Contents

Original Papers

Desensitization to Fear-Inducing COVID-19 Health News on Twitter: Observational Study (e26876)
Hannah Stevens, Yoo Oh, Laramie Taylor. ................................................................. 2

Monitoring Depression Trends on Twitter During the COVID-19 Pandemic: Observational Study (e26769)
Yipeng Zhang, Hanjia Lyu, Yubao Liu, Xiyang Zhang, Yu Wang, Jiebo Luo. ......................................................... 14

The Role of the Canadian Media During the Initial Response to the COVID-19 Pandemic: A Topic Modelling Approach Using Canadian Broadcasting Corporation News Articles (e25242)
Janhavi Patel, Harsheev Desai, Ali Okhovat. ................................................................. 32

Infodemic Signal Detection During the COVID-19 Pandemic: Development of a Methodology for Identifying Potential Information Voids in Online Conversations (e30971)
Tina Purnat, Paolo Vacca, Christine Czerniak, Sarah Ball, Stefano Burzo, Tim Zecchin, Amy Wright, Supriya Bezbaruah, Faizza Tanggol, Éve Dubé, Fabienne Labbé, Maude Dionne, Jaya Lamichhane, Avichal Mahajan, Sylvie Briand, Tim Nguyen. ................................................................. 44

Examining the Public’s Most Frequently Asked Questions Regarding COVID-19 Vaccines Using Search Engine Analytics in the United States: Observational Study (e28740)
Nicholas Sajjadi, Samuel Shepard, Ryan Ottwell, Kelly Murray, Justin Chronister, Micah Hartwell, Matt Vassar. ................................................................. 59

Corrigenda and Addenda

Metadata Correction: “Desensitization to Fear-Inducing COVID-19 Health News on Twitter: Observational Study” (e32231)
Hannah Stevens, Yoo Oh, Laramie Taylor. ................................................................. 43
Desensitization to Fear-Inducing COVID-19 Health News on Twitter: Observational Study

Hannah R Stevens¹, BA; Yoo Jung Oh¹, MA; Laramie D Taylor¹, PhD
Department of Communication, University of California, Davis, Davis, CA, United States

Corresponding Author:
Hannah R Stevens, BA
Department of Communication
University of California, Davis
1 Shields Ave
Davis, CA, 95616
United States
Phone: 1 530 752 1011
Email: hrstevens@ucdavis.edu

Related Article:
This is a corrected version. See correction statement: https://infodemiology.jmir.org/2021/1/e32231

Abstract

Background: As of May 9, 2021, the United States had 32.7 million confirmed cases of COVID-19 (20.7% of confirmed cases worldwide) and 580,000 deaths (17.7% of deaths worldwide). Early on in the pandemic, widespread social, financial, and mental insecurities led to extreme and irrational coping behaviors, such as panic buying. However, despite the consistent spread of COVID-19 transmission, the public began to violate public safety measures as the pandemic got worse.

Objective: In this work, we examine the effect of fear-inducing news articles on people’s expression of anxiety on Twitter. Additionally, we investigate desensitization to fear-inducing health news over time, despite the steadily rising COVID-19 death toll.

Methods: This study examined the anxiety levels in news articles (n=1465) and corresponding user tweets containing “COVID,” “COVID-19,” “pandemic,” and “coronavirus” over 11 months, then correlated that information with the death toll of COVID-19 in the United States.

Results: Overall, tweets that shared links to anxious articles were more likely to be anxious (odds ratio [OR] 2.65, 95% CI 1.58-4.43, P<.001). These odds decreased (OR 0.41, 95% CI 0.2-0.83, P=.01) when the death toll reached the third quartile and fourth quartile (OR 0.42, 95% CI 0.21-0.85, P=.01). However, user tweet anxiety rose rapidly with articles when the death toll was low and then decreased in the third quartile of deaths (OR 0.61, 95% CI 0.37-1.01, P=.06). As predicted, in addition to the increasing death toll being matched by a lower level of article anxiety, the extent to which article anxiety elicited user tweet anxiety decreased when the death count reached the second quartile.

Conclusions: The level of anxiety in users’ tweets increased sharply in response to article anxiety early on in the COVID-19 pandemic, but as the casualty count climbed, news articles seemingly lost their ability to elicit anxiety among readers. Desensitization offers an explanation for why the increased threat is not eliciting widespread behavioral compliance with guidance from public health officials. This work investigated how individuals’ emotional reactions to news of the COVID-19 pandemic manifest as the death toll increases. Findings suggest individuals became desensitized to the increased COVID-19 threat and their emotional responses were blunted over time.

(JMIR Infodemiology 2021;1(1):e26876) doi:10.2196/26876

KEYWORDS
desensitization; death toll; pandemic; fear-inducing; fear; health news; anxiety; COVID-19; mass media; public health; behavior change; coronavirus
Introduction

Background

The COVID-19 outbreak has spread worldwide, affecting most countries. Since the outbreak of COVID-19, the number of confirmed cases and the death toll have steadily risen. According to Johns Hopkins University, as of May 9, 2021, more than 157.9 million cases of COVID-19 and 3.2 million deaths have been reported worldwide [1]. Among the countries affected by COVID-19, the United States has had 32.7 million cases (23.5% of confirmed cases worldwide) and 580,000 deaths (17.7% of deaths worldwide). The overabundance of information, misinformation, and disinformation surrounding COVID-19 on social media in the United States has fueled a COVID-19 infodemic, which has jeopardized public health policy aimed at mitigating the pandemic [2], raising questions about the cognitive processes underlying public responses to COVID-19 health information.

Extreme safety precautions (eg, statewide lockdowns, travel bans) have impacted individuals’ physical and mental health in the United States. People experienced intense psychological frustration and anxiety regarding the virus and strict safety measures (eg, stay-at-home measures), especially during the early stages of the COVID-19 pandemic [3-5]. Social, financial, and mental insecurities have even led to extreme and irrational coping behaviors, such as panic buying from January to March 2020 [5]. However, throughout the pandemic, the public became desensitized to reports of COVID-19’s health threat, and the rising number of confirmed cases and death toll began to lose impact [6,7]. As a result, segments of the public began violating public safety measures as the pandemic progressed, despite the consistent spread of COVID-19 [8-10].

From such observations, two key considerations arise. First, fear-eliciting health messages have a significant effect on eliciting motivation to take action to control the threat. However, repeated exposure to these messages over long periods results in desensitization to those stimuli [6,18]. Desensitization refers to the process by which cognitive, emotional, and physiological responses to a stimulus are reduced or eliminated over protracted or repeated exposure [19]. It can play an important adaptive role in allowing individuals to function in difficult circumstances that might otherwise result in overwhelming and persistent anxiety or fear. For example, one analysis of Twitter messages from a region of Mexico with then-rising violence found the expressions of negative emotions declined [20]. Although increasing anxiety and fear might prompt security-seeking behavior, these emotions may also be paralyzing; some measure of desensitization can facilitate continuing with necessary everyday tasks.

Numerous studies have demonstrated desensitization to media content. Research has often focused on fictional depictions of violence [21,22]; however, desensitization has also been demonstrated in response to repeated exposure to violent news stories [23], hate speech [24], and sexually explicit internet content [25], although this last finding has mixed support [26]. Researchers studying social media data have explored the possibility that news messages can result in desensitization. Li and colleagues [27] analyzed a large sample of Twitter data, examining posts linked to guns and shootings for emotional language. They observed that across 3 years of mass shootings and school shootings in the United States, the frequency of negative emotional words used in shooting-related tweets declined; they argued that this reflected desensitization to gun violence.

Effect of Fear-Inducing Messages on Public Anxiety

The current pandemic has fueled rapidly evolving news cycles and shaped public sentiment [11,12]. Public health experts’ recommendations to mitigate the COVID-19 threat, including widespread business shutdowns and physical distancing guidelines, have proven psychologically and emotionally taxing [13], inducing intense psychological frustration and anxiety among the public [3-5]. Previous literature suggests that fear-inducing messages influence emotions and behaviors when individuals perceive the message to be relevant (ie, they feel susceptible to the threat) and serious (ie, the threat is severe). That is, the heightened threat induces fear and anxiety that, in turn, motivate people to take action [14-17].

In the context of the COVID-19 pandemic, the efficacy of fear-inducing messages on behavioral compliance with public health officials is consequential. Reports of increased COVID-19 transmission and the rising death toll may elicit anxiety about the virus, consequently motivating behaviors intended to manage the problem. For instance, a national survey examining the mental health consequences of COVID-19 fear among US adults during March 2020 found that respondents generally expressed moderate to high COVID-19 fear and anxiety (7 on a scale of 10), and increased anxiety was most prevalent in areas with the highest reported COVID-19 cases [3]. Subsequently, the fear and anxiety induced by COVID-19–related threats can lead people to seek more health-related protective strategies. For example, one study found that as the threat of COVID-19 increased, people expressed more fear-related emotions and they were subsequently increasingly motivated to search for preventative behaviors and information online [5].

Desensitization to Fear-Inducing Messages

Although fear-based health messages have been shown to motivate changes in behavior, repeated exposure to even highly arousing stimuli—such as news of the rising death toll from COVID-19—may eventually result in desensitization to those stimuli [6,18]. Desensitization refers to the process by which cognitive, emotional, and physiological responses to a stimulus are reduced or eliminated over protracted or repeated exposure [19]. It can play an important adaptive role in allowing individuals to function in difficult circumstances that might otherwise result in overwhelming and persistent anxiety or fear. For example, one analysis of Twitter messages from a region of Mexico with then-rising violence found the expressions of negative emotions declined [20]. Although increasing anxiety and fear might prompt security-seeking behavior, these emotions may also be paralyzing; some measure of desensitization can facilitate continuing with necessary everyday tasks.
threatening COVID-19 information and experienced diminished anxiety over time, even in the face of an increasing threat.

**Rationale and Aims**

The public relies heavily on news disseminated through social media for information about the spread of the virus [29]. Twitter, in particular, is a popular outlet for sharing news [30] and has become a forum for individuals to communicate their feelings about COVID-19 [11]. Social media text analysis has emerged as a particularly effective way to assess sentiment dynamics surrounding public health crises; consider, for example, the Zika outbreak [31]. This study uses social media text analysis to examine the anxiety levels in news articles and related tweets over 11 months, then considers those levels in the context of deaths from COVID-19 on the day the post was shared [32].

The general hypothesis guiding this research is that audiences will have become desensitized to COVID-19 deaths over the course of the pandemic, decreasing the level of anxiety elicited by fearful COVID-19 health information reported in the news. To the best of our knowledge, ours is the first study to investigate whether, as the objective threat and harm of COVID-19 has increased, individuals have become desensitized to news reports of cautionary COVID-19 health information.

**Methods**

**Overview**

This study examined how anxiety levels in news articles predicted users’ tweet anxiety levels over 11 months, then correlated that information with the total death toll of COVID-19 in the United States as reported to the Centers for Disease Control and Prevention (CDC) on the day the post was shared [32]. Employing semantic analysis procedures to analyze anxiety in the full news articles and their corresponding user tweets allowed us to examine how fear elicited by COVID-19 health news manifests as individuals become desensitized to news of COVID-19–related deaths.

**Data Collection**

The sample comprises content shared to Twitter, a popular social media platform used for sharing news [30]. The text of 1465 news articles and corresponding posts by users were collected from tweets containing the terms “COVID-19,” “COVID,” “pandemic,” and “coronavirus” from January 1 to December 2, 2020. For an overview of the data collection process, see Figure 1.

The Python programming language was used to extract posts sharing news reports of COVID-19 health information. We collected a quota sample of 32,000 US tweets containing one of four key terms (ie, COVID, COVID-19, coronavirus, pandemic) each week from January 1 to December 2, 2020. The GetOldTweets3 Python3 library was used to scrape tweets for the months of January-July 2020 [33]. Twitter’s application programming interface (version 2) was used to collect tweets from August-December 2020 [34].

Human coders then filtered through the sample of 1,410,901 tweets to randomly extract a quota of 8 original tweets per key term from each week sharing a news report about COVID-19. Data collection resulted in thousands of tweets containing links per week. To facilitate the representativeness of the news articles, 32 tweets were drawn from each week from a shuffled list of tweets containing hyperlinks. Since we aimed to assess users’ reactions to the text of the article they read, without the confounding textual framing of other peoples’ commentary about an article, retweets were excluded from the analysis. If a quota of 32 tweets each week (8 per key term) was not met, additional tweets were sampled for that week. Notably, the disease and pandemic were not commonly referred to as COVID-19 in early January; accordingly, three weeks did not have 8 tweets with the terms “COVID” and “COVID-19” per week.

The news articles were collected from links shared by Twitter users in general, regardless of who posted the tweet. We only included users sharing links to news articles regarding COVID-19 in the United States; all other content was excluded (eg, news about the rock band Pandemic Fever). If all posts for that week were excluded, another sample from that week was drawn. If a tweet linked to a news article that had been taken down, a replacement post was sampled from the same week. We then extracted the text from the news articles and their corresponding tweets. The final sample was comprised of n=1465 news-sharing tweets.
Linguistic Inquiry and Word Count Sentiment Analysis

Once the final sample was collected (n=1465), we analyzed articles and tweets using the Linguistic Inquiry and Word Count (LIWC) program [35]. The body text of the news articles was analyzed to measure article anxiety, while the tweet text was analyzed to measure tweet anxiety. LIWC is a natural language processing text analysis program that classifies texts by counting the percentage of words in a given text that fall into prespecified categories, such as a linguistic category (eg, prepositions) or psychological processes (eg, anxiety, sadness). In this study, we focused on the percentage of LIWC anxiety lexicon words in news articles and tweets because this psychological process is germane to the efficacy of fear-based news messaging [14,36]. LIWC calculates the percentage of anxiety words relative to all words contained in a text to account for long versus short text classification. For example, we might discover that 15/745 (2.04%) words in a given article were anxiety lexicon words. The LIWC output would then assign that particular article an anxiety score of 2.04 (see Figure 2 for an example).

Figure 1. Flowchart of data collection process.
Figure 2. Sample text from a COVID-19 news article shared to Twitter [37]. The words highlighted in red are LIWC anxiety words. Since this article contains 15 anxiety words out of 745 words total (2.4%), this article is assigned a LIWC anxiety score of 2.4. LIWC: Linguistic Inquiry and Word Count.

How Global Corruption Threatens the U.S. Pandemic Response

Abigail Bellows
Abigail Bellows is a nonresident scholar in the Democracy, Conflict, and Governance Program at the Carnegie Endowment for International Peace
@ABIGAIL_BELLOWS

As Ebola began to rage across the Democratic Republic of the Congo in 2018, the disease had a powerful accomplice: corruption. The country’s health minister and his financial adviser embezzled $400,000 in relief funds—crimes for which they were recently sentenced to five years of forced labor. Yet the systemic vulnerabilities that enable this type of fraud persist around the world. How is the U.S. government assisting partners in Africa, Latin America, and South Asia in addressing corruption vulnerabilities before they are hit by full-blown outbreaks of the new coronavirus?

IGNORING CORRUPTION IS RISKY

Corruption is like gasoline poured on the flames of a pandemic. Healthcare systems already debilitated by graft will struggle to address the most basic of needs during the crisis. Citizens who can’t afford to pay bribes may be locked out of access to testing and treatment, a problem that would accelerate the virus’s spread. Those who can bribe their way out of quarantines will probably do so, as has been reported already in Cameroon and Uganda. And government attempts to convey public health messages are likely to fall flat in places where decades of corruption have deeply undermined trust in the state.

At the elite level, the pandemic is setting off a flurry of public procurement spending, which faces serious risks for diversion, especially since traditional watchdog groups are also scrambling to adapt. Foreign assistance pouring in from the United States and other countries is also vulnerable to leakage. In normal times, various sources estimate that more than 10 percent of global healthcare spending is siphoned off by corruption, amounting to losses of more than $500 billion annually—and these risks are only heightened during a disaster. Meanwhile, oligarchs may be using the proceeds of corruption to buy up ventilators and arrange for private healthcare, as seen in Russia, a practice that drains resources from the public health system.

The pandemic’s further spread around the globe, fueled by corruption, could cause serious harm to U.S. interests and foreign policy objectives. Public anger at government malfeasance could topple regimes, embolden antistablishment populists, and provide openings for terrorist recruitment. Long-standing allies may turn away from the United States and toward China, desperately applying authoritarian measures in hopes of containing the virus. If corruption becomes more entrenched overseas, U.S. businesses will struggle to compete.

CHARTING A DIFFERENT COURSE

The good news is that the U.S. government can take action to avoid this dark prognosis. The biggest near-term step would be inclusion of the Countering Russian and Other Overseas Kleptocracy (CROOK) Act (H.R. 3843/S. 3026) in Congress’s planned fourth coronavirus-related spending package. The bill would form an Anti-Corruption Action Fund to surge support to countries eager to take rapid action against corruption, as the current crisis demands. The proposal, which is budget-neutral and enjoys bipartisan support, has already passed the House Foreign Affairs Committee and has been introduced in the Senate.

The next step would be for relevant committee leaders and the bill’s bipartisan sponsors—Representatives Bill Keating and Brian Fitzpatrick alongside Senators Roger Wicker and Ben Cardin—to fold the measure into upcoming legislation. This would fill a glaring gap in congressional action to date on the coronavirus and signal that U.S. decisionmakers recognize the links between international corruption, public health, and U.S. national security.

Alongside congressional leadership, the State Department and the United States Agency for International Development (USAID) could take important steps to address the corruption-coronavirus nexus. Global health assistance should include strong anticorruption safeguards and support for emergency procurement mechanisms that are rigorous and transparent—in both U.S. and multilateral assistance. Diplomats should reinforce the need to maintain anticorruption law enforcement to deter crime during the pandemic. The United States can also celebrate local officials who act with integrity and urge civil society, media outlets, and whistleblowers to keep playing their vital roles in spite of rising repression.

The virus has yet to become a full-scale disaster in the most corruption-prone parts of the world—but time is running short. If the United States seeks to avoid a replay of the Ebola epidemic—on a far graver scale—it must act now to address global corruption risks.

Statistical Analysis

We paired the final sample with the CDC’s aggregate death toll on the day the tweet was posted. Contextualizing the articles and tweets allowed us to examine how fear elicited by COVID-19 news manifests as individuals become desensitized to news of COVID-19–related deaths.

The outcome of interest was tweet anxiety. Note that the distribution of count data outcome variables (in our case, LIWC tweet anxiety) often contains excess zeros; this result is known as zero inflation. The positive values are skewed, and a considerable “clumping at zero” is trailed by a bump representing positive values [38]. In our specific distribution, the “clumping at zero” represents texts containing zero anxiety...
lexicon terms. Generalized linear models are not appropriate for zero-inflation data. As all observed zeros are unambiguous, they are best analyzed separately from the nonzeros.

Two distinct distributions generally characterize zero-inflation data; thus, a zero-inflated model, which separates the zero and nonzero counts, is appropriate [39,40]. In zero-inflated models, the distribution of positive count values depends on the probability of exceeding the hurdle and reaching the distribution of positive values. In other words, it considers the odds of having any anxiety in a tweet versus none at all. For tweets that clear the hurdle, it then considers how much anxiety will be in a tweet on a continuous distribution.

We employed a zero-inflated model using a gamma distribution with a log link to examine any association between article anxiety and death toll, along with their interaction with subsequent tweet anxiety for all values of tweet anxiety greater than zero. We paired that with a model that used a binomial distribution with a logit link to determine zero anxiety versus nonzero anxiety in tweets. We recoded the death toll into categories reflecting the death count at the second quartile, the third quartile, and the fourth quartile relative to the first quartile of the total death count (see Figure 3 for a breakdown). This was necessitated by the skewed and logarithmic character of the distribution. These values were then used in place of the continuous variable to model the interaction. We used R statistical software for data analysis (version 3.6.2; The R Foundation for Statistical Computing).

Figure 3. Distribution of death toll quartiles over time.

Ethics Statement
This study only used information available in the public domain. No personally identifiable information was included in this study. Ethical review and approval was not required for this study because the institutional review board recognizes that the analysis of publicly available data does not constitute human subjects research.

Results
Results suggest that as the death toll increased over time, the baseline level of anxiety lexicon words in articles decreased; this was evidenced by our finding that when the pandemic’s severity and threat increased, individuals shared less news coverage containing COVID-19 anxiety words (eg, “risk,” “worried,” “threatens”). When assessing the odds of a tweet having no anxiety versus anxiety, we found that the baseline odds of not having anxiety in a tweet were 0.11; the odds of having anxiety in a tweet increased (odds ratio [OR] 2.65, 95% CI 1.58-4.43, P<.001) with each unit increase in anxiety within an article. The odds of tweet anxiety decreased as paired with CDC total deaths in the third quartile (OR 0.41, 95% CI 0.21-0.85, P=.01) and fourth quartile (OR 0.42, 95% CI 0.21-0.85, P=.01), respectively (see Table 1 and Figure 4).
Table 1. The odds of a tweet containing anxiety language versus no anxiety language, as determined using a zero-inflated model with categorical death.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.11 (0.07-0.16)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Anxiety in article</td>
<td>2.65 (1.58-4.43)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Second quartile (22,253-133,665 deaths)</td>
<td>0.76 (0.41-1.41)</td>
<td>.39</td>
</tr>
<tr>
<td>Third quartile (133,666-193,321 deaths)</td>
<td>0.41 (0.2-0.83)</td>
<td>.01</td>
</tr>
<tr>
<td>Fourth quartile (≥193,322 deaths)</td>
<td>0.42 (0.21-0.85)</td>
<td>.02</td>
</tr>
<tr>
<td>Interaction anxiety in article by second quartile deaths (22,253-133,665 deaths)</td>
<td>0.71 (0.34-1.48)</td>
<td>.36</td>
</tr>
<tr>
<td>Interaction anxiety in article by third quartile deaths (133,666-193,321 deaths)</td>
<td>1.32 (0.54-3.24)</td>
<td>.55</td>
</tr>
<tr>
<td>Interaction anxiety in article by fourth quartile deaths (≥193,322 deaths)</td>
<td>1.9 (0.75-4.83)</td>
<td>.18</td>
</tr>
</tbody>
</table>

*This table reports the odds of no tweet anxiety versus tweet anxiety. Deaths were categorized based on the second, third, and fourth quartiles relative to the first quartile.*

Figure 4. Article anxiety predicting the odds of tweet anxiety versus no tweet anxiety at the first, second, third, and fourth quartiles of the COVID-19 death toll.

