Review article

DIGITAL HEALTH

Digital Health
Volume 9: 1–16

DOI: 10.1177/20552076231155682
journals.sagepub.com/home/dhj
SAGE

A comparison between online social media discussions and vaccination rates: A tale of four vaccines

Qingqing Chen[1], Arie Croitoru[2] and Andrew Crooks[1]

Abstract

The recent COVID-19 pandemic has brought the debate around vaccinations to the forefront of public discussion. In this discussion, various social media platforms have a key role. While this has long been recognized, the way by which the public assigns attention to such topics remains largely unknown. Furthermore, the question of whether there is a discrepancy between people's opinions as expressed online and their actual decision to vaccinate remains open. To shed light on this issue, in this paper we examine the dynamics of online debates among four prominent vaccines (i.e., COVID-19, Influenza, MMR, and HPV) through the lens of public attention as captured on Twitter in the United States from 2015 to 2021. We then compare this to actual vaccination rates from governmental reports, which we argue serve as a proxy for real-world vaccination behaviors. Our results demonstrate that since the outbreak of COVID-19, it has come to dominate the vaccination discussion, which has led to a redistribution of attention from the other three vaccination themes. The results also show an apparent discrepancy between the online debates and the actual vaccination rates. These findings are in line with existing theories, that of agenda-setting and zero-sum theory. Furthermore, our approach could be extended to assess the public's attention toward other health-related issues, and provide a basis for quantifying the effectiveness of health promotion policies.

Keywords

agenda-setting, COVID-19, influenza, MMR, HPV, social media, vaccination

Submission date: 7 August 2022; Acceptance date: 20 January 2023

Introduction

The recent coronavirus (COVID-19) pandemic has brought considerable attention to vaccinations. Today, the discussion on vaccinations cuts across a multitude of issues, ranging from topics such as efficacy, safety, and trust, to topics related to ethical, religion, and personal liberties to name a few.[1–4] It has also attracted participation among diverse audiences such as policymakers and government agencies, health professionals, scientists, various advocacy and interest groups, and the public at large. Although the current discussion around vaccinations may seem a relatively recent phenomenon, it is, to a large extent, but one chapter in a longstanding public discussion spanning over more than two centuries, which revolved around "pro–" and "anti–" vaccination debates and social movements.[5,6] The contemporary public discussion about vaccinations occurs largely online,[7] and in particular on social media.[8] This use of online platforms not only impacts the vaccination discussion itself but also more broadly the way by which vaccine related information is consumed and produced. While the use of online platforms in the context of vaccinations has attracted considerable attention, the interplay between participation in vaccination discussion and actual vaccination-related behaviors is not yet fully understood.[7] This paper aims to explore this interplay through the lens of public attention as it is captured by online social media. It is necessary to point out that the term

[1]Department of Geography, University at Buffalo, Buffalo, NY, USA
[2]Department of Computational & Data Sciences, George Mason University, Fairfax, Virginia, USA

Corresponding author:
Qingqing Chen, Department of Geography, University at Buffalo, Buffalo, NY, USA.
Email: qchen47@buffalo.edu

"interplay" here refers to the potential impact of social media on human actions, which has been studied elsewhere (e.g.,[9,10]). In other words, our basic premise is that social media can impact human behavior. Based on this premise, we examine how public attention is allocated across several vaccination types, and then compare it with actual vaccination rates. Our analysis takes advantage of the recent COVID-19 pandemic which has evidently attracted much attention in the online public vaccination-related discourse, thus providing a unique opportunity to explore how the public assigns attention to vaccinations in the presence of a major disruptive public health event.

The overall approach in investigating these questions is by focusing on four distinct vaccinations, namely COVID-19, Influenza, Measles, Mumps, and Rubella (MMR), and Human Papillomavirus (HPV). While the COVID-19 vaccines were chosen due to their impacts on the current world order, the other three (i.e., Influenza, MMR, and HPV) stem from the robust nature of the debate around them in recent years (e.g.,[11,12]) before the emergence of COVID-19. A second, equally important reason for this choice was the public availability of authoritative data on their respective vaccination rates. However, it should be noted that the MMR and HPV vaccinations are geared towards children, while the other two are designed for the majority of the population. Together, this set of vaccinations allows us to compare public engagement (and attention) and vaccination rates across the different vaccines and assess the impact of the COVID-19 vaccination debate on public engagement.

Our assessment of public engagement and attention distribution among the various vaccines in social media focuses on Twitter due to its prominent role in vaccine-related debates (e.g.,[13,14]). While public attention is often difficult to quantify, one approach to assessing it is through monitoring and analyzing the ebb and flow of online discussion on social media platforms (e.g., Twitter), an area that has received increasing attention in recent years in public health studies (e.g.,[15,16]). To that end, we analyzed a data set covering 2015 to 2021 in terms of tweet message volumes that is used as a proxy for public engagement and attention.

As the Greek philosopher Aristotle said, "Man is by nature a social animal,"[17] and it is through the interaction of various media that we communicate and disseminate information. Over time, the medium of dissemination has evolved, from handwritten papal bulls, the printing press, to wireless communication technologies, such as radio and television broadcasting. Such evolution enabled the dissemination of audio and visual content to a more and more broader audience both nationally and internationally. For example, by the late twentieth century, around 98% of households owned at least one television.[18]

Today printed and broadcast media, often referred to as "vertical media,"[19] aims to cover potential interests and topics, and has become an integral part of modern society. Consequently, considerable research has been dedicated to the study of mass communication and the underlying principles and theories that govern it. In the context of our research, there are two key related theories that are of particular importance for our analysis: the agenda-setting theory and the zero-sum theory

Agenda-setting theory,[20] generally examines the relationship between the media and the public agendas with respect to a specific topic, highlighting that the media can affect the salience of issues in the public agenda. In earlier agenda-setting studies, it was argued that the media could not only influence what people think about, but also how they think about it.[21,22] This early work has spurred research investigating the impacts of the media on the public's perceptions with respect to politics, economics, and public health.[23–26] In the context of public health, several authors have noted the role of agenda setting played by the media in influencing health-promoting change.[27–31]

The rise of Web 2.0[32] has sprung renewed interest in agenda-setting in the context of social media due to the ability to facilitate "two-way communication". Such technologies, also called "horizontal media,"[19] emphasize the importance of the individual and reduce the barriers to user participation. Various studies have explored the applicability of agenda-setting theory in the digital age, as well as how social media shape public discourse and attitudes towards specific topics.[19,22,33–35] One example of this is how the agenda can be influenced by the networked association of issues through social media.[36–38]

A second, related theory, that is of particular importance to our research is the zero-sum theory. This theory, which can be seen as complementing the agenda-setting theory, is based on the premise that human (and by extension the public's) attention is finite, which leads to a competition among the various issues in the media in which increased attention to one issue will lead to a decrease in attention given to other issues.[39] Like the agenda-setting theory, the zero-sum theory was originally proposed in the context of vertical media (i.e., "one-way" communication). More recent studies which have examined this theory in the context of social media suggested its applicability is given the inherent finite nature of human attention regardless of the subject matter at hand (e.g.,[40–42]).

Turning back to the study of social media on health-related topics, we are especially interested in the online debates regarding vaccination under this context. Looking back at disease outbreaks over the past two decades, such as the Ebola and Zika viruses and the current COVID-19 pandemic, they all have substantially impacted, to varying degrees, on our environment, economy, and society as a whole.[43–50] Therefore, developing vaccines, especially for vaccine-preventable diseases, has become crucial to preventing the spread and thereupon helping the world return to somewhat normalcy.

As mentioned above, the discussion regarding vaccination has been longstanding but with disparate voices. Pro-vaccination supporters urge the importance of vaccination for public health, whereas anti-vaccination groups argue their concerns about safety, religious and philosophical beliefs, which has made the vaccination debate polarized over time.[2,4,51,52] This is especially the case today with mass communication through social media. Social media enables individuals to share their opinions about vaccinations online, but at the same time puts them in an information "dye vat," in the sense there are all sorts of words (either positive or negative) about vaccinations that can become mixed together and alter people's perceptions of vaccination.[53] It is through such exposure that an individual's perception towards vaccination might change under the impacts of various information traffic, resulting in diverse vaccination sentiments, and thereafter, making vaccination campaigns more challenging.[54,55] In this regard, it is crucial to understand how the public's attention maps to various vaccination-related themes on social media, especially with respect to their limited attention capacity, and how such attention compares to the reported vaccination rates in reality. These questions have direct practical applications, such as topics around improving public health awareness and the effectiveness of health promotion policies.

In the remainder of the paper, we will demonstrate the methodology used, including data collection, sentiment analysis, and comparisons between online vaccination debates and offline vaccination rates, to investigate the dynamic attention of individuals on various vaccination-related themes on social media (section "Methodology"). After that, we will discuss and conclude the key findings (section "Results and Discussion") and how our work contributes to the current literature in the Conclusion Section.

Methodology

Studies focusing on improving public health awareness have increasingly relied on leveraging social media data (e.g., Twitter). These studies have indicated that public attention can be reshaped, to a certain degree, when different public health issues are prominent on social media. In this regard, we would argue that understanding the dynamics of public attention among competing themes for attention in the context of vaccination could have the potential to provide informed insights for improving vaccination-related health outcomes.

To illustrate this potential, we took the United States as a case study to quantify public attention to four vaccines through monitoring and analyzing the ebb and flow of their online discussion on Twitter and identifying the corresponding vaccine sentiments. In what follows, we will introduce the datasets used in this study (section "Data Collection"), and describe how we classified this data by vaccine sentiment (section "Sentiment Analysis and Emotion Detection"). We will then compare four different online vaccination discussions with actual vaccination rates to uncover potential links between them (section "Online/Offline Comparison"). These steps are illustrated in Figure 1 and are further elaborated in the subsequent subsections.

