wenliang	**li wenliang**	jumps outbreak	chinese rage	whistleblo wer	death coronavirus
7	Feb-10~Feb-16	ship	cruise	**cruise ship**	covid	reports new	spike	**trump**	seemed leveling	outside china	princess
8	Feb-17~Feb-23	**iran**	**italy**	covid	ship	cruise	cruise ship	outside china	spike	reports new	diamond princess
9	Feb-24~Mar-01	italy	iran	covid	trump	outside china	many cases	china real	coronavirus many	countries affected	updates number
10	Mar-02~Mar-08	**covid**	italy	iran	trump	york	**new york**	covid cases	cases covid	cruise	declares
11	Mar-09~Mar-15	covid	italy	trump	covid cases	cases covid	iran	coronavirus covid	**spain**	spread covid	community
12	Mar-16~Mar-22	covid	italy	trump	covid cases	chloroquine	cases covid	covid patients	spread covid	**coronavirus pandemic**	old
13	Mar-23~Mar-29	covid	**trump**	covid cases	italy	covid patients	chloroquine	cases covid	york	new york	**hydroxychl oroquine**
14	Mar-30~Apr-05	covid	trump	covid cases	covid patients	hydroxychlo roquine	italy	york	new york	spain	cases covid
15	Apr-06~Apr-12	covid	trump	**hydroxychl oroquine**	covid cases	covid patients	covid deaths	york	new york	italy	hours
16	Apr-13~Apr-19	covid	trump	covid cases	covid patients	covid deaths	hydroxychlo roquine	cases covid	homes	york	new york
17	Apr-20~Apr-26	covid	trump	covid cases	covid patients	hydroxychlo roquine	covid deaths	cases covid	york	new york	nursing
18	Apr-27~May-03	covid	covid cases	trump	covid patients	covid deaths	**remdesivir**	cases covid	homes	**nursing**	new deaths
19	May-04~May-10	covid	trump	covid cases	covid patients	covid deaths	highest	nursing	homes	cases covid	new deaths
20	May-11~May-17	covid	covid cases	trump	covid patients	nursing	covid deaths	homes	nursing homes	hydroxychlo roquine	cases covid
21	May-18~May-24	covid	trump	hydroxychlo roquine	covid cases	covid patients	nursing	homes	covid deaths	nursing homes	cases covid
22	May-25~May-31	covid	covid cases	trump	covid patients	covid deaths	hydroxychlo roquine	nursing	homes	nursing homes	highest
23	Jun-01~Jun-07	covid	covid cases	covid patients	trump	hydroxychlo roquine	covid deaths	new covid	**george floyd**	hours	spike
24	Jun-08~Jun-14	covid	covid cases	trump	covid patients	spike	active cases	highest	new covid	cases covid	covid deaths
25	Jun-15~Jun-21	covid	covid cases	trump	**dexametha sone**	florida	covid patients	spike	record	new covid	cases covid

Figure 6.

The influence of scientific and non-scientific events

The list of events that we collected is shown on the right-hand side of *Figure 7*, while the trends of perceived benefits and barriers are shown on the left-hand side. We observed that both scientific and non-scientific events are associated with fluctuations in health beliefs. The scales of the fluctuations observed vary over time. There are more non-scientific events around the two most massive spikes, but scientific events are majorly distributed along the timeline where many gentle fluctuations can be found. The Kruskal-Wallis test showed no significant difference for between the influence of scientific and non-scientific events for both perceived benefits and barriers (H = 0.078, *P-value* = .78 and H = 0.002, *P-value* = .92, respectively).

