COVID-19 Vaccine Perceptions in the Initial Phases of US Vaccine Roll-out: An Observational Study on Reddit

Navin Kumar (navin.kumar@yale.edu)  
Yale University

Isabel Corpus  
Yale University

Meher Hans  
Yale University

Nikhil Harle  
Yale University

Nan Yang  
Yale University

Curtis McDonald  
Yale University

Shinpei Nakamura Sakai  
Yale University

Kamila Janmohamed  
Yale University

Keyu Chen  
Yale University

Frederick L. Altice  
Yale University

Weiming Tang  
University of North Carolina Project-China

Jason L. Schwartz  
Yale University

S. Mo Jones-Jang  
Boston College

Koustuv Saha  
Microsoft Research – Montréal

Shahan Ali Memon  
New York University Abu Dhabi

Chris T. Bauch
COVID-19 vaccine perceptions in the initial phases of US vaccine roll-out: An observational study on Reddit

Navin Kumar1*, Isabel Corpus2, Meher Hans2, Nikhil Harle2, Nan Yang3, Curtis McDonald4, Shinpei Nakamura Sakai4, Kamila Janmohamed2, Keyu Chen1, Frederick L. Altice1,19, Weiming Tang5,6,7, Jason L. Schwartz8, S. Mo Jones-Jang9, Koustuv Saha10, Shahan Ali Memon11, Chris T. Bauch12, Munmun De Choudhury13, Orestis Papakyriakopoulos14, Joseph D. Tucker5,15,16, Abhay Goyal17, Aman Tyagi18, Kaveh Khoshnood19 and Saad Omer20

*Correspondence: Navin.kumar@yale.edu

1Section of Infectious Diseases, Yale School of Medicine, New Haven, CT, USA
2Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
3Department of Statistics, Yale University, New Haven, CT, USA
4University of North Carolina Project-China, Guangzhou, China
5Social Entrepreneurship to Spur Health (SESH) Global, Guangzhou, China
6University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
7Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA
8Department of Communications, Boston College, Boston, MA, USA
9Microsoft Research Lab, Montreal, Quebec, Canada
10New York University, Abu Dhabi, UAE
11Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada
12School of Interactive Computing, Georgia Tech, Atlanta, GA, USA
13Center for Information Technology Policy, Princeton University, Princeton, NJ, USA
14School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
15Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, USA
16Department of Computer Science, Stony Brook University, New York, NY, USA
17Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA
18Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA
19Yale Institute for Global Health, New Haven, CT, USA
20Full list of author information is available at the end of the article
Abstract

Background: Open online forums like Reddit provide an opportunity to quantitatively examine COVID-19 vaccine perceptions early in the vaccine timeline. We examine COVID-19 misinformation on Reddit following vaccine scientific announcements, in the initial phases of the vaccine timeline.

Methods: We collected all posts on Reddit (reddit.com) from January 1 2020 - December 14 2020 (n=266,840) that contained both COVID-19 and vaccine-related keywords. We used topic modeling to understand changes in word prevalence within topics after the release of vaccine trial data. Social network analysis was also conducted to determine the relationship between Reddit communities (subreddits) that shared COVID-19 vaccine posts, and the movement of posts between subreddits.

Results: There was an association between a Pfizer press release reporting 90% efficacy and increased discussion on vaccine misinformation. We observed an association between Johnson and Johnson temporarily halting its vaccine trials and reduced misinformation. We found that information skeptical of vaccination was first posted in a subreddit (r/Coronavirus) which favored accurate information and then reposted in subreddits associated with antivaccine beliefs and conspiracy theories (e.g. conspiracy, NoNewNormal).

Conclusions: Our findings can inform the development of interventions where individuals determine the accuracy of vaccine information, and communications campaigns to improve COVID-19 vaccine perceptions, early in the vaccine timeline. Such efforts can increase individual- and population-level awareness of accurate and scientifically sound information regarding vaccines and thereby improve attitudes about vaccines, especially in the early phases of vaccine roll-out. Further research is needed to understand how social media can contribute to COVID-19 vaccination services. Text for this section.