Data collection

In order to understand how public attention maps to various vaccination-related themes on social media, we extracted vaccine-related tweets sent in the United States from January 2015 to July 2021, which consists of around 11.7 million Twitter messages sent by approximately 2.6 million distinct users. The data were collected by utilizing Geosocial Gauge system[56] using the keywords "vaccination" and its derivatives, such as "vaccine," "vaccines," "vax," "vaxine," "vaxx," as well as "immunization" and "immunisation" as search terms (these terms were found to frequently turn up in the online narrative about vaccines on Twitter.[57] Table 1 provides an overview of the distribution of distinct tweets and users by years. It is necessary to point out that the study period used here was constrained by the available data at hand, and we will come to this issue in the Discussion section. Nonetheless, the Twitter dataset used here still contains a considerable long period for investigating changes in public attention over time in the subsequent analysis.

In order to understand how public attention compares to reported vaccination rates, we collected actual vaccination rates for four distinct vaccines, namely COVID-19, Influenza, HPV, and MMR vaccines, from the United States Centers for Disease Control and Prevention (CDC). More specifically, the COVID-19 vaccination dataset[58] records the total number of people with at least one dose based on the jurisdiction where a recipient lives from Jan 2021 to July 2021. The influenza vaccination dataset[59] consists of estimated seasonal influenza vaccination coverage from 2015 to May 2021, which is measured based on the National Immunization Survey-Flue (NIS-Flu) and the Behavioral Risk Factor Surveillance System (BRFSS). It is necessary to mention that the influenza vaccination coverage collected was measured based on different socio-demographic groups (e.g., "> 6 Months", "5–12 Years", "13–17 Years", "> 18 Years"), but they all have a very similar pattern (see Appendix A1). As we are focusing on vaccination rates and people's choices to become vaccinated at large, the age of legal majority (i.e., 18 years) was selected to represent the decision-making of whether to be vaccinated or not. Turning to MMR, the vaccination dataset documents the estimated MMR vaccination among adolescents from 2015 to 2021, and the HPV dataset records estimated HPV vaccination among adolescents with individuals at least one dose from 2016 to

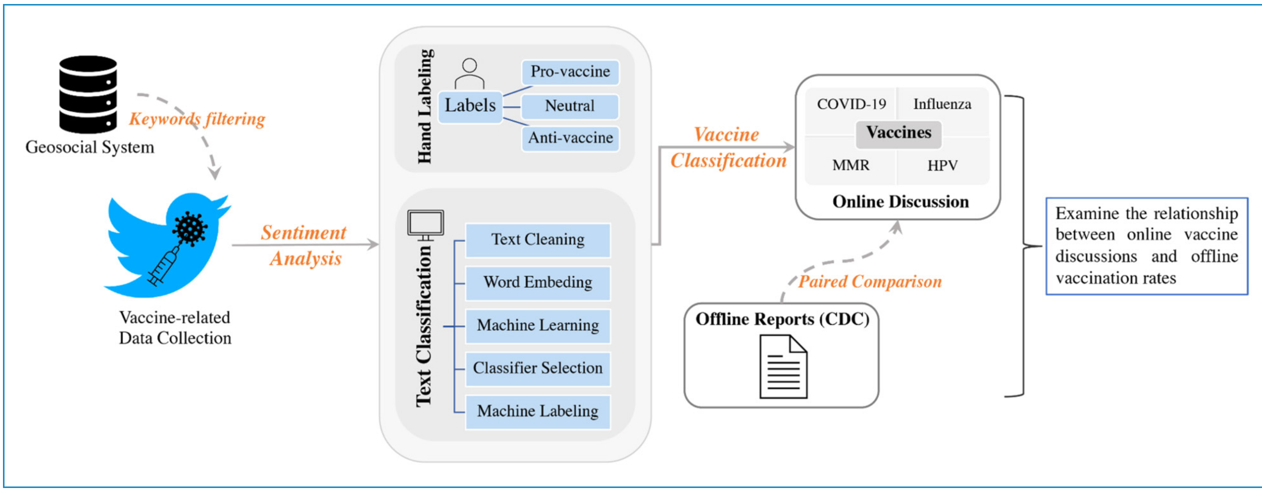

Figure 1. The workflow for comparing between online social media discussion and vaccination rates.

Table 1. The distribution of distinct tweets and users by years from 2015 to 2021.

Year	Unique tweets	Unique users
2015	597,488 (5.12%)	166,985 (5.02%)
2016	477,445 (4.09%)	146,958 (4.42%)
2017	460,788 (3.95%)	150,203 (4.52%)
2018	375,970 (3.22%)	153,637 (4.62%)
2019	887,571 (7.60%)	295,239 (8.88%)
2020	3,285,936 (28.15%)	1,090,222 (32.77%)
2021	5,589,713 (47.88%)	1,323,331 (39.78%)

2021.[60] Table 2 gives a summary of all datasets aforementioned.

Sentiment analysis and emotion detection

To extract information about the population's attitude towards vaccination from the collected Twitter data corpus, sentiment analysis was performed. In the context of this paper, we refer to sentiment analysis as Natural Language Processing (NLP) based techniques that focus on unraveling individuals' opinions, attitudes, or emotions from text messages regarding to a specific topic—in our case, the online social media discussion about vaccination.[61] In this study, we aim to classify tweets into one of three classes—"*Pro-vaccine*," "*Neutral*," and "*Anti-vaccine*", and do so by integrating supervised machine learning with word embedding techniques.

Ground Truth Collection In order to build a supervised machine learning model (i.e., a classifier), the first and most critical step is to collect labeled samples (referred to as "ground truth" in the remainder of the paper), which serve as the foundation for the model training. For this purpose, we used a dynamic online survey approach proposed by Chen and Crooks[62] for hand-labeling tweets. More specifically, each tweet in a survey is manually assigned a label from three possible labels: "pro-vaccine," "neutral," and "anti-vaccine." Table 3 lists examples of manually labeled tweets of the three types of sentiment. Each generated survey contained a sample of 100 tweets randomly drawn from a sample pool that consisted of 15,000 randomly selected tweets from our dataset. Both random samples were generated using a random uniform distribution. The advantage of this design is that it allows some overlap across different survey instances while enabling us to perform quality assessment of the labeled tweets. The generated surveys were sent to volunteers for labeling, resulting in 96 responses and a total number of 5959 unique hand-labeled tweets. We then conducted a reliability measurement on these labeled tweets to examine the internal consistency and the inter-rater reliability. Specifically, we measured the percentage of agreement of tweets that were duplicated between each participant and only kept those tweets with higher than 50% agreement of themselves. We then used Cohen's kappa statistic[63] to examine the inter-rater reliability between participants and the researchers and excluded participants with scores below 0.4 (i.e., moderate agreement). This resulted in a set of 2032 reliable hand-labeled tweets and we further expanded this corpus with other tweets from the same data corpus using the same labeling procedure[64] to increase sample size while preserving the sample quality. As such, we got a total of 7086 hand-labeled tweets as the ground truth for the subsequent analysis and modeling.

Table 2. Description of datasets used in the study.

Dataset	Time periods	Description	Source
Twitter	Jan 2015–Jul 2021	Vaccine-related tweets sent in the United States.	Geosocial Gauge system (56)
COVID-19 vaccination	Jan 2021–Jul 2021	Total number of people with at least one dose based on the jurisdiction where the recipient lives.	Data.CDC.gov (58)
Influenza vaccination	Jul 2015–May 2021	Estimated seasonal influenza vaccination coverage with people who are over 18 years old	Data.CDC.gov (59)
MMR[a] vaccination	Jan 2015–Dec 2021	Estimated MMR vaccination coverage among adolescents	Data.CDC.gov (60)
HPV[b] vaccination	Jan 2016–Dec 2021	Estimated MMR vaccination coverage among adolescents with at least one dose	Data.CDC.gov (60)

[a]MMR refers to Measles, Mumps, and Rubella; [b] HPV refers to Human Papillomavirus.

Preprocessing The ground truth data was then cleaned by a sequence of preprocessing tasks to remove noise in the texts that are ineffective for the classification process. As in other studies (e.g.,[65]), we first removed characters that are consecutively duplicated and deleted Unicode. In addition, we also removed hashtags and URLs. Second, we replaced emojis with the corresponding emotion by utilizing the "*demoji*" package.[65] However, unlike other studies (e.g.,[64]), we did not remove stop words or punctuation for each tweet during the preprocessing step. Our rationale for not doing so is that it has been shown that the removal of these elements can influence the performance of the sentiment analysis based on the nature of tweets, such as ironic words or phrases encountered in them.[66,67]

Classifier Training and Machine Labeling After preprocessing the labeled tweets, we turned them into machine-readable features before applying machine learning models. We did so by converting words into a vector representation based on word embedding techniques,[68] considering its advantage in capturing non-trivial relationships among words while preserving their contexts. To operationalize this, we utilized a 300-dimensional word2vec embeddings trained on Google News[69] to compute the numeric representation of words for every single tweet, which were then used as the input in different machine learning algorithms for classification. Specifically, we first split the hand-labeled tweets as discussed in the Ground Truth Collection section to training and test sets based on stratified sampling. The training set was used for model training on five classifiers, including Naive Bayes, support vector machine (SVM), logistic regression, and extreme gradient boosting (XGBoost), followed by fivefold cross-validation and hyperparameter tuning. The test set was then used for comparing model performance. In the end, the XGBoost classifier, which demonstrated the best performance (accuracy = 0.74), was selected to label the sentiments of all tweets in the rest data corpus. The performance metrics of the XGBoost classifier can be found in Table 4.

Table 3. Examples of hand-labeled tweets of each type of sentiment.