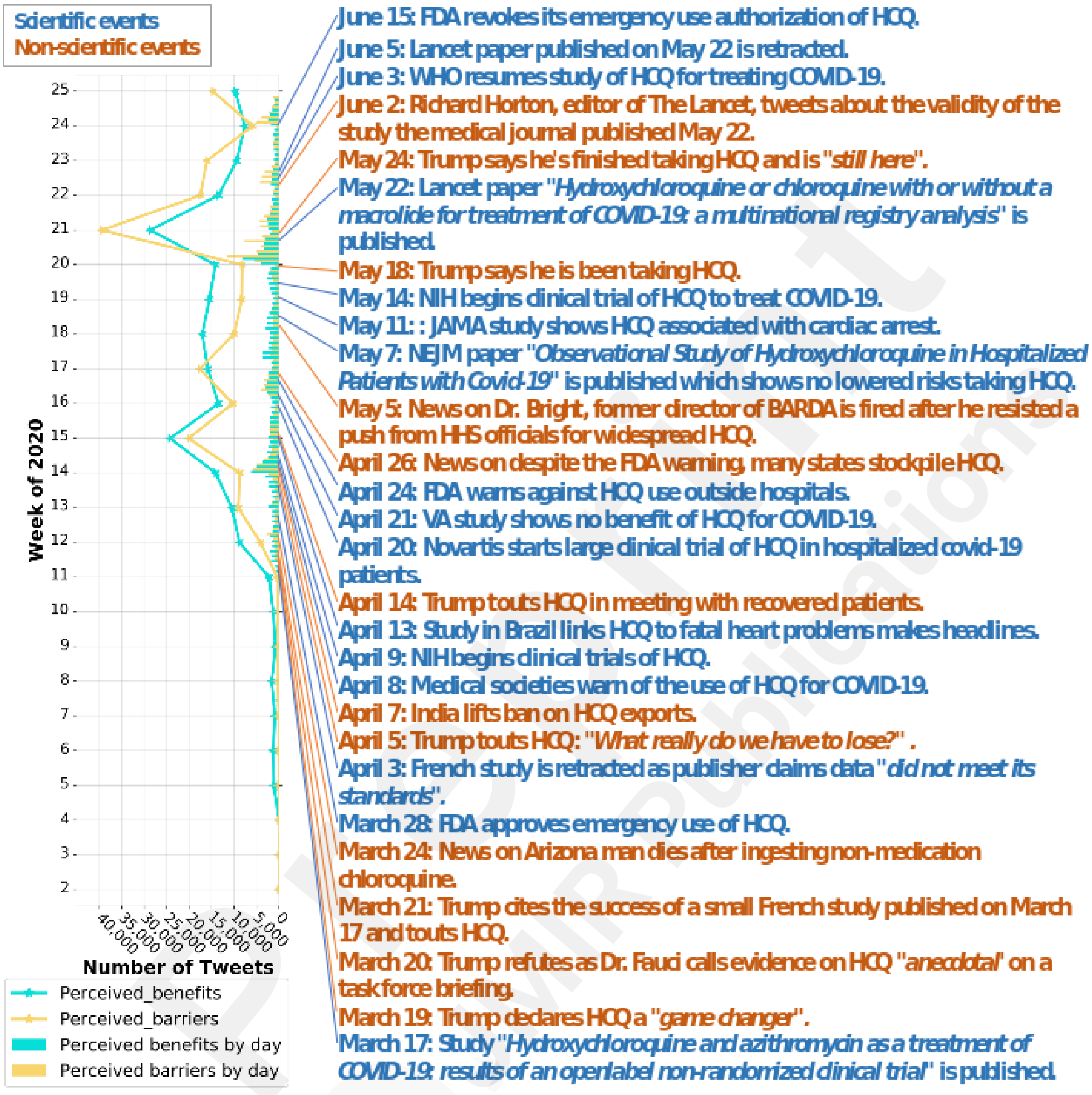


Figure 7.

Discussions

Through the utilization of natural language processing and machine learning, we employed the health belief model (HBM) to identify tweets associated with health beliefs. Through further evaluation of HBM related tweets, our findings demonstrate that trends in health beliefs are correlated with dynamics in positive cases and mortality rates. Additionally, we observe a decline in perceived disease susceptibility during government-issued lockdowns, while perceived severity appears unaltered. Lastly, our study identifies top news events, scientific and non-scientific, that may play a role in altering health beliefs. These findings lay the groundwork to better understand how the general public's COVID-19 related health beliefs are influenced by case and mortality rates, government policies, current news, and significant events. Through careful study of these observations, we may better implement management strategies to combat the pandemic and infodemic.

In commonly used models for infectious diseases, infection is considered to be spreading in a population when $R_0 > 1$ and the epidemic is harder to control with a larger value of R_0. Therefore, given the R_0 of 7.62, it is reasonable to conclude that an infodemic is ongoing in our study population.

It is interesting that health beliefs involving perceived susceptibility increased in advance of the actual evolvement of the pandemic. Because we observed the basic reproduction number R_0 of 7.62, suggestive of an infodemic, these findings may suggest that the volume of information regarding COVID-19 affects the Twitter users' perspectives regarding the risk of infection. In the early stages of the pandemic, before mortalities were observed, it is possible that less severity was assumed. Overtime, perceived severity may have increased as the number of deaths cumulated. Strong correlations between perceived susceptibility and perceived severity regarding the case and death dynamics may suggest that the ongoing situation of the pandemic is a significant impact factor on health beliefs.

From the line chart in Figure 5, we observed a dramatic increase in daily cases between week 11 and week 14. There is also an upward trend in perceived susceptibility starting from week 11, which began decreasing by week 13. This phenomenon is interesting when we take the lockdown situation into consideration, as starting from week 13 was when most of the states were under the government issued lockdown. Thus, official interventions are observed to potentially mitigate the general public's perceived susceptibility of COVID-19. Meanwhile, we see the growth of the number of confirmed cases slowing down during this same period. Interestingly, we fail to see a decline in perceived severity even when almost all the states are under quarantine. Previously, we showed that perceived severity was found to most strongly correlate with mortality; thus, it is reasonable that lockdown policies did not ease such health concerns, perhaps owing to the fact that while lockdowns slow down the spread of infection, they do not offer complete protection, especially in the absence of viable medications or treatment strategies.

As shown in Figure 6, for the first three to four weeks, topics predominately covered confirmed cases worldwide when the global pandemic was not yet affirmed by the authorities. In proceeding weeks, *cruise ship* [28] and *li wenliang* [29] came into the spotlight. In early February, a large and notable cluster of COVID-19 cases occurred on the Diamond Princess cruise ship. Dr. Wenliang Li, the Chinese doctor who tried to alarm a possible outbreak of a disease that resembled severe acute respiratory syndrome (SARS) in Wuhan, China, died of the infection on February 7. During weeks 8 and 9, when the COVID-19 outbreak heightened in Italy [30] and Iran [31], topics related to these countries began trending. On March 18 (week 12), President Trump announced that he was taking HCQ as prophylaxis for COVID-19 [32] and triggered massive discussions. In fact, starting week 13, discussions involving Hydroxychloroquine (HCQ) / Chloroquine (CQ) begin to dominate. Lastly, we were initially surprised to observe other topics like *george floyd* [33] in the health belief related tweets. However, this topic is related to many gathering events that happened while many states were still under lockdown, possibly provoking health concerns. Through this analysis, we suspect the news of all sources may penetrate into discussions regarding health beliefs and, thus, may influence health beliefs. Therefore, the news that we consume everyday inadvertently may be a substantial factor that affects our health beliefs, which may contribute to and even exacerbate the infodemic on social media.

We observed that speeches of politicians could have dramatic impacts on the health beliefs of the general public who read the news. However, politicians' speeches do not necessarily recapitulate scientific facts or evidence and could sometimes be misleading [34]. Thus, we expect to rely more on scientific sources, such as publications with scientific evidence or announcements made by health authorities, for more accurate and reliable information regarding the pandemic. However, it is uncertain whether scientific events or non-scientific events have a more profound influence on altering the health beliefs of the general public.
The results from the Kruskal-Wallis test implies that scientific events and non-scientific events did not significantly differ from one another in regard to their effect on health beliefs within the given period (January 6 to June 21). We found it surprising that scientific events did not appear to be significantly associated with altering the health beliefs towards potential treatments in our dataset. This might be due to the public's distrust in science arising from the many uncertainties involving the pandemic or the instances of being delivered conflicting information such as *"Don't wear masks"* to *"Wear masks all the time"*. To better cope with COVID-19 circumstances, everyone in society, online and offline, should be aware of the overabundance of information and its potential impact on health beliefs. Thus, it is essential to be prudent to screen the authenticity of each piece of information.