Keywords: COVID-19; Vaccine; Reddit; Computational; Misinformation

Introduction

The first vaccine against COVID-19 from Pfizer-BioNTech received emergency use authorization (EUA) on December 11, 2020 [1]. On August 23, 2021, The US FDA issued a Letter of Approval to Pfizer for full use of the vaccine, under the name COMIRNATY for use in persons aged 16 years and older and delayed pediatric approval for those younger than 12 years [2]. From early 2021 on, COVID-19 vaccines have helped alleviate the pandemic’s burden on society by mitigating contagion, protecting the population against severe disease, and allowing for less restrictive measures, especially in countries with high vaccine uptake and availability [3]. Since mid-2021, vaccine availability does not pose a problem in high-income countries. Instead, these countries face the challenge that vaccine-hesitant and vaccine-denial individuals pose to the timely completion of vaccination programs [4]. There are a diverse range of individuals who are skeptics of vaccination including those who are antivaccine or antivaxxers (individuals who are opposed to vaccination or laws that require vaccination) and those who are vaccine hesitant (those who delay in acceptance or refusal of vaccination) [5]. The diverse groups who are skeptical of vaccines may react to information in different ways [6]. Often, the lack of individuals
willing to receive the vaccine at a given moment has caused the expiration and discard of available vaccine doses [7]. Therefore, it is crucial for these countries, especially the US, to understand the drivers of vaccine hesitancy [8] and implement timely initiatives for re-selling or donating surplus doses to countries where they are needed. More recently, a new wave of COVID-19 cases caused by the highly transmissible delta and omicron variants is exacerbating the worldwide public health crisis, and has led to consideration of the potential need for, and optimal timing of, booster doses for vaccinated populations [9].

Thus, for these vaccines to be successful, they not only need to be deemed safe and effective by scientists, but also widely accepted by the public [10]. Effective health communication is key to vaccine acceptance, but is a complex task given widespread vaccine hesitancy, rapidly changing vaccine information [11], and vaccine misinformation [12]. Misinformation is defined as information that has the features of being false, determined based on expert evidence, but shared with no intention of harm [13]. Such information may worsen existing fear around a vaccine and limit public uptake of a COVID-19 vaccine and its boosters [12]. With low willingness to vaccinate globally [14], and substantial COVID-19 misinformation [15], achieving sufficient vaccination coverage to reach population-level benefits will be challenging.

Reduced vaccine uptake may impinge on population-level impact [16], and COVID-19 control at the population level [4]. For example, reduced vaccine uptake may increase the mortality cost of COVID-19 [17] and create clusters of non-vaccinators that disproportionately increase pandemic spread [18]. In addition, willingness to accept a COVID-19 vaccine seems to be fluctuating in the US [19]. Thus, vaccine acceptance is not constant or uniform, and likely affected by several factors, such as being responsive to information and perceptions regarding the vaccine, and the state of the pandemic and economy.

Several studies have detailed the relationship between exposure to COVID-19 misinformation and vaccine acceptance [20], as well as COVID-19 vaccine perceptions assessed via Twitter [21, 22] and online surveys [23, 24]. To our knowledge, relatively few studies have focused on Reddit (reddit.com), a social news aggregation and discussion website. Registered Reddit members submit posts (text, images, videos) to the site, which are then voted up or down by other members. Posts are organized by subject into user-created boards called communities or subreddits, which cover a large range of topics. Reddit may be a useful setting for examining vaccine perceptions because similar topics have been discussed before [25], including topics related to COVID-19 vaccine development [26]. Moreover, as seen with the recent GameStop trading event, Reddit is increasingly important in online conversations [27]. We note that Reddit and similar online sources are not necessarily representative of what the overall US general public feels [28]. However, Reddit provides insights on highly shared news, and can rapidly transmit both misinformation and accurate information [29–31].

Recent work observed how the HPV vaccine is characterized on Reddit over time and by user gender. Findings demonstrated that women and men both discussed HPV, highlighting that Reddit users do not perceive HPV as an issue that only pertains to women [32]. A similar study indicated that Reddit users perceived the HPV vaccine domain from a virus-framed perspective that could impact their lifestyle
choices and that their awareness of the HPV vaccine for cancer prevention is also lacking [33]. Regarding COVID-19, researchers used sentiment analysis and topic modeling on data collected from Reddit communities focusing on the COVID-19 vaccine from Dec 1, 2020, to May 15, 2021, finding that sentiments expressed in these communities are overall more positive than negative and have not meaningfully changed since December 2020 [34]. Another study used topic modelling to generate latent topics from user generated Reddit corpora on reasons for vaccine hesitancy, finding factors such as fear of risks and side effects, and lack of trust in policymakers [35]. A study using COVID-19 Reddit data and topic modelling found that during the pandemic, the proportion of Reddit comments predominated by conspiracy theories outweighed that of any other topics [36]. However, limited research has explored how online vaccine perceptions are associated with major events in the early in the vaccine development and implementation timeline (e.g. major pharmaceutical firms halting vaccine trials or publishing results on vaccine effectiveness) and how online vaccine discussions move across arenas that have different baseline vaccine perceptions.