Sentiment	Examples
Pro-vaccine	This baby got measles because of anti-vaxers.
Neutral	Amid measles concerns, debater ages in Texas over vaccination requirements.
Anti-vaccine	MMR vaccine failure covers up disclosure more evidence of MMR vaccine failure: university Mumps outbreak among vaccinated.

Table 4. Performance metrics of the XGBoost classifier.

	Anti-vaccine	Neutral	Pro-vaccine
Precision	0.78	0.71	0.74
Recall	0.61	0.59	0.87
F1 Score	0.69	0.65	0.80
Accuracy		0.74	

Emotion Detection In addition to sentiment analysis, we also applied the National Research Council (NRC) Word-Emotion Association Lexicon (aka EmoLex),[70,71] which consists of more than 10k word-sense pairs, to detect eight emotions (i.e., anger, fear, anticipation, trust,

Table 5. Number of users by quarter in four different vaccination discussions on twitter from 2015 to 2021.

Year Quarter	COVID-19	Influenza (Flu)	MMR	HPV
2015 Q1	0.0% (0)	15.7% (5413)	69.6% (24,035)	14.8% (5102)
2015 Q2	0.0% (0)	25.5% (5926)	47.7% (11,098)	26.8% (6237)
2015 Q3	0.0% (0)	53.2% (11,859)	18.4% (4109)	28.4% (6342)
2015 Q4	0.0% (0)	74.2% (22,667)	8.7% (2656)	17.1% (5232)
2016 Q1	0.0% (0)	34.9% (7643)	16.5% (3608)	48.6% (10,652)
2016 Q2	0.0% (0)	42.0% (6142)	26.8% (3929)	31.2% (4566)
2016 Q3	0.0% (0)	45.7% (9188)	12.4% (2492)	41.9% (8416)
2016 Q4	0.0% (0)	63.8% (13,486)	11.4% (2418)	24.7% (5226)
2017 Q1	0.0% (0)	41.8% (7355)	27.1% (4770)	31.1% (5468)
2017 Q2	0.0% (0)	26.8% (5677)	45.5% (9636)	27.7% (5879)
2017 Q3	0.0% (0)	50.3% (8590)	18.2% (3111)	31.5% (5387)
2017 Q4	0.0% (0)	73.4% (20,787)	10.0% (2840)	16.6% (4701)
2018 Q1	0.0% (0)	81.3% (14,690)	5.7% (1035)	12.9% (2338)
2018 Q2	0.0% (0)	39.0% (3057)	14.9% (1167)	46.1% (3611)
2018 Q3	0.0% (0)	44.9% (10,671)	26.2% (6233)	28.8% (6851)
2018 Q4	0.0% (0)	66.8% (28,262)	15.3% (6460)	18.0% (7598)
2019 Q1	0.0% (0)	25.3% (14,833)	62.1% (36,438)	12.6% (7375)
2019 Q2	0.0% (0)	12.1% (7260)	74.8% (45,050)	13.1% (7892)
2019 Q3	0.0% (0)	43.1% (13,461)	32.7% (10,202)	24.2% (7537)
2019 Q4	0.0% (0)	59.6% (28,851)	26.6% (12,884)	13.8% (6661)
2020 Q1	73.5% (135,566)	19.1% (35,224)	3.5% (6480)	3.8% (7061)
2020 Q2	86.5% (301,884)	11.5% (40,283)	1.3% (4556)	0.7% (2432)
2020 Q3	88.1% (8346)	10.1% (961)	1.0% (96)	0.8% (75)
2020 Q4	95.1% (681,480)	4.4% (31,297)	0.4% (2690)	0.2% (1398)
2021 Q1	97.4% (1,061,667)	2.1% (23,030)	0.2% (2551)	0.2% (2651)
2021 Q2	96.3% (644,512)	3.0% (19,862)	0.5% (3045)	0.5% (2022)
2021 Q3	96.1% (36,797)	3.2% (1211)	0.4% (161)	0.3% (127)

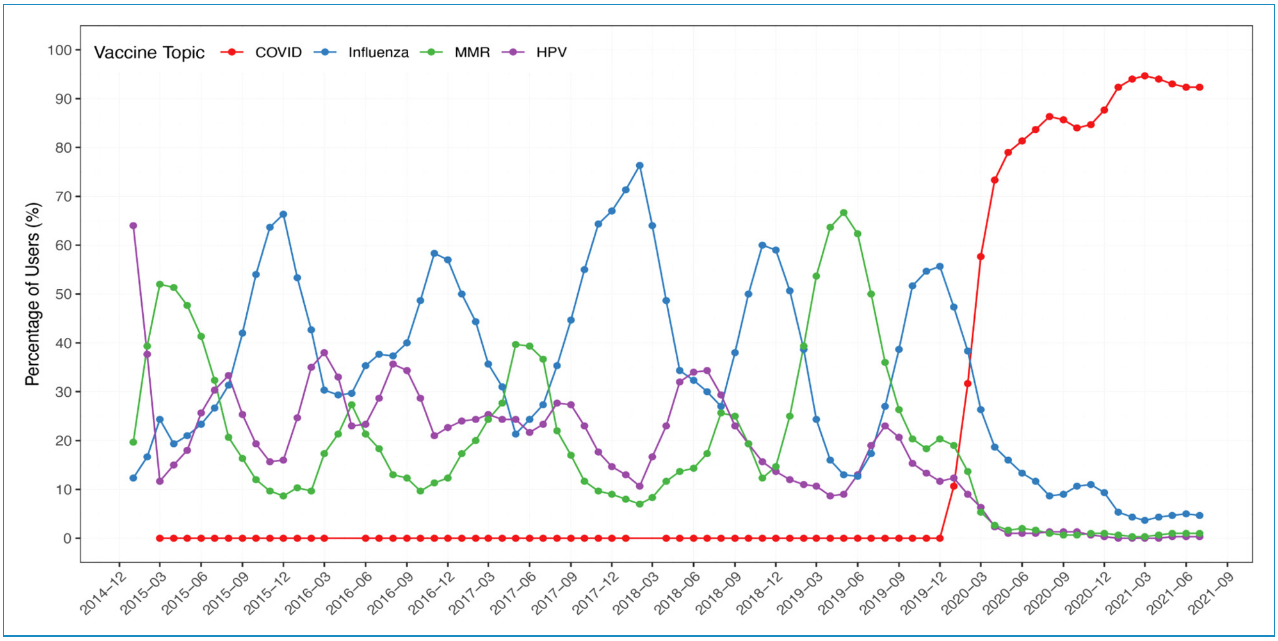

Figure 2. The quarterly distribution of percentage of users by different vaccine discussion from 2015 to 2021.

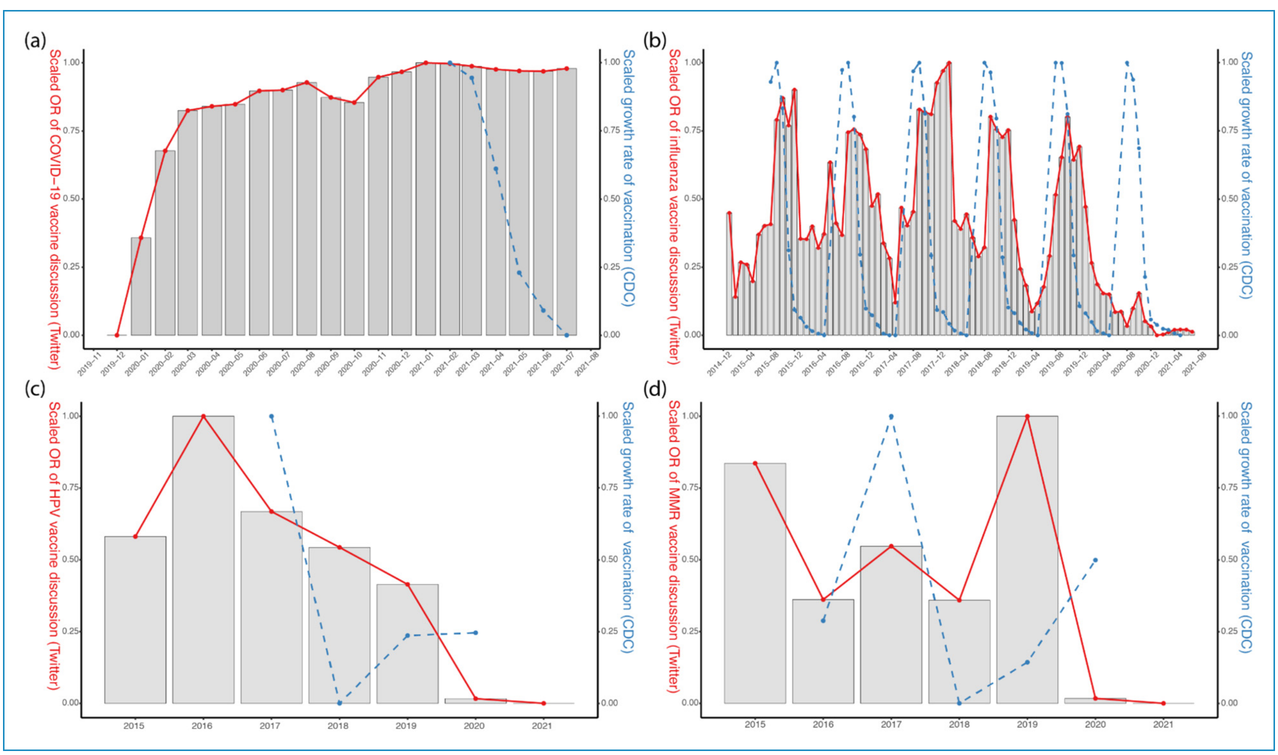

Figure 3. The comparison between different vaccine discussions on Twitter and growth rate of the actual vaccination rate collected from the CDC (a) COVID-19; (b) Influenza; (c) HPV; (d) MMR.

surprise, sadness, joy, and disgust) for the four vaccination themes, and we do so for each year. As such, we are able to compare how people's emotions with respect to different vaccines have changed over time, which can help us understand the underlying mechanism of how social media is shaping and influencing people's emotions.