Limitations and future works

We identify some limitations in the current study. The tweets analyzed in this study covered the English language, and there might be divergences across different languages that are not addressed. Although English tweets takes the largest proportion among all the tweets, the number of tweets in other languages or undefined languages is still considerable [35]. We hope to expand the analysis to a multilingual setting in future works. Additionally, although we did not cover every potential treatment at this stage, our framework is extensible to assess the influence on health beliefs of additional treatments or interventions, such as vaccines. In fact, we plan to apply similar approaches to investigate health beliefs in COVID-19 vaccines once available to the general public. Furthermore, we likely have not considered other factors that may contribute to alterations in health beliefs.
The analysis used data extracted from one of the social media platforms, Twitter, which may also introduce bias. Users may not represent the health beliefs of the entire population since not everyone uses Twitter. More social media platforms will be incorporated in future works, such as Facebook, Instagram, and Reddit. Additionally, it would be interesting to compare our crowdsourcing results with health beliefs obtained through hospital administrated surveys from patients with COVID-19 and their caretakers.

Technically, this study employed the very classic text classification methods, which use a combination of the Bag-of-Words model and machine learning classifiers. Yet, the experiments showed that they worked well (AUC over 0.9) on the given data. Deep learning architectures were not discussed in this study. Majorly because there is no guarantee that deep learning models always work better than simple machine learning classifiers. Meanwhile, deep learning models bear higher technical barriers which compromise the accessibility for people from other domains. Deep learning models are also known to demand considerable energy [36], so we were also trying to trade off the energy-performance balance. But it is definitely worth a whole other study discussing various NLP techniques for the classification task. In future studies, we are also interested in investigating the performance of various natural language processing techniques on the current text classification tasks.

Conclusion

Our data suggest that we are not only fighting a pandemic but also an infodemic. The excessive information disseminated on social media platforms and other sources is closely related to the dynamics of the general public's health beliefs. The dynamics of the pandemic, news, scientific and non-scientific events, and even the related tweets already published on social media platforms may influence the health beliefs of the general public on social media to some extent. Our findings provide clues and evidence for more effective management of the infodemic associated with the COVID-19 pandemic.

Conflicts of Interest

The authors declare no competing interests.

Abbreviations

HBM: health belief model
NLP: natural language processing
AUC: area under receiver operating characteristic
WHO: World Health Organization
API: application programming interface
SMMT: Social Media Mining Toolkit
HCQ: hydroxychloroquine
CQ: chloroquine
tf-idf: term frequency-inversed document frequency
NLTK: Natural Language Toolkit
SARS: severe acute respiratory syndrome
FDA: U.S. Food & Drug Administration
NIH: National Institutes of Health
JAMA: The Journal of the American Medical Association
NEJM: The New England Journal of Medicine
BARDA: Biomedical Advanced Research and Development Authority
HHS: United States Department of Health and Human Services
VA: United States Department of Veterans Affairs

Acknowledgment

H.Wang, Y.Li, and Y.Luo conceived of the presented idea. H.Wang, Y.Li, and M.Hutch contributed to the data annotation and validation. H.Wang carried out all the experiments and conducted all the data analyses. H.Wang wrote the manuscript and designed the tables and figures with supports from M.Hutch and Y.Li. A.Naidech provided clinical insights and

contributed to the result interpretation. Y.Luo advised and supervised the entire project.

We thank Twitter for providing the application programming interface (API) for developers to retrieve tweets for research.

Multimedia Appendix 1

The 5,000 manually annotated data for the health belief model and machine learning classifier training code snips have been deposited in the GitHub https://github.com/HanyinWang/CovidHealthBeliefTweets.

Reference

1. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*. 2020;395(10223):507-513.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*. 2020;395(10223):497-506.
3. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*. 2020.
4. World Health Organization. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. Published 2020. Accessed Oct-17, 2020.
5. Zarocostas J. How to fight an infodemic. *The Lancet*. 2020;395(10225):676.
6. Eysenbach G. Infodemiology: The epidemiology of (mis) information. *The American journal of medicine*. 2002;113(9):763-765.
7. World Health Organization. 1st WHO Infodemiology Conference. https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference. Published 2020. Accessed Oct-17, 2020.
8. Cinelli M, Quattrociocchi W, Galeazzi A, et al. The covid-19 social media infodemic. *arXiv preprint arXiv:200305004*. 2020.
9. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. *Journal of medical Internet research*. 2020;22(6):e21820.
10. Hua J, Shaw R. Corona virus (Covid-19)“infodemic” and emerging issues through a data lens: The case of china. *International journal of environmental research and public health*. 2020;17(7):2309.
11. Janz NK, Becker MH. The health belief model: A decade later. *Health education quarterly*. 1984;11(1):1-47.
12. Rosenstock IM. Historical origins of the health belief model. *Health education monographs*. 1974;2(4):328-335.
13. Strecher VJ, Rosenstock IM. The health belief model. *Cambridge handbook of psychology, health and medicine*. 1997;113:117.
14. Rosenstock IM, Strecher VJ, Becker MH. The health belief model and HIV risk behavior change. In: *Preventing AIDS*. Springer; 1994:5-24.
15. Du J, Cunningham RM, Xiang Y, et al. Leveraging deep learning to understand health beliefs about the Human Papillomavirus Vaccine from social media. *npj Digital Medicine*. 2019;2(1):27.
16. Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisie F. Gender differences in food choice: the contribution of health beliefs and dieting. *Annals of behavioral medicine*. 2004;27(2):107-116.
17. Czeisler MÉ, Tynan MA, Howard ME, et al. Public attitudes, behaviors, and beliefs related to COVID-19, stay-at-home orders, nonessential business closures, and public health guidance—United States, New York City, and Los Angeles, May 5–12, 2020. *Morbidity and Mortality Weekly Report*. 2020;69(24):751.
18. Guidry JPD, Carlyle KE, LaRose JG, Perrin P, Messner M, Ryan M. Using the Health