We thus propose a study to detail the behavior of top Reddit users, posts’ relationship with events in the initial phases of vaccine timeline, and the relationship between subreddits that shared COVID-19 vaccine posts. We provide an overview of Reddit conversations around the COVID-19 vaccine from January 1 2020 - December 14 2020, to give understanding of vaccine narratives when vaccines were first trialed and introduced. It is important to understand the behavior of top users, how vaccine perceptions are related to events in the vaccine timeline and how vaccine discussion on Reddit migrates across subreddits that differ in their vaccine perceptions, to mitigate vaccine misinformation early in the vaccine development timeline. Most users of online platforms are passive or participate with a very low frequency. A small number of Reddit users are hyperactive and may over-proportionally influence vaccine perceptions online [37]. Thus, describing the behavior of hyperactive users is key to understanding shifts in vaccine perceptions, early in the vaccine timeline. Understanding how perceptions are related to initial vaccine-related events may allow stakeholders to better design communication and education campaigns [38, 39] in response to early vaccine distribution setbacks. Given the range of vaccine-related viewpoints online, greater insight on how discussions move across Reddit communities will allow stakeholders to better disseminate evidence-based information on Reddit. The purpose of this analysis was to detail the behavior of top Reddit users, posts’ relationship with events early in the vaccine timeline, and the relationship between subreddits that shared COVID-19 vaccine posts. Our findings hope to inform stakeholders on how to manage online narratives around vaccines early in the vaccine timeline, to mitigate misinformation as it arises. Developing vaccine misinformation mitigation techniques early in the vaccine timeline is critical to managing misinformation before it proliferates later in the vaccine timeline.

**Methods**

**Data acquisition and processing**

Using the Pushshift API and the Python Reddit API Wrapper [40, 41], we collected all posts on the entire Reddit (reddit.com), across all subreddits from January 1
December 14 2020 that contained both COVID-19 and vaccine keywords (see Supplement, only posts that had COVID-19 AND vaccine-related keywords were collected) derived from systematic reviews on the topic. Reddit is a publicly available website. We also collected metadata for each post e.g. the username, ID, subreddit. We then preprocessed our data as follows: 1) removed duplicate entries; 2) filtered out entries <50 characters as these generally do not provide enough information for meaningful analysis [42, 43]; 3) filtered the content using a curated set of search terms (as shown in the Supplement) to retain only COVID-19 vaccine-related content; 4) removed text in non-English languages, URLs, emojis, and punctuation.

Hyperactive users
To better understand the possibly outsize influence of some individuals, we provided a descriptive overview of the behavior of top 10 users, focusing on content and number of posts.

Topic modeling
We used topic modeling to understand changes in word prevalence within topics around COVID-19 vaccines (see Supplement for additional detail). Topic modeling is a computer-aided content analysis technique through which texts are organized into themes known as “topics” [44, 45]. We used an approach to topic modeling known as Structural Topic modeling (STM) [46, 47]. STMs [46, 47] enable the generation of topics with regards to document metadata such as date and source and other covariates relevant to the research question, such as new COVID-19 cases, and thus was used instead of other topic modelling methods. We used the following metadata covariates for the STM model: date (1 was denoted for the first day and numbered sequentially after), new COVID-19 cases per day worldwide, new COVID-19 deaths per day worldwide (publicly available and both obtained from COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University [48]), S and P 500 opening score (publicly available from the Wall Street Journal), post type (comment or post), score (upvotes - downvotes). We used worldwide cases and deaths instead of US cases/deaths as Reddit COVID-19 discussion centers on pandemic progression both globally and in the US, despite most users being from the US. These control variables may address underlying factors possibly influencing vaccine perceptions. By considering a broader picture of what may influence topic proportions around vaccine discussion, we can better test the claims relation to the association between specific events and topic proportions.

March 11 2020 was denoted as the start date for our analysis, the date the World Health Organization declared COVID-19 a pandemic [49].

As STM is an unsupervised approach, the number of topics (k) to estimate is key to the analysis. We first estimated several models ranging from 5 to 30 topics. These models were then evaluated qualitatively by two authors (IC, AG) independently for 1) their ability to produce coherent topics and 2) appropriately capture topics regarding COVID-19 vaccination [50]. The two authors agreed on the same topic solution (k=20). Topic interpretation was influenced by authors' first reading the top 100 most-cited COVID-19 peer-reviewed research articles and the top 10 most cited peer-reviewed research articles around topic modeling. Two authors assigned topics
(IC, AG) [Cohen’s kappa (k) >0.8] and a third author (NK) resolved disagreements when they arose [Cohen’s kappa (k) >0.8].

We also detailed how events in the vaccine timeline (described in following section) were associated with topic prevalence. We generated linear regression models with expected topic proportions for each topic as dependent variables and vaccine events as main explanatory variables, with the following additional covariates: new COVID-19 cases per day worldwide, new COVID-19 deaths per day worldwide, SP 500 opening score, post type. We first conducted a visual examination on the pattern of the time series by plotting them and generating auto-correlation and partial correlation plots. No seasonal patterns were identified. Auto-correlation was tested with the Durbin-Watson test. Nonstationarity was identified using the augmented Dickey-Fuller test and corrected through differencing. To validate regression analyses in Figure 1, we undertook a close reading of the 100 most representative text fragments for exemplar topics. We found that these topics varied in line with the indicated events.