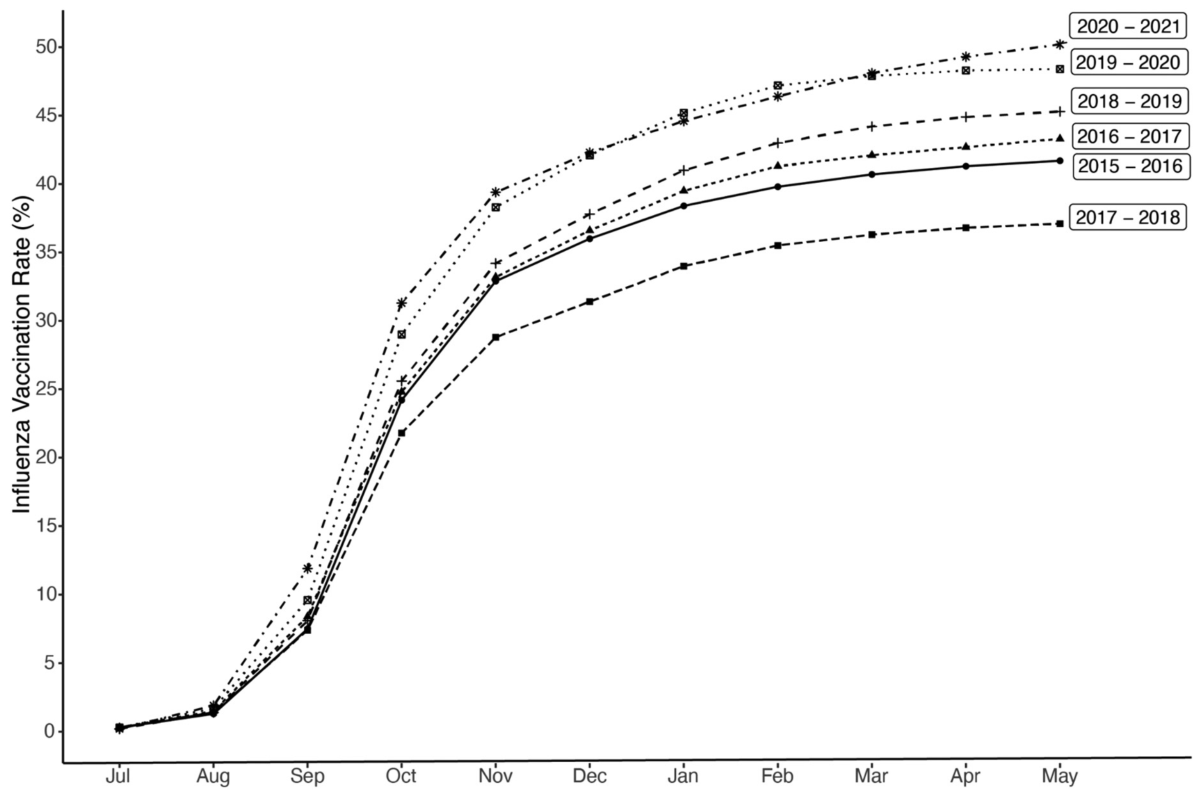

Figure 4. The distribution of absolute influenza vaccination rate by month recorded by the CDC (*Note: there is no data available for June*).

Online/offline comparison

After all the tweets were labeled, we separated the tweets into the four different vaccination themes aforementioned (i.e., COVID-19, Influenza, MMR, and HPV). Specifically, we used the keyword "coronavirus" and its derivatives, such as "covid," "covvax," "covvaax," to extract tweets that are related to COVID-19 vaccination theme. Similarly, we utilized keywords of "influenza," "mmr," "hpv" and their derivatives to extract tweets related to Influenza, MMR, and HPV vaccination themes, separately. It is crucial to point out that although some tweets may consist of multiple themes, we only focused on tweets that had one specific theme discussed. The reason for this relatively stringent condition is to prevent inferring type I errors in sentiment classification. In this regard, tweets that had more than one vaccination theme mentioned were removed from the corpus, resulting in a total of 3,720,721 unique tweets for subsequent analysis and comparison. In this form, the distinct vaccination themes allowed us to make a cross-comparison of public engagement and attention as well as the corresponding vaccine sentiments among different themes in social media.

To quantify changes in the activity levels of vaccination discussions on Twitter, we used a quarterly time interval as the analysis unit and measured the relative percentage of users who participated in each vaccination discussion. Subsequently, we compared trends in the Twitter traffic of the four vaccination discussions to trends in the corresponding vaccination rates over time. In doing so, we are able to better understand how public engagement and attention to different vaccination-related themes in social media compare to actual vaccination rates.

Results and discussion

Table 5 displays both absolute value (i.e., number of unique users involved in each vaccination discussion) and relative value (i.e., percentage of users involved in each vaccination discussion relative to total users for the quarter) while Figure 2 shows the quarterly percentage of users from 2015 to 2021. Generally speaking, these show how the public's attention, while being finite due to the zero-sum theory,[39] switches from one vaccination to another over time. For example, in the first half of 2015, the vaccination conversation about the MMR dominated the online discussion which can be attributed to the measles outbreaks in California,[57] but the percentage of users gradually declined in the second half of 2015 as the public's attention switched to focusing largely on influenza during the winter period. More recently, they show the substantial redirection of

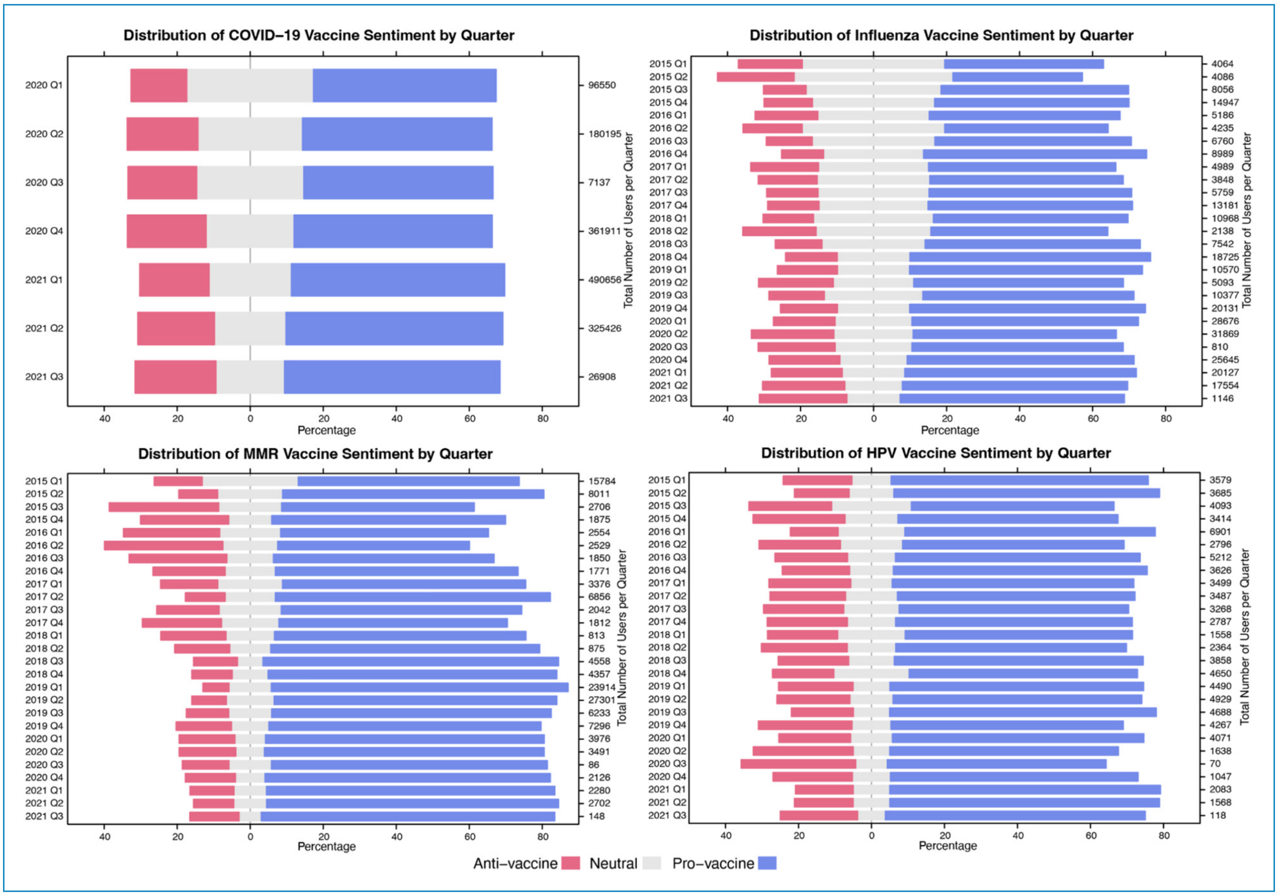

Figure 5. Distribution of different vaccine sentiments by quarter (a) COVID-19; (b) Influenza; (c) MMR; (d) HPV.

public attention towards the COVID-19 vaccine in 2020. Since then, the dominance of the COVID-19 vaccine discussion has drawn the most public attention and has maintained its dominance until the end of the study period. Specifically, the number of users of who posted COVID-19 vaccine tweets accounted for 73.5% in 2020 Q1 (see red highlights in Table 5) and gradually increased afterward, reaching its highest point in 2021 Q1 (97.4%). Although there was a subtle decrease (around 1%) after 2021 Q1, the COVID-19 vaccine discussion overshadows other vaccine discussions. One possible explanation for this finding is that generally, people tend to be more interested in recent happenings that are more relevant to them and that the uncertainty of new things stimulates their curiosity, prompting them to engage in the discussion of more contemporary issues in order to gain an understanding of the current situation. This observed online behavior is in line with the agenda-setting and the zero-sum theories, which state that the media has the capability in shaping people's agenda or priority of issues, that public attention is finite, and that the public is generally uncomfortable in new settings until they achieve some degree of orientation to their new surroundings.[72]

Another noteworthy finding is that the influenza vaccine discussion demonstrates a cyclical pattern, with peaks generally occurring during the winter flu seasons before the COVID-19 outbreak (see blue highlights in Table 5). Although the phenomenon diminished significantly with the rise of the COVID-19 vaccine discussion, we still observed a small peak of flu vaccine discussion during winter flu season in 2020, while the public's attention to other vaccines has diminished (see Figure 2). This suggests a potential association between COVID-19 and flu vaccines that may result from the perceived similarity between the two illnesses (e.g.,[73]). For example, both COVID-19 and flu viruses are contagious respiratory illnesses that can spread from person to person, and people who have COVID-19 or flu usually have several similar symptoms.[74,75] Such an association between COVID-19 and influenza was also recently reported by Bruine de Bruin et al.[76] across demographic groups in the United States.