We then examined the actual estimated linguistic anxiety of tweets, looking only at all of the values in a continuous distribution, excluding those values with zero anxiety (i.e., the tweet did not contain any anxiety lexicon words). Although not statistically significant at $P<.05$, the results illuminate an emerging yet meaningful trend. The baseline level of anxiety in a tweet was 3.45. The tweet anxiety level trend increased (OR 1.25, 95% CI 0.99-1.59, $P=.068$) with each unit increase of article anxiety. Overall, tweets that shared links to more anxious articles expressed more anxious terms (e.g., “avoid,” “uncertain,” “paranoid”). Notably, the interaction between article anxiety and deaths was not found to be a significant predictor of tweet anxiety level. Tweet anxiety rose rapidly with articles when the death toll was low and then decreased in the third quartile of deaths (OR 0.61, 95% CI 0.37-1.01, $P=.06$). As predicted, in addition to the increasing death toll being matched by a lower level of article anxiety, the extent to which article anxiety elicited tweet anxiety decreased when the death count reached the second quartile (see Table 2 and Figure 5).
Table 2. Actual anxiety expressed in tweets, as predicted by article anxiety and COVID-19 death toll: gamma regression model with categorical death\(^a\).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (95% CI)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.45 (2.77-4.28)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Anxiety in article</td>
<td>1.25 (0.99-1.59)</td>
<td>.07</td>
</tr>
<tr>
<td>Second quartile (22,253-133,665 deaths)</td>
<td>1.53 (1.1-2.15)</td>
<td>.01</td>
</tr>
<tr>
<td>Third quartile (133,666-193,321 deaths)</td>
<td>1.47 (0.97-2.22)</td>
<td>.17</td>
</tr>
<tr>
<td>Fourth quartile ((\geq)193,322 deaths)</td>
<td>1.21 (0.83-1.75)</td>
<td>.32</td>
</tr>
<tr>
<td>Interaction anxiety in article by second quartile deaths (22,253-133,665 deaths)</td>
<td>0.78 (0.56-1.08)</td>
<td>.14</td>
</tr>
<tr>
<td>Interaction anxiety in article by third quartile deaths (133,666-193,321 deaths)</td>
<td>0.61 (0.37-1.01)</td>
<td>.06</td>
</tr>
<tr>
<td>Interaction anxiety in article by fourth quartile deaths ((\geq)193,322 deaths)</td>
<td>0.78 (0.5-1.22)</td>
<td>.28</td>
</tr>
</tbody>
</table>

\(^a\)Deaths were categorized based on the second, third, and fourth quartiles relative to the first quartile. This table reports the actual estimated anxiety in the tweet, looking only at all of the values in a continuous distribution, excluding those with zero anxiety.

Figure 5. Article anxiety predicting nonzero tweet anxiety at the first, second, third, and fourth quartiles of the COVID-19 death toll.

Discussion
Principal Findings
This study reports exploratory findings on the effects of fear-inducing news messages during a pandemic. Most importantly, we demonstrated a link between the anxiety expressed in news articles and the odds of anxiety being expressed by those who shared the articles to Twitter. This likely reflects the ability of pandemic-related news messages to elicit a measure of fear in their readers, consonant with public health goals. However, likely as a function of the rising COVID-19 threat over time (as indicated by LIWC news article anxiety) and a low perceived ability to prevent the rapid spread of the virus, anxiety did not increase in response to climbing death tolls over time. Instead, anxiety in tweets increased sharply in response to article anxiety early on in the pandemic, but as the death toll climbed, it flattened out, and news articles seemingly lost their ability to elicit anxiety among readers.

Such findings from this study provide several insights and directions for future research. Our findings reveal that responses to COVID-19 news as well as the rising death toll are increasingly bland. Growing desensitization in the face of threatening pandemic information impedes public health experts’ efforts to mitigate the COVID-19 crisis [41]. Therefore, future research should investigate how to “resensitize” the public and motivate them to take active roles in COVID-19–related responses (eg, wearing masks, washing hands, vaccination). Here, literature on behavioral theories may be helpful in implementing effective resensitization tactics. For instance, the transtheoretical model [42,43], which explains behavior change
through stages of change, suggests that to initiate and maintain health behaviors, it is important to have supportive relationships and motivate one another to share successes and experiences related to engaging in certain behaviors. In addition, it is suggested that reinforcement management—such as getting rewards from behavioral engagement—can be effective. In the context of COVID-19, health care providers can apply these tactics (ie, social support, reward) to motivate people to adhere to public health measures such as vaccination.

Second, since extant research shows that both statistics (eg, percentage of deaths) and cognitive dissonance can elicit desensitization [44,45], scholars should investigate the role of additional psychological processes in desensitization to the COVID-19 threat. Third, as self-disclosure varies by platform [46], more work is needed to explore how anxiety manifests on other platforms for discussing COVID-19 news. Finally, our findings suggest that health care practitioners should be prepared for public desensitization to future global pandemic scenarios. More specifically, it would be important to carefully monitor the public’s level of desensitization to health news and implement appropriate resensitization strategies based on different stages in the pandemic.

Limitations

Our findings illuminate desensitization to fear-inducing news messages during the pandemic; however, this study is not without limitations. By focusing on Twitter, we neglected to explore how anxiety manifests on other platforms for sharing news (eg, the comments section of digital news sites). As different platforms have different community norms [46], it is reasonable to expect manifestations of anxiety to vary by platform. Furthermore, Twitter users are younger, more democratic, and wealthier than the general population of Americans [47]. Acknowledging the biases associated with using computational social media data [48], our findings should be interpreted as representing a subset of the US population (ie, Twitter users), not all US residents. Second, among 1.4 million tweets collected, only a small number of tweets were sampled in this study. Therefore, our study may lack generalizability. Additionally, the LIWC computerized coding tool does not allow for the nuanced coding that could be achieved with human coders. Although we have attempted to minimize this potential bias using a well-validated sentiment analysis procedure, LIWC [35], this study is limited in its use of anxiety in text as a measure of user anxiety.

Conclusions

This work investigates how individuals’ emotional reactions to news of the COVID-19 pandemic manifest as the death toll increases. Individuals become desensitized to an increased health threat and their emotional responses are blunted over time. Our results suggest desensitized public health reactions to threatening COVID-19 news, which could affect the propensity of individuals to adopt recommended health behaviors. Public health agencies made recommendations to slow the pandemic’s spread, including physically distancing from others when appropriate, wearing masks, engaging in frequent handwashing, and disinfecting frequently touched surfaces. The consequences of ignoring these guidelines initially incited widespread fear and anxiety around contracting the virus or having family and friends contract it and fall ill. Social scientists have tried to inform interventions aimed at promoting compliance with public health experts [49]. The results of this study suggest the increased threat conveyed in COVID-19 news has, however, diminished public anxiety, despite an increase in COVID-19–related deaths. Desensitization offers one way to explain why the increased threat is not eliciting widespread compliance with guidance from public health officials. This work sheds light on both the effectiveness and shortcomings of fear-based health messages during the pandemic, as well as the utility of natural language processing to gain an understanding of public responses to emerging health crises.

Conflicts of Interest

None declared.

References


41. Fitzgerald M. Desensitization and the Coronavirus: How mass shootings have molded the American psyche. Medium. 2020. URL: https://medium.com/@mqfitzgeraldesensitization-and-the-coronavirus-how-mass-shootings-have-molded-the-american-psyche-be5217f52a6 [accessed 2021-02-04]


Abbreviations

CDC: Centers for Disease Control and Prevention
LIWC: Linguistic Inquiry and Word Count
OR: odds ratio
Monitoring Depression Trends on Twitter During the COVID-19 Pandemic: Observational Study

Yipeng Zhang1, BSc; Hanjia Lyu1*, MSc; Yubao Liu1*, MSc; Xiyang Zhang2, MA; Yu Wang1, PhD; Jiebo Luo1, PhD

1University of Rochester, Rochester, NY, United States
2University of Akron, Akron, OH, United States
*these authors contributed equally

Corresponding Author:
Jiebo Luo, PhD
University of Rochester
500 Joseph C Wilson Blvd
Rochester, NY
United States
Phone: 1 585 276 5784
Email: jluo@cs.rochester.edu

Abstract

Background: The COVID-19 pandemic has affected people’s daily lives and has caused economic loss worldwide. Anecdotal evidence suggests that the pandemic has increased depression levels among the population. However, systematic studies of depression detection and monitoring during the pandemic are lacking.

Objective: This study aims to develop a method to create a large-scale depression user data set in an automatic fashion so that the method is scalable and can be adapted to future events; verify the effectiveness of transformer-based deep learning language models in identifying depression users from their everyday language; examine psychological text features’ importance when used in depression classification; and, finally, use the model for monitoring the fluctuation of depression levels of different groups as the disease propagates.

Methods: To study this subject, we designed an effective regular expression-based search method and created the largest English Twitter depression data set containing 2575 distinct identified users with depression and their past tweets. To examine the effect of depression on people’s Twitter language, we trained three transformer-based depression classification models on the data set, evaluated their performance with progressively increased training sizes, and compared the model’s tweet chunk-level and user-level performances. Furthermore, inspired by psychological studies, we created a fusion classifier that combines deep learning model scores with psychological text features and users’ demographic information, and investigated these features’ relations to depression signals. Finally, we demonstrated our model’s capability of monitoring both group-level and population-level depression trends by presenting two of its applications during the COVID-19 pandemic.

Results: Our fusion model demonstrated an accuracy of 78.9% on a test set containing 446 people, half of which were identified as having depression. Conscientiousness, neuroticism, appearance of first person pronouns, talking about biological processes such as eat and sleep, talking about power, and exhibiting sadness were shown to be important features in depression classification. Further, when used for monitoring the depression trend, our model showed that depressive users, in general, responded to the pandemic later than the control group based on their tweets (n=500). It was also shown that three US states—New York, California, and Florida—shared a similar depression trend as the whole US population (n=9050). When compared to New York and California, people in Florida demonstrated a substantially lower level of depression.

Conclusions: This study proposes an efficient method that can be used to analyze the depression level of different groups of people on Twitter. We hope this study can raise awareness among researchers and the public of COVID-19’s impact on people’s mental health. The noninvasive monitoring system can also be readily adapted to other big events besides COVID-19 and can be useful during future outbreaks.

(2021;1(1):e26769) doi:10.2196/26769

KEYWORDS
mental health; depression; social media; Twitter; data mining; natural language processing; transformers; COVID-19
Introduction

Background

COVID-19 is an infectious disease that has been spreading rapidly worldwide since early 2020. It was first identified on December 31, 2019, and was officially declared as a pandemic by the World Health Organization on March 11, 2020 [1]. As of September 15, 2020, COVID-19 has infected 216 countries, areas, or territories with over 29 million confirmed cases and 930,000 confirmed deaths [1]. In response to the pandemic, over 190 countries have issued nationwide closures of educational facilities [2], and many governments have issued flight restrictions and stay-at-home-orders, affecting the everyday lives of people worldwide.

Mental disorders were affecting approximately 380 million people of all ages worldwide before COVID-19 [3]. Previous psychological studies have shown that mental disorders lead to many negative outcomes including suicide [4,5]. However, these studies face two challenges. First, it is known that individuals with mental disorders are sometimes unwilling or ashamed to seek help [6]. Second, it is oftentimes infeasible for psychological studies to obtain and track a large sample of diagnosed individuals and perform statistically significant numerical analysis.

Multiple studies have investigated the economic and social impacts of COVID-19 [7,8], and various studies have shown that COVID-19 has greatly impacted people’s mental health worldwide. These studies found that there are higher rates of depression, anxiety, posttraumatic stress disorder (PTSD), and stress symptoms reported during COVID-19 than before [9]. Females, young age groups, students, and low education groups are especially susceptible to depression during the pandemic [9]. The pandemic negatively affected individuals’ mental health because of the changes that it brought to life. For example, it has been shown that after nationwide lockdowns people experienced high levels of stress because of social isolation [10]; the fact that a large proportion of the population is not wearing masks also makes people experience high levels of anxiety and depression [11]. For individuals with mental disorders, their need is amplified; the study by Hao et al [12] suggests that, during the pandemic, psychiatric patients reported more moderate to severe anger and impulsivity as well as concerns about their physical health, as opposed to the healthy controls, and that ideal remote mental health services such as telepsychiatry consultation and home delivery of medications could not be fully established due to the sudden lockdown [12].

Given this pressing situation, we would like to quantify mental health conditions of the general population during the pandemic. Nevertheless, the data source selection is critical for overcoming the two challenges mentioned previously. In the past decade, people have been increasingly relying on social media platforms such as Facebook, Twitter, and Instagram to express their feelings. Social media can thus serve as a resourceful medium for mining information about the public’s mental health conditions [13-17]. The public have long been known to search online for information about diseases and medical issues [18]. COVID-19 is no exception. Indeed, using social media, public opinions on personal face mask use [19] and COVID-19 vaccine uptake [20,21] have been investigated. Existing research has also studied the predictive power of online medical consultation, online medical appointment, and online medical search in forecasting regional outbreaks and found online medical consultation to be the most predictive [22]. Furthermore, a recent longitudinal study on the mental health of the Chinese population during the pandemic has found that dissemination of health information via radio was associated with higher levels of anxiety and depression, and suggested television and the internet as alternatives [23]. Therefore, we believe social media platforms like Twitter offer a solution to the challenges, as they enable us to perform a large-scale quantitative study on mental disorders in a noninvasive fashion.

As shown in Figure 1, we used data from the ForSight by Crimson Hexagonplot [24] to plot the word frequencies of several mental disorders on Twitter, including “depression,” “PTSD,” “bipolar disorder,” and “autism,” from January 1 to May 4, 2020. Note that we excluded false-positive tweets that contained misleading phrases such as “economic depression” or “great depression.” We noticed a rapid growth of the word frequencies of these mental disorders starting from March 17, when the pandemic spread across most of the world. Past research has suggested that depression is more pervasive than other psychological disorders during the COVID-19 period [9]. Similarly, we found that the word “depression” occurs substantially more frequently on Twitter compared to the other three mental disorders. Accordingly, depression is likely to be triggered most frequently by COVID-19, and we focused on understanding COVID-19’s impact on depression in this study.
Prior Work

The potential of machine learning models for identifying Twitter users who have been diagnosed with depression was pioneered by De Choudhury et al [25], who analyzed how features obtained by Linguistic Inquiry and Word Count (LIWC) were related to depression signals on social media and how that can be used for user-level classification on a data set containing 171 depression users. The data was collected by designing surveys for volunteers through crowdsourcing. Following this work, Coppersmith et al [26] used LIWC, 1-gram language model, character 5-gram model, and user’s engagement on social media (user mention rate, tweet frequency, etc) to perform tweet-level classification on a data set containing 441 depression users.

The CLPsych 2015 Shared Task data set containing 447 diagnosed depression users [27] was published in 2015 and was favored by a wide range of studies [28-30]. The data was gathered by regular expression search in tweets in combination with manual annotation. Among these studies, the performance of traditional machine learning classification algorithms (decision trees, support vector machines [SVMs], naive Bayes, logistic regression) on 1-grams and 2-grams was investigated by Nadeem [30]; Jamil et al [28] used SVM on bag of words (BOW) and depression word count along with LIWC features and NRC sentiment features; Orabi et al [29] explored the performance of small deep neural network architectures—one-dimensional convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) with context-aware attention—and achieved the best performance (87% accuracy) on the task.

The CLPsych 2019 Shared Task [31] focused on evaluating Reddit users’ suicide risk based on their posts, for which Matero et al [32] applied a pretrained Bidirectional Encoder Representations from Transformers (BERT) [33] embedding to encode the data. Suicide risk assessment on Spanish tweets was also studied by Ramírez-Cifuentes et al [34]. We argue that our task is different since few detected depressive Twitter users express suicide intent, while all the positive suicidal users in the suicide risk data sets should be viewed as in late stages of depression [35,36]. There are also some studies that performed depression detection on Reddit users [37-39] with sample sizes of less than 1300 Reddit posts. By contrast, we used the transformer-based models in our study, which have been shown to achieve state-of-the-art results in a wide range of natural language processing problems [33,40,41].

In addition to these two challenge data sets, several studies attempted to gather their own data of various forms. Tsugawa et al [42] performed analysis of models using BOW, latent Dirichlet allocation (LDA) [43], and social media engagement features on a data set containing 81 Japanese-speaking depression Twitter users collected by crowdsourcing. Zhou et al [44] used ubiquitous multimodal sensors and performed in-depth analysis on users’ social media content, social network, webcam video, and user interaction on a sample of 5 depression users. Detecting depression from Spanish tweets using sentiment and emotion lexicons was used by Leis et al [45]. Zhang et al [46] performed observational analysis of the relationship between deteriorating depression and behavior changes when engaging with Google search and YouTube on 49 depressive college students. Shen et al [47] proposed a multimodal dictionary learning method that used topic, social media

Figure 1. Density of Twitter coverage regarding “depression,” “ptsd,” “bipolar disorder,” and “autism.” ptsd: posttraumatic stress disorder.
engagement, profile image, and emotional features to learn a latent feature dictionary that performed well on a data set of 1402 users with depression, the largest Twitter depression data set used to the best of our knowledge. Given the skyrocketing word density of “depression” in Figure 1, we show that a substantially larger depression data set can be quickly constructed from the COVID-19–related tweets within several months.

Goal of the Study

Although the time series plots of keyword frequencies in Figure 1 offer an intuitive reading of depression’s general trend in the population, they are apparently filled with noise and lack plausible explanation to be an accurate representation. To generalize beyond keywords, we would like to train machine learning–based models to identify depression on social media. Reddit automatically gathers posts of the same topic into “subreddits”; however, as pointed out by Pirina and Çöltekin [38], labeling posts completely according to subreddit names causes categories to be topically specific and cannot be generalized to regular social media text. Moreover, depression prediction models can potentially be used on the population level [48], but none of the work mentioned in the previous section applied their models to the general Twitter population on the fly.

Therefore, the main objectives of this study are to develop a method to create a large-scale depression user data set in an automatic fashion so that the method is scalable and can be adapted to future events; to verify the effectiveness of transformer-based deep learning language models in identifying depression users from their everyday language; to further improve the depression classification model using explainable psychological text features and to examine their importance in classification; and, finally, to use the model for monitoring the fluctuation of depression levels of different groups as the disease propagates.

Methods

Data Collection

First, we identified users with depression from 41.3 million COVID-19–related tweets posted by about 36.6 million users from March 23 to April 18, 2020. We collected the COVID-19–related tweets using the keywords “corona,” “covid19,” “covid19,” “coronavirus,” “#Corona,” “#Covid_19,” and “#coronavirus.” From these tweets, we looked for signals that can tell whether the user has depression from both the text and the user profile description.

Empirically, we observed that many Twitter users with depression described themselves as “depression fighters” in their descriptions. Some of them may also post relevant tweets to declare that they have been diagnosed with depression. Inspired by Coppersmith et al [26], we used regular expressions to find these authors by examining their tweets and descriptions. Building upon their method, we further extended our regular expression search based on some patterns we noticed on manually identified depression users, in pursuit of efficacy. In tweets, we searched for phrases such as “I have/developed/got/suffer(ed) from X depression,” “my X depression,” “I’m healing from X depression,” and “I’m diagnosed with X depression,” where X is a descriptive word such as “severe” and “major” (X can be empty as well). In descriptions, we further added phrases such as “depression fighter/sufferer/survivor” to the regular expression list; we removed users that had “practitioner” and “counselor” in their descriptions to exclude mental health practitioners. The remaining users captured by the regular expressions were considered to have depression.

In the end, 2575 distinct Twitter users were classified into the depression group. Of 200 randomly sampled users in the depression set, 86% were labeled positive by human annotators. We randomly selected another 2575 distinct users so that depression-related terms did not appear in their past 200 tweets or descriptions as our control group. Users in this group were not considered to have depression (nondepression group). Once we found the targeted Twitter users, we used the Tweepy application programming interface (API) to retrieve the public tweets posted by these users within the last 3 months since the time of posting the depression-related tweet, with a maximum of 200 tweets per user. We chose 200 tweets because, on average, it is roughly the number of tweets posted by an individual within a 3-month time span, which is the length commonly adopted by previous work [25,26]. If a user was identified from the description, we limited the time scope from January 18 to April 18, 2020.

Data Analysis

Personality

Previous psychological research has shown that the big five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) are related to depression [49,50]. In particular, low extraversion, high neuroticism, and low conscientiousness were associated with depressive symptoms [50]. We estimated individuals’ personality scores using IBM’s Personality Insights service [51]. For each individual, we aggregated all their tweets into a single textual input and used the Personality Insights API to obtain the scores. The minimum number of words for using the API was 100, and we were able to retrieve 4697 (91.2%) of the 5150 users’ scores. Summary statistics are shown in Multimedia Appendix 1.

Sentiments

Besides personality, we hypothesized that individuals’ sentiments and emotions could also reflect whether they were experiencing depression or not. Sentiment analysis is widely-used in deciphering people’s health and well-being from text data [52]. We estimated individuals’ sentiments using the Valencia Aware Dictionary and Sentiment Reasoner (VADER). VADER is a lexicon and rule-based model developed by researchers from the Georgia Institute of Technology [53]. We aggregated a user’s tweets into a single chunk, applied VADER, and retrieved its scores for positive and negative emotions. In Figure 2, we reported the VADER score distributions of positive emotions and negative emotions among the depression and nondepression groups. Compared with individuals with no
depression, those with depression tended to exhibit both stronger positive and negative emotions.