Selecting events of interest
We first assembled a preliminary list of COVID-19 vaccine-related events based on a review of online news sites and peer-reviewed vaccine research articles and consultation with experts on vaccination, resulting in a list of six events (see Supplement). We then conducted preliminary analyses with remaining events to determine the ones that were associated with the greatest shift in topic proportions for each topic. The three events below were selected as our final list, as these were associated with a shift in topic proportions for most topics. Events as follows: 1) AstraZeneca halts Phase 3 vaccine trial (September 8 2020); 2) Johnson Johnson temporarily halts vaccine trial (October 12 2020); 3) Pfizer announces preliminary vaccine clinical trial results showing 90% efficacy (November 9 2020).

Social network analysis
Next, we conducted social network analysis to provide insights on how vaccine discussion on Reddit migrates across subreddits that differ in their vaccine perceptions, and the relationship between these subreddits [51]. While standard social networks tend to assess relationships between people, we used a network to describe relationships between subreddits, studying the connections between people as mediated by the subreddits they were in and the posts shared between these subreddits. We used node sizes to represent number of users in a subreddit, node edges to indicate shared COVID-19 vaccine posts between subreddits (an edge was indicated if there was >one shared post), and node labels to detail the subreddit name. Edge direction was based on whether a node had >50% of its posts made prior to its adjacent connecting node e.g. A->B if >50% of A’s shared posts were made before B. When analyzing the trajectory of posts from one subreddit to another, we assumed that posts moved from A->B->C if a post was made first in A, then followed by B and C. This may allow us to see how posts moved from one subreddit to another. We used the fast-greedy algorithm for cluster identification. We focused on the main social network in our data (largest component subgraph) and excluded all edges with a weight of one (i.e. all connections between subreddits that had only one
post in common) and all clusters that had <15 vertices and whose vertices had a betweenness centrality <20 (we used a range of network characteristics to yield an easy to understand social network and the above measures yielded the clearest output).

**Results**

Post-processing, we had 266,840 documents (25,400,556 words).

**Overview of hyperactive users**

We reviewed the posts for the top 10 users who posted the most in our dataset, ranging from 159 - 278 posts/person. Six of these users posted evidence-based information (e.g. Effectiveness of the COVID-19 vaccine: real-world evidence from healthcare workers, Vaccine linked to reduction in risk of COVID-19 admissions to hospitals), but four users (one of these users was suspended from Reddit at time of writing) seemed to be skeptical of vaccination (e.g. Hell Gates says Vaccines are Americans’ only hope to return to Normal Life!, Doctors Around the World Issue Dire Warning: DO NOT get the experimental covid vaccine, At What Point Do We Realize Bill Gates Is Dangerously Insane?). Individuals skeptical of vaccination were common among those who posted the most frequently in our data.

**Topic modeling**

Table 1 indicated the topics in the dataset, their proportions, and the top 10 words for each topic (see Table 1). Broadly, our data centered on the severity of the pandemic (Topic 13), hope for a swift end to the pandemic (Topic 10), and suspicion of science and mainstream media (Topic 5). The severity of the pandemic topic focused on death, risk and sickness in relation to the pandemic. The hope for a swift end to the pandemic topic was about hope, safety and the length of the pandemic. Finally, the suspicion of science topic was around reduced trust and belief in the media and science. We also noted several other topics, such as evidence-based COVID-19 discussion (exploring factually sound and true evidence about COVID-19) (Topic 20), COVID-19 transmission rates and patterns (Topic 12), the effect of the virus on humans (Topic 2), COVID-19 vaccine conspiracy theories and misinformation (e.g. Bill Gates-related vaccine conspiracy theories) (Topic 19), and racism on social media (Topic 15).

We then explored how various events in the vaccine timeline were related to topic prevalence (see Figure 1). We observed an association between AstraZeneca temporarily halting its vaccine trials, and increased discussion around government expenses ($\beta_{\text{intercept}} = 5.019e-02, p < 0.001$), such as funds spent on businesses, and supporting the economy. Similarly, we found an association between Johnson and Johnson temporarily halting its vaccine trial, and increased discussion of federal policies ($\beta_{\text{intercept}} = -1.08e-2$) and then-US President Donald Trump’s leadership ($\beta_{\text{intercept}} = 4.330e-02, p < 0.001$). We found an association among Johnson and Johnson temporarily halting its vaccine trial and greater discussion around suspicion of science and mainstream media ($\beta_{\text{intercept}} = 7.536e-02, p < 0.001$). Similarly, there was an association between Pfizer announcing preliminary Phase 3 results showing 90% vaccine efficacy and reduced discussion about suspicion of
science and mainstream media ($\beta_{intercept} = 1.027e-01, p < 0.001$). We detailed an association between Johnson and Johnson temporarily halting its vaccine trial and reduced discussion around COVID-19 vaccine conspiracy theories and misinformation ($\beta_{intercept} = 2.619e-02, p < 0.05$). We found an association between Pfizer announcing preliminary vaccine clinical trial results, an increase in discussion around COVID-19 vaccine conspiracy theories and misinformation ($\beta_{intercept} = -1.11e-3$), and a corresponding decrease in evidence-based COVID-19 discussion ($\beta_{intercept} = 5.291e-02, p < 0.001$). We also found that new COVID-19 deaths and cases were positively associated with increased discussion around COVID-19 vaccine conspiracy theories and misinformation ($\beta_{intercept} = 6.517e-07, p < 0.01$), and suspicion of science and mainstream media ($\beta_{intercept} = 2.99e-7$), highlighting the relationship between COVID-19 progression and similar rises in misinformation (See Supplement for full results).