The second part of our analysis focused on examining the relationship between the online vaccination debates and the public's actual vaccination behavior over time. To that end, we compared trends in Twitter traffic for the

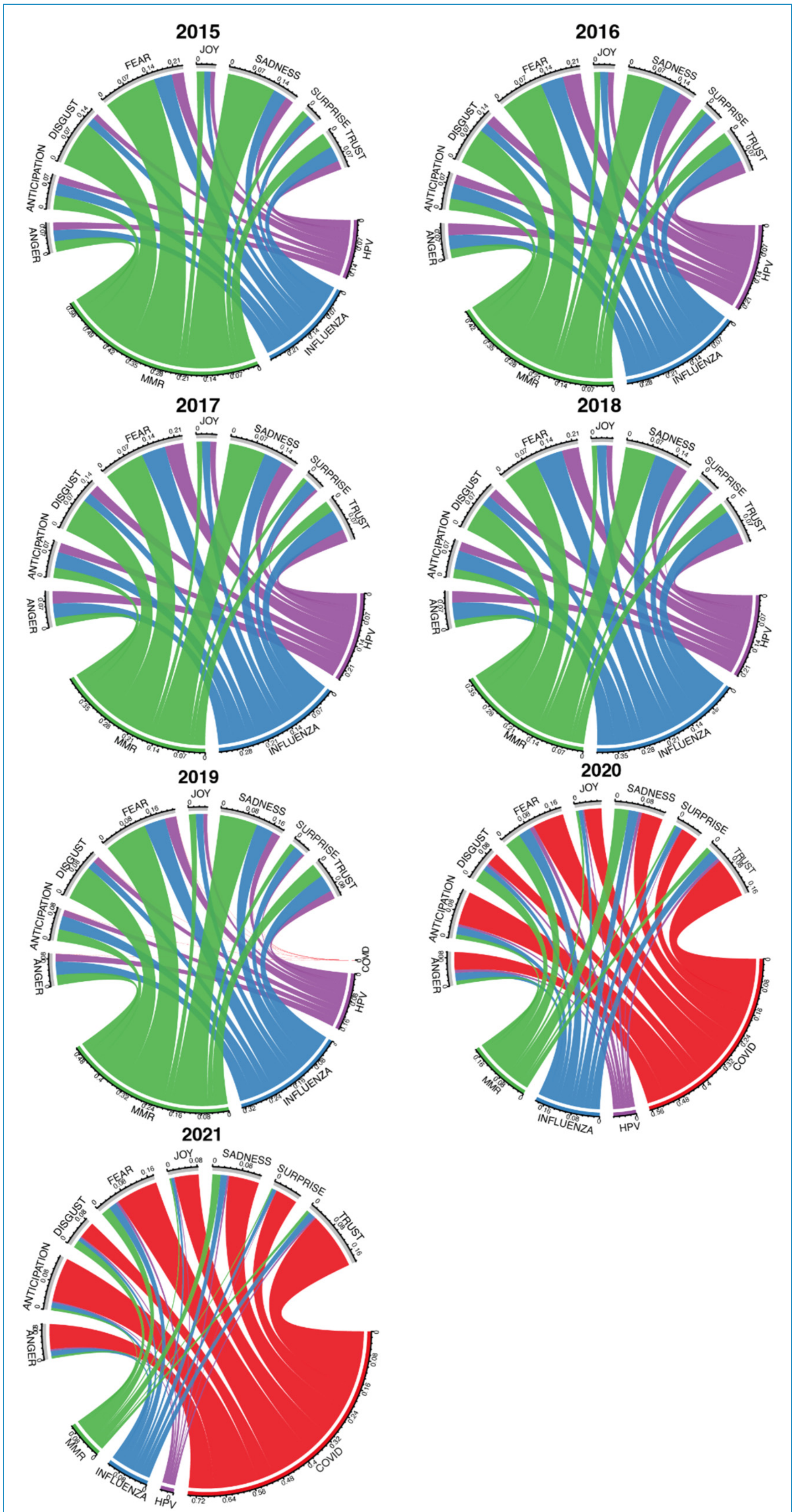

Figure 6. The changes of emotion over time for different vaccines.

four different vaccines with trends in the actual vaccination rates. Due to the limited time granularity of the CDC data, we measured the monthly vaccination rates for COVID-19 and influenza vaccines and annual rates for MMR and HPV vaccines, respectively. To mitigate the impact of the size-related issue (e.g., unequal distributed population) on representing the Twitter traffic, we calculated the odds ratio of users by normalizing the corresponding Twitter population (see Eq. 1).

$$OR_{ij} = \frac{\dfrac{N_{\text{vaccine theme } i \text{ in time } j}}{N_{\text{vaccine theme } i \text{ in entire study period}}}}{\dfrac{N_{\text{time } j}}{N_{\text{entire study period}}}} \quad (1)$$

where N refers to the number of users, time i refers to the i^{th} month or year according to the corresponding vaccine theme j.

Figure 3 displays the temporal distribution of scaled odds ratio of vaccine discussion online (red color) as well as the scaled growth rate of actual vaccination offline (blue color) for the four different vaccines. Figure 3 also shows how the HPV and MMR vaccine rates are rather volatile at the yearly time granularity, without any apparent patterns in their trends in Twitter and actual vaccination rates. Nonetheless, we do observe a periodic change in the influenza vaccine, where the peak rate of flu vaccinations emerges close to the peak of the flu vaccine discussion on Twitter, and the peak regularly appeared during the winter period (i.e., flu season) (see Figure 3(b)). Exploring this phenomenon more, we compared the absolute flu vaccination rate at each month and found that more people took the flu vaccine during the COVID-19 pandemic (flu vaccination rate increased from around 40% to 50% after the outbreak) as shown in Figure 4. This finding coincides with the work of Roman et al.,[77] indicating that the prominence of an issue (here referring to COVID-19 vaccine discussion) on social media has the potential to affect the public's behavior on another issue (here refers to the uptake of flu vaccine) due to the public associating the two themes. This is also in line with the concept of the Network Agenda Setting introduced by Guo and McCombs,[38] who pointed out the role of cognitive components in the process of representing reality. For instance, with respect to COVID-19 (i.e., the reality) the cognitive components could be considered as information that describes the characteristics of the virus (e.g., the symptoms) or the vaccine. To this end, the more frequently two issues are associated within the media, the more likely they are to be perceived as interdependent on the public agenda. Moreover, because COVID-19 and influenza share certain aforementioned similarities, the public's uptake of influenza vaccine (as shown in Figure 4) was influenced by online COVID-19 vaccine discussion, whereas the same phenomenon was not observed in MMR or HPV vaccines.

To further help understand the underlying mechanism of how social media is shaping and influencing people's perception and emotion towards vaccination, we classified users' attitudes into three sentiments, namely pro-vaccine, neutral and anti-vaccine, and monitored the shifts of emotion with respect to different vaccines over time. Figure 5 presents the distribution of these sentiments with respect to the four different vaccines on a quarterly basis. The results show that in general, a positive vaccine sentiment is dominant, and the pro-vaccine users grow faster than the anti-vaccine group, indicating that the online vaccine discussion has the potential to enhance public awareness of the importance of vaccination. Digging more into this, Figure 6 displays the changes in people's emotions regarding the four vaccines over time. This figure clearly demonstrates the gradual shift from influenza, MMR, and HPV to COVID-19 after 2019 (the COVID-19 outbreak). Moreover, besides the feelings of fear, anger, and sadness concerning the COVID-19 vaccine, we observed more positive emotions in the online discussion, such as anticipation and trust. This again implies the positive nature of the social media discourse, and by extension the public awareness, of vaccination campaigns.

It is also interesting to examine the relationship between the public's online attention to COVID-19 and influenza, as shown in Figure 2, and the actual influenza vaccination rates as shown in Figure 4. Specifically, while most online public attention was diverted toward the debate around COVID-19, and despite a mental association between COVID-19 and influenza,[76] there was a relatively moderate increase in the influenza vaccination rates in 2020 and 2021 (10%) which is consistent with the variations observed between years prior to COVID-19 (e.g., when comparing 2017–2018 and 2018–2019). A possible explanation for this is the existence of a psychological omission bias in the decision whether to vaccinate for influenza or not, i.e., a preference toward inaction that may result in harm greater than the potential harm that may be caused by action.[78,79] More recently, however, it was suggested that the choice between action and inaction is more nuanced and involves a judgment of how effective taking action might be when faced with an undesirable situation.[80]

Conclusion

The unprecedented COVID-19 pandemic has taken the discussion around vaccinations to new heights in the modern information environment, where dissemination and communication of information have become much more convenient and rapid than ever before. This environment, along with the finite nature of human attention creates a perpetual challenge of choosing how attention should be distributed across several competing vaccination themes (i.e., zero-sum dynamics).