Figure 2. Distributions of positive and negative emotion scores among the depression and nondepression groups. VADER: Valence Aware Dictionary for Sentiment Reasoning.

Demographics

Previous psychological studies have shown differences in depression rates among people of different ages and of different genders [54-56]. Research has shown a U-shaped relationship between age and depression, with depression reaching its lowest level around the age of 45 years [54]. Women are known to be substantially more likely to have depression [57]. To estimate the age and gender of the user, we adopted the M3-inference model proposed by Wang et al [58]. The M3 model performs multimodal analysis on a user’s profile image, username, and description. Following M3’s structure, we labeled each user with a binary gender label (as approximation) and a one-hot age label among four age intervals (≤18 years, 19-29 years, 30-39 years, ≥40 years), which were then used in our fusion model. Of the 5150 users, we were able to retrieve 5059 (98.2%) users’ demographic information.

Linguistic Inquiry Word Count

We used LIWC—a well-validated psycholinguistic dictionary [59]—to capture people’s psychological states by analyzing the contents of their tweets. LIWC is a dictionary-based linguistic analysis tool that can count the percentage of words that reflect different emotions, thinking styles, and social concerns, and captures people’s psychological states. Zhang et al [60] applied LIWC to the tweets of US working adults to analyze the influence of COVID-19 on their well-being; some LIWC features in college students’ YouTube and Google search logs have been shown to correlate with their Patient Health Questionnaire-9 depression scores [46]; Coppersmith et al [26] showed the relationship between the use of the first person pronoun (which is one of the LIWC features) and depression [26].

We chose 8 features that were analyzed in previous works [26,61,62] and 7 other features that we found relevant to our study. Similar to the methods of Chen et al [63], we then applied LIWC to the concatenated tweets of individuals. Figure 3 shows the linguistic profiles for the tweets of the depression and nondepression groups. Both the depression and nondepression groups exhibited slightly positive tones, with negligible differences. The tweets of the nondepression group showed more analytical thinking, more clout, and less authentic expression than those of the depression group. The tweets of the depression group scored higher in both positive and negative emotion categories than the ones of the nondepression groups, which suggests a higher degree of immersion [64]. Moreover, the tweets of the depression group also showed more anxiety and anger emotions, and included more swear words—the anxiety, anger, and swear scores of the depression group were 50%, 22%, and 45% higher than that of the nondepression group, respectively—which is consistent with the findings of Coppersmith et al [26]. Death-related words appeared more frequently in the tweets of the depression group, which echoes Stirman and Pennebaker [62]. Similar to these 2 studies, we found more first person singular usage in the tweets of the depression group.

We also found that the tweets of the depression group expressed more sadness emotion and used words related to the biological process more frequently. Although there is no clear link between biological process–related words and depression, this finding shows that people with depression may pay more attention to their biological statuses. The power score for the tweets of the nondepression group was higher, which reflects a higher need for the power according to the findings of McClelland [65]. By comparing the work scores of the depression and nondepression groups, we found that the users of the nondepression group paid more attention to work-related issues as well.
Social Media Engagement
We used the proportion of tweets with mentions, number of responses, unique user mentions, user mentions, and tweets to measure the social media engagement of each user, as did Coppersmith et al [26]. To better understand the difference of social media engagement between the depression and nondepression groups, we added 0.1 to the number of responses, unique users mentions, users mentions, and tweets, and took the logarithm. By applying the Mann-Whitney rank test, we found that, except for the number of unique user mentions, other features were statistically different (\(P<.05\)) between the depression and nondepression groups. The users of the depression group posted more tweets and replied more. They tended to post fewer tweets with mentions, while the number of mentions for the depression group was larger, which suggests that when users of the depression group posted tweets to interact with other users, it involved more users.

Modeling
Task Definition
We formulated our task as a classification task, where the model was trained to predict whether a particular tweet or a chunk of tweets comes from a user from the depression set. Note that not all tweets by people in the depression set were explicitly referring to depression per se. By definition, though, they were all posted by users with depression and were thus labeled true. To help improve the model’s generalizability, during training and testing, we excluded all the tweets used to identify the users with depression by regular expressions that contained trivial patterns and keywords. We assumed there were subtle differences in the language used between the depression and nondepression groups. Our goal was to build a model capable of capturing these subtleties and classifying users correctly.

Tweet Chunking and Preprocessing
We performed stratified random sampling on our data set. We first sampled 500 users to form our testing set. On the rest of the users, we progressively added users to the training sets and recorded the performance of the models trained on sets of 1000, 2000, and 4650 users. All the training and testing sets have a 1:1 (depression:nondepression) ratio.

Jamil et al [28] have shown that one single tweet does not contain enough signals to determine whether a user has depression. Thus, we concatenated consecutive tweets of the same user together to create tweet chunks of 250 words and labeled the chunks based on the user’s label. Given an input sentence, the transformer tokenizer first splits each word from the input sentence into word-pieces and then vectorizes them for computation. The 250 words roughly corresponded to the maximum 512 input word-pieces allowed by transformer-based language models including BERT [33] and Robustly Optimized BiLSTM Memory Pretraining Approach (RoBERTa) [40]. This limitation is due to the self-attention mechanism in the transformer, whose time complexity scales quadratically with the input sequence length.

We preprocessed the text using the tweet preprocessing pipeline proposed by Baziotis et al [66]. We adopted this method especially due to its capability of marking Twitter-specific text habits and converting them to special tokens such as “<allcaps>” (capitalized words), “<elongated>” (repeated letters), “<repeated>” (repeated words), etc. For example, “YESSSSS, I love it so much!!!” after preprocessing will be in the form of “Yes <allcaps><elongated>, I love it so <repeated> much! <elongated>.”

After chunking and preprocessing, on average, each user had 6-7 text chunks, making the actual sizes of the 4650-user train-validation set and the 500-user testing set to be 29,315.
and 3105, respectively. The preprocessed tweet chunk data sets were then passed to deep learning models for training.

**Deep Learning Models**

We used deep learning models to perform chunk-level classification. We set up two baseline models, multi-channel CNN and BiLSTM with context-aware attention (attention BiLSTM), as described in Orabi et al [29], which achieved the best performance on the CLPsych 2015 data set. We used the pretrained GloVe embedding (840B tokens, 300d vectors) [67] augmented with the special tokens added during preprocessing. The embedding weights were further trained jointly with the model. Recently, transformer-based deep learning language models have achieved state-of-the-art performance in multiple language modeling tasks. We trained three representative transformer-based sequence classification models—BERT [33], RoBERTa [40], and XLNet [41]—with their own pretrained tokenizers augmented with the special tokens for tokenization. We chose to use the base models for all of them since we found no noticeable performance gains using their larger counterparts.

**Signal Fusion**

We ran the models on all the tweet chunks of the same user and took the average of the confidence scores to get the user-level confidence score. There were 4163 (89.5%) out of 4650 users remaining in the training set and 446 (89.2%) out of 500 users in the testing set whose entire features were retrievable. We then passed different combinations of user-level scores (personality, VADER, demographics, engagement, LIWC, and average confidence) to machine learning classification algorithms including random forest, logistic regression, and SVM provided by the scikit-learn library [68]. We only used the explainable LIWC features mentioned in the data collection section for training the classifiers.

**Training Details**

During training, we randomly split the train-validation set to training and validation sets with a ratio of 9:1. We used Adam optimizer with a learning rate of 7e-3 and weight decay of 1e-4 for training attention BiLSTM. We used Adam optimizer with a learning rate of 5e-4 for training CNN. We used AdamW optimizer with a learning rate of 2e-5 for training BERT and RoBERTa, and 8e-6 for training XLNet. We used the cross-entropy loss for all our models during training. We used the stochastic gradient descent optimizer with adaptive learning rate, with initial learning rate as 0.1 for training SVM and logistic regression classifier. We recorded the models’ performances on the validation set after each epoch and kept the model with the highest accuracy and F1 scores while training until convergence. We manually selected the hyperparameters that gave the best accuracy and F1 scores on the deep learning models.

**Results**

**Chunk-Level Classification**

In Table 1, we report our classification results at the chunk level on the testing set. Our evaluation metrics included accuracy, F1 score, area under the receiver operating characteristic curve (AUC), precision, and recall. One immediate observation was that, regardless of the model type, the classification performance improved as we increased the size of our train-validation set. This shows that for building depression classification models it is imperative to have a large number of training samples. At the same time, it also confirms that the larger number of training samples in our experiments was indeed an advantage.

Another observation was the performance gain of transformer-based models over BiLSTM and CNN models. The CNN model slightly outperformed BiLSTM, which replicated the findings of Orabi et al [29]. We observed that BERT, RoBERTa, and XLNet invariably outperformed BiLSTM and CNN regardless of the size of our training set. In particular, the XLNet model recorded the best AUC and accuracy of all the models when trained with our full training set.
Table 1. Chunk-level performance (%) of all 5 models on the 500-user testing set using training-validation sets of different sizes.\(^a\)

<table>
<thead>
<tr>
<th>Model and training-validation set</th>
<th>Accuracy</th>
<th>F1</th>
<th>AUC(^b)</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention BiLSTM(^c)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 users</td>
<td>70.7</td>
<td>69.0</td>
<td>76.5</td>
<td>70.9</td>
<td>67.3</td>
</tr>
<tr>
<td>2000 users</td>
<td>70.3</td>
<td>68.3</td>
<td>77.4</td>
<td>70.7</td>
<td>66.1</td>
</tr>
<tr>
<td>4650 users</td>
<td>72.7</td>
<td>71.6</td>
<td>79.3</td>
<td>72.1</td>
<td>71.1</td>
</tr>
<tr>
<td><strong>CNN(^d)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 users</td>
<td>71.8</td>
<td>72.6</td>
<td>77.4</td>
<td>72.7</td>
<td>72.6</td>
</tr>
<tr>
<td>2000 users</td>
<td>72.8</td>
<td>74.5</td>
<td>80.3</td>
<td>72.2</td>
<td>76.9</td>
</tr>
<tr>
<td>4650 users</td>
<td>74.0</td>
<td>70.9</td>
<td>81.0</td>
<td>77.4</td>
<td>68.9</td>
</tr>
<tr>
<td><strong>BERT(^e)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 users</td>
<td>72.7</td>
<td>74.4</td>
<td>79.8</td>
<td>72.0</td>
<td>76.9</td>
</tr>
<tr>
<td>2000 users</td>
<td>75.7</td>
<td>76.3</td>
<td>82.9</td>
<td>76.1</td>
<td>75.7</td>
</tr>
<tr>
<td>4650 users</td>
<td>76.5</td>
<td>77.5</td>
<td>83.9</td>
<td>76.3</td>
<td>78.8</td>
</tr>
<tr>
<td><strong>RoBERTa(^f)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 users</td>
<td>74.4</td>
<td>75.7</td>
<td>82.0</td>
<td>74.2</td>
<td>77.3</td>
</tr>
<tr>
<td>2000 users</td>
<td>75.9</td>
<td>77.9</td>
<td>83.2</td>
<td>73.8</td>
<td>82.5(^g)</td>
</tr>
<tr>
<td>4650 users</td>
<td>76.2</td>
<td>78.0</td>
<td>84.1</td>
<td>74.4</td>
<td>81.9</td>
</tr>
<tr>
<td><strong>XLNet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 users</td>
<td>73.7</td>
<td>75.1</td>
<td>80.7</td>
<td>73.2</td>
<td>77.2</td>
</tr>
<tr>
<td>2000 users</td>
<td>74.6</td>
<td>76.8</td>
<td>82.6</td>
<td>72.6</td>
<td>81.5</td>
</tr>
<tr>
<td>4650 users</td>
<td>77.1</td>
<td>77.9</td>
<td>84.4</td>
<td>77.5</td>
<td>78.3</td>
</tr>
</tbody>
</table>

\(^a\)We used 0.5 as the threshold when calculating the scores.
\(^b\)AUC: area under the receiver operating characteristic curve.
\(^c\)BiLSTM: bidirectional long short-term memory.
\(^d\)CNN: convolutional neural network.
\(^e\)BERT: Bidirectional Encoder Representations from Transformers.
\(^f\)RoBERTa: Robustly Optimized BiLSTM Pretraining Approach.
\(^g\)Italics indicate the best performing model in each column.

User-Level Classification

Next, we report our experiment results at the user level. Since XLNet trained on the 4650-user data set outperformed the other models, we took it for user-level performance comparison. Our experimental results demonstrated a substantial increase on the user-level scores of XLNet shown in Table 2 compared to the chunk-level score shown in Table 1. This indicates that more textual information of a user yields more reliable results on determining whether the user has depression. Building on the user-level XLNet scores, we further included VADER, demographic, engagement, personality, and LIWC scores as signals. We first used all features and compared the performance of random forest, logistic regression, and SVM. We noticed that SVM achieved the best scores on accuracy and F1, slightly surpassing logistic regression. Thus, we used SVM for testing the performance when using part of the features collected. The results are shown in Table 2. The results have shown that using VADER, demographics, and social media engagement features alone does not help the classification by much. Classifiers using personality features and LIWC features perform relatively better. We then used these five feature groups and obtained a better result (accuracy 71.5%; F1 score 72.0%). However, the classifier was still outperformed by XLNet, showing that the transformer-based models indeed worked better on depressive Twitter text modeling compared with other approaches. We further increased the classifier’s performance by using all the features, namely, VADER, demographics, engagement, personality, and LIWC features, and the averaged XLNet confidence score; the performance of the three machine learning algorithms did not vary much, and the SVM classifier achieved the best accuracy (78.9%) and F1 (79.2%) scores.

In an attempt to investigate what specific textual features besides those extracted by XLNet have the most impact on depression classification, we calculated the permutation feature importance [69] on the trained random forest classifier using the VADER, engagement, personality, and LIWC features with 10 repeats. The importance scores of individual features are shown in Figure...
Among the LIWC features, “i,” “bio,” “power,” “sad,” and “authentic” are shown to be important in classification. Among the five personality features, “conscientiousness” and “neuroticism” were shown to be closely related to depression cues. We did not observe a strong relation between VADER sentiment features or social media engagement features and the depression signals. As for the LIWC sentiment features, only “sad” and “anxiety” were shown to be relatively important. It is worth noting that LIWC’s “sad” and “anxiety” categories each referred to about 150 words. By contrast, more than 7500 words or features fell in to the negative category in VADER. The insignificance of VADER features can be attributed to the more focused nature of LIWC.

Table 2. User-level performance (%) using different features.

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
<th>F1</th>
<th>AUC&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>VADER&lt;sup&gt;c&lt;/sup&gt;</td>
<td>54.9</td>
<td>61.7</td>
<td>54.6</td>
</tr>
<tr>
<td>Demographics</td>
<td>58.7</td>
<td>56.0</td>
<td>61.4</td>
</tr>
<tr>
<td>Engagement</td>
<td>58.7</td>
<td>62.3</td>
<td>61.7</td>
</tr>
<tr>
<td>Personality</td>
<td>64.8</td>
<td>67.8</td>
<td>72.4</td>
</tr>
<tr>
<td>LIWC&lt;sup&gt;d&lt;/sup&gt;</td>
<td>70.6</td>
<td>70.8</td>
<td>76.0</td>
</tr>
<tr>
<td>V + D + E + P + L&lt;sup&gt;e&lt;/sup&gt;</td>
<td>71.5</td>
<td>72.0</td>
<td>78.3</td>
</tr>
<tr>
<td>XLNet</td>
<td>78.1</td>
<td>77.9</td>
<td>84.9</td>
</tr>
<tr>
<td>All (random forest)</td>
<td>78.4</td>
<td>78.1</td>
<td>84.9</td>
</tr>
<tr>
<td>All (logistic regression)</td>
<td>78.3</td>
<td>78.5</td>
<td>86.4&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>All (SVM&lt;sup&gt;g&lt;/sup&gt;)</td>
<td>78.9</td>
<td>79.2</td>
<td>86.1</td>
</tr>
</tbody>
</table>

<sup>a</sup>We used SVM for classifying individual features.
<sup>b</sup>AUC: area under the receiver operating characteristic curve.
<sup>c</sup>VADER: Valence Aware Dictionary and Sentiment Reasoner.
<sup>d</sup>LIWC: Linguistic Inquiry and Word Count.
<sup>e</sup>V + D + E + P + L: VADER + demographics + engagement + personality + LIWC.
<sup>f</sup>Italics indicate the best performing model in each column.
<sup>g</sup>SVM: support vector machine.

Figure 4. Permutation importance of different features. LIWC: Linguistic Inquiry and Word Count; VADER: Valence Aware Dictionary for Sentiment Reasoning.

Application Results

In this section, we report two COVID-19–related applications of our XLNet based depression classifier: (1) monitoring the evolution of depression levels among the depression group and the nondepression group, and (2) monitoring the depression level at the US country level and state level during the pandemic.
We chose to use XLNet because of its simplicity as a stand-alone model, as it performed comparably to the fusion model.

**Depression Monitoring on Depression and Nondepression Groups**

We took the 500 users from the testing set (n=500), along with their tweets from January 1 to May 22, 2020. We concatenated a user’s tweets consecutively from January 1 one by one until reaching 250 words and labeled this chunk’s date as the date of the author posting the tweet that was in the middle of the chunk. We grouped 3 days into a bin from January 1 and assigned the chunks to the bins according to the labeled date. We ran the XLNet model on the preprocessed tweet chunks and recorded the confidence scores. We trimmed the upper and lower 10% of the data to reduce the skew in the score distribution. We then took the mean of the scores for each time bin and plotted the depression trend shown in Figure 5. We further took a moving average of 5 time bins to smooth the curves.

**Figure 5.** Aggregated depression level trends of the depression and nondepression groups from January 1 to May 22, 2020. Since users with depression have a substantially higher depression level, we used different y-axes for the 2 groups’ depression levels to compare them side by side.

Two immediate observations followed. First, depression level among users in the depression group was substantially higher than that in the nondepression group. This held across the entire observation period from early January to late May 2020. Second, and more importantly, the depression levels shared a strikingly similar trend among the two groups.

Delving deeper into these curves, we marked three important time points on the plot—the first confirmed case of COVID-19 in the United States (January 21, 2020), the US National Emergency announcement (March 13), and the last stay-at-home order issued (South Carolina, April 7). In January, both groups experienced a drop in depression scores. This may be caused by the fact that people’s mood usually hits its lowest in winter [70]. From the day when there was the first confirmed case in the United States to the day of the announcement of the US National Emergency, the trends of the depression and nondepression groups were different. The depression level of the depression group went down slightly, while the depression level of the nondepression group went up. Aided by psychological findings, we hypothesized that depressive users were less affected by negative events happening in the outside world because they focused on their own feelings and life events, since they were mostly affected by negative events that threatened them directly [71] and more interactions with the outside world gave them more negative feedback [72]. Moreover, the depression levels of the depression and nondepression groups both increased after the announcement of the US National Emergency.

To better understand the trend, we applied the LDA model to retrieve the topics before and after the announcement of the US National Emergency. Each chunk of the tweets was assigned 5 weights for each of the 5 topics. We labeled the topic of the highest weight as the dominant topic of this chunk of the tweets and counted the frequency of each topic shown in Figure 6. Details about the keywords of the topics are reported in Multimedia Appendix 1. Before the announcement, the two most frequent topics of the depression and nondepression groups were the discussions about US President Donald Trump and...
about school and work. The third most frequent topic of the nondepression group was about health while that of depression group was about entertainment. This supports the difference of the depression level trends of the two groups. After the announcement of the US National Emergency, the most frequent topic of the depression group was depression and anxiety during COVID-19, while this was the third most frequent topic of the nondepression group. Further, all 5 topics of each group were about COVID-19. This shows that, when people mostly talk about COVID-19, depression signals rise for both groups.

**Figure 6.** Topic distributions of depression and nondepression groups before and after the announcement of the US National Emergency.

### Aggregated Depression in COVID-19

To investigate country-level and state-level depression trends during COVID-19, we randomly sampled users who had US state locations stated in their profiles and crawled their tweets between March 3 and May 22, 2020, the period right before and after the US announced a National Emergency on March 13. Using the same logic as in the previous section, we plotted the change of depression scores of 9050 geolocated users (n=9050) sampled from the 36.6 million users mentioned, excluding those used for training, as the country-level trend. For state-level comparison, we plotted the aggregated scores of three representative states—economical center New York on the East Coast that was highly affected by the virus, tech center California on the West Coast that was also struck hard by the virus, and the less affected tourism center Florida in the southeast. Each selected state had at least 550 users in the data set to validate our findings. Their depression levels are shown in **Figure 7**.

The first observation of the plot is that depression scores of all three states and the United States behaved similarly during the pandemic; they experienced a decrease right before the National Emergency; a steady increase after that; a slight decrease past April 23, 2020; and another sharp increase after May 10. We also noticed that the overall depression score of Florida was substantially lower than the US average and the other two states. Since Florida had a lower score both before and after the virus outbreak, we hypothesized that it has a lower depression level overall compared to the average US level irrespective of the pandemic.

We calculated the topics at the state level after the announcement of the US National Emergency. As shown in **Figure 8**, the most frequent topic was the government’s policy on COVID-19. California and Florida were the states that paid relatively more attention to this topic compared to the US average and New York State. Florida also talked more about the life change during COVID-19. Another finding was that people in New York talked more about the hospital news, likely because the state contained the majority of cases in the country by May 22, 2020 [73].
Figure 7. Aggregated depression level trends of the United States, New York, California, and Florida after the announcement of the US National Emergency.