**Subreddit networks**

To understand the relationship between subreddits that shared COVID-19 vaccine posts, we analyzed the greatest component subgraph, as this was substantially larger than all other subgraphs which had 1-5 nodes and did not provide for meaningful conclusions (see Figure 2). The largest node/subreddit (r/Coronavirus, the official community for COVID-19 on Reddit) had 2.4 million users. Nodes connected by an edge shared two to 41 posts. We found nine posts that were first posted in r/Coronavirus and then subsequently posted in at least one subreddit. Posts were reposted one to 10 times. Eight of these posts concerned evidence-based information (e.g. COVID-19 timeline, Vaccine development timeline) and were reposted in other subreddits favoring evidence-based information (e.g. AmericanPolitics, worldnews). However, one post (COVID-19 much milder than believed) was aligned with vaccination skepticism and subsequently posted in subreddits favoring vaccine skeptic narratives (e.g. conspiracy, NoNewNormal - we read through the first 50 posts in these subreddits and verified they were largely around disagreement with evidence-based measures to mitigate the pandemic). This suggests that misinformation is present in some subreddits which generally feature accurate information. This may also indicate that most posts which start in the main COVID-19 subreddit (r/Coronavirus) and then re-posted in other subreddits tend to be evidence-based. However, a minority of posts in r/Coronavirus are skeptical of vaccination, but then do not get reposted in evidence-based subreddits, but instead in subreddits broadly skeptical of vaccination.

**Discussion**

Our analysis of 266,840 posts on COVID-19 vaccines between March 11 2020 - December 14 2020 generated several key findings, useful for understanding the early stages of the COVID-19 vaccine timeline. First, there was a relationship between interim positive announcements followed by increased vaccine misinformation, and a relationship between halting vaccine trials and reduced misinformation discussion. Past research has indicated shifts in vaccine perceptions with time [52, 53]. We expand on that work, suggesting an association between events early in the
vaccine timeline and vaccine perceptions. Information skeptical of vaccination may flow from a regulated and legitimate source to avenues centering on misinformation and distrust in science. Previous research indicated how antivaccine posts travel online, with users largely moving from one antivaccine post to another [54, 55]. Building on this work, we propose that individuals skeptical of vaccination may selectively highlight posts from legitimate online environments and then forward these posts in arenas aligned with vaccine-skeptic narratives, moving information that was previously under the purview of a more neutral, science-trusting audience to individuals skeptical of vaccination - perhaps providing opportunities to engage such individuals and reduce misinformation. The strength of our work is the use of computational methods to explore how Reddit vaccine perceptions are associated with events early in the vaccine timeline and how posts move among environments with differing vaccine perceptions. Such outcome measurement is central to understanding how vaccine perceptions shift early in the vaccine timeline. Findings may allow for accurate public health messaging when vaccines are first announced, capable of improving COVID-19 vaccine perceptions in the critical initial periods of the vaccine timeline.

There was an association between positive vaccine developments and an increase in discussion of COVID-19 vaccine misinformation, and a relationship between development setbacks and reduced misinformation discussion. Past research has indicated shifting vaccine perceptions over time [53], but there is limited research on specific events, especially early vaccine trials and their relationship with vaccine perceptions. Previous work also indicated that COVID-19 misinformation can be remedied with scientific facts [56], but we highlight the complexity of the phenomenon. The spread and production of misinformation can sometimes be due to confirmation bias, where individuals consume, interpret, and favor information that supports their beliefs [57]. For true antivaxxers, COVID-19 vaccine successes may be interpreted as attempts by Bill Gates to track the population through microchips, and thus news around vaccine successes may be interpreted in a misinformation framework, perhaps explaining the relationship between vaccine success and increased discussion around COVID-19 vaccine conspiracy theories. Similarly, when vaccine trials are halted, such news may cohere with antivaxxers, who may have no interest in engaging with news that possibly demonstrates the failure of medical science - given antivaxxers’ distrust of medical experts [58], perhaps explaining the reduced misinformation discussion. Thus, simply presenting scientific data to antivaxxers [56] may not be effective, as demonstrated in a study where presenting some antivaxxers with facts made them more antivaccine [59].