In this context, we have proposed a way to study and compare the public vaccination discussion and actual vaccination-related behaviors based on social media data. We did so by analyzing and monitoring the ebb and flow of online debates regarding four prominent vaccines (i.e., COVID-19, Influenza, MMR, and HPV vaccines) on Twitter in the United States from 2015 to 2021, and compared them with actual vaccination rates from governmental reports, which are considered as "real-world" vaccination related behaviors. Moreover, in order to uncover the underlying mechanism of social media in shaping and influencing people's vaccination attitude and emotions, sentiment analysis based on word-embedding techniques and machine learning algorithms, as well as emotion detection based on the NRC Word-Emotion Association Lexicon were conducted to classify online users' attitudes and emotions towards different vaccinations over time. In doing so, we were able to capture how the public's attention maps to various vaccination-related themes and how such attention relates to the actual vaccination behaviors.

Using the United States as a case study, we found COVID-19 vaccination drove a sharp drop in online communication about other vaccination topics (i.e., Influenza, MMR, and HPV). In other words, the COVID-19 vaccination debate surged in the wake of the COVID-19 outbreak, reallocating attention from other vaccination-related themes. However, a more intriguing insight is that we still can observe an associated conversation about influenza under the dominance of COVID-19 vaccination debate, while other vaccination discussions have almost diminished. Part of the reason could be the underlying similarity (e.g., similar symptoms) between the COVID-19 and influenza. Furthermore, our results show that positive vaccine sentiment is dominant across different vaccination themes throughout the years, implying the positive nature of social media and the public awareness of vaccination campaigns. Our findings also highlight the discrepancy between the online debate as captured on Twitter and the actual vaccination rates. Specifically, while more recently most attention was given to COVID-19 over influenza, the actual influenza vaccination rates remain relatively constant. This apparent discrepancy can be explained, at least in part, by the role omission bias plays in people's decision to vaccinate or not.

As the modern information landscape is highly complex and diverse, it is important to consider the findings described here in that context. While our study focused on Twitter as a representative of the public discussion around vaccinations, it is but one of several prominent social media platforms on which the vaccination debate is carried on. Thus, additional research is needed in order to examine if, and if so how our findings would compare with a similar analysis on other social media platforms. Building on this, this study focused on the United States, and from this point of view, future work is needed on applying the proposed method to other countries. In other words, the methodology is informed by the nature of the data and the academic context, so is contingent by definition. The methodology proposed may then need to adjustment for studying other countries. Such comparative research using different data sources or different case studies, however, does not currently exist to the authors' knowledge. At a more local scale, another possible extension of this work could focus on a more nuanced analysis to explore if there are any differences across different geographical regions. This expansion could be further complemented by considering changes in scientific evidence, guidelines, systemic features, such as access, debates around possible mandates, vaccine shortages, to name a few, to improve the analysis accuracy. Another area of further inquiry relates to the fact that while our study only used the vaccination rate as a simple proxy for people's vaccination behavior, it should be recognized that vaccination rates are the product of many factors, such as demographic characteristics, socio-economic status, and accessibility to health resources. Additional research is therefore needed in order to examine the relationships between these factors and the public's overall vaccination behavior. In terms of analysis, tweets that contained more than one vaccine were removed as mentioned in the Methodology to prevent inferring type I error in sentiment classification. However, in future work more analysis is needed to better understand such kinds of tweets, for instance using aspect extraction to identify opinion targets in texts. Moreover, the emotion measures in this study may reduce our understanding of specific contents, such as discussion of more marginal debates both nationally and globally. Therefore, more concrete sentiment analysis needs to be considered in future work, for example, the debate regarding vaccine equity, and the common challenge of sarcasm.[81]

Nonetheless, analyzing and comparing the four different online vaccination debates as well as linking observed findings to existing theories in mass communication (e.g., agenda-setting and zero-sum theory) and psychology (e.g., omission bias) provide insightful information for understanding the underlying engagement and attention competition among different vaccination themes. The approach presented in this paper takes advantage of the recent availability of social media data, which enables, for the first time, to examine the vaccination debate and its relation to actual health outcomes over long periods of time. With the increased availability of such data sources, the method presented here could be expanded to assess the public's attention to other health-related issues and provide a basis for quantifying the effectiveness of health promotion policies.

Acknowledgements: The authors would like to thank the anonymous reviewers who provided valuable feedback which helped improve the quality of the paper.

Contributorship: QC: Methodology, Software, Formal analysis, Visualization, Writing - Original Draft, Review & Editing; AC[2]: Conceptualization, Writing – Original Draft, Review & Editing, AC[1]: Supervision, Conceptualization, Methodology, Writing - Review & Editing.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval and consent to participate: Not applicable, as this article utilized publicly available tweets, which is not considered as human subjects research. It did not entail the collection of data through interaction with individuals, nor any private information. Moreover, we adhered to Twitter's developer agreement and policy.

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Guarantor: Not applicable.

ORCID iD: Qingqing Chen https://orcid.org/0000-0002-3787-3134

References

1. Miyake E and Martin S. Long COVID: online patient narratives, public health communication and vaccine hesitancy. *Digit Health* 2021 Jan 1 [cited 2022 Aug 1]; 7: 20552076211059650. Available from: https://doi.org/10.1177/20552076211059649.
2. Phipps H. Vaccines: personal liberty or social responsibility? Academic Festival [Internet]. 2020 Apr 24; Available from: https://digitalcommons.sacredheart.edu/acadfest/2020/all/13.
3. Steffens MS, Dunn AG, Leask J, et al. Using social media for vaccination promotion: practices and challenges. *Digit Health* 2020 Jan 1 [cited 2022 Aug 1]; 6: 2055207620970785. Available from: https://doi.org/10.1177/2055207620970785.
4. Tram KH, Saeed S, Bradley C, et al. Deliberation, dissent, and distrust: understanding distinct drivers of coronavirus disease 2019 Vaccine Hesitancy in the United States. *Clin Infect Dis* 2022 Apr 15 [cited 2022 May 1]; 74: 1429–1441. Available from: https://doi.org/10.1093/cid/ciab633.
5. Stern AM and Markel H. The history of vaccines and immunization: familiar patterns, new challenges. *Health Aff* 2005 May 1 [cited 2022 Jan 26]; 24: 611–621. Available from: https://www.healthaffairs.org/doi/10.1377/hlthaff.24.3.611.
6. Wolfe RM and Sharp LK. Anti-vaccinationists past and present. *BMJ* 2002 Aug 24 [cited 2022 May 16]; 325: 430–432. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1123944/.
7. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm – An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine* 2012 May 28 [cited 2022 Jan 26]; 30: 3778–3789. Available from: https://www.sciencedirect.com/science/article/pii/S0264410X11019086.
8. Betsch C, Brewer NT, Brocard P, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. *Vaccine* 2012 May 28 [cited 2022 May 24]; 30: 3727–3733. Available from: https://www.sciencedirect.com/science/article/pii/S0264410X12002083.
9. Alonzo R, Hussain J, Stranges S, et al. Interplay between social media use, sleep quality, and mental health in youth: a systematic review. *Sleep Med Rev* 2021 Apr 1 [cited 2022 Nov 30]; 56: 101414. Available from: https://www.sciencedirect.com/science/article/pii/S108707922030157X.
10. Laranjo L. Social media and health behavior change. In: Syed-Abdul S, Gabarron E and Lau AYS (eds) *Participatory health through social media [internet].* London: Academic Press, 2016 [cited 2022 Nov 30], pp.83–111. Available from: https://www.sciencedirect.com/science/article/pii/B9780128092699000062.
11. Braeye T, Emborg HD, Llorente-García A, et al. Age-specific vaccination coverage estimates for influenza, human papillomavirus and measles containing vaccines from seven population-based healthcare databases from four EU countries - The ADVANCE project. *Vaccine* 2020 Apr 3; 38: 3243–3254.
12. Principi N and Esposito S. Vaccine-preventable diseases, vaccines and Guillain-Barre' syndrome. *Vaccine* 2019 Sep 3; 37: 5544–5550.
13. Gunaratne K, Coomes EA and Haghbayan H. Temporal trends in anti-vaccine discourse on Twitter. *Vaccine* 2019 Aug 14; 37: 4867–4871.
14. Keim-Malpass J, Mitchell EM, Sun E, et al. Using Twitter to understand public perceptions regarding the #HPV Vaccine: opportunities for public health nurses to engage in social marketing. *Public Health Nurs* 2017 Jul; 34: 316–323.
15. Jordan SE, Hovet SE, Fung ICH, et al. Using Twitter for public health surveillance from monitoring and prediction to public response. *Data* 2019 Mar [cited 2022 Mar 29]; 4: 6. Available from: https://www.mdpi.com/2306-5729/4/1/6.
16. Walsh L, Hyett N, Juniper N, et al. The use of social media as a tool for stakeholder engagement in health service design and quality improvement: a scoping review. *Digit Health* 2021 Jan 1 [cited 2022 Aug 1]; 7: 2055207621996870. Available from: https://doi.org/10.1177/2055207621996870.
17. Ameriks K and Clarke DM. *Aristotle: Nicomachean ethics*. Cambridge University Press, 2000, 268 p.
18. US Census Bureau. History - U.S. Census Bureau [Internet]. 2015 [cited 2022 Apr 16]. Available from: https://www.census.gov/history/www/homepage_archive/2015/october_2015.html.
19. Ragas MW and Roberts MS. Agenda setting and agenda melding in an age of horizontal and vertical media: a new theoretical lens for virtual brand communities. *Journal Mass Commun Q* 2009 Mar 1 [cited 2022 Apr 28]; 86: 45–64. Available from: https://doi.org/10.1177/107769900908600104.
20. McCombs ME and Shaw DL. The agenda-setting function of mass media. *Public Opin Q* 1972 Jan 1 [cited 2022 Apr 19]; 36: 176–187. Available from: https://doi.org/10.1086/267990.
21. McCombs M. Agenda setting function of mass media. *Public Relat Rev* 1977 Dec 1 [cited 2022 Apr 20]; 3: 89–95. Available from: https://www.sciencedirect.com/science/article/pii/S0363811177800088.