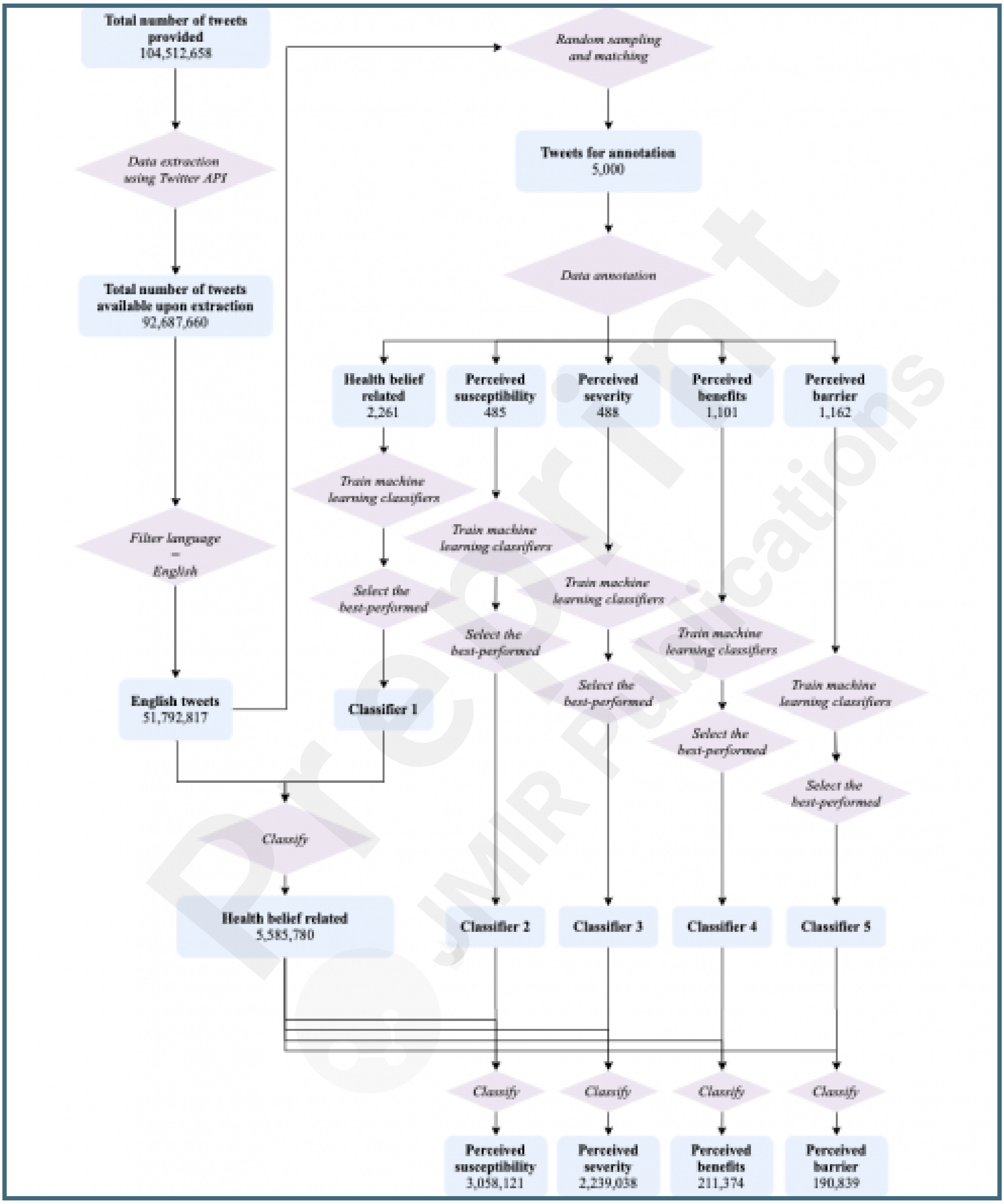
Belief Model to Analyze Instagram Posts about Zika for Public Health Communications. *Emerg Infect Dis*. 2019;25(1):179-180.
19. Sesagiri Raamkumar A, Tan SG, Wee HL. Use of Health Belief Model-Based Deep Learning Classifiers for COVID-19 Social Media Content to Examine Public Perceptions of Physical Distancing: Model Development and Case Study. *JMIR Public Health Surveill*. 2020;6(3):e20493.
20. Sheppard J, Thomas CB. Community pharmacists and communication in the time of COVID-19: applying the health belief model. *Research in Social and Administrative Pharmacy*. 2020.
21. Banda JM, Tekumalla R, Wang G, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research--an international collaboration. *arXiv preprint arXiv:200403688*. 2020.
22. Tekumalla R, Banda JM. Social Media Mining Toolkit (SMMT). *Genomics Inform*. 2020;18(2):e16.
23. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*. 2012;22(3):276-282.
24. Bird S, Klein E, Loper E. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc."; 2009.
25. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011;12:2825-2830.
26. Bailey NT. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.; 1975.
27. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. *Signal Processing*. 2020;167:107299.
28. Moriarty LF. Public health responses to COVID-19 outbreaks on cruise ships—worldwide, February–March 2020. *MMWR Morbidity and mortality weekly report*. 2020;69.
29. Green A. Li wenliang. *The Lancet*. 2020;395(10225):682.
30. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *The Lancet*. 2020.
31. Abdi M. Coronavirus disease 2019 (COVID-19) outbreak in Iran: Actions and problems. *Infection Control & Hospital Epidemiology*. 2020;41(6):754-755.
32. Berkeley Lovelas Jr., Kevin Breuninger. Trump says he takes hydroxychloroquine to prevent coronavirus infection even though it's an unproven treatment. https://www.cnbc.com/2020/05/18/trump-says-he-takes-hydroxychloroquine-to-prevent-coronavirus-infection.html. Published 2020. Updated May 18. Accessed Aug 27, 2020.
33. Wikipedia. George Floyd protests. https://en.wikipedia.org/wiki/George_Floyd_protests. Published 2020. Updated Aug 27. Accessed Aug 27, 2020.
34. Lewandowsky S, Jetter M, Ecker UKH. Using the president's tweets to understand political diversion in the age of social media. *Nature Communications*. 2020;11(1):5764.
35. Pobiruchin M, Zowalla R, Wiesner M. Temporal and Location Variations, and Link Categories for the Dissemination of COVID-19–Related Information on Twitter During the SARS-CoV-2 Outbreak in Europe: Infoveillance Study. *J Med Internet Res*. 2020;22(8):e19629.

36. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:190602243.* 2019.

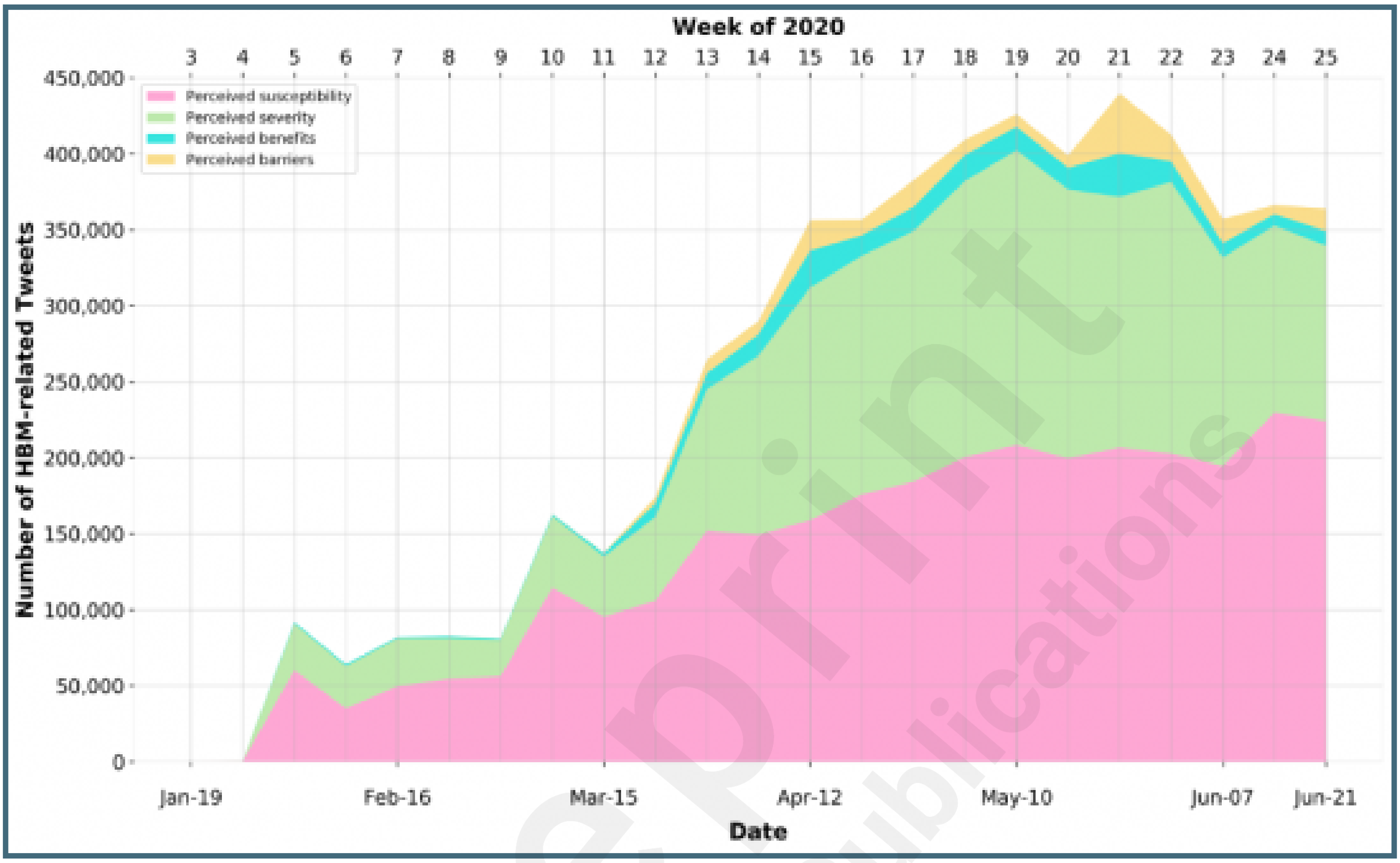
Supplementary Files

Figures

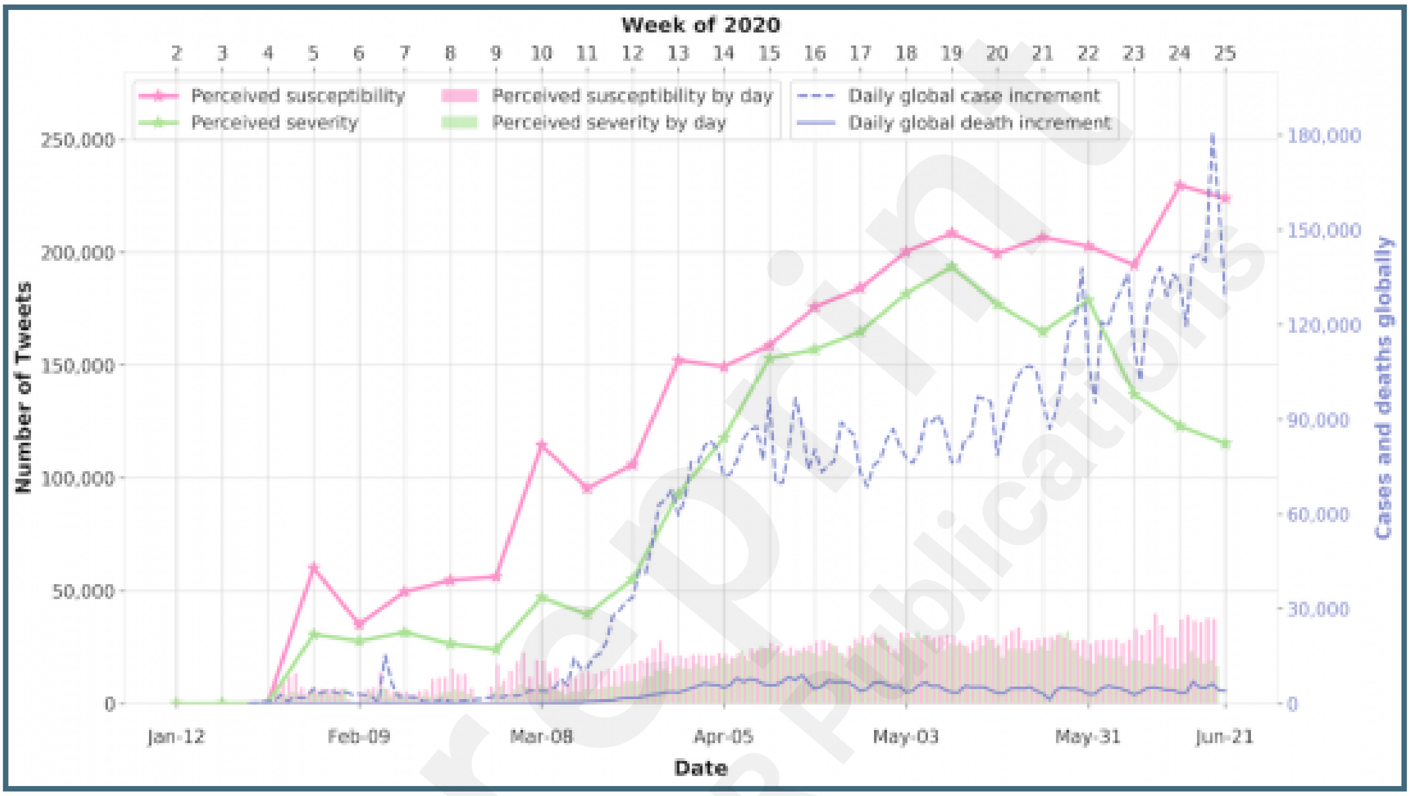
The workflow of data selection, data annotation, and machine learning classifier training and implementation.