There was a relationship between an early vaccine trial halting and increased discussion around suspicion of science and mainstream media, and a vaccine trial being effective and reduced discussion around suspicion of science and mainstream media. Factors such as political conservatism and lower levels of education may be associated with lack of trust in science [60], and we build on such research by suggesting that news around science successes and setbacks is associated with trust in science. In an environment where individuals are unsure what to believe around vaccines [61], we propose that early vaccine successes build faith in science, and vaccine setbacks erodes this trust.
We also documented how posts skeptical of vaccination, early in the vaccine timeline, may move from more legitimate avenues to arenas where vaccine-skeptical narratives are more popular. In addition, such posts were popular among some highly active users in our dataset. COVID-19 misinformation is present in mainstream environments and does not always get fact-checked [62] and Reddit is no different. Individuals with largely antivaccine beliefs seek out information that coheres with their views [54]. We build on this work and suggest that individuals skeptical of vaccination, early in the vaccine timeline, also look for information from venues that tend to have evidence-based discussion, but then may interpret such information in line with their views and moral foundations, later sharing such information in forums more skeptical of vaccination. This may indicate that skeptics of vaccination do venture out of their echo chambers to enter spaces where accurate information is the norm - presenting attractive opportunities for constructive intervention.

To improve COVID-19 vaccine perceptions, especially early in the vaccine timeline, minimize misinformation, and increase vaccination rates, public health authorities should conduct tailored interventions and communications campaigns to counter the rhetoric of vaccine misinformation [63, 64]. An example intervention could ask respondents to determine information accuracy around vaccines [65, 66] nudging individuals through the design of these programs toward accurate vaccine information. It is possible that interventions of this sort could shift the beliefs of the vaccine hesitant and thereby boost vaccine uptake, despite potentially little or no effect on committed opponents of vaccination. The concomitant spread of misinformation about COVID-19 vaccines and scientific implications provides insights about the mechanism of misinformation spread. Given our findings around a vaccine trial halting and increased discussion around suspicion of science, we suggest that scientists be more communicative on the difficulties they face in creating vaccines to mitigate science mistrust. Communications campaigns can harness these findings and forward evidence-based posts in subreddits where misinformation is common, when vaccine trial data is released. Given the possibility that individuals seemingly more interested in antivaccine narratives may sometimes venture into more evidence-based environments, interventions can target skeptics or critics of vaccination who sometimes enter more mainstream spaces, engaging them with more evidence-based information, keeping in mind how antivaxxers may deal with such information. Similarly, as legitimate online spaces contain COVID-19 vaccine misinformation, more effective moderation policies can be enacted in these and similar environments e.g. perhaps including a "verified" tag to a post if it comes from a credible source. Such measures may augment health outcomes through several modes. For example, improved vaccine perceptions, especially early in the vaccine development timeline, may lead to reduced vaccine hesitancy and thereby increase vaccine acceptance and COVID-19 vaccination rates. Reduced vaccine misinformation may also improve trust in science and health systems, more broadly, enhancing larger efforts to address health disparities observed in vaccination coverage and many other areas [67].

Limitations
Our findings relied on the validity of data collected with our search terms. We searched all of Reddit for COVID-19 vaccine posts, and our data contained text
fragments representative of vaccine perceptions. We are thus confident in the comprehensiveness of our data. Any use of Reddit data presents several challenges and limitations. As no personal information is collected on Reddit, the demographic makeup of users is unknown [68]. Some evidence suggests that Reddit users are likely to be male, younger than the general population and mostly based in the US [69]. Our results should be interpreted in line with this probable gender and age skew of our data. We note the rapidly changing situation of the COVID-19 pandemic where our data does not reflect the latest situation of the pandemic. We instead provide a cross-sectional overview of the pandemic when vaccine developments were first reported, supplying information stakeholders can utilize for future vaccine roll-outs. We note that the time period of analysis witnessed major US political polarization, major economic shifts in economy, and changes in social lives which may explain some of the variation in our results. Future work will attempt to control for these factors.

It was not possible to determine what posts were viewed by skeptics of vaccination in more legitimate subreddits, but subsequently not reposted in subreddits more supportive of antivaccine narratives, thereby providing more support for our suggestion around confirmation bias. It is possible that posts were made in one subreddit before another purely due to chance, and that the directionality assumed is due to coincidence. We cannot be certain why individuals created the text in our data, the processes behind the shift in narratives, and why individuals shared the same post in more than one subreddit, and we cannot address these mechanisms with our data. Future work can address these questions and explore the motivations of those creating and sharing such text. We conducted a retrospective and observational study, and thus cannot draw causal conclusions regarding vaccine perceptions. It is possible that other vaccine-related events may have caused the observed changes, and that vaccine success stimulated debate that brought to the surface existing antivaccine discussion, instead of causing it.

**Conclusion**

There was a relationship between perceived vaccine efficacy and increased discussion around vaccine misinformation, early in the vaccine development timeline, and an association between halting of vaccine trials and reduced misinformation discussion. Posts skeptical of vaccination can move from more legitimate avenues to arenas where antivaccine narratives are more popular, suggesting that individuals skeptical or opposed to vaccines do venture out of their online echo chambers, providing potential opportunities to engage such individuals and reduce misinformation. Given the time period of our data extends to the period immediately prior to the launch of large-scale vaccination among the US public, stakeholders may utilize our findings to better design and inform upcoming vaccination communication efforts, for a range of future vaccination roll-outs.