22. McCombs M. A look at agenda-setting: past, present and future. *Journal Stud* 2005 Nov 1 [cited 2022 Apr 20]; 6: 543–557. Available from: https://doi.org/10.1080/14616700500250438.
23. Green-Pedersen C and Wilkerson J. How agenda-setting attributes shape politics: basic dilemmas, problem attention and health politics developments in Denmark and the US. *J Eur Public Policy* 2006 Sep 1 [cited 2022 Apr 21]; 13: 1039–1052. Available from: https://doi.org/10.1080/13501760600924092.
24. Hester JB and Gibson R. The economy and second-level agenda setting: a time-series economy. *Journal Mass Commun Q* 2003 Mar 1 [cited 2022 Apr 21]; 80: 73–90. Available from: https://doi.org/10.1177/107769900308000106.
25. Kiousis S, Mitrook M, Wu X, et al. First- and second-level agenda-building and agenda-setting effects: exploring the linkages among candidate news releases, media coverage, and public opinion during the 2002 Florida gubernatorial election. *J Public Relat Res* 2006 Jul 1 [cited 2022 Apr 20]; 18: 265–285. Available from: https://doi.org/10.1207/s1532754xjprr1803_4.
26. Tedesco JC. Issue and strategy agenda setting in the 2004 presidential election: exploring the candidate–journalist relationship. *Journal Stud* 2005 May 1 [cited 2022 Apr 20]; 6: 187–201. Available from: https://doi.org/10.1080/14616700500057270.
27. Flay BR. Mass media and smoking cessation: a critical review. *Am J Public Health* 1987 Feb [cited 2022 Apr 21]; 77: 153–160. Available from: https://ajph.aphapublications.org/doi/10.2105/AJPH.77.2.153.
28. Pierce JP, Macaskill P and Hill D. Long-term effectiveness of mass media led antismoking campaigns in Australia. *Am J Public Health* 1990 May [cited 2022 Apr 21]; 80: 565–569. Available from: https://ajph.aphapublications.org/doi/10.2105/AJPH.80.5.565.
29. Meyer AJ. Skills training in a cardiovascular health education campaign. *J Consult Clin Psychol* 1980; 48: 129–142.
30. Redman S, Spencer EA and Sanson-fisher RW. The role of mass media in changing health-related behaviour: a critical appraisal of two models. *Health Promot Int* 1990 Jan 1 [cited 2022 Apr 21]; 5: 85–101. Available from: https://doi.org/10.1093/heapro/5.1.85.
31. Hacker JS. National health care reform: an idea whose time came and went. *J Health Polit Policy Law* 1996; 21: 647–696.
32. O'Reilly T. What is Web 2.0: Design patterns and business models for the next generation of software [Internet]. Rochester, NY: Social Science Research Network; 2007 Aug [cited 2022 Apr 27]. Report No.: 1008839. Available from: https://papers.ssrn.com/abstract=1008839.
33. Carazo-Barrantes C. Agenda-setting in a social media age: exploring new methodological approaches. *Agenda Sett J* 2021 Mar 31 [cited 2022 Apr 28]; 5: 31–55. Available from: https://www.jbe-platform.com/content/journals/10.1075/asj.20006.car.
34. Russell Neuman W, Guggenheim L, Mo Jang S, et al. The dynamics of public attention: agenda-setting theory meets big data. *J Commun* 2014 Apr 1 [cited 2022 Apr 28]; 64: 193–214. Available from: https://doi.org/10.1111/jcom.12088.
35. McCombs ME, Shaw DL and Weaver DH. New directions in agenda-setting theory and research. *Mass Commun Soc* 2014 Nov 2 [cited 2022 Apr 27]; 17: 781–802. Available from: http://www.tandfonline.com/doi/abs/10.1080/15205436.2014.964871.
36. Guo L, Vu H and McCombs M. An expanded perspective on agenda-setting effects: Exploring the third level of agenda setting. [Internet]. 2012 [cited 2022 Apr 28]. Available from: https://www.semanticscholar.org/paper/An-Expanded-Perspective-on-Agenda-Setting-Effects%3A-Guo-Vu/e39087404907426389a5322d41648d0e34e09d9c.
37. Guo L. The application of social network analysis in agenda setting research: a methodological exploration. *J Broadcast Electron Media* 2012 Oct 1 [cited 2022 Apr 28]; 56: 616–631. Available from: https://doi.org/10.1080/08838151.2012.732148.
38. Guo L and McCombs M. Network agenda setting: A third level of media effects. 2011;20.
39. Zhu JH. Issue competition and attention distraction: a zero-sum theory of agenda-setting. *Journal Mass Commun Q* 1992 Dec [cited 2022 May 31]; 69: 825–836. Available from: https://scholars.cityu.edu.hk/en/publications/issue-competition-and-attention-distraction(1e7c3461-ef7d-4650-9e12-d0a1a5790921).html.
40. Jang SM and Pasek J. Assessing the carrying capacity of Twitter and online news. *Mass Commun Soc* 2015 Sep 3 [cited 2022 May 31]; 18: 577–598. Available from: https://doi.org/10.1080/15205436.2015.1035397.
41. Lukito J and Pevehouse JC. Competing for attention on Twitter during the 2012 and 2016 U.S. presidential debates. *J Inf Technol Politics* 2022 Apr 19 [cited 2022 May 31]; 20: 1–14. Available from: https://doi.org/10.1080/19331681.2022.2063216.
42. Mo Jang S and Park YJ. Redirecting the focus of the agenda: testing the zero-sum dynamics of media attention in news and user-generated media. *Int J Commun* 2017 Jan [cited 2022 May 31]; 11: 3998–4017. Available from: https://gate.lib.buffalo.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=126813118&site=eds-live&scope=site.
43. Bibby K, Casson LW, Stachler E, et al. Ebola virus persistence in the environment: state of the knowledge and research needs. *Environ Sci Technol Lett* 2015 Jan 13 [cited 2022 May 1]; 2: 2–6. Available from: https://doi.org/10.1021/ez5003715.
44. Ali S, Gugliemini O, Harber S, et al. Environmental and social change drive the explosive emergence of Zika virus in the Americas. *PLOS Negl Trop Dis* 2017 Feb 9 [cited 2022 May 1]; 11: e0005135. Available from: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0005135.
45. Kunelius R. On the overlap of systemic events: COVID-19, climate, and journalism. *Soc Media Soc* 2020 Jul 1 [cited 2022 Jun 4]; 6: 2056305120948197. Available from: https://doi.org/10.1177/2056305120948197.
46. Lowe R, Barcellos C, Brasil P, et al. The Zika virus epidemic in Brazil: from discovery to future implications. *Int J Environ Res Public Health* 2018 Jan [cited 2022 May 1]; 15: 96. Available from: https://www.mdpi.com/1660-4601/15/1/96.
47. Menéndez C, Lucas A, Munguambe K, et al. Ebola crisis: the unequal impact on women and children's health. *Lancet Glob Health* 2015 Mar 1 [cited 2022 May 1]; 3: e130. Available from: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(15)70009-4/fulltext.
48. Saadat S, Rawtani D and Hussain CM. Environmental perspective of COVID-19. *Sci Total Environ* 2020 Aug 1 [cited 2021 Dec 17]; 728: 138870. Available from: https://www.sciencedirect.com/science/article/pii/S0048969720323871.