Figure 8. Distributions of the top 5 topics (state level) after the announcement of the US National Emergency.
Discussion

Principal Results

In this study, we developed a practical pipeline that included first gathering and cleaning a large-scale Twitter depression classification data set quickly in response to an outbreak, then training an accurate depression signal detection model on this data set, and finally applying the model to monitoring public depression trends. We analyzed the depression level trends during the COVID-19 pandemic, which shed light on the psychological impacts of the pandemic. Our main results were fourfold and corresponded to the four objectives listed in the Goal of the Study section.

First, using a stringent yet effective regular expression-based search method, we constructed by far the largest data set with 5150 Twitter users, including half identified as depression users and half as control users, along with their tweets within the past 3 months and their Twitter activity data.

Second, we developed a chunking and regrouping method to construct 32,420 tweet chunks, with 250 words each in the data set. We progressively added data to our training set and showed experimentally that the performance of deep learning models improves as the size of the training set grows, which validates the importance of our data set size. We compared the models’ performances at the chunk level with the user level and observed further performance gain, which added credibility to our chunking method.

Third, we built a more accurate classification model (with 78.9% accuracy on n=449) upon the deep learning models along with linguistic analysis of dimensions including personality, LIWC, sentiment features, and demographic information. A permutation importance test showed that conscientiousness, neuroticism, appearance of first person pronouns, talking about biological processes such as eating and sleeping, talking about power, and exhibiting sadness are closely related to depression cues.

Finally, we showed the feasibility of the two proposed methods for monitoring the change of public depression levels as the disease propagates by aggregating individuals’ past tweets within a time frame. Our method can target different groups of people, and we showed the depression trends of identified depression and nondepression groups (n=500), and of groups at different geolocations (n=9050). The temporal trends showed that the nondepression group’s depression level rose earlier than that of the depression group, which we explained by psychological theories and LDA topics extracted from key time points. We also found that New York, California, Florida, and the United States in total all shared a similar depression trend, with Florida having a substantially lower depression level, which was also verified by LDA topic analysis.

Practical Implications

Our study has practical implications. For example, upon detecting a rise in depression levels in a certain area, internet-based intervention services can be recommended by the social media platforms to the users. An intervention for depression commonly recommended is cognitive behavioral therapy (CBT), which is a type of therapy that targets one’s irrational thinking patterns and unadaptable behavioral patterns [74]. During the COVID-19 period, digital-based CBT can be adopted. It has shown to be effective in reducing symptoms of mental disorders [75,76]. At the same time, it is also cost-effective and practical during the pandemic [75]. In addition to digital-based CBT, social media–based suicide prevention messages have also shown to be effective [77] and can be sent to individuals at risk.

Limitations

Although our data collection method is fast and fully automatic, we acknowledge that the same limitations exist as noted in detail by Coppersmith et al [26]. Specifically, the users with depression captured by us can only represent a subpopulation (those who use Twitter and are willing to disclose their conditions) of the general depression population, and we cannot guarantee that the control group was not contaminated.

Comparison With Prior Work

The data set used in this study containing 2575 depression users was much larger than those used previously, which contained 1402 depression users at most. De Choudhury et al [48] demonstrated that depression prediction models can potentially be used at the population level. However, to the best of our knowledge, all Twitter user depression identification studies reviewed in the introduction section focus on either tweet-level or user-level classification rather than applying the model to analyzing the mental health trends of a large population. To our knowledge, we were also the first to apply the transformer-based models (BERT, RoBERTa, XLNet) to identifying depression users on Twitter using a large-scale data set and to monitor the public depression trend.

Conclusions

COVID-19 has infected over 100 million people worldwide [1], virtually bringing the whole world to a halt. During this period, social media witnessed a spike in depression terms. Against this backdrop, we have developed transformer-based models trained with by far the largest data set on depression. We have analyzed our models’ performance in comparison to existing models and verified that the large training set we compiled was beneficial to improving the models’ performance. We further showed that our models can be readily applied to the monitoring of stress and depression trends of targeted groups over geographical entities such as states. We noticed substantial increases in depression signals as people talked more about COVID-19. We hope researchers and mental health practitioners find our models useful and that this study raises awareness of the mental health impacts of the pandemic.
Authors' Contributions
YZ and JL conceived and designed the study. YZ performed regular expression search and preprocessing, examined feature importance, and wrote the majority of the manuscript. HL performed data collection and applied the LDA models. HL and YZ analyzed the data and wrote part of the manuscript. YZ and YL trained the models and performed depression monitoring. XZ analyzed the findings using psychological theories. All authors helped design the study and edit the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental data statistics and tables.

References
2. Education: from disruption to recovery. UNESCO. 2020 Mar 04. URL: https://en.unesco.org/covid19/educationresponse [accessed 2020-12-21]


68. scikit-learn. URL: https://scikit-learn.org/stable/ [accessed 2020-12-21]


Abbreviations

API: application programming interface
AUC: area under the receiver operating characteristic curve
BERT: Bidirectional Encoder Representations from Transformers
BiLSTM: bidirectional long short-term memory
BOW: bag of words
CBT: cognitive behavioral therapy
CNN: convolutional neural network
LDA: latent Drichlet allocation
LIWC: Linguistic Inquiry and Word Count
PTSD: posttraumatic stress disorder
RoBERTa: Robustly Optimized Bidirectional Long Short-Term Memory Pretraining Approach
SVM: support vector machine
VADER: Valence Aware Dictionary and Sentiment Reasoner

Please cite as:
The Role of the Canadian Media During the Initial Response to the COVID-19 Pandemic: A Topic Modelling Approach Using Canadian Broadcasting Corporation News Articles

Janhavi Patel¹*, BMSc; Harsheev Desai²*, BS; Ali Okhowat³, MD

¹Michael G DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada
²Faculty of Engineering and Architectural Science, Ryerson University, Toronto, ON, Canada
³UBC Global Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
*these authors contributed equally

Corresponding Author:
Janhavi Patel, BMSc
Michael G DeGroote School of Medicine
McMaster University
1280 Main Street West
Hamilton, ON, L8S 4L8
Canada
Phone: 1 905 525 9140
Email: patelj10@mcmaster.ca

Abstract

Background: Beginning as a local epidemic, COVID-19 has since rapidly evolved into a pandemic. As countries around the world battle this outbreak, mass media has played an active role in disseminating public health information.

Objective: The aim of this study was to get a better understanding of the role that the Canadian media played during the pandemic and to investigate the patterns of topics covered by media news reporting.

Methods: We used a data set consisting of news articles published on the Canadian Broadcasting Corporation (CBC) website between December 2019 and May 2020. We then used Python software to analyze the data using Latent Dirichlet Allocation topic modelling. Subsequently, we used the pyLDAvis tool to plot these topics on a 2D plane through multidimensional scaling and divided these topics into different themes.

Results: After removing articles that were published before the year 2019, we identified 6771 relevant news articles. According to the CV coherence value, we divided these articles into 15 topics, which were categorized into 6 themes. The three most popular themes were case reporting and testing (n=1738), Canadian response to the pandemic (n=1259), and changes to social life (n=1171), which accounted for 25.67%, 18.59%, and 17.29% of the total articles, respectively.

Conclusions: Understanding the Canadian media’s reporting on the COVID-19 pandemic shows that the Canadian pandemic response is a product of consistent government communication, as well as the public’s understanding of and adherence to protocols.

(JMIR Infodemiology 2021;1(1):e25242) doi:10.2196/25242

KEYWORDS
COVID-19; topic modelling; LDA; health communication; mass media; coronavirus; media; dissemination; online health information; public health

Introduction

COVID-19, which started as a local epidemic, evolved into a pandemic in a matter of months [1]. Countries around the world are battling the spread of this disease and the unfortunate consequences of COVID-19–related mortality and morbidity, resource limitations, and severe economic burden [1,2]. Canada is no different and continues to observe a rising number of COVID-19 cases [3].

Due to the initial lack of vaccines and knowledge about the disease and its treatment, countries were forced to take unique approaches to combat the spread of the virus. Canada’s response has been widely reported as being adequate, though much more could have been and remains to be done in tackling the spread
of COVID-19 [4]. The Canadian government website for its COVID-19 response highlights the measures Canada has taken, including the creation of the COVID Alert app, an ethics framework for policy makers, and economic support for Canadians. Support involves both financial measures and safety, including for Canadians abroad and vulnerable populations in Canada [3]. Additionally, there has been an emphasis on public education, collaboration, and guidance for researchers and frontline health workers [3].

With the uncertainty surrounding this novel coronavirus, the media—especially online news sources—have played a key role in informing the public about events related to the pandemic. Mass media has been successfully used for decades to increase public health awareness. News media outlets have been used across the globe for addressing public health issues like reducing tobacco use, participating in screening for cancer, and cardiovascular disease prevention [5]. The Eat Well campaign, which was advertised through a combination of news and commercial media outlets, increased awareness about meal prepping and healthy food choices in the Canadian population [6]. A postcampaign evaluation showed that low-resource communities had a greater uptake of information, thus highlighting the need to better understand the impact of different information dissemination campaigns to better cater to the target population [6].

The Canadian Broadcasting Corporation (CBC) is a daily source of local and national information for many Canadians [7]. The CBC’s digital offering sees an average of 16.1 million new monthly visits [7] and continues to grow every month. Assessing the content of CBC articles can therefore provide insights into the information delivered to Canadians about the pandemic. Given that success in the fight against the pandemic greatly depends on the support of the public (eg, maintaining appropriate social distance and taking proper precautions), the information that media outlets report is important as it provides the public with up-to-date guidance.

The aim of this study was to better understand the role that news articles played in disseminating public health information, by specifically focusing on the topics reported and frequency of each topic reported regarding the COVID-19 pandemic. The methods used and results from this study could be relevant when reporting future events related to health care and national safety, which rely heavily on public support and awareness.

### Methods

#### Data Collection

The data set was collected from the CBC website using a Python programming language script [8]. The script was used to extract information from over 6700 news articles, including the title, article summary, and main text for each article, using the term “coronavirus” as the search word. The extracted news articles were published between December 2012 and May 2020; however, only the articles published in 2019 and 2020 were included in this study.

We used Latent Dirichlet Allocation (LDA) to analyze these news articles. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modelled as a finite mixture over an underlying set of topics. The basic idea behind LDA is that documents can be represented as arbitrary mixtures over latent topics, which in turn are characterized by a distribution over words [9]. LDA has been extensively used and evaluated for its applicability in topic modelling research [10,11]. Moreover, Lancichinetti [12] showed that LDA has high reproducibility and accuracy for topic classification.

According to LDA, there are diverse topics in each news article, and the words in these articles can be allotted to one of these topics. However, LDA only groups inputs (ie, news articles in this case) based on the abovementioned distribution over words and it is subjective how these groups are interpreted as topics. To facilitate accurate representation, randomly selected articles from each topic were manually checked to make sure they were consistent with the interpreted topic.

#### Data Processing

There were a total of 6771 news articles remaining after removing the articles published before 2019. These remaining articles were dated between December 22, 2019, and May 3, 2020.

Before moving forward with topic modelling, we used Python along with libraries, including the Pandas and Natural Language Toolkit (NLTK) libraries [13], to clean the data. The detailed process for this is displayed in Figure 1. We used the English language stop words provided by NLTK to remove common words such as “an,” “all,” “and,” “for,” and “from” as they hold no semantic value for our analysis. URLs and social media mentions consisting of “@” were also removed. The two primary inputs to the LDA model are the dictionary and the corpus, which were created using the Gensim library [14].
We used the CV coherence score to evaluate models with different numbers of topics and selected the one with the highest CV score. This approach mitigates one of LDA's limitations—the need to know the number of topics ahead of time. According to Röder et al [15], the CV coherence score is one of the fastest measures of coherence, and the most accurate. Henri Trenquier defines coherence as the human's semantic appreciation of a topic represented by its N top words [16]. We chose the top 15 (N) words in each topic to calculate the coherence.

As evident from Figure 2, the highest CV coherence score was achieved at 15 topics. This means that the top words in each topic were most closely related semantically when the news article data set was divided into 15 different topics using LDA.
We then used the pyLDAvis tool [17] and Python to further analyze the 15 topics to extract valuable insights from the articles. The 15 topics were represented on an intertopic distance map, which is an interactive representation offered by the pyLDAvis tool (Figure 3). The topics are plotted as circles in a 2D plane whose centers are determined by computing the distance between topics [16].

Figure 3. Intertopic distance map.
The weight parameter $\lambda$ was adjusted to find the theme for each topic based on the top words in the topic. Setting $\lambda=1$ ranks the words in a topic by frequency, while setting $\lambda=0$ ranks the words based on uniqueness to that topic [17]. We used the interactive bar provided by the pyLDAvis tool to adjust $\lambda$ and understand the theme for each of the 15 topics. To use topic 9 as an example, Figure 4 shows the top 30 most frequently occurring words in topic 9. As multiple topics might have similar words that occur frequently, we need to adjust the $\lambda$ value to better gauge what topic an article might be about. For instance, when we set $\lambda$ to 0.04, the terms most unique to topic 9 are captured, and presented in descending order in Figure 5. This analysis identifies words like “test,” “positive,” and “spread” as being unique to topic 9. Using the keywords, we identified that the general theme of articles in topic 9 is “testing.” This process was repeated for each topic (Table 1).

**Figure 4.** Top 30 most relevant terms ($\lambda=1.0$).
Figure 5. Top 30 most relevant terms ($\lambda=0.4$).
Table 1. Themes and topics (N=6771).

<table>
<thead>
<tr>
<th>Themes and topics</th>
<th>Number of news articles, n (%)</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theme 1: Case reporting and testing (n=1738)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 9: Testing</td>
<td>876 (12.97)</td>
<td>Coronavirus, spread, public, positive, novel, official, people, health, test, covid</td>
</tr>
<tr>
<td>Topic 14: Case reporting</td>
<td>862 (12.73)</td>
<td>Province, number, confirm, total, people, report, health, death, case, covid</td>
</tr>
<tr>
<td><strong>Theme 2: Canadian response to pandemic (n=1259)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 5: General response</td>
<td>295 (4.37)</td>
<td>Pandemic, nation, member, Windsor, covid, community, family, first, local, want</td>
</tr>
<tr>
<td>Topic 6: Health care/hospital response</td>
<td>267 (3.93)</td>
<td>Emergency, temporary, staff, state, hospital, Sudbury, worker, declare, general, covid</td>
</tr>
<tr>
<td>Topic 10: Vaccine research</td>
<td>399 (5.88)</td>
<td>Ottawa, world, around, Canada, global, point, latest, covid, coronavirus, point</td>
</tr>
<tr>
<td>Topic 1: Medical supplies and resources</td>
<td>298 (4.40)</td>
<td>Canadian, ventilator, doctor, could, Canada, happening, mask, province, available, covid</td>
</tr>
<tr>
<td><strong>Theme 3: Changes to everyday life (n=1171)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 2: Social gathering cancellations</td>
<td>322 (4.76)</td>
<td>Summer, pandemic, coronavirus, cancel, festival, university, event, season, covid, plant</td>
</tr>
<tr>
<td>Topic 8: School closure/virtual learning</td>
<td>397 (5.88)</td>
<td>Parent, school, family, child, learning, student, covid, equipment, worker, pandemic</td>
</tr>
<tr>
<td>Topic 12: General lifestyle changes</td>
<td>452 (6.68)</td>
<td>People, avoid, coming, together, change, normal, covid, province, pandemic, government</td>
</tr>
<tr>
<td><strong>Theme 4: Communication from the government (n=1002)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 11: Public health announcements</td>
<td>481 (7.04)</td>
<td>Medical, chief, public, officer, health, people, province, covid, Friday</td>
</tr>
<tr>
<td>Topic 3: Prime Minister’s addresses</td>
<td>521 (7.68)</td>
<td>Minister, prime, Justin, Trudeau, worker, pandemic, essential, coronavirus, health, covid</td>
</tr>
<tr>
<td><strong>Theme 5: International news (n=826)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 4: Articles related to news in the United States</td>
<td>388 (5.73)</td>
<td>Trump, unite, outbreak, state, president, country, cruise, coronavirus, Canada, Canadian</td>
</tr>
<tr>
<td>Topic 7: Articles related to news in China</td>
<td>438 (6.48)</td>
<td>Chinese, china, outbreak, answer, Canadian, flight, morning, expert, question, expert</td>
</tr>
<tr>
<td><strong>Theme 6: Government initiatives (n=775)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 13: Initiatives for vulnerable populations</td>
<td>343 (5.08)</td>
<td>Shelter, homeless, social, distance, encourage, people, pandemic, covid, measure, physical</td>
</tr>
<tr>
<td>Topic 15: Economy and business</td>
<td>432 (6.39)</td>
<td>Business, economy, government, federal, pandemic, support, covid, million, premier Canada</td>
</tr>
</tbody>
</table>

*The percentages have been calculated using the N value (ie, 6771).

**Results**

Using the pyLDAvis tool, we grouped the 15 topics into 6 themes as shown in Table 1. Theme 1 (case reporting and testing) had the greatest number of articles (n=1738), while theme 6 (government initiatives) represented the lowest number of news reports (n=775). The trend in the total frequency of articles related to COVID-19 over our study time period is shown in Figure 6.
The most frequent theme, theme 1 (case reporting and testing), consisted primarily of topics that covered articles related to information about testing (12.97%) and case reporting (12.73%). The information in the articles related to testing focused on information regarding the tests being conducted to assess the spread of the virus, whereas the articles related to case reporting primarily focused on reporting the number of confirmed cases and deaths around the country, with words like “number,” “report,” “confirm,” and “death” being frequently used. Similar to the trend in frequency of total articles about COVID-19, Theme 1 saw a sudden increase in the number of articles published, starting from the month of February and increasing throughout March and April (Figure 7).

The Canadian media’s focus in relation to the outbreak and Canada’s response is highlighted in themes 2 and 3. Theme 2 (Canadian response to pandemic) includes topics like general response (topic 5, n=295, 4.37%), health care/hospital response (topic 6, n=267, 3.93%), vaccine research (topic 10, n=399, 5.88%), and medical supplies and resources (topic 1, n=298, 4.40%). Theme 3 (changes to everyday life) discusses the changes resulting from the pandemic and includes topics like social gathering cancellations (topic 2, n=322, 4.76%), school closure/virtual learning (topic 8, n=397, 5.88%), and general...
lifestyle changes (topic 12, n=452, 6.68%). Both these themes saw a steep increase in the number of reported articles after the month of February.

Theme 5 (international news) consisted primarily of topics related to the United States (n=388, 5.73%) and China (n=438, 6.48%). The information in articles related to the United States, Canada’s geographic neighbor and largest trading partner, focused more on political relations, with frequently used words like “president,” “Trump,” and “state.” On the contrary, the articles about China, the place of origin of the coronavirus, primarily focused on information that would enable a better understanding of the outbreak, with words like “outbreak,” “question,” and “answer” used more frequently.

Themes 4 and 6, although not high in frequency, focused on important themes like communication from the government and government initiatives. Communication from the government included both public health announcements as well as the Prime Minister’s addresses to the public. The government initiatives theme included topics that discussed initiatives for vulnerable populations, specifically people experiencing homelessness (topic 13, n=343, 5.008%), as well as the economy and business (topic 15, n=432, 6.39%). The articles about government announcements had a steep increase leading to the declaration of pandemic in March 2020 but the slope reduced from March to April. Contrastingly, theme 6 saw a steep, consistent increase from February to April 2020.

Discussion

Principal Results

The COVID-19 pandemic has been a steep learning curve for all countries worldwide. Dissemination of information in a timely manner across communities and countries was crucial to limit the spread of COVID-19 and determine the efficacy of different treatment and management interventions. With the ensuing social isolation, online media took over as an important source of information available to the public; thus, understanding the role that the media played highlights key aspects of the challenges faced by Canada and its response to the pandemic. We used topic modelling using articles collected from the CBC’s online platform and identified different themes reported through the articles.

Even though some reports about a fatal pneumonia of unknown cause had started coming out from China in early January, it was not until after the World Health Organization declared COVID-19 as a global health emergency that articles about the virus started increasing in Canada. The number of articles about COVID-19 showed a sharp increase starting February 2020 for most themes, after the World Health Organization declared it a global health emergency on January 30. Resource shortages and panic buying have been an issue in many countries battling COVID-19 [18]. Our study identified that there were about 300 articles focused on resources. It is postulated that anxiety around sudden lockdowns and uncertainty about the duration of the pandemic might have contributed to the response of preservation of self and family [18]. Retrospectively, it would be beneficial for health care professionals, the government, and the media to work closely together to provide better guidelines and policies for the public, both to reduce anxiety and ensure more equitable distribution of resources.

Additionally, our topic modelling showed that a considerable proportion of news articles in the study focused on the conditions of marginalized populations, such as people experiencing homelessness. Many people experiencing homelessness did not receive timely shelter and space to self-isolate, putting their lives and the lives of others at risk. Thus, our study results highlight the importance of creating an equitable response strategy during future pandemics.

Throughout the course of the pandemic, the most reported information was regarding testing and case reporting. This is consistent with any communicable disease, wherein proper testing, contact tracing, and case reporting are crucial to control the spread of the disease [19]. This information can contribute to increased anxiety, as witnessed by people’s fear of acquiring the disease from health care facilities, and thus being reluctant to access care for other acute illnesses, including heart attacks and strokes [20]. On the other hand, having this information could make people feel more accountable for their actions and encourage them to be more socially responsible. Although the neglect of other conditions was an unintended, unfortunate consequence of pandemic-related public health measures, for future events, more holistic communication from health care professionals (ie, about considering other acute illnesses in times of crisis) and reporting from the media on this topic could aid in better management of people with acute and chronic illnesses.