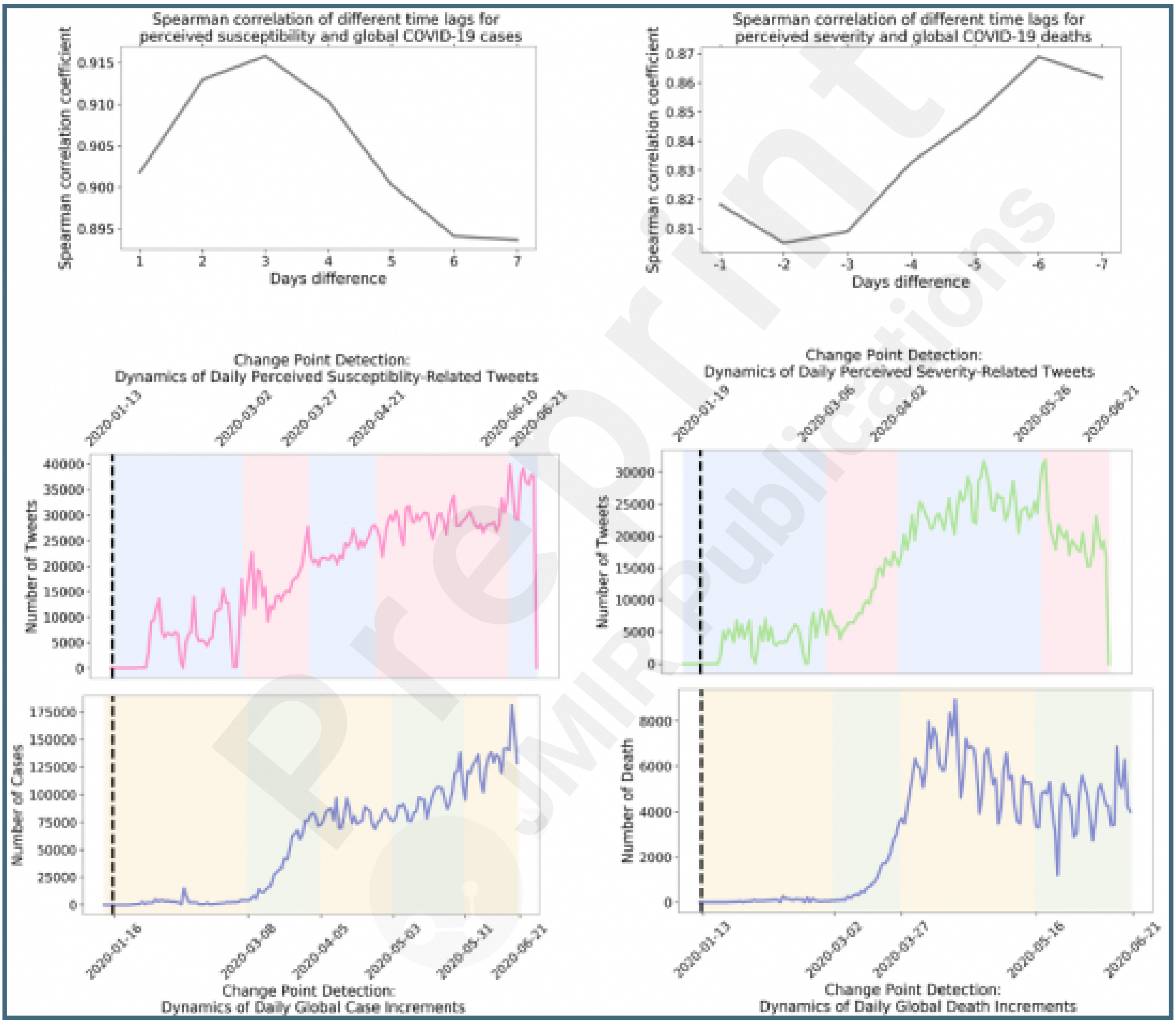
Stacked area chart for the four core constructs of health belief model (HBM) from January 19, 2020 to June 21, 2020. The upper x-axis shows the week number while the lower x-axis shows the end date of the corresponding week.