49. Saladino V, Algeri D and Auriemma V. The psychological and social impact of COVID-19: New perspectives of well-being. *Front Psychol* 2020 [cited 2022 Jan 26]; 11: 2550. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2020.577684.
50. Zhu Y, Wang R and Pu C. "I Am Chatbot, Your virtual mental health adviser." What drives citizens' satisfaction and Continuance intention toward mental health chatbots during the COVID-19 pandemic? An empirical study in China. *Digit Health* 2022 Jan 1 [cited 2022 Aug 1]; 8: 20552076221090030. Available from: https://doi.org/10.1177/20552076221090031.
51. Saad E, Sadiq S, Jamil R, et al. Predicting death risk analysis in fully vaccinated people using novel extreme regression-voting classifier. *Digit Health* 2022 Jan 1 [cited 2022 Aug 1]; 8: 20552076221109530. Available from: https://doi.org/10.1177/20552076221109530.
52. Blume S. Anti-vaccination movements and their interpretations. *Soc Sci Med* 2006 Feb; 62: 628–642.
53. Vanderslott S. Exploring the meaning of pro-vaccine activism across two countries. *Soc Sci Med* 2019 Feb 1 [cited 2022 Nov 29]; 222: 59–66. Available from: https://www.sciencedirect.com/science/article/pii/S0277953618307160.
54. Finney Rutten LJ, Zhu X, Leppin AL, et al. Evidence-based strategies for clinical organizations to address COVID-19 vaccine hesitancy. *Mayo Clin Proc* 2021 Mar; 96: 699–707.
55. Evrony A and Caplan A. The overlooked dangers of anti-vaccination groups' social media presence. *Human Vaccines Immunother* 2017 Apr 13 [cited 2022 Jan 26]; 13: 1475–1476. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5489284/.
56. Croitoru A, Crooks A, Radzikowski J, et al. Geosocial gauge: a system prototype for knowledge discovery from social media. *Int J Geogr Inf Sci* 2013 Dec 1 [cited 2021 Nov 10]; 27: 2483–2508. Available from: https://doi.org/10.1080/13658816.2013.825724.
57. Radzikowski J, Stefanidis A, Jacobsen KH, et al. The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Health Surveill* 2016 Jan 4 [cited 2022 Jan 26]; 2: e5059. Available from: https://publichealth.jmir.org/2016/1/e1.
58. CDC. COVID-19 Vaccination in the United States [Internet]. Centers for Disease Control and Prevention. 2022 [cited 2021 Dec 6]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/safety-of-vaccines.html.
59. CDC F vaccination. Influenza vaccination coverage for all ages [Internet]. Data.CDC.gov. 2022 [cited 2022 Mar 31]. Available from: https://data.cdc.gov/Flu-Vaccinations/Influenza-Vaccination-Coverage-for-All-Ages-6-Mont/vh55-3he6.
60. Teen Vaccination. Vaccination coverage among adolescents (13-17 Years) [Internet]. 2022 [cited 2022 Mar 31]. Available from: https://data.cdc.gov/Teen-Vaccinations/Vaccination-Coverage-among-Adolescents-13-17-Years/ee48-w5t6.
61. Zhang L and Liu B. Sentiment analysis and opinion mining. In: Encyclopedia of Machine Learning and Data Mining. 2017.
62. Chen Q and Crooks A. Analyzing the vaccination debate in social media data pre- and post-COVID-19 pandemic. *Int J Appl Earth Obs Geoinf* 2022 Jun 1 [cited 2022 Jun 1]; 110: 102783. Available from: https://www.sciencedirect.com/science/article/pii/S030324342200109X.
63. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012 Oct 15 [cited 2021 Nov 17]; 22: 276–282. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/.
64. Yuan X, Schuchard RJ and Crooks AT. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Soc Media + Soc* 2019 Apr 1 [cited 2021 Nov 17]; 5: 2056305119865465. Available from: https://doi.org/10.1177/2056305119865465.
65. Solomon B. demoji: Accurately remove and replace emojis in text strings [Internet]. 2021 [cited 2021 Dec 5]. Available from: https://github.com/bsolomon1124/demoji.
66. Effrosynidis D, Symeonidis S and Arampatzis A. A comparison of pre-processing techniques for Twitter sentiment analysis. In: Kamps J, Tsakonas G, Manolopoulos Y, Iliadis L and Karydis I (eds) *Research and advanced technology for digital libraries*. Cham: Springer International Publishing, 2017, pp.394–406.
67. Saif H, Fernandez M, He Y, et al. On stopwords, filtering and data sparsity for sentiment analysis of Twitter. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation [Internet]. Reykjavik, Iceland; 2014 [cited 2022 Jan 27]. p. 810–7. Available from: http://www.lrec-conf.org/proceedings/lrec2014/pdf/292_Paper.pdf.
68. Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. 2013 Jan 16 [cited 2021 Nov 28]; Available from: https://arxiv.org/abs/1301.3781v3.
69. Rezaeinia SM, Rahmani R, Ghodsi A, et al. Sentiment analysis based on improved pre-trained word embeddings. *Expert Syst Appl* 2019 Mar 1 [cited 2021 Nov 28]; 117: 139–147. Available from: https://www.sciencedirect.com/science/article/pii/S095741741830558X.
70. Mohammad S and Turney P. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text [Internet]. Los Angeles, CA: Association for Computational Linguistics; 2010 [cited 2022 Aug 2]. p. 26–34. Available from: https://aclanthology.org/W10-0204.
71. Mohammad SM and Turney PD. Crowdsourcing a word–emotion association lexicon. *Comput Intell* 2013 [cited 2022 Aug 2]; 29: 436–465. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8640.2012.00460.x.
72. McCombs M and Weaver D. Voters' need for orientation and use of mass communication. 1973 Apr [cited 2022 May 19]; Available from: https://eric.ed.gov/?id=ED077061.
73. CDC. Similarities and differences between Flu and COVID-19 [Internet]. Centers for Disease Control and Prevention. 2022 [cited 2022 Jun 2]. Available from: https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm.
74. Larsen JR, Martin MR, Martin JD, et al. Modeling the onset of symptoms of COVID-19. *Front Public Health* 2020 [cited 2022 May 16]; 8, Available from: https://www.frontiersin.org/article/10.3389/fpubh.2020.00473.
75. Monto AS, Gravenstein S, Elliott M, et al. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med* 2000 Nov 27; 160: 3243–3247.
76. Bruine de Bruin W, Carman KG and Parker AM. Mental associations with COVID-19 and how they relate with self-reported protective behaviors: a national survey in the United States. *Soc Sci Med* 2021 Apr 1 [cited 2022 Jun 2];

275: 113825. Available from: https://www.sciencedirect.com/science/article/pii/S027795362100157X.
77. Roman PC. Influenza vaccinations during the COVID-19 pandemic — 11 U.S. Jurisdictions, September–December 2020. MMWR Morb Mortal Wkly Rep [Internet]. 2021 [cited 2022 May 24];70. Available from: https://www.cdc.gov/mmwr/volumes/70/wr/mm7045a3.htm.
78. Asch DA, Baron J, Hershey JC, et al. Omission bias and Pertussis vaccination. *Med Decis Making* 1994 Apr 1 [cited 2022 Jun 3]; 14: 118–123. Available from: https://doi.org/10.1177/0272989X9401400204.
79. Ritov I and Baron J. Reluctance to vaccinate: omission bias and ambiguity. *J Behav Decis Making* 1990 [cited 2022 Jun 3]; 3: 263–277. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.3960030404.
80. Sweeny K, Shepperd JA and Howell JL. Do as I Say (Not as I Do): inconsistency between behavior and values. *Basic Appl Soc Psychol* 2012 Mar 1 [cited 2022 Jun 3]; 34: 128–135. Available from: https://doi.org/10.1080/01973533.2012.655998.
81. Ebrahimi M, Yazdavar AH and Sheth A. Challenges of sentiment analysis for dynamic events. *IEEE Intell Syst* 2017 Sep; 32: 70–75.

Appendix

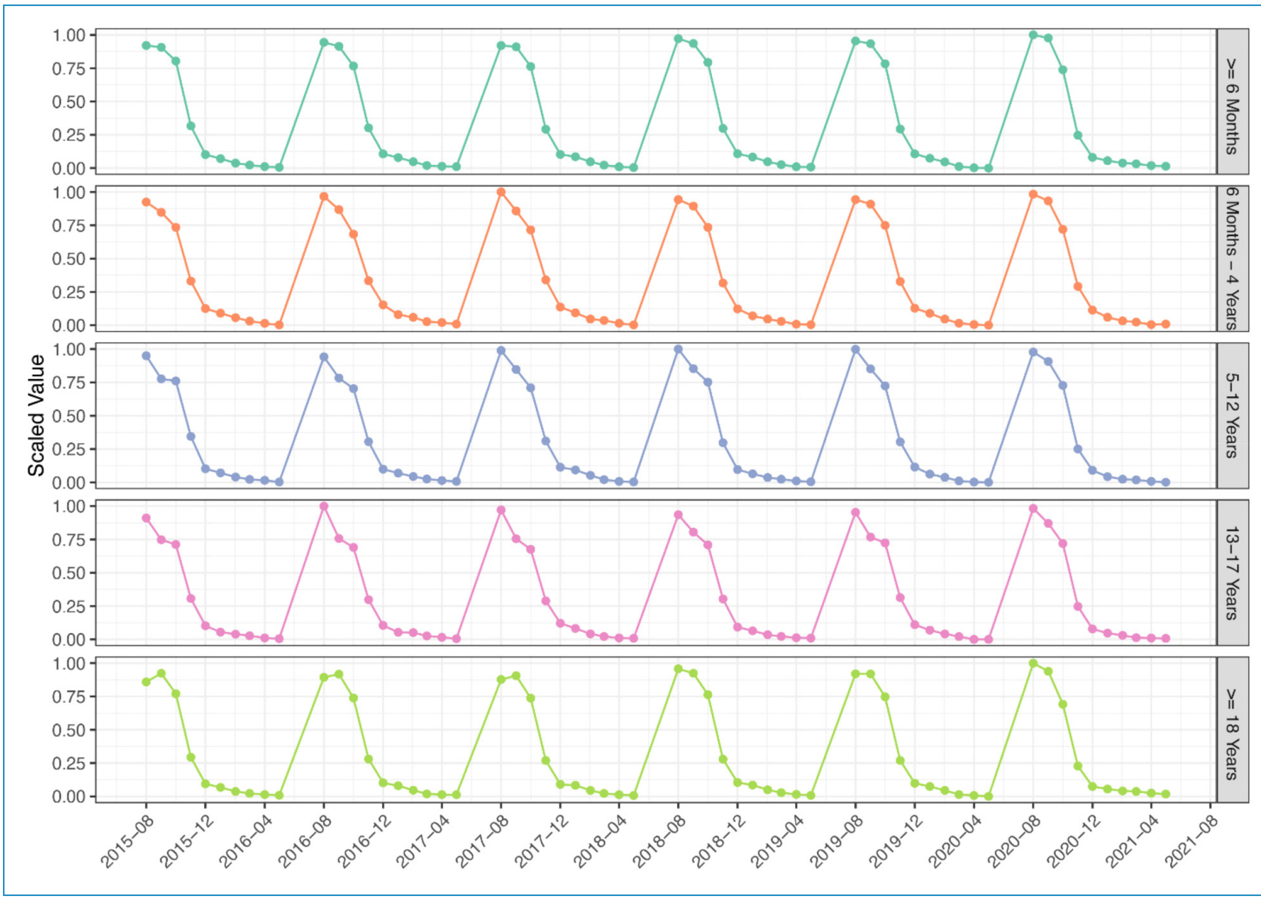

Figure A1. Scaled influenza vaccination coverage by different ages.