Limitations

This study only contains news articles published on CBC’s online platform that were tagged with the term “coronavirus.” There are several other sources of media available in Canada and future studies can focus on including multiple different sources of both digital and print media. The pandemic is ongoing and Canada’s response and policies are constantly changing. Thus, doing a long-term study and constantly monitoring multiple media outlets’ efforts will be helpful for future studies. This study nonetheless provides a glimpse of the Canadian media’s role in the communication and dissemination of information. The LDA model has certain limitations; for example, the different topics need to be manually interpreted and are open to misinterpretation or overinterpretation. Some of its other limitations include the inability to capture correlations between topics and the use of a fixed number of topics, which must be known ahead of time.

Comparison With Prior Work

Our study identified several similar and unique themes compared to the themes identified by another similar study on Chinese media reporting [21]. Topics like case reporting, disease spread, medical supplies and resources, and research and development were similarly observed in media in both studies. However, the Chinese study did not identify any themes related to communication from the government or the country’s response regarding vulnerable populations. In contrast, although lower in frequency compared to other topics, Canada’s media and response focused on ensuring proper communication from the
government and support for vulnerable populations. The government actively communicated with the public, not only through public health officials but also via regular addresses from the country’s prime minister during the pandemic. Studies have shown that a leader's address to the public is very effective in reassuring people during times of crisis [22] and the media reports suggest that it played a big part in Canada’s response to the pandemic.

Compared to initial communication during the H1N1 pandemic in 2009, which involved the dissemination of misinformation, leading to widespread panic, the slow dissemination of public health information by media outlets initially led to panic early in the COVID-19 pandemic. After the H1N1 pandemic, the Centers for Disease Control and Prevention conducted an audit on public health information dissemination and provided several guidelines for communications in future pandemics [23]. In line with the guidelines, our study topics found that the Canadian response had consistent messaging from federal government and public health officials; however, Canada’s response still fell short with regard to prioritizing marginalized populations and reducing the public’s initial stress stemming from widespread misinformation.

Conclusions

Our study highlights that, based on the topical analysis of CBC news articles, the Canadian response to the COVID-19 pandemic was a joint effort guided by government policies and communications in conjunction with people’s response and adherence to protocol.

One of the most important factors in preventing the spread of COVID-19 is to empower the public with accurate information [24].

The media plays an important bridging role by relaying information from the government to the public. Thus, by understanding and analyzing the extent to which certain events and policies affect public sentiment and response, policy makers can proactively improve communication for any similar future events, including pandemics, natural disasters, or issues related to national safety.

Conflicts of Interest
None declared.

References


Abbreviations

CBC: Canadian Broadcasting Corporation
LDA: Latent Dirichlet Allocation
NLTK: Natural Language Toolkit

©Janhavi Patel, Harsheev Desai, Ali Okhowat. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 18.06.2021. 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.
Corrigenda and Addenda

Metadata Correction: “Desensitization to Fear-Inducing COVID-19 Health News on Twitter: Observational Study”

Hannah R Stevens1, BA; Yoo Jung Oh1, MA; Laramie D Taylor1, PhD
Department of Communication, University of California, Davis, Davis, CA, United States

Corresponding Author:
Hannah R Stevens, BA
Department of Communication
University of California, Davis
1 Shields Ave
Davis, CA, 95616
United States
Phone: 1 530 752 1011
Email: hrstevens@ucdavis.edu

Related Article:
Correction of: https://infodemiology.jmir.org/2021/1/e26876/

(JMIR Infodemiology 2021;1(1):e32231) doi:10.2196/32231

In “Desensitization to Fear-Inducing COVID-19 Health News on Twitter: Observational Study” (JMIR Infodemiology 2021;1(1):e26876) the authors noted one error.

One author’s name was displayed as follows:

Laramie R Taylor

The middle initial in the name has now been corrected as follows:

Laramie D Taylor

The correction will appear in the online version of the paper on the JMIR Publications website on July 28, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 19.07.21; this is a non–peer-reviewed article; accepted 20.07.21; published 28.07.21.

Please cite as:
Stevens HR, Oh YJ, Taylor LD
Metadata Correction: “Desensitization to Fear-Inducing COVID-19 Health News on Twitter: Observational Study”
JMIR Infodemiology 2021;1(1):e32231
URL: https://infodemiology.jmir.org/2021/1/e32231
doi:10.2196/32231
PMID:

©Hannah R Stevens, Yoo Jung Oh, Laramie D Taylor. Originally published in JMIR Infodemiology (https://infodemiology.jmir.org), 28.07.2021. 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.
Infodemic Signal Detection During the COVID-19 Pandemic: Development of a Methodology for Identifying Potential Information Voids in Online Conversations

Tina D Purnat1*, MSc; Paolo Vacca2*, BSt; Christine Czerniak3*, PhD; Sarah Ball2*, BA; Stefano Burzo4*, MA; Tim Zecchin2*, BA; Amy Wright2*, BA; Supriya Bezbaruah5*, PhD; Faizza Tanggol6*, BA; Ève Dubé7*, PhD; Fabienne Labbé7*, PhD; Maude Dionne7*, MSc; Jaya Lamichhane3*, MA, MBA; Avichal Mahajan3*, PhD; Sylvie Briand3*, MPH, MD, PhD; Tim Nguyen3*, MSc

1Digital Health and Innovation, Science Division, World Health Organization, Geneva, Switzerland
2Media Measurement Ltd, London, United Kingdom
3Emergency Preparedness, World Health Organization, Geneva, Switzerland
4Department of Political Science, University of British Columbia, Vancouver, BC, Canada
5Health Emergencies Programme, World Health Organization Regional Office for South East Asia, New Delhi, India
6World Health Organization Country Office Malaysia, Brunei Darussalam and Singapore, Putrajaya, Malaysia
7Institut national de santé publique du Québec, Montreal, QC, Canada

* these authors contributed equally

Corresponding Author:
Christine Czerniak, PhD
Emergency Preparedness
World Health Organization
20 Avenue Appia
Geneva, 1211
Switzerland
Phone: 41 (0)227912111
Email: czerniakc@who.int

Abstract

Background: The COVID-19 pandemic has been accompanied by an infodemic: excess information, including false or misleading information, in digital and physical environments during an acute public health event. This infodemic is leading to confusion and risk-taking behaviors that can be harmful to health, as well as to mistrust in health authorities and public health responses. The World Health Organization (WHO) is working to develop tools to provide an evidence-based response to the infodemic, enabling prioritization of health response activities.

Objective: In this work, we aimed to develop a practical, structured approach to identify narratives in public online conversations on social media platforms where concerns or confusion exist or where narratives are gaining traction, thus providing actionable data to help the WHO prioritize its response efforts to address the COVID-19 infodemic.

Methods: We developed a taxonomy to filter global public conversations in English and French related to COVID-19 on social media into 5 categories with 35 subcategories. The taxonomy and its implementation were validated for retrieval precision and recall, and they were reviewed and adapted as language about the pandemic in online conversations changed over time. The aggregated data for each subcategory were analyzed on a weekly basis by volume, velocity, and presence of questions to detect signals of information voids with potential for confusion or where mis- or disinformation may thrive. A human analyst reviewed and identified potential information voids and sources of confusion, and quantitative data were used to provide insights on emerging narratives, influencers, and public reactions to COVID-19–related topics.

Results: A COVID-19 public health social listening taxonomy was developed, validated, and applied to filter relevant content for more focused analysis. A weekly analysis of public online conversations since March 23, 2020, enabled quantification of shifting interests in public health–related topics concerning the pandemic, and the analysis demonstrated recurring voids of verified health information. This approach therefore focuses on the detection of infodemic signals to generate actionable insights to rapidly inform decision-making for a more targeted and adaptive response, including risk communication.
Conclusions: This approach has been successfully applied to identify and analyze infodemic signals, particularly information voids, to inform the COVID-19 pandemic response. More broadly, the results have demonstrated the importance of ongoing monitoring and analysis of public online conversations, as information voids frequently recur and narratives shift over time. The approach is being piloted in individual countries and WHO regions to generate localized insights and actions; meanwhile, a pilot of an artificial intelligence–based social listening platform is using this taxonomy to aggregate and compare online conversations across 20 countries. Beyond the COVID-19 pandemic, the taxonomy and methodology may be adapted for fast deployment in future public health events, and they could form the basis of a routine social listening program for health preparedness and response planning.

(JMIR Infodemiology 2021;1(1):e30971) doi: 10.2196/30971

KEYWORDS
infodemic; COVID-19; infodemic management; social listening; social monitoring; social media; pandemic preparedness; pandemic response; risk communication; information voids; data deficits; information overload

Introduction

Background

Since the beginning of the COVID-19 pandemic, digital communication and social networking have supported the rapid growth of real-time information sharing about the virus that causes COVID-19 (SARS-CoV-2) and the disease in the public domain and across borders. The breadth of conversation, diversity of sources, and polarity of opinions have sometimes resulted in excessive information, including false or misleading information, in digital and physical environments during an acute public health event: this can lead to confusion and risk-taking behaviors that can harm health, trust in health authorities, and the public health response [1]. The excess of information can amplify and protract outbreaks, and it can reduce the effectiveness of pandemic response efforts and interventions.

To address this challenge, the World Health Organization (WHO) Information Network for Epidemics (EPI-WIN), in collaboration with digital research partners, developed a methodology for weekly analysis of digital social media data to identify, categorize, and understand the key concerns expressed in online conversations [2]. The application of this methodology provided the WHO with week-on-week analysis for the prioritization of actions to address online information voids and sources of confusion using verified health information as part of ongoing emergency response planning. When there is a lack of quality information about topics of concern for online users, these topics can be quickly filled with conjecture, low-quality health information, and viral misleading content [3,4], thus potentially causing harm to communities. This approach therefore focuses on the detection of infodemic signals—identifying or predicting rising areas of concern and information voids in the online information ecosystem on a weekly basis to generate actionable insights to rapidly inform decision-making for a more effective response, including adapting risk communication [5].

Infodemic Management During a Health Emergency

Previous research has explored the use of data produced and consumed on the web to inform public health officials, agencies, and policy—a concept known as infodemiology [6]. Initially, the concept of infodemiology aimed to identify the gap between expert knowledge and public practice [7], and it has since evolved to detect and analyze health information on the web through publicly shared search queries, blogs, websites, and social media posts.

The design of interventions for infodemic response must account for an ecosystem where information flow online can cause public health harm offline. Metrics and frameworks related to digital information flows and online behavior are most useful to practitioners when they can be coupled with other online and offline sources of public health data that inform public health decision-making. The WHO has therefore expanded the concept of infodemiology into a multidisciplinary scientific field that amalgamates cross-disciplinary and mixed methods approaches designed to inform the health emergency response [8].

Health emergencies give rise to information overload, which has been shown to influence people’s risk perceptions and protective actions during health emergencies [9]. Overload of information of variable quality, timeliness, and relevance is strongly associated with people’s experience of information anxiety, which in turn can give rise to information avoidance. Recent examples, from HIV to Ebola virus to Zika virus to polio, have demonstrated the high cost to public health and health systems when misinformation sows distrust, exacerbated by ineffective public health communication and community engagement [3,10]. A lack of active community collaboration in the health response early on deepened distrust, especially as these epidemics unfolded. Currently, most emergency and outbreak recommendations emphasize the value of listening to communities, involving them early in the response, and communicating clearly with them in a timely manner [11,12].

Health authorities therefore not only face the challenge of providing relevant, high-quality health information but also must provide it at the right time, in the right format, and with collaborative engagement of communities [13]. Social listening can help overcome barriers to acceptance of high-quality health information and enactment of healthy behaviors by enabling better understanding of community questions, confusion, information seeking, or intensified attention for given topics. Critical information voids can be identified and characterized in both the online and offline information ecosystems. Our research focuses on the identification and characterization of points of confusion, harmful narratives, and key questions that can reveal information voids in the online social media space.
during a health emergency, thereby adding analytical methods to the field of infodemiology that are practical and can directly inform the public health response during a health emergency.

**Analytical Approaches and Metrics To Date**

The rise of social media platforms has generated a readily available source of real-time data related to what people express and share in online communities. The 2009 H1N1 influenza pandemic was the first pandemic to occur in the era of social media and was one of the earliest outbreaks informed by analysis of online conversations and information-seeking behaviors. The previous pandemic offers a case study that evidences how online social listening has been used to follow rapidly evolving public sentiment, track actual disease activity, and monitor the emergence of misinformation [14-16]. Although social media platforms have been used to quantify public concerns and sentiment and to monitor real-time pandemic data, they have also been identified as a medium that can enable the spread of low-quality information. For example, within health emergencies, false information has been shown to be posted twice as frequently as evidence-based information, although it is retweeted less frequently [17]. Provision of targeted, relevant, timely, understandable, and resonant health information can therefore benefit from upstream infodemic management activities of public health authorities, including more robust social listening programs.

The onset of the COVID-19 pandemic has exacerbated concerns about misinformation. Throughout the pandemic, there has been a demand for information; at first, this demand was for information about the origin of the virus, and now it is focused mainly on the response to the virus, particularly vaccination and wider public health and social measures. Similar to information voids [5], COVID-19 misinformation trackers have defined the concept of data deficits in the online space when there are “high levels of demand for information about a topic, but credible information is in low supply” [3]. The issue with a lack of quality information is that the conversation space can be much more readily filled by misinformation, which may be easier to create and share, more emotive (resonant), and better promoted by content promotion algorithms than factual health information.

Despite the influx of studies as to how information is being spread and shared in the era of COVID-19 and how information is influencing people’s health practices, major gaps in knowledge remain as to how best to monitor, understand, and respond to it [8]. Among many possible solutions, social listening, content pretesting, and other computational social science methods have been identified as ways to detect and analyze information voids and viral misinformation narratives [13]. Misinformation research has focused on social media platforms with easier access to data, such as Twitter and YouTube [18,19]; however, misinformation is prevalent across the digital ecosystem (as well as offline). Culture and access to the internet can also affect the nature of misinformation and how it spreads [20]. Beyond identifying what misinformation looks like, studies have also attempted to identify how it emerges [21], aligning with the concept of information voids. Although social listening has tended to focus on spotting myths and rumors, as well as content items with high engagement and reshare rates, the methodology introduced in this paper expands the scope of social listening and positions it as a core practice of emergency response. This includes prioritizing detection of information voids for more proactive infodemic management before these gaps in understanding are filled with more speculation, misleading information, and counterproductive narratives.

Detecting viral misinformation narratives and information voids in real-time data is crucial to a rapid, comprehensive response by authorities for effective delivery of health information to populations during a health emergency, although this does not ensure that people will necessarily act in accordance with that health information. Previous research has evaluated the correction of misinformation and the role of individuals versus organizations in using real-time data [22-24]. The pathway from receiving information, to intent, to action is understudied and a priority area for future research [8]. Evaluation of intervention impact is challenging [4], but evaluation of interventions must be integrated as part of adaptive infodemic management, including social listening.

Interventions need to address the different aspects of the information ecosystem that influence the spread and health impact of an infodemic. For platforms, content moderation policies, modification of content promotion algorithms, and designing for friction can discourage sharing of misinformation and unverified information [25], while supporting literacies such as health, media, information, digital, and data literacies can promote resilience [26]. The literature highlights the value of a multilayered approach for addressing infodemics at various levels in the digital information ecosystem. However, although public health authorities can influence and interact with the other participants in this space, there is a need to suggest immediate and practical tools that public health authorities can deploy within their mandate in a health emergency context in support of their health operations and communication activities [8].

**A Need for Practical Tools for Health Authorities**

Research is ongoing to assist policy makers in understanding public concerns and sentiment around the pandemic as well as in tracking information outbreaks and the emergence of misinformation. However, there is little to no empirical evidence on how this research can be used to develop practical tools for an outbreak response by public health authorities. More collaborations between researchers and public health practitioners are needed to fill this gap. As a contribution to the infodemic response toolbox, the taxonomy and methodology in this study offer a practical, structured approach for identifying information voids and narratives of concern that warrant attention and action. This approach has already provided actionable data to help the WHO focus its efforts for the COVID-19 pandemic response.

https://infodemiology.jmir.org/2021/1/e30971

Purnat et al

JMIR Infodemiology 2021 | vol. 1 | iss. 1 | e30971 | p.46

(page number not for citation purposes)
Methods

Development of a Public Health Taxonomy for Social Listening

A social listening taxonomy for COVID-19 conversations was developed specifically for this analysis. It was designed to filter digital content referring to COVID-19 (and synonyms) for items of relevance in a public health context and to classify that content into categories. The taxonomy consisted of 35 keyword-based searches (one set of searches for each of two languages, namely English and French) which were grouped into 5 overarching topic categories representing thematic areas in which people were engaging, writing, or searching for information.

The 5 top-level categories and corresponding 35 subcategories of this social listening taxonomy for COVID-19 conversations were defined based on established epidemic management and public health practices during an outbreak of infectious disease [27] (Figure 1). The first 4 categories refer to the focus of epidemic management activities during the pandemic: (1) the cause of the disease—what do we know about the virus, and how is it spreading? (2) the illness—what are the symptoms, and how is it transmitted? (3) the treatment—how can it be cured? and (4) the interventions—what is being done by authorities and institutions? In addition, a fifth category was included to examine public perceptions on circulating information (ie, metaconversations about evidence and statistics, mis- and disinformation, successful and harmful content, or key influencers who have been actively amplifying information on COVID-19). This category was designed because misinformation, rumors, and polarization of factual versus misleading narratives are common challenges in epidemic management.

![Figure 1. Structure of the social listening taxonomy for COVID-19 conversations.](https://infodemiology.jmir.org/2021/1/e30971)

Each of these 5 categories were segmented into subcategory levels that are familiar to the epidemiologist’s investigation and management of the outbreak, resulting in a total of 35 taxonomy subcategory levels (Figure 1). For example, the taxonomy category about the illness was further defined by subcategories to identify conversations, questions, or confusion about the symptoms of the illness, how it transmits, and what populations may be affected by it (across demographics, vulnerable populations, and people with underlying conditions). By defining the social listening taxonomy across the investigation areas of epidemic management [27], the resulting infodemic insights can be more quickly evaluated by public health professionals and turned into actionable recommendations to inform the epidemic response.

Each of the 35 taxonomy subcategories encompassed a list of topics that captured different aspects of that segment of the online conversation on COVID-19. Keywords for the 35 subcategory searches were generated based on expert knowledge from the WHO EPI-WIN team and translated into Boolean search strings to identify topic-related language for review of relevant social media posts and news content. The keywords of the taxonomy are available on request at the contact address listed in the Acknowledgments section.
In addition to the taxonomy subcategory levels, the keyword-based Boolean search string was created to also identify posts containing a question; this enabled analysis of categories for which people were seeking information and, therefore, potential information voids. The question search string was designed to be paired with each of the 35 taxonomy Boolean search strings to identify posts referring to the topic and containing question words, verb-subject inversions, and auxiliaries.

Finally, the sum of the total volume of the social media conversation (on all topics) was estimated by monitoring the number of posts mentioning at least one of the most commonly used words in English (eg, the, and, or, I) and French (eg, le, la, ou, et). The data were collected via a Boolean search string comprised of these most commonly used words.

Data Sources and Data Collection

The analysis was based on the weekly aggregation of publicly available social media data in English and French using Meltwater Explore. Institutional Review Board review was not sought, as the analysis used large bodies of text written by humans on the internet and on some social media platforms. The analysis and resulting reports focused on the identification of conversation narratives and thematic questions instead of on individual statements and users.

The Meltwater social listening platform was configured to collect verbatim mentions of keywords associated with the 35 predefined taxonomy category Boolean searches from 9 open data sources and fora (Twitter, blog entries, Facebook, Reddit posts and comments, other unspecified message boards or fora, comments under news articles and blog entries, Instagram posts, product reviews, and YouTube video titles and comments). A total of 87.02% of the resulting analysis data set was sourced from Twitter. Blogs (5.34%) and, specifically, the Reddit platform (4.34%) were the next most prominent sources in the data set. These were followed by message boards (2.14%), comments under news articles (0.89%), online review websites (0.13%), Instagram (0.12%), and Facebook (0.03%).

For each of the 35 taxonomy subcategories, the daily total volume of posts, and the volume of posts posing a question, were recorded on a weekly basis. Tracking changes in volume from week to week also enabled determination of the velocity for a given subcategory.

Testing and Validation of the Taxonomy

The methodology used to test and validate the retrieval and classification in this study used both retrieval precision and retrieval recall, which are related to how much retrieved data is relevant and how much relevant data is retrieved, respectively [28,29]. These validation metrics are useful for assessing the performance of machine learning models in information retrieval and have been used for metrics on content retrieved and classified via Boolean searches for news media content [28] and Twitter data [29].

To test whether the taxonomy categories captured the intended information (retrieval precision), a random sample of content captured by each of the 35 Boolean searches was human-coded for relevance (10,500 posts in total) by a single reviewer, with a second reviewer validating the coding. The post was coded as either relevant to the search subcategory (1) or not relevant (0).

The aim of the coding was to determine the proportion of relevant (R; also, “true positive” [TP]) results as a percentage of the retrieved sample. The coders judged whether a post was relevant according to the intended definition of the specific subcategory search for which the post was returned. For example, if a post had been returned for the “The Illness – Confirmed Symptoms” search, the coder would check if the post referred to a confirmed symptom of COVID-19 (TP) or whether the matched keywords were mentioned in a different context (irrelevant [I]; also, false positive [FP]). For instance, if a post had been returned by the Boolean search for COVID-19 vaccines, did the post refer to COVID-19 vaccinations? If yes, the post was coded as a TP. If the post in question mentioned COVID-19, but the part of the post mentioning vaccines was about the influenza vaccine, the post would be coded as an FP.

The initial retrieval precision testing showed an average result of 82% for the 35 taxonomy subcategory searches. The retrieval precision rate was calculated as precision = [TP ÷ (TP + FP)] x 100%.