Dynamics of perceived susceptibility and severity with case and death trends. The upper x-axis displays the week number from the start of 2020, while the dates on the lower x-axis indicate the date of the end of the corresponding week. The y-axis on the left shows the number of tweets for the "star" marked pink and green lines, as well as the pink and green bars. The pink and green lines with "star" marks reflect the weekly cumulative number of tweets for perceived susceptibility and severity, while the pink and green bars on the x-axis indicate the daily number of tweets related to perceived susceptibility and severity. The blue y-axis on the right is for the global cases and deaths daily increments. The solid blue line reflects the daily increment of deaths globally, while the blue dashed line reflects the daily increment of cases globally. The global cases and deaths dynamics are available since January 22.



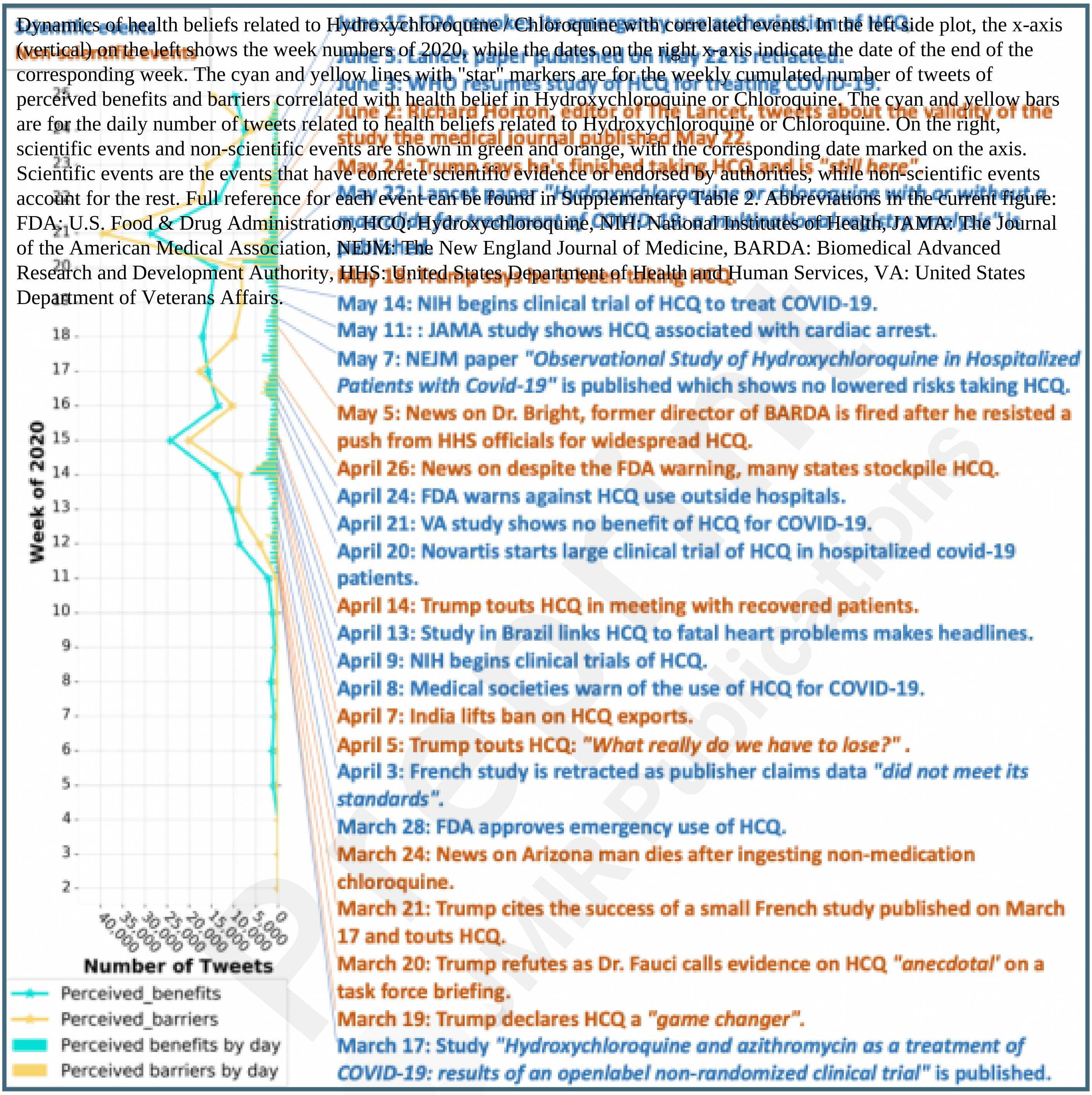
Correlation between perceived susceptibility-related tweets and COVID-19 cases dynamics. The top graphs display Spearman's correlation coefficient for the number of perceived susceptibility-related tweets and global case dynamics (left), perceived severity-related tweets, and global deaths dynamics (right). The x-axes show the time difference while the y-axes show the Spearman's correlation coefficient calculated under the given time lag. Middle graphs display the trend of perceived susceptibility (pink) trend and perceived severity trend (green). Bottom graphs are the trend of global cases (left) dynamics and deaths (right) dynamics. Each pair are staggered according to the time differences that achieve the highest correlation in the top graphs (3 days and -6 days, respectively). The pink/blue and yellow/green shades depict the change points detected from the change point analysis.