**Declarations**

**Acknowledgements**

Study was funded by the Yale Institute for Global Health and the Whitney and Betty MacMillan Center for International and Area Studies at Yale University. The funding bodies had no role in the design, analysis or interpretation of the data in the study.
Funding
Study was funded by the Yale Institute for Global Health and the Whitney and Macmillan Center for International and Area Studies at Yale University. The funding bodies had no role in the design, analysis or interpretation of the data in the study.

Availability of data and materials
The datasets used and analyzed during the current study available from the corresponding author on reasonable request. The data used in the study is available at this link: drive.google.com/drive/folders/19ZZB5w0pF8LKGIB7-PUwxpVv4angmuQ?usp=sharing

Ethics approval and consent to participate
Approval and informed consent were not needed as we used an anonymized dataset. Yale University IRB committee guidelines waived the need for informed consent and ethical approval. Research was performed in accordance with the Declaration of Helsinki. This study was pre-registered on the Open Science Framework (OSF.IO/urp2a).

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable

Authors’ contributions
All authors made significant contributions to the manuscript. The following were the respective roles for each author: NK, AG, CM, IC, KJ, MH, NY, NH, SNS, KK, SO contributed to the study design, hypothesis generation, data collection, data analysis, data interpretation, and manuscript write-up and review. MDC, JDT, OP, WT, CB, AT, JLS, SMJ, SAM, KC, FLA, KS contributed to the manuscript write-up and review.

Author details
1 Section of Infectious Diseases, Yale School of Medicine, New Haven, CT, USA.
2 Yale College, New Haven, CT, USA.
3 Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA.
4 Department of Statistics, Yale University, New Haven, CT, USA.
5 University of North Carolina Project-China, Guangzhou, China.
6 Social Entrepreneurship to Spur Health (SESH) Global, Guangzhou, China.
7 University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
8 Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA.
9 Department of Communications, Boston College, Boston, MA, USA.
10 Microsoft Research Lab, Montreal, Quebec, Canada.
11 New York University, Abu Dhabi, UAE.
12 Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada.
13 School of Interactive Computing, Georgia Tech, Atlanta, GA, USA.
14 Center for Information Technology Policy, Princeton University, Princeton, NJ, USA.
15 School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
16 Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, USA.
17 Department of Computer Science, Stony Brook University, New York, NY, USA.
18 Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA.
19 Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA.
20 Yale Institute for Global Health, New Haven, CT, USA.