A total of 7 searches returned content below the target minimum retrieval precision rate of 70%, with a range of 42% to 100% (Table 1). To reduce the rate of false positives, the keywords for the 7 searches that performed below the target minimum rate were subsequently reviewed and updated to exclude keywords and phrases returning irrelevant content. On retesting, the average retrieval precision rate for the 35 searches was 87%, with a range of 72% to 100%. The full results of the retrieval precision testing and subsequent retesting can be seen in Table 1.

To spot-check the coding for reliability, we deployed a second reviewer to analyze 10% of the posts (30 per taxonomy category search, 1500 in total). We calculated the Cohen kappa to determine intercoder reliability, which was found to be high (κ=0.81, observed agreement [p_o]=0.95, expected agreement [p_e]=0.76).

A further test was performed to assess retrieval recall: whether content of relevance to the research aims failed to be retrieved by the taxonomy searches. To test this, a random sample of 1000 items of content, mentioning COVID-19 (and synonyms) but excluding the taxonomy category keywords (the “not retrieved” sample in Table 2), was human-coded for relevance from a public health perspective. Posts in this sample were determined by the coder to be relevant (R) to the aims of the public health research (false negative [FN]), or irrelevant (I) to the research aims (true negative [TN]). Coding was performed by the same reviewer and was binary; content was irrelevant (I, and therefore also TN) or was relevant (R, and therefore also TP) or was relevant (R) and therefore also TN) or was relevant (R, and therefore also TP) or was relevant (R, and therefore also TN) or was relevant (R, and therefore also TN).
Table 1. Results of retrieval precision testing and retesting with a sample size of 300 posts analyzed per subcategory.

<table>
<thead>
<tr>
<th>Subcategory</th>
<th>Posts retrieved by the taxonomy category search human-coded as true positives and retrieval precision rate, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Cause – The Cause</td>
<td>217 (72.3)</td>
</tr>
<tr>
<td>The Cause – Further Spread – Stigma</td>
<td>260 (86.7)</td>
</tr>
<tr>
<td>The Cause – Further Spread – Immunity</td>
<td>245 (81.7)</td>
</tr>
<tr>
<td>The Illness – Confirmed Symptoms</td>
<td>189(a) (63)/239(b) (79.7)</td>
</tr>
<tr>
<td>The Illness – Other Discussed Symptoms</td>
<td>141(a) (47)/218(b) (72.7)</td>
</tr>
<tr>
<td>The Illness – Asymptomatic</td>
<td>300 (100)</td>
</tr>
<tr>
<td>The Illness – Presymptomatic</td>
<td>300 (100)</td>
</tr>
<tr>
<td>The Illness – Means of Transmission</td>
<td>295 (98.3)</td>
</tr>
<tr>
<td>The Illness – Protection From Transmission</td>
<td>299 (99.7)</td>
</tr>
<tr>
<td>The Illness – Underlying Conditions</td>
<td>238 (79.3)</td>
</tr>
<tr>
<td>The Illness – Demographics – Sex</td>
<td>215 (71.7)</td>
</tr>
<tr>
<td>The Illness – Demographics – Age</td>
<td>215 (71.7)</td>
</tr>
<tr>
<td>The Illness – Vulnerable People</td>
<td>287 (95.7)</td>
</tr>
<tr>
<td>The Illness – Vulnerable Communities</td>
<td>269 (89.7)</td>
</tr>
<tr>
<td>Treatment – Vaccines</td>
<td>300 (100)</td>
</tr>
<tr>
<td>Treatment – Current Treatment</td>
<td>144(a) (48)/224(b) (74.7%)</td>
</tr>
<tr>
<td>Treatment – Research &amp; Development</td>
<td>290 (96.7)</td>
</tr>
<tr>
<td>Treatment – Nonproven Treatment (Nutrition)</td>
<td>245 (81.7)</td>
</tr>
<tr>
<td>Treatment – Myths</td>
<td>126(a) (42)/221(b) (73.7)</td>
</tr>
<tr>
<td>Interventions – Measures in Public Settings</td>
<td>243 (81)</td>
</tr>
<tr>
<td>Interventions – Testing</td>
<td>280 (93.3)</td>
</tr>
<tr>
<td>Interventions – Supportive Care – Equipment</td>
<td>204(a) (68)/257(b) (85.7%)</td>
</tr>
<tr>
<td>Interventions – Supportive Care – Health Care</td>
<td>289 (96.3)</td>
</tr>
<tr>
<td>Interventions – Personal Measures</td>
<td>298 (99.3)</td>
</tr>
<tr>
<td>Interventions – Reduction of Movement</td>
<td>256 (85.3)</td>
</tr>
<tr>
<td>Interventions – Protection</td>
<td>276 (92)</td>
</tr>
<tr>
<td>Interventions – Technology</td>
<td>278 (92.7)</td>
</tr>
<tr>
<td>Interventions – Travel</td>
<td>250 (83.3)</td>
</tr>
<tr>
<td>Interventions – Faith</td>
<td>201(a) (67)/269(b) (89.7)</td>
</tr>
<tr>
<td>Interventions – Unions and Industry</td>
<td>223 (74.3)</td>
</tr>
<tr>
<td>Interventions – The Environment</td>
<td>183(a) (61)/290(b) (96.7)</td>
</tr>
<tr>
<td>Interventions – Inequalities</td>
<td>280 (93.3)</td>
</tr>
<tr>
<td>Interventions – Civil Unrest</td>
<td>280 (93.3)</td>
</tr>
<tr>
<td>Information – Misinformation</td>
<td>273 (91)</td>
</tr>
<tr>
<td>Information – Statistics</td>
<td>244 (81.3)</td>
</tr>
</tbody>
</table>

\(a\)Indicates a taxonomy subcategory search that performed below minimum requirements and was subsequently updated and retested to yield better performance.

\(b\)Number and percentage of posts in the sample coded as true positives in the retesting of the taxonomy subcategory search following the update.
The results of the coding of the “not retrieved” sample indicated the proportion of TN results as a proportion of the sample; 70% of content was judged not to be relevant to the research aims, and therefore it was deemed correct that this content was not retrieved by our taxonomy. From the data, we also calculated the retrieval recall rate as \( \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \).

The overall retrieval recall rate was 74%. This coding process enabled identification of areas where the existing Boolean string could be expanded to include more relevant keywords to retrieve more relevant content, or where the taxonomy could be expanded to include new and emerging issues. From the content that was not retrieved but was judged to be of potential relevance to the research aims (false negatives, FNs), 3 topics were identified that will be added to the taxonomy in a pending update: mutations/variants of the COVID-19 virus; “long covid” (long-term symptoms of COVID-19); and the impact of the pandemic on mental health and well-being.

To validate the coding of the sample of “not retrieved” content for reliability, we deployed a second reviewer to analyze 10% of the posts (100 posts). We calculated the Cohen kappa to determine intercoder reliability, which was found to be high (\( \kappa = 0.86, p < 0.01 \), \( p_c = 0.93, p_c = 0.50 \)).

From the results of the coding of the retrieved and unretrieved samples, we calculated an F1 score and an F0.5 score with the following formulas: \( F_1 = \frac{(2 \times \text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \) and \( F_{0.5} = \frac{(1.25 \times \text{precision} \times \text{recall})}{(0.25 \times \text{precision} + \text{recall})} \).

The F1 score (harmonic mean of precision and recall) for the searches was 0.80, and the F0.5 score was 0.84. F1 and F0.5 scores range from 0 to 1, with 1 representing perfect performance. A higher F1 or F0.5 score is considered reasonable, with a score closer to 1 indicating stronger performance of a retrieval and classification approach. The inclusion of the F0.5 measure reflects the greater importance of retrieval precision in this study: given the vast number of potentially relevant pieces of content, it is more important to the aims of this project to correctly classify the retrieved posts than to collect every possibly relevant post. Therefore, we consider it a positive result to achieve a higher F0.5 score than F1 score. This is because in this study, it is more important that the results are not impacted by a high number of false positives and that the true positives are classified into the correct subcategory. The retrieval recall testing is also helpful because it enables identification of new or changing pandemic issues, such as new terminology being used that can be added to the taxonomy category search language over time.

### Quantitative Data Analysis

Potential information voids were identified based on 3 parameters within the weekly data set: the volume (ie, how many social media items referred to topic X?), the velocity (ie, the rate of increase of the number of social media items that have engaged with topic X over the course of the past week), and the presence of questions about the topic. The volume was the sum of the online items that mentioned COVID-19 together with a keyword related to each tracked topic. Velocity was determined as the percentage increase of the volume of content items aggregated under each topic from week to week, where velocity = (current week’s total number of mentions – previous week’s total number of mentions) \( \div \) (previous week’s total number of mentions) \( \times \) 100%.

Starting in late March 2020, weekly global analysis reports were produced that supplied the EPI-WIN team with early warnings of points of concern expressed in public comments by online users [2,4]. By May 4, 2021, the data sample consisted of a sum of 1.02 billion unique social media posts. This was a subset of the larger pool of 1.3 billion total public social media posts in English and French mentioning COVID-19 gathered by the data aggregator. The sample of 1.02 billion posts consisted of approximately 3% of the pool of all public social media posts written in English and French that had been gathered by the data aggregator since March 2020. The data set of total public social media posts gathered by the aggregator was verified through the automated search of mention of the most common words in English and French (eg, the, le, and, et).

Each week, social media conversations were segmented based on levels of velocity and quantitatively examined for public engagement (eg, likes, shares, poll votes, reactions), hashtags, and most-used keywords and phrases. From this weekly quantitative data, up to 10 topics with high velocity and/or a large proportion of social media posts expressing a question, and/or with high levels of engagement, were identified as potential priority information voids or sources of confusion or concern.

The identified issues on social media were then further evaluated using engagement data and Google search trends to determine whether a significant number of online users had also been looking for information on these topics to help determine whether the information void was more widespread.

---

**Table 2. Results of human coding of retrieved and unretrieved samples for calculation of retrieval recall and F-scores.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Coded relevant Samples, n</th>
<th>Description</th>
<th>Coded irrelevant Samples, n</th>
<th>Description</th>
<th>Total coded sample, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieved(^a)</td>
<td>875</td>
<td>True positive</td>
<td>125</td>
<td>False positive</td>
<td>1000a</td>
</tr>
<tr>
<td>Not retrieved</td>
<td>304</td>
<td>False negative</td>
<td>696</td>
<td>True negative</td>
<td>1000</td>
</tr>
<tr>
<td>Total</td>
<td>1179</td>
<td>N/A(^b)</td>
<td>821</td>
<td>N/A</td>
<td>2000</td>
</tr>
</tbody>
</table>

\(^a\)The “retrieved” sample size was downweighted to equal the “not retrieved” sample size.

\(^b\)N/A: not applicable.
Qualitative Analysis

Each week, we used the quantitative analysis to identify up to 10 topics reflecting potential information voids and areas of concern. These topics were then examined in more detail via qualitative analysis to understand the context and identify where action may need to be taken in line with a sequential explanatory design approach [30]. The qualitative analysis involved ad hoc human-led review of the key narratives, influencers, and public reactions as reflected in the content.

This analysis prioritized the flagging of widespread confusion or frequently asked questions, the rapid amplification of misinformation, or ad hoc aspects of the conversation that were particularly relevant to public health, such as vaccine questioning ahead of and during a vaccination campaign.

Reporting

The quantitative data were compiled in a web-based dashboard accessible to the emergency responders in the EPI-WIN team, and insights were discussed with EPI-WIN emergency responders on a weekly basis. The dashboard was updated weekly to allow investigation of short- and long-term trends in volumes, changes in velocity, and the volume of questions for each topic.

Weekly written reports outlined quantitative and qualitative findings about the 5 to 10 topics of concern, included visualizations from the dashboard, and summarized recommendations for action when needed [31].

Results

Quantitative analysis of the volume changes indicated that the narratives and questions in the online conversations shifted as the pandemic evolved over the course of 2020 and into 2021 (Table 3). Based on the average weekly rises of the topics within each of the 5 taxonomy categories in the yearly quarters between March 23, 2020, and March 31, 2021, it was observed that the second quarter (Q2) and third quarter (Q3) of 2020 were characterized by a steady increase in conversations about “the interventions.” Although discussion of “the illness” decreased in 2020, it surged again in the first quarter (Q1) of 2021. In the fourth quarter (Q4) of 2020, “the treatment” had the highest velocity in digital conversations, while the metaconversation on COVID-19 information experienced the greatest velocity in Q1 of 2021.

Table 3. Most discussed topics by month and results of the pivoted data set by month, sorted by volume of social media mentions.

<table>
<thead>
<tr>
<th>Year and month</th>
<th>Most discussed topic</th>
<th>Volume (millions of social media mentions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March</td>
<td>Interventions – Testing</td>
<td>37</td>
</tr>
<tr>
<td>April</td>
<td>Interventions – Testing</td>
<td>18</td>
</tr>
<tr>
<td>May</td>
<td>Interventions – Measures in Public Settings</td>
<td>12</td>
</tr>
<tr>
<td>June</td>
<td>Interventions – Testing</td>
<td>11</td>
</tr>
<tr>
<td>July</td>
<td>Interventions – Testing</td>
<td>14</td>
</tr>
<tr>
<td>August</td>
<td>Interventions – Testing</td>
<td>9</td>
</tr>
<tr>
<td>September</td>
<td>Interventions – Testing</td>
<td>8</td>
</tr>
<tr>
<td>October</td>
<td>Interventions – Testing</td>
<td>17</td>
</tr>
<tr>
<td>November</td>
<td>Interventions – Testing</td>
<td>8</td>
</tr>
<tr>
<td>December</td>
<td>Treatment – Vaccines</td>
<td>15</td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>January</td>
<td>Treatment – Vaccines</td>
<td>15</td>
</tr>
<tr>
<td>February</td>
<td>Treatment – Vaccines</td>
<td>12</td>
</tr>
<tr>
<td>March</td>
<td>Treatment – Vaccines</td>
<td>15</td>
</tr>
<tr>
<td>April</td>
<td>Treatment – Vaccines</td>
<td>15</td>
</tr>
</tbody>
</table>

At the same time, topics re-emerged periodically in terms of popularity. All 35 categories of topics that were tracked resumed a higher velocity throughout the reporting period for an average of 18 weeks combined (Table 4). The 2 topics that attracted increasing interest most frequently were “myths” and “risk based on age demographics” (rising for 26 and 24 weeks, respectively) followed by “the cause” of the virus and “reduction of movement” (both 23 weeks) “vaccines” and “stigma” (both 22 weeks), and “other discussed symptoms” (21 weeks). Digital conversations on “the cause” of the epidemic, “misinformation” as a phenomenon, and “immunity” had the longest continuous periods of surge in volume of social media posts discussing these topics in the context of the COVID-19 pandemic; the conversation on “the cause” increased in both the first and second half of the analysis period for 7 continuous weeks during the first half of the reporting period, while the metaconversation about misinformation increased for 6 consecutive weeks. Conversations about “immunity” increased for 5 consecutive weeks in June-July 2020 and in November-December 2020.
Table 4. Frequency of weekly velocity growth (number of weeks in which a topic experienced positive velocity) and average weekly increase rate (or decrease, when a negative value is returned) by topic.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Number of weeks in which a topic experienced positive velocity (increase of social media mentions since previous week)</th>
<th>Average weekly increase in number of social media mentions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment – Myths</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td>The Illness – Demographics – Age</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>Interventions – Reduction of Movement</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>The Cause – The Cause</td>
<td>23</td>
<td>7</td>
</tr>
<tr>
<td>Vaccines</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>The Cause – Further Spread: Stigma</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Interventions – Faith</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>The Illness – Other Discussed Symptoms</td>
<td>21</td>
<td>52</td>
</tr>
<tr>
<td>Treatment – Current Treatment</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Interventions – Travel</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Interventions – The Environment</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>The Illness – Confirmed Symptoms</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>The Illness – Asymptomatic transmission</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>The Illness – Means of Transmission</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>Interventions – Measures in Public Settings</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>The Cause – Further Spread: Immunity</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>Treatment – Nonproven Treatment (Nutrition)</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Information – Statistics and Data</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Interventions – Technology</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>Information – Misinformation</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>The Illness – Vulnerable Communities</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>Interventions – Testing</td>
<td>17</td>
<td>-1</td>
</tr>
<tr>
<td>Interventions – Supportive Care – Health Care</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>Interventions – Protection</td>
<td>17</td>
<td>-3</td>
</tr>
<tr>
<td>The Illness – Presymptomatic</td>
<td>16</td>
<td>40</td>
</tr>
<tr>
<td>The Illness – Underlying Conditions</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>Interventions – Supportive Care – Equipment</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>The Illness – Protection From Transmission</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Interventions – Personal Measures</td>
<td>15</td>
<td>-3</td>
</tr>
<tr>
<td>The Illness – Demographics – Sex</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Treatment – Research &amp; Development</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Interventions – Unions and Industry</td>
<td>15</td>
<td>-1</td>
</tr>
<tr>
<td>Interventions – Inequalities</td>
<td>14</td>
<td>-1</td>
</tr>
<tr>
<td>The Illness – Vulnerable People</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>Interventions – Civil Unrest</td>
<td>14</td>
<td>32</td>
</tr>
</tbody>
</table>

Analysis of the peaks in discussion of 2 of the leading recurring topics, “risk related to age demographics” and “the cause,” provided insight into how narratives around these topics were fueled by real-life events. The conversation on “risk related to age demographics” increased in velocity 24 times throughout the period studied. A total of 3 million public social media posts engaged with the topic: 64% of these posts were focused on children, whereas 30% focused on older people. The volume of conversation on children and COVID-19 risk increased above the yearly average for 133 days. Speculation about the severity of COVID-19 infection in children was raised consistently throughout the evaluation period, and it represented fertile
ground for confusion and potential misinformation. Major triggers included news reports of child deaths (560,000 public posts discussed children and mortality), reports of symptoms observed in children in particular (300,000 public posts discussed children and symptoms), the debate over school reopenings, particularly with regard to transmission (818,000 public posts) and, most recently, COVID-19 immunization (656,000 public posts). In relation to this topic, doubts resurfaced repeatedly about the threat of COVID-19 to children; however, there was a diversity of narrative foci for these doubts, linked to changing events during the pandemic.

By contrast, public discussion on the possible origins of the pandemic (“the cause”) had a persistent narrative throughout the evaluation period. “The cause” of the epidemic was a focus of 3.26 million public social media posts throughout the period monitored. The size of the conversation was most prominent at the beginning of the pandemic and diminished as of June 2020, but with periodically recurring peaks in the number of posts. Conspiracy theories suggesting the artificial origin of the virus as a bioweapon were persistent in online discussion, and prominent influencers operating in the conspiracy theory space were often linked to resurgence peaks in public online discussion. The phrase “biological weapon” was mentioned 326,000 times in the public social media space (cf. 141 million mentions of COVID-19 vaccines in the same period). The rate of mentions decreased by 65% from Q2 to Q3 2020 (as it decreased to 34,000 mentions globally), but it surged to 110,000 in Q4 as the theory regained prominence in the public discourse, in part driven by the release of a preprint paper claiming that the virus was an “unrestricted bioweapon” [31,32]. In Q4 2020, 16% of posts referring to the virus as a bioweapon referenced the authors of that paper. Although the nature of the narrative around COVID-19 as a “bioweapon” was relatively constant, our findings indicate that existing conspiracy theories can be fueled with new details in debates about science [32], underscoring a need to improve science literacy and communication.

**Discussion**

**Principal Findings**

The insights obtained in this study have afforded public health experts the opportunity for a more rapid and targeted assessment of a subsample of narratives across the English and French languages using public digital sources. These insights can be combined with others to better understand whether and how people are understanding public health and social measures and putting them into practice to protect themselves and their communities.

The application of this taxonomy to successive weekly online social listening analysis has resulted in a better understanding of the evolution and dynamics of high-velocity conversations about COVID-19 worldwide in English and French during the pandemic. The taxonomy also provides a quantifiable approach to support more adaptive and targeted planning and prioritization of health response activities. For example, monitoring and characterizing re-emerging topics can guide re-evaluation and updating of risk communication and community engagement initiatives to improve understandability and resonance, or highlight where adjustments in technical guidance, public health policy, and social measures may be needed. In addition, the fact that narratives discussed online often overlap across different categories reveals the breadth of this taxonomy, and this overlap enables emerging narratives and potential information voids to be picked up through velocity alerts raised in different elements of the taxonomy.

The testing process described in this article forms the basis of the taxonomy review and maintenance process. Updates to the taxonomy are also informed by observations from the weekly analysis and reporting of the data, and public health expert knowledge via WHO, the wider news agenda, and epidemic management context of the pandemic. The taxonomy has been updated twice since its creation in March 2020, with a third update forthcoming in 2021. The aim of the taxonomy updates is to ensure that important new and emerging topics are captured as the pandemic evolves (as in the examples of variants/mutations and “long covid” above) and that the taxonomy includes the latest language and terms being used by the public [29]. For example, as the pandemic progressed, members of the public increasingly dropped the use of formal terms, such as referring to the virus as “Covid” rather than “COVID-19”; therefore, the taxonomy keywords were expanded to reflect this change. When the taxonomy was updated and validated, the database was also updated back to the start date of the research to ensure consistency in the analysis data set and to allow for analysis of long-term trends.

There is added value in using a common social listening taxonomy for integration of insights from a variety of data sources and research methods in online and offline communities. This can provide a more systematic way to integrate analysis of different data sources and facilitate complementarity of digital social listening data with other data such as knowledge, attitudes, and practices research to help uncover drivers of online discussion, and to support social listening in vulnerable or more marginalized communities, including those with limited access to online platforms.