Dynamics of health beliefs related to COVID-19 and the trend of cases/deaths fluctuation in the US with lockdown status. The upper x- displays the week numbers of 2020, while the dates on the lower x-axis indicate the date of the end of the corresponding week. The left y-axis is for the number of weekly perceived susceptibility (pink) and severity (green) tweets in the US, respectively. The y-axis on the right is for the daily cases and deaths increment in the US, the solid purple line indicates daily deaths increment, and the dashed purple line indicates the daily cases increment. The lower half of the figure shows the official lockdown circumstances in each state of the US. The red font indicates a lockdown decision announced by a Republican governor, while the blue font indicates a Democratic governor. The lines corresponding to each state represents the start and end date of the official lockdowns. South Dakota (Republican), North Dakota (Republican), Iowa (Republican), Nebraska (Republican), and Arkansas (Republican) did not announce official lockdown and are not shown in this figure.

Top 10 topics of each week. Top phrases for each week are organized horizontally in each row. The blue shade in each cell indicates the term frequency – inversed document frequency (tf-idf) score of the phrase, the higher the tf-idf score, the darker the shade. The phrases likely associated with news are highlighted in purple.

Week	Data interval	1	2	3	4	5	6	7	8	9	10
2	Jan-06~Jan-12	advisory chinese	experts search	low experts	via tldr	viral nucleic	difficult confirm	search answers	notify health	weeks returning	outbreak ask
3	Jan-13~Jan-19	closed known	market infection	turn lethal	lethal likely	spread seafood	coronavirus rpts	ncov japan	report laboratory	major transport	dies novel
4	Jan-20~Jan-26	spreads multiple	france confirms	confirmed chicago	texas student	publishes early	warns grave	alarming consequences	consequenc es	points alarming	exposed lancet
5	Jan-27~Feb-02	outside china	philippines	spreads multiple	germany	confirmed uk	spreads regions	russia	outbreak spreads	rises outbreak	jumps outbreak
6	Feb-03~Feb-09	cruise	ship	cruise ship	li	wenliang	li wenliang	jumps outbreak	chinese rage	whistleblo wer	death coronavirus
7	Feb-10~Feb-16	ship	cruise	cruise ship	covid	reports new	spike	trump	seemed leveling	outside china	princess
8	Feb-17~Feb-23	iran	italy	covid	ship	cruise	cruise ship	outside china	spike	reports new	diamond princess
9	Feb-24~Mar-01	italy	iran	covid	trump	outside china	many cases	china real	coronavirus many	countries affected	updates number
10	Mar-02~Mar-08	covid	italy	iran	trump	york	new york	covid cases	cases covid	cruise	declares
11	Mar-09~Mar-15	covid	italy	trump	covid cases	cases covid	iran	coronavirus covid	spain	spread covid	community
12	Mar-16~Mar-22	covid	italy	trump	covid cases	chloroquine	cases covid	covid patients	spread covid	coronavirus pandemic	old
13	Mar-23~Mar-29	covid	trump	covid cases	italy	covid patients	chloroquine	cases covid	york	new york	hydroxychl oroquine
14	Mar-30~Apr-05	covid	trump	covid cases	covid patients	hydroxychlo roquine	italy	york	new york	spain	cases covid
15	Apr-06~Apr-12	covid	trump	hydroxychl oroquine	covid cases	covid patients	covid deaths	york	new york	italy	hours
16	Apr-13~Apr-19	covid	trump	covid cases	covid patients	covid deaths	hydroxychlo roquine	cases covid	homes	york	new york
17	Apr-20~Apr-26	covid	trump	covid cases	covid patients	hydroxychlo roquine	covid deaths	cases covid	york	new york	nursing
18	Apr-27~May-03	covid	covid cases	trump	covid patients	covid deaths	remdesivir	cases covid	homes	nursing	new deaths
19	May-04~May-10	covid	trump	covid cases	covid patients	covid deaths	highest	nursing	homes	cases covid	new deaths
20	May-11~May-17	covid	covid cases	trump	covid patients	nursing	covid deaths	homes	nursing homes	hydroxychlo roquine	cases covid
21	May-18~May-24	covid	trump	hydroxychlo roquine	covid cases	covid patients	nursing	homes	covid deaths	nursing homes	cases covid
22	May-25~May-31	covid	covid cases	trump	covid patients	covid deaths	hydroxychlo roquine	nursing	homes	nursing homes	highest
23	Jun-01~Jun-07	covid	covid cases	covid patients	trump	hydroxychlo roquine	covid deaths	new covid	george floyd	hours	spike
24	Jun-08~Jun-14	covid	covid cases	trump	covid patients	spike	active cases	highest	new covid	cases covid	covid deaths
25	Jun-15~Jun-21	covid	covid cases	trump	dexametha sone	florida	covid patients	spike	record	new covid	cases covid



Dynamics of health beliefs related to Hydroxychloroquine / Chloroquine with correlated events. In the left side plot, the x-axis (vertical) on the left shows the week numbers of 2020, while the dates on the right x-axis indicate the date of the end of the corresponding week. The cyan and yellow lines with "star" markers are for the weekly cumulated number of tweets of perceived benefits and barriers correlated with health belief in Hydroxychloroquine or Chloroquine. The cyan and yellow bars are for the daily number of tweets related to health beliefs related to Hydroxychloroquine or Chloroquine. On the right, scientific events and non-scientific events are shown in green and orange, with the corresponding date marked on the axis. Scientific events are the events that have concrete scientific evidence or endorsed by authorities, while non-scientific events account for the rest. Full reference for each event can be found in Supplementary Table 2. Abbreviations in the current figure: FDA: U.S. Food & Drug Administration, HCQ: Hydroxychloroquine, NIH: National Institutes of Health, JAMA: The Journal of the American Medical Association, NEJM: The New England Journal of Medicine, BARDA: Biomedical Advanced Research and Development Authority, HHS: United States Department of Health and Human Services, VA: United States Department of Veterans Affairs.

Multimedia Appendixes

Supplementary information.
URL: https://asset.jmir.pub/assets/51277e0bb6e178b192b0a2536b54e721.docx

Tracked change version of the updated manuscript.
URL: https://asset.jmir.pub/assets/329ccaf7d3d9ba2675d3871471976737.docx

TOC/Feature image for homepages

Untitled.