A challenge of this analysis approach is the need for human analysts to continuously monitor and evolve the taxonomy in line with the developing narratives and emerging topics as well as the changing language used in discussions of the COVID-19 pandemic. Ideally, taxonomies would be tested, reviewed, and updated frequently, particularly when a new stage of the pandemic begins (eg, when the vaccine rollout started), as such events in the pandemic timeline can generate new topics of discussion and new terminology (eg, “Covid passports”). However, the benefit of more frequent updates is balanced by the need for comparability of data across time as well as by the fact that this analytical method needs to be rapidly reproducible, including in more resource-constrained environments, to have real, practical use week-on-week during the pandemic to inform the immediate needs of the health authority response, including risk communication and community engagement in any country context.

To help identify actionable insights, the weekly analysis was focused less on exact counts of mentions and more on relative changes, narratives, and topic signals to evaluate and
contextualize infodemic signals. When rapidly identifying up to 10 information voids in large weekly data sets, absolute precision was less important than the early detection of an actionable signal to help trigger a timely response. For example, if there was a sudden rise in online narratives expressing concern over a treatment, coupled with other information available from the emergency response, the exact number of mentions was less important than signal detection, analysis, and recommendations for possible action. Despite this, more research is needed to refine and streamline the process for rapidly updating and publishing such taxonomies, especially in protracted epidemics, where shifts in concerns and conversations are bound to occur.

A key takeaway from the analysis that can be applied during the current pandemic is the frequent recurrence of topics of concern and its implications for communication. Public health authorities, governments, and nongovernmental organizations must be prepared to communicate repeatedly on the same issue, adapting frames, approaches, and content as public perceptions of issues and topics shift. Our analysis shows that areas of concern wax and wane, with confusion disappearing and re-emerging as new information comes to light or new events occur. Monitoring the changing narratives on a weekly basis and over time using a taxonomy, such as the one used in this study, can enable health authorities to assess longer-term trends and to be more nimble in adapting approaches to respond effectively to topics of concern and to counter misinformation. Further research can help to adapt these digital social listening approaches to provide metrics for evaluation of infodemic management interventions.

The taxonomy has been adapted, translated, and applied in a number of country-level studies in Mali, the Philippines, and Malaysia [33-35]. Applying the approach at the country level included the localization of keywords and their validation. Once this work was completed, the taxonomy and methodological approach proved to be a useful tool for generating insights into narratives in public discourse and potential information voids at the national level. Furthermore, the research framework is now being applied in Canada by the National Institute of Public Health of Québec as an input into the public health response and risk communication in that province, showing that the taxonomy is also applicable at the subnational level [36]; Institut national de santé publique du Québec [forthcoming].

A pilot project by WHO EPI-WIN and research partners, Early Artificial intelligence–supported Response with Social Listening (EARS) [37], also built on the taxonomy from this research and applied it to an automated classification of content and analysis of publicly shared opinions and concerns in 20 countries. The EARS project is enabling both country-level analysis and cross-country comparisons of themes in online conversations, although obtaining in-depth contextual insights still requires human-led analysis of potential information voids and sources of confusion. Therefore, more investment in analytical capacities in social listening at the country level is needed to provide more contextual analysis, interpretation of infodemic insights, and formulation of recommendations for action, as well as to build capacities for using social listening for health response evaluation and adaptation.

There is an opportunity to apply the taxonomy and methodology described in this paper to detect information voids during future, as yet unknown, pandemics and other public health crises. The 5 top-level categories and some of the 35 sub-categories are relevant to social listening in any outbreak but would need to be adapted to the type of pathogen. If, for example, the HIV/AIDS epidemic had started in the digital, connected world of 2020 rather than in the 1980s, the online social listening taxonomy structure would have needed some adjustment to filter and segment public discourse related to the epidemic and identify information voids. For example, a “Demographics – Men Who Have Sex With Men” topic could be added under the category “the illness” to better hear questions and concerns from this particular demographic group. This approach could also include adjustments to subcategories under “the intervention” to remove irrelevant subcategories of “Reduction of Movement” and “Unions and Industry.” After such a taxonomy review and adjustment, the keywords used to capture content related to each category and subcategory would also need to be systematically reviewed to ensure they were appropriate to the narratives in relation to specific illness in question. For example, terms relating to injected drug use, sex between men, sex between a man and woman, and mother-to-child transmission could be added under “The Illness – Modes of Transmission.” Having a taxonomy structure and methodology already in place as a starting point would enable faster deployment of digital social listening activities in a future outbreak.

Limitations

Interpretation of the analysis must account for the limitations of the data sources included in the content aggregator. During health emergencies, health authorities require surge support in social listening, response, and evaluation functions. Analysis services from a central analytics unit or from commercial or academic institutions need to be set up quickly to use a systematic approach to detect and understand people’s changing concerns, questions, and possible areas of confusion shared publicly online. The overhead in management of data from open sources can be high, and in settings where the social listening analytics capacity is not yet in place for routine analysis, content aggregators can be used to rapidly set up an analysis workflow. The media content aggregation platform used for this study offers firehose access to Twitter, ensuring a complete set of data for analysis, subject to privacy limitations. Other sources in the platform are either sampled from or limited to public posts only [38]. This is a limitation that applies to most analytics of this type, as Facebook and other social media platforms set limitations on the data they make available due to their privacy policies and commercial interests. As a result, there is an overrepresentation of Twitter content in this analysis [39,40]. The use of private data aggregators may lead to the use of unconventional, uncontrolled samples whose breadth and comprehensiveness are constrained by practical and legal limitations. Other methods would be required to characterize conversations in hidden online communities, closed groups, and closed messaging apps, and thorough consideration of the ethics of social listening would be warranted in such contexts.

This research is global and is limited to two major languages (English and French). As a result, only major online narrative
themes and information voids were identified, and the resulting interpretations may not be representative of trends and patterns that could be observed in digital communities for other languages. Moreover, in a global weekly analysis, smaller or more localized conversations may go undetected. One of the aims of this work is to apply and advance the methods to develop taxonomies that can be rapidly applied to any linguistic context for different geographies and public health events.

It has also been observed that the global English-language data set is prone to overrepresent the voice of social media in geographic regions or communities that are more digitally active than others. A key challenge in this study was the digital amplification of discourse pertaining to US politics, the elections, and the digital prominence of US civil society thereof [41]. In such situations, exclusion keywords may be used to exclude major events or large-scale media coverage from analysis so that they do not mask citizens’ publicly shared narratives that are more relevant for public health authorities. This can also be addressed when presenting the analysis results. For example, the weekly reports presented analysis of the narratives from the United States and the United Kingdom separately from the analysis of data from other countries where English was the language of online conversation. This helped to uncover previously undetected narratives outside the United States and the United Kingdom. Future research is needed to assess how results may vary in different linguistic communities and to evaluate the effects of geographies that may be superinfluencers of global discourse.

Another limitation of this research is the start date of the project, March 23, 2020, which is several weeks after COVID-19 was declared a public health emergency of international concern; however, data prior to this date (back to January 2020) have been retrieved and stored for future analysis, ensuring that it is possible to analyze a longer timeline. Adaptation and application of the taxonomy structure in future outbreaks must also take into account validation of information retrieval and recall. The test scores referenced in the taxonomy testing and validation section should be taken as estimates of the accuracy of the retrieval process by the taxonomy category searches, and function most effectively as a tool for identifying areas for improvement. A key limitation of the test results is that human coders can make errors [29]. The human coders involved in the testing and validation were highly experienced in coding and highly familiar with the topic in question, which can help minimize the incidence of coding errors. Future applications of this validation approach could also deploy more coders in an effort to remove potential bias introduced by reliance on a small number of coders.

**Conclusions**

This research focuses on the identification of potential information voids and sources of confusion in online social conversations to provide actionable insights for risk communication and community engagement and other health response activities. While it can provide insight into the opinions expressed online, integration with other analyses, including from listening to offline communities is needed. Applying this methodology globally has provided the added and needed insight, inspiring new ways of thinking and use of information in support of risk communication during health emergencies. Much of the value of the taxonomy we developed is in the capacity to rapidly deploy and provide ongoing insights about information voids during an outbreak, which then allows a health authority to take evidence-informed action and course-correct risk communication during an epidemic. The application of the taxonomy and methodology for social listening at regional, country, and subnational levels in the COVID-19 pandemic—which is already being tested—offers possibilities for more actionable insights that must increasingly support a localized response. Moreover, this method offers an approach for monitoring of concerns, questions, and information voids in future outbreaks, enabling a faster response by the health authorities in affected countries during the next acute health event.

**Data Availability**

The listing of keywords and search terms per taxonomy subcategory is available upon request by contacting enquiry@mediameasurement.com.

**Acknowledgments**

The authors would like to thank the EPI-WIN team members for their contributions to the ideation on this work, and Bernardo Mariano, WHO Director of Digital Health and Innovation, for supporting and advocating for this digital analytics innovation work. S Bezbaruah, S Briand, CC, JL, AM, TN, TDP, and FT are staff of the WHO. These authors are alone responsible for the views expressed in this paper, and these views do not represent those of their organization.

**Conflicts of Interest**

S Ball, PV, AW, and TZ are employed by a media monitoring company that provides a wide range of services to clients in media monitoring and listening, including the WHO. The work described in this paper was part of the contractual service to the WHO. The other authors have no conflicts to declare.

**References**


13. Purnat T, Wilhelm E. Building systems for respond to infodemics and build resilience to misinformation. LinkedIn. 2020 Dec 02. URL: https://www.linkedin.com/pulse/building-systems-respond-infodemics-build-resilience-tina-d-purnat/ [accessed 2021-07-16]


37. Early AI-supported Response with Social Listening. World Health Organization. URL: https://apps.who.int/ears [accessed 2021-07-16]


**Abbreviations**

**EARS:** Early Artificial intelligence–supported Response with Social Listening  
**EPI-WIN:** World Health Organization Information Network for Epidemics  
**FN:** false negative  
**FP:** false positive  
**pe:** expected agreement  
**po:** observed agreement  
**Q1:** first quarter  
**Q2:** second quarter  
**Q3:** third quarter  
**Q4:** fourth quarter  
**TN:** true negative  
**TP:** true positive  
**WHO:** World Health Organization
Examining the Public’s Most Frequently Asked Questions Regarding COVID-19 Vaccines Using Search Engine Analytics in the United States: Observational Study

Abstract

Background: The emergency authorization of COVID-19 vaccines has offered the first means of long-term protection against COVID-19–related illness since the pandemic began. It is important for health care professionals to understand commonly held COVID-19 vaccine concerns and to be equipped with quality information that can be used to assist in medical decision-making.

Objective: Using Google’s RankBrain machine learning algorithm, we sought to characterize the content of the most frequently asked questions (FAQs) about COVID-19 vaccines evidenced by internet searches. Secondarily, we sought to examine the information transparency and quality of sources used by Google to answer FAQs on COVID-19 vaccines.

Methods: We searched COVID-19 vaccine terms on Google and used the “People also ask” box to obtain FAQs generated by Google’s machine learning algorithms. FAQs are assigned an “answer” source by Google. We extracted FAQs and answer sources related to COVID-19 vaccines. We used the Rothwell Classification of Questions to categorize questions on the basis of content. We classified answer sources as either academic, commercial, government, media outlet, or medical practice. We used the Journal of the American Medical Association’s (JAMA)’s benchmark criteria to assess information transparency and Brief DISCERN to assess information quality for answer sources. FAQ and answer source type frequencies were calculated. Chi-square tests were used to determine associations between information transparency by source type. One-way analysis of variance was used to assess differences in mean Brief DISCERN scores by source type.

Results: Our search yielded 28 unique FAQs about COVID-19 vaccines. Most COVID-19 vaccine–related FAQs were seeking factual information (22/28, 78.6%), specifically about safety and efficacy (9/22, 40.9%). The most common source type was media outlets (12/28, 42.9%), followed by government sources (11/28, 39.3%). Nineteen sources met 3 or more JAMA benchmark criteria with government sources as the majority (10/19, 52.6%). JAMA benchmark criteria performance did not significantly
differ among source types ($\chi^2=7.40; P=.12$). One-way analysis of variance revealed a significant difference in mean Brief DISCERN scores by source type ($F_{4,25}=10.27; P<.001$).

Conclusions: The most frequently asked COVID-19 vaccine–related questions pertained to vaccine safety and efficacy. We found that government sources provided the most transparent and highest-quality web-based COVID-19 vaccine–related information. Recognizing common questions and concerns about COVID-19 vaccines may assist in improving vaccination efforts.

**KEYWORDS**
content; COVID-19; frequently asked questions; internet; machine learning; natural language processing; quality; question; SARS-CoV-2; search analytics; search engine; transparency; vaccine hesitancy; vaccine; web-based health information

**Introduction**
As of August 01, 2021, COVID-19 has affected over 198 million people and has been responsible for over 4.2 million deaths worldwide [1,2]. In response to the pandemic, the US Food and Drug Administration issued emergency use authorizations for 2 COVID-19 vaccines in late 2020, 1 manufactured by Pfizer-BioNTech and the second by Moderna [3,4]. Overcoming logistical barriers will be crucial for enabling successful vaccine campaigns. Additionally, addressing the public’s perception of COVID-19 vaccines and the quality of available information is vital for promoting positive public reception and reducing vaccine hesitancy. Vaccine hesitancy, which refers to reluctance or refusal to receive vaccines, is complex and is determined by numerous factors such as trust in vaccine safety and efficacy, perceived risk of receiving or refusing a vaccine, and accessibility to and affordability of vaccines [5]. Hesitancy toward COVID-19 vaccines may hinder successful vaccination efforts.

The pace of vaccine development, misinformation, and overall growth in vaccine hesitancy are factors potentially contributing to COVID-19 vaccine refusal [5,6]. Identifying factors associated with COVID-19 vaccine refusal may assist in developing strategies to reduce vaccine hesitancy. To identify demographic factors associated with COVID-19 vaccine acceptance, Lazarus et al [7] surveyed individuals in 19 countries and reported that individuals who reported a high degree of trust in the government were more likely to report vaccine acceptance than those with low trust. In the United States, a survey study by the US Census Bureau showed that 49% of respondents were reluctant to receive a COVID-19 vaccine. Of those reluctant to receive COVID-19 vaccines, the most common reason for reluctance was concern for side effects. The second most common reason was planning to wait and see [8]. A US survey conducted early in the pandemic sought to predict COVID-19 vaccine acceptance in the United States and found that several vulnerable populations reported low willingness [9]. The growing prevalence of vaccine hesitancy highlights the importance of clinician preparedness to address patients’ concerns as access to COVID-19 vaccines grows. Health care professionals should serve as reliable sources of vaccine information, instilling confidence in patients and potentially enhancing vaccine acceptance [10], especially for COVID-19 vaccines [11].

Apart from consulting health care professionals, individuals frequently use the internet when seeking health care information; some use the internet as their primary source for health information [12]. In the United States, 61% of adults have searched the internet for medical information [13]. Searching the internet for medical information simultaneously presents benefits and challenges regarding patient-provider interactions [14]. The increasingly common practice of using the internet to obtain health care information makes it possible to study commonly held medical concerns by examining searching patterns and behaviors. Previous studies have documented the prevalence of COVID-19 vaccine hesitancy in the United States [8,15] and globally [7], but none of these studies explored the content of COVID-19 vaccine concerns evidenced by internet searching. Moreover, the quality of COVID-19 vaccine information resulting from internet searching has yet to be investigated. Thus, the primary objective of this study was to use Google’s RankBrain machine learning algorithm to characterize the content of the most frequently asked questions (FAQs) about COVID-19 vaccines in the United States. Secondarily, we sought to grade the transparency and quality of suggested information regarding COVID-19 vaccines. We aim to equip health care professionals and researchers with information about the common concerns regarding COVID-19 vaccines, possibly supporting more successful vaccination efforts. We hypothesize that most COVID-19 vaccine–related FAQs in the United States will pertain to safety and efficacy, as survey studies have indicated these concerns as the most important driver of COVID-19 vaccine hesitancy in the United States.

**Methods**

**Background**
We used Google to perform our search as it is the most frequently used search engine globally as of 2015 [16]. Moreover, Google’s search engine uses a powerful machine learning system called RankBrain [17] alongside the natural language processing technology known as Bidirectional Encoder Representations from Transformers [18] to detect patterns from large volumes of search queries. Google assesses the intent of a search query using rigorous language processing algorithms to sort through billions of indexed webpages and to suggest the ones most relevant to the search [19]. The resulting patterns and data are used to formulate lists of FAQs related to the original search contents. FAQs are found in boxes labeled
“People also ask” or “Common questions.” Google assigns each FAQ a link to information that “answers” the question [20]. Google uses its webmaster guidelines to remove low-quality spam websites from search results and prioritize high-quality sources using a system called PageRank [19]. Taken together, these FAQs represent millions of common inquiries regarding medical information. Linked answers to each FAQ reveal which information sources individuals are likely to encounter when searching Google for medical information. Our methodology was adapted from a study by Shen et al [21], who used Google FAQs to reliably reveal common concerns about orthopedic procedures and to assess the transparency of the suggested information.

Systematic Search
On January 23, 2021, using a newly installed web browser to minimize personalized advertisement algorithms, we separately searched Google [22] for the following three terms: “covid 19 vaccine,” “pfizer covid vaccine,” and “moderna covid vaccine.” We selected these terms to capture the most likely general inquiries concerning the only 2 COVID-19 vaccines available at the time of our search. For each inquiry, we refreshed the list of FAQs found in the “Common questions” or “People also ask” box generated by Google. By expanding the tab on a FAQ, additional FAQs appear. We repeated this process until reaching a minimum of 150 FAQs for each search, as studies using similar methodology have recommended using 50-150 sources [21]. We used the high end of the recommended number of sources (150) for two reasons: to increase the likelihood of encountering an FAQ that would be pertinent to the current study and to reflect the precedent set in the literature. Since query results are tailored to the user’s location, search history, and search settings, we used clean browsers to minimize any influence of history and settings while allowing results to reflect queries from the United States [19].

Data Extraction
Of the resultant FAQs, we extracted only those directly pertaining to or mentioning COVID-19 vaccines along with their answer links. In a masked duplicated fashion, investigators NS and SS extracted these data using a Google Form on January 23, 2021. FAQ data extraction was completed on January 23, 2021. After extraction, any duplicate FAQs from the individual searches were removed, followed by the removal of any duplicate FAQs among the 3 searches. After the screening and reduction process, our searches resulted in a compilation of unique FAQs regarding COVID-19 vaccines.

Question Classification and Answer Source Type
Applying methodology adapted from previous studies [16,21], we first used the Rothwell Classification of Questions [23] to categorize FAQs under three broad categories: fact, policy, and value. Fact questions were further subclassified into four groups: safety and efficacy, vaccine administration schedule, cost, and technical details. Policy questions were subclassified into two groups: indications and complications. Value questions were subclassified into two groups: evaluation of credibility and appraisal of risk or benefit. Next, we categorized answer sources as either commercial, academic, medical practice, government, or media outlet according to previously established classification schemes [21,24]. Table 1 shows the Question Classification and Answer Source Type definitions. For each answer source, we extracted the country of origin.

https://infodemiology.jmir.org/2021/1/e28740

JMIR Infodemiology 2021 | Vol. 1 | Iss. 1 | e28740 | p.61

(page number not for citation purposes)
Table 1. The Rothwell Classification of Questions, Question Classification by Topic, and Answer Source Type.

<table>
<thead>
<tr>
<th>Rothwell classification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact</td>
<td>Asks objective, factual information regarding COVID-19 vaccines (ie, “How long does it take the vaccine to work?”)</td>
</tr>
<tr>
<td>Policy</td>
<td>Asks information on a specific course of action under given circumstances related to COVID-19 (ie, should people on immunosuppressants get the vaccine?)</td>
</tr>
<tr>
<td>Value</td>
<td>Asks to conceptually evaluate COVID-19 vaccines (ie, “Will the COVID-19 vaccine work better than masks?”)</td>
</tr>
</tbody>
</table>

Question subclassification by topic

**Fact**
- Safety and efficacy: Questions about vaccine safety including side effects and how well the vaccine works
- Vaccine administration schedule: Specific questions about the vaccine schedule, number of shots, and vaccine distribution
- Cost: Cost of the vaccine, whether it is free, or who is paying for it
- Technical Details: Mechanism by which the vaccine works, including specific questions about immunologic responses

**Policy**
- Indications: Who should or should not receive a COVID-19 vaccine
- Complications: Questions about specific complications after being vaccinated

**Value**
- Evaluation of Credibility: Seeking authoritative approval from a trustworthy source; seeking ethos
- Appraisal of Risk or Benefit: Necessity of preventive measures after vaccination (ie, “Is getting vaccinated worth it?”)

Answer source type

- Commercial: Organization that publishes medical information that is not otherwise associated with an academic institution, government agency, health care system, or nonmedical news outlet such as WebMD and Healthline
- Academic: Institution with clear academic affiliations, as evidenced by information on the website that did not better meet criteria for another classification or website ending in “.edu,” such as Mayo Clinic and Harvard University
- Medical practice: Affiliation with a health care system or individual health care professional who did not explicitly state a commercial, academic, or government affiliation, such as private practice and a hospital system
- Government: Websites hosted by government organizations or sources from websites ending in “.gov,” such as the Centers for Disease Control and the US Food and Drug Administration
- Media outlet: Nonmedical organizations or social media pages claiming to publish news-related stories for the purpose of information-sharing in the form of interviews, blog posts, or articles, such as the National Public Radio, Wall Street Journal, and USA Today

### Information Transparency and Quality

The Journal of the American Medical Association’s (JAMA’s) benchmark criteria [25] was then used to assess information transparency for each answer source. JAMA benchmark criteria have been used to effectively screen web-based information for fundamental aspects of information transparency [21,26-28]. JAMA benchmark criteria were also used to characterize web-based misinformation regarding COVID-19 in early 2020 [29]. Sources meeting 3 or more criteria are considered to have high transparency, while sources meeting less than 3 criteria have poor transparency. Table 2 lists the JAMA benchmark criteria definitions.

Table 2. Journal of the American Medical Association’s benchmark criteria.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authorship</td>
<td>Clearly identifiable author and contributors with affiliations and relevant credentials present.</td>
</tr>
<tr>
<td>Attribution</td>
<td>References and sources clearly listed with any copyright information disclosed.</td>
</tr>
<tr>
<td>Currency</td>
<td>Clearly identifiable posting date of any content as well as the date of any revisions.</td>
</tr>
<tr>
<td>Disclosure</td>
<td>Website ownership clearly disclosed along with any sponsorship, advertising, underwriting, and financial support.</td>
</tr>
</tbody>
</table>

The information quality was assessed using the Brief DISCERN information quality assessment tool. DISCERN is a series of questions originally developed by Charnock et al [30] as a means for patients and providers to quickly and reliably ascertain the
quality of written health care information regarding medical treatments. The DISCERN quality assessment tool has been used to assess the quality of internet sources in a variety of medical fields [31-33]. Khazaal et al [34] developed an abbreviated 6-item version (Brief DISCERN) with comparable reliability and validity, which preserves the advantages of the original tool while affording a potentially more user-friendly format. Thus, we used the Brief DISCERN quality assessment tool, which has been previously used [35,36]. Sources are scored from 1 to 5 based on the criteria listed in Table 3.

Authors NS and SS applied the JAMA benchmark criteria and the Brief DISCERN tool in a masked duplicate fashion, and author MH resolved any discrepancies. This protocol was submitted to the institutional review board of Oklahoma State University Center for Health Sciences and was determined to be non–Human Subjects Research.

**Table 3.** Brief DISCERN questions and scoring.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Score</th>
<th>Moderate (2-4) “Partially”</th>
<th>High (5) “Yes”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it clear what sources of information were used to compile the publication (other than the author or producer)?</td>
<td>No sources of evidence for the information are mentioned</td>
<td>The sources are clear to some extent and are referenced in the text or in a bibliography</td>
<td>The sources are very clear and are referenced in text and in a bibliography</td>
</tr>
<tr>
<td>Is it clear when the information used or reported in the publication was produced?</td>
<td>No dates have been given</td>
<td>Only the date of the publication itself is clear, or dates for some of but not all acknowledged sources are given</td>
<td>Dates for all acknowledged sources are clear</td>
</tr>
<tr>
<td>Does it describe how each treatment works?</td>
<td>None of the descriptions about treatments include details of how it works</td>
<td>Descriptions of some but not all treatments are given or the details provided are unclear or incomplete</td>
<td>The description of treatment includes details of how it works</td>
</tr>
<tr>
<td>Does it describe the benefits of each treatment?</td>
<td>No benefits are described</td>
<td>A benefit is described for some but not all treatments</td>
<td>A benefit is described for each treatment</td>
</tr>
<tr>
<td>Does it describe the risk of each treatment?</td>
<td>No risks are described for any of the treatments.</td>
<td>A risk is described for some but not all treatments.</td>
<td>A risk is described for each treatment.</td>
</tr>
<tr>
<td>Does it describe how the treatment choices affect overall quality of life?</td>
<td>There is no reference to overall quality of life in relation to treatment choices.</td>
<td>The publication includes a reference to overall quality of life in relation to treatment choices, but the information is unclear or incomplete.</td>
<td>The publication includes a clear reference to overall quality of life in relation to any of the treatment choices mentioned.</td>
</tr>
</tbody>
</table>

**Analyses**

Frequencies and percentages were reported for each FAQ’s classification. Chi-square tests were used to determine associations between JAMA benchmark criteria by source type. One-way analysis of variance was used to determine whether the mean Brief DISCERN score differed by source type. Post hoc comparisons, performed using t tests with Bonferroni correction, were used to identify mean differences between source type categories. Interrater agreement for each assessment was determined using intraclass correlation coefficients.

**Results**

A total of 467 FAQs were generated from all 3 searches: 161 from “covid 19 vaccine,” 155 from “moderna covid vaccine,” and 151 from “pfizer covid vaccine.” Of these, “covid 19 vaccine” yielded 5 vaccine-related FAQs, “moderna covid vaccine” yielded 22, and “pfizer covid vaccine” yielded 14. After removing duplicates, our searches yielded a total of 28 unique FAQs regarding COVID-19 vaccines (Table 4).
### Table 4. List of the 28 unique frequently asked questions regarding COVID-19 vaccines.

<table>
<thead>
<tr>
<th>Frequently asked questions</th>
<th>Rothwell classification</th>
<th>Subclassification</th>
<th>Answer source</th>
<th>JAMA benchmark criteria (≥3)</th>
<th>Brief DISCERN score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are both Covid vaccines 2 doses?</td>
<td>Fact</td>
<td>Vaccine administration schedule</td>
<td>Commercial</td>
<td>No</td>
<td>15</td>
</tr>
<tr>
<td>Are you immune to Covid after vaccine?</td>
<td>Fact</td>
<td>Safety and efficacy</td>
<td>Media outlet</td>
<td>No</td>
<td>21</td>
</tr>
<tr>
<td>Can I get COVID-19 right after being vaccinated?</td>
<td>Fact</td>
<td>Technical details</td>
<td>Government</td>
<td>Yes</td>
<td>29</td>
</tr>
<tr>
<td>Can the COVID-19 vaccine make you sick?</td>
<td>Fact</td>
<td>Safety and efficacy</td>
<td>Government</td>
<td>Yes</td>
<td>29</td>
</tr>
<tr>
<td>Can you still get Covid after first vaccine?</td>
<td>Fact</td>
<td>Technical details</td>
<td>Media outlet</td>
<td>No</td>
<td>18</td>
</tr>
<tr>
<td>Can you test positive for Covid after vaccine?</td>
<td>Fact</td>
<td>Technical details</td>
<td>Media outlet</td>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Do COVID-19 vaccines require more than one shot?</td>
<td>Fact</td>
<td>Vaccine administration schedule</td>
<td>Government</td>
<td>Yes</td>
<td>29</td>
</tr>
<tr>
<td>Do you have to wait 90 days after Covid to get the vaccine?</td>
<td>Fact</td>
<td>Vaccine administration schedule</td>
<td>Media outlet</td>
<td>Yes</td>
<td>15</td>
</tr>
<tr>
<td>Do you have to wear mask after Covid vaccine?</td>
<td>Value</td>
<td>Risk/benefit appraisal</td>
<td>Media outlet</td>
<td>Yes</td>
<td>28</td>
</tr>
<tr>
<td>Does Covid vaccine Stop Spread?</td>
<td>Value</td>
<td>Risk/benefit appraisal</td>
<td>Media outlet</td>
<td>Yes</td>
<td>22</td>
</tr>
<tr>
<td>Has the Pfizer-BioNTech COVID-19 vaccine been authorized by the FDA?</td>
<td>Fact</td>
<td>Safety and efficacy</td>
<td>Government</td>
<td>Yes</td>
<td>30</td>
</tr>
<tr>
<td>How does the COVID-19 mRNA vaccine work?</td>
<td>Fact</td>
<td>Technical details</td>
<td>Government</td>
<td>No</td>
<td>25</td>
</tr>
<tr>
<td>How effective is the Pfizer COVID-19 vaccine?</td>
<td>Fact</td>
<td>Safety and efficacy</td>
<td>Media outlet</td>
<td>No</td>
<td>17</td>
</tr>
<tr>
<td>How long do you have to wait between Covid vaccines?</td>
<td>Fact</td>
<td>Vaccine administration schedule</td>
<td>Media outlet</td>
<td>No</td>
<td>16</td>
</tr>
<tr>
<td>How many shots of Moderna COVID-19 vaccine should I get?</td>
<td>Fact</td>
<td>Vaccine administration schedule</td>
<td>Government</td>
<td>Yes</td>
<td>29</td>
</tr>
<tr>
<td>Is it safe to take the COVID-19 vaccine?</td>
<td>Fact</td>
<td>Safety and efficacy</td>
<td>Government</td>
<td>Yes</td>
<td>29</td>
</tr>
<tr>
<td>Is the Moderna vaccine for COVID-19 approved by the FDA?</td>
<td>Fact</td>
<td>Safety and efficacy</td>
<td>Academic</td>
<td>Yes</td>
<td>30</td>
</tr>
<tr>
<td>Should you get the Covid vaccine if you were previously infected with Covid?</td>
<td>Policy</td>
<td>Indications</td>
<td>Media outlet</td>
<td>Yes</td>
<td>15</td>
</tr>
<tr>
<td>What are some common side effects of the COVID-19 vaccine?</td>
<td>Fact</td>
<td>Safety and efficacy</td>
<td>Government</td>
<td>Yes</td>
<td>29</td>
</tr>
</tbody>
</table>

### Question Classification

Using the Rothwell classification system, the majority of FAQs were seeking factual information (22/28; 78.6%). Among these factual questions, the most common topic was safety and efficacy (9/22, 40.9%) followed by technical details (6/22, 27.3%), vaccine administration schedule (6/22, 27.3%), and cost (1/22, 4.5%) (Table 4).

### Answer Sources

The most common answer source type overall was media outlets (12/28, 42.9%), followed by government sources (11/28, 39.3%), commercial sources (3/28, 10.7%), academic sources (1/28, 3.55%), and medical practice (1/28, 3.55%). FAQs classified as technical details were most frequently answered by a media outlet (4/6, 66.7%). Of FAQs classified as fact, most were answered by government sources (11/22, 50%). Government sources also most commonly answered FAQs related to safety and efficacy (5/9, 55.6%), cost (1/1, 100%), and vaccine administration schedule (3/6, 50%) (Table 4). In total, 26 of 28 (92.8%) answer sources were from the United States, 1 was from the United Kingdom (3.6%), and 1 was from Australia (3.6%).

### Information Transparency

In total, 19 sources met 3 or more JAMA benchmark criteria, of which government sources were the majority (10/19, 52.6%), followed by media outlets (7/19, 36.8%), commercial sources (1/19, 5.3%), and academic sources (1/19, 5.3%). Among sources meeting less than 3 criteria, media outlets were the most common (5/9, 55.6%), followed by commercial sources (2/9, 22.2%), medical practice (1/9, 11.1%), and government sources (1/9, 11.1%). Approximately 92.7% (11/12) of government sources met 3 or more JAMA benchmark criteria, whereas 58.3% (7/12) of media outlets met 3 or more criteria. The overall JAMA Benchmark Criteria performance did not significantly
differ among source types ($χ^2=7.40; P=.12$); however, we found significant associations between individual source’s performance on meeting JAMA benchmark criteria for authorship and the source type ($χ^2=18.03, P<.001$), with 11/28 (39.3%) media outlet sources meeting authorship criteria compared to 10/28 (35.7%) government sources not meeting the authorship criteria. We also found a similar but negative relationship with JAMA benchmark criteria’s disclosure criteria and source type ($χ^2=15.36; P=0.004$) with 10/28 (35.7%) government sources meeting these criteria compared to 9/28 (32.1%) media outlets not meeting these criteria (Tables 5 and 6).

### Table 5. Journal of the American Medical Association’s benchmark criteria and by source type.

<table>
<thead>
<tr>
<th>Source type, n (%)</th>
<th>Total</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of the American Medical Association’s benchmark criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3+ Academic</td>
<td>1 (3.6)</td>
<td>1 (3.6)</td>
<td>10 (35.7)</td>
</tr>
<tr>
<td>&lt;3 Academic</td>
<td>0 (0.0)</td>
<td>2 (7.1)</td>
<td>1 (3.6)</td>
</tr>
<tr>
<td>Authorship</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Academic</td>
<td>1 (3.6)</td>
<td>2 (7.1)</td>
<td>10 (35.7)</td>
</tr>
<tr>
<td>Yes Academic</td>
<td>0 (0.0)</td>
<td>1 (3.6)</td>
<td>1 (3.6)</td>
</tr>
<tr>
<td>Attribution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Academic</td>
<td>0 (0.0)</td>
<td>2 (7.1)</td>
<td>1 (3.6)</td>
</tr>
<tr>
<td>Yes Academic</td>
<td>1 (3.6)</td>
<td>1 (3.6)</td>
<td>10 (35.7)</td>
</tr>
<tr>
<td>Currency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Academic</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (3.6)</td>
</tr>
<tr>
<td>Yes Academic</td>
<td>1 (3.6)</td>
<td>3 (10.7)</td>
<td>10 (35.7)</td>
</tr>
<tr>
<td>Disclosure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Academic</td>
<td>0 (0.0)</td>
<td>3 (10.7)</td>
<td>1 (3.6)</td>
</tr>
<tr>
<td>Yes Academic</td>
<td>1 (3.6)</td>
<td>0 (0.0)</td>
<td>10 (35.7)</td>
</tr>
</tbody>
</table>

### Table 6. Brief DISCERN scores by source type.

<table>
<thead>
<tr>
<th>Source type</th>
<th>Mean (SD)</th>
<th>Average (SD)</th>
<th>F value (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brief DISCERN score, mean (SD)</td>
<td>30.0 (0.0)</td>
<td>17 (2.6)</td>
<td>28.6 (1.4)</td>
<td>18.0 (0.0)</td>
</tr>
</tbody>
</table>

### Information Quality

ANOVA revealed significant differences in mean Brief DISCERN scores by source type ($F_{4,23}=10.27; P<.001$), suggesting important differences in quality among the different source types. Post hoc analysis with Bonferroni correction revealed significant differences in Brief DISCERN scores between government and commercial sources ($P=.002$) and between government sources and media outlets ($P=.001$). Mean (SD) values of Brief DISCERN scores by source are provided in Table 6. Interrater agreement for our analyses was high (interclass correlation=0.96; 95% CI 0.95-0.97).

### Discussion

#### Principal Findings

Using Google and its search analytics, we were able to identify the most frequently asked questions regarding COVID-19 vaccines in the United States. Google generated these FAQs by using millions of search queries nationwide. Additionally, we evaluated the assigned “answer” source for each FAQ, assessing each source’s information transparency and quality. To our knowledge, this study is the first of its kind to evaluate the public’s most frequently asked questions concerning the COVID-19 vaccines in the United States using Google search analytics. Our study is also the first of its kind to identify common answer sources used to address COVID-19 vaccine–related concerns and to assess their transparency and quality. In the following discourse, we discuss the importance of knowing COVID-19 FAQs in the context of the current COVID-19 vaccination campaigns while also providing recommendations for improving the public’s confidence and willingness to be vaccinated.
FAQs

The most popular COVID-19 vaccine–related questions sought factual information regarding safety and efficacy, indicating greater public concern regarding these topics. Consistent with our findings, survey studies found that safety and efficacy were among the most common COVID-19 vaccine concerns reported by the public and health care workers [37-40]. Additionally, studies have identified safety concerns as being one of the most common reasons for COVID-19 vaccine hesitancy [8,38-42]. In the United States, surveys indicate that 10% to 20% of adults and an estimated 8% of health care workers will refuse COVID-19 vaccines [8,37,39,43]. While the willingness to receive the COVID-19 vaccines has increased, the alarmingly high percentage of adults refusing vaccination creates a significant barrier to protecting our most vulnerable populations [43-45]. The potential cost of vaccine hesitancy and refusal in the United States is not exclusive to the COVID-19 pandemic. For example, an outbreak of measles virus, a pathogen for which vaccines effectively control outbreaks, occurred in Clark County, Washington, in 2019 [46]. Of 71 individuals involved, 61 (86%) were unvaccinated and 52 (73%) were children [46,47]. Moreover, vaccination rates in Clark County have been 10%-14% below the national average (88%) since 2013. The measles outbreak in 2019 was estimated to cost US $3.3 million to $3.5 million in labor, direct medical costs, and productivity losses [48]. It is likely that the cost of the Clark County measles outbreak could have been mitigated or reduced with adequate vaccination [47]. Thus, to prevent similar, but likely far worse, outcomes with COVID-19, effectively educating the public on the safety of COVID-19 vaccines is paramount for enhancing COVID-19 vaccine acceptance [49].

Answer Sources

Overall, COVID-19 vaccine FAQs were most often answered by media outlets, followed by government sources. FAQs about safety and efficacy were answered more often by government sources, while media outlets frequently answered FAQs about technical details. The answer sources linked to each FAQ are found in “People also ask” or “Common concerns” boxes and are direct answers generated by Google [50]. These direct answers are supplied from Google’s “trusted entities” database and are based on relational topics and machine learning [50]. While “trusted entities” seems rather vague, it appears that Google considers direct answers to be “trusted” based on clarity, completeness, and the lack of excessive promotional jargon. With the public’s trust and willingness to accept the vaccine being a key element in a successful vaccination campaign [44,51-53], it may be more appropriate for direct answers addressing COVID-19 vaccine FAQs to be based on scientific integrity, objectivity, and transparency.

Transparency and Quality of the Answer Source

The FAQs with direct answers from government sources were more likely to meet 3 or more JAMA benchmark criteria, indicating that government answers were more transparent. Additionally, government and academic sources were found to be of significantly higher quality. While media outlets are unquestionably an important source of health information to the public, these findings suggest that government sources may be better for addressing the public’s COVID-19 vaccine concerns. Although media outlets had moderate transparency and quality, there are notable reasons to use more reliable and objective sources. Generally, COVID-19 misinformation is rampant and the public opinion can be easily manipulated [29,45]. Indeed, media outlets are a frequent source of COVID-19 misinformation, and false claims are amplified by widespread news coverage [29,54]. For example, news stories early in the pandemic touting hydroxychloroquine as a “cure” perpetuated this misinformation in the absence of evidence [55]. More recently and more specifically related to the COVID-19 vaccines, rumors that COVID-19 vaccines cause infertility in women have circulated on social media [56]. Lastly, the politicization and polarization of news coverage surrounding the COVID-19 pandemic heavily influenced the public’s attitude to COVID-19 response policies [55,57-60]. Taken together, trouble with media outlets as trustworthy sources further supports the use of unbiased answer sources such as government agencies.

Recommendations

Above all, we recommend that individuals consider health care professionals as the primary source of information regarding COVID-19 vaccines. However, in cases where access to a health care professional is limited, web-based sources unquestionably present opportunities to quickly provide high-quality and accurate information regarding COVID-19 vaccines. We agree with Mills and Sivelä [61] that a successful COVID-19 vaccination campaign depends on gaining the public’s trust in health care systems and government agencies, such as the Centers for Disease Control and Prevention and the World Health Organization, while also minimizing vaccine misinformation. Additionally, government sources must strive to translate scientifically dense literature into easily understandable information that answers widespread concerns. Therefore, the dissemination of this study’s findings may promote the public’s trust in these institutions as we have shown that government and academic sources provided the most transparent and highest-quality information addressing COVID-19 vaccine–related concerns.

Google recently demonstrated their willingness to support these COVID-19 vaccination campaigns by collaborating with Ohio State University to combat COVID-19 misinformation [62]. This partnership aims to ensure that people receive accurate information about COVID-19 vaccines to increase the public’s confidence and willingness to be vaccinated. Thus, in alignment with Google’s current intentions, we recommend that all COVID-19 vaccine FAQs be linked to government and academic answer sources; this would provide people with transparent and quality vaccine information. At a minimum, FAQs on safety and efficacy should be answered by government sources, as safety and efficacy concerns are among the primary drivers of COVID-19 vaccine hesitancy [39-42].

Strength and Limitations

Our study’s primary strength is the incorporation of Google FAQs as a novel source of insight regarding millions of individual inquiries about COVID-19 vaccines, which is an application of methodology adapted from the published literature.
and improved upon herein. Using FAQs generated by Google to explore the content of concerns regarding COVID-19 vaccine safety and efficacy may prevent common limitations of survey studies such as low response rates, reporting biases, and selection bias. Additionally, Google’s large data set is continuously analyzed in real time and may offer improved and more specific targets when approaching the public’s medical concerns. All classifications and assessments were performed in a masked duplicate fashion in accordance with standards set by the Cochrane Review and experts in the meta-research field [63,64] with high interrater reliability between investigators.

Our study is not without limitations though, such as those due to the dynamic nature of Google’s search outputs. As searching for COVID-19 vaccine–related information continues, new and updated FAQs will be generated, limiting the generalizability of our study to the time when our search was performed. Additionally, the transparency and quality assessments we used do not check for information accuracy, as this would require source-by-source comparison to generally accepted truths regarding COVID-19 vaccines, rendering our assessments as gauges of information transparency and not of information accuracy. Lastly, the categorizing of FAQs and answer sources was limited owing to their subjectivity. Although the categories were developed in line with previous reports and had high interobserver reliability, there is still potential for overlap between categories.

**Conclusions**

The expedient development and approval of COVID-19 vaccines is the culmination of the world’s greatest scientific achievements; however, without positive public reception and adequate counseling and education, COVID-19 vaccination efforts may be hindered. Using Google allowed us to obtain a list of FAQs based on millions of searches for content related to COVID-19 vaccines, which reflected widespread and common concerns. We found that the most common COVID-19 vaccine–related questions pertained to vaccine safety and efficacy, which is supported by the findings of survey studies. We found that government and academic sources provided the most transparent and highest-quality web-based information for answering the public’s most frequently asked questions about COVID-19 vaccines. Recognizing common concerns about COVID-19 vaccines may better assist health care professionals, researchers, and government agencies in improving vaccination efforts. Ensuring a successful vaccination campaign requires the public’s trust, which may be enhanced through the availability of high-quality and transparent COVID-19 vaccine information, such as that provided by government sources.

**Conflicts of Interest**

Author KM declares that her college received research grant funding as a part of a study funded by Eli Lilly and the National Institute of Allergy and Infectious Diseases of which she was a sub-investigator. The other authors declared no conflicts of interest.

**References**


22. Google. URL: http://google.com [accessed 2021-03-12]


**Abbreviations**

FAQ: frequently asked question  
JAMA: Journal of the American Medical Association