References
1. Gee, J., Marquez, P., Su, J., Calvert, G.M., Liu, R., Myers, T., Nair, N., Martin, S., Clark, T., Markowitz, L., et al.: First month of covid-19 vaccine safety monitoring—united states, december 14, 2020–january 13, 2021. Morbidity and Mortality Weekly Report 70(8), 283 (2021)
2. Tanne, J.H.: Covid-19: FDA approves Pfizer-BioNTech vaccine in record time. British Medical Journal Publishing Group (2021)
3. Bauer, S., Contreras, S., Dehning, J., Linden, M., Iftekhar, E., Mohr, S.B., Olivera-Nappa, A., Priesemann, V.: Relaxing restrictions at the pace of vaccination increases freedom and guards against further covid-19 waves. arXiv preprint arXiv:2103.06228 (2021)
4. Aw, J., Seng, J.J.B., Seah, S.S.Y., Low, L.L.: Covid-19 vaccine hesitancy—a scoping review of literature in high-income countries. Vaccines 9(8), 900 (2021)
5. MacDonald, N.E., et al.: Vaccine hesitancy: Definition, scope and determinants. Vaccine 33(34), 4161–4164 (2015)
6. Dubé, E., Vivion, M., MacDonald, N.E.: Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert review of vaccines 14(1), 99–117 (2015)
7. Feinmann, J.: How the world is (not) handling surplus doses and expiring vaccines. bmj 374 (2021)
8. Wirtz, K.: Changing readiness to mitigate sars-cov-2 steered long-term epidemic and social trajectories. Scientific Reports 11(1), 1–11 (2021)
9. Callaway, E., et al.: Covid vaccine boosters: the most important questions. Nature 596(7871), 178–180 (2021)
10. Troiano, G., Nardi, A.: Vaccine hesitancy in the era of covid-19. Public Health (2021)
11. Machingaidze, S., Wiysonge, C.S.: Understanding covid-19 vaccine hesitancy. Nature Medicine 27(8), 1338–1339 (2021)
12. Hotez, P., Batista, C., Ergoul, O., Figueredo, J.P., Gilbert, S., Gursel, M., Hassanain, M., Kang, G., Kim, J.H., Lall, B., et al.: Correcting covid-19 vaccine misinformation: Lancet commission on covid-19 vaccines and therapeutics task force members. EClinicalMedicine 33 (2021)
13. Vraga, E.K., Bode, L.: Correction as a solution for health misinformation on social media 110, 278–280 (2020)
14. Sallam, M.: Covid-19 vaccine hesitancy worldwide: a concise systematic review of vaccine acceptance rates. Vaccines 9(2), 160 (2021)
15. Jemielniak, D., Krempovych, Y.: An analysis of astrazeneca covid-19 vaccine misinformation and fear mongering on twitter. Public Health 200, 4–6 (2021)
16. Donzelli, G., Palomba, G., Fedegari, I., Aquino, F., Cioni, L., Verani, M., Carducci, A., Lopalco, P.: Misinformation on vaccination: A quantitative analysis of youtube videos. Human vaccines & immunotherapeutics 14(7), 1654–1659 (2018)
17. Martin, C.A., Marshall, C., Patel, P., Goss, C., Jenkins, D.R., Ellwood, C., Barton, L., Price, A., Brunskill, N.J., Khunti, K., et al.: Association of demographic and occupational factors with sars-cov-2 vaccine uptake in a multi-ethnic uk healthcare workforce: a rapid real-world analysis. MedRxiv (2021)
18. Salathé, M., Bonhoeffer, S.: The effect of opinion clustering on disease outbreaks. Journal of The Royal Society Interface 9(9), 1505–1508 (2008)
19. Wagner, A.L., Sheinfeld Gorin, S., Boulton, M.L., Glover, B.A., Moreno, J.D.: Effect of vaccine effectiveness and safety on covid-19 vaccine acceptance in detroit, michigan, july 2020. Human Vaccines & Immunotherapeutics, 1–6 (2021)
20. Loomba, S., de Figueiredo, A., Platek, S.J., de Graaf, K., Larson, H.J.: Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. Nature human behaviour 5(3), 337–348 (2021)
21. Malova, E.: Understanding online conversations about covid-19 vaccine on twitter: vaccine hesitancy amid the public health crisis. Communication Research Reports 38(5), 346–356 (2021)
22. Lyu, J.C., Le Han, E., Luli, G.K.: Covid-19 vaccine-related discussion on twitter: topic modeling and sentiment analysis. Journal of medical Internet research 23(6), 24435 (2021)
23. Lazarus, J.V., Ratzan, S.C., Palayew, A., Gostin, L.O., Larson, H.J., Rabin, K., Kimball, S., El-Mohandes, A.: A global survey of potential acceptance of a covid-19 vaccine. Nature medicine 27(2), 225–228 (2021)
24. Acheampong, T., Akorsikumah, E.A., Osae-Kwapong, J., Khud, M., Appiah, A., Amuasi, J.H.: Examining vaccine hesitancy in sub-saharan africa: a survey of the knowledge and attitudes among adults to receive covid-19 vaccines in ghana. Vaccines 9(8), 814 (2021)
25. Manikonda, L., Beigi, G., Liu, H., Kambhampati, S.: Twitter for sparking a movement, reddit for sharing the moment: #metoo through the lens of social media. arXiv preprint arXiv:1803.08022 (2018)
26. Priya, S., Sequeira, R., Chandra, J., Dandapat, S.K.: Where should one get news updates: Twitter or reddit. Online Social Networks and Media 9, 17–29 (2019)
27. Chohan, U.W.: Counter-hegemonic finance: The gamestop short squeeze. Available at SSRN (2021)
28. Glenski, M., Pennycuff, C., Weninger, T.: Consumers and curators: Browsing and voting patterns on reddit. IEEE Transactions on Computational Social Systems 4(4), 196–206 (2017)
29. Tomeny, T.S., Vargo, C.J., El-Toukhy, S.: Geographic and demographic correlates of autism-related anti-vaccine beliefs on twitter, 2009-15. Social science & medicine 191, 168–175 (2017)
30. Love, B., Himelboim, I., Holton, A., Stewart, K.: Twitter as a source of vaccination information: content drivers and what they are saying. American journal of infection control 41(6), 566–570 (2013)
31. Salathé, M., Khadwal, S.: Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 7(10), 1002199 (2011)
32. Lama, Y., Hu, D., Jamison, A., Quinn, S.C., Broniatowski, D.A.: Characterizing trends in human distributional semantics and pathfinder networks. Cancer Control 18(1), 12480 (2020)
33. Saha, K., Torous, J., Ernala, S.K., Rizuto, C., Stafford, A., De Choudhury, M.: A computational study of mental health awareness campaigns on social media. Translational behavioral medicine 9(6), 1197–1201 (2021)
34. Duraivel, S., Lavanya, R.: Understanding vaccine hesitancy with application of latent dirichlet allocation to reddit corpora (2021)
35. Wu, W., Luy, H., Luo, J.: Characterizing discourse about covid-19 vaccines: A reddit version of the pandemic story. arXiv preprint arXiv:2101.06321 (2021)
36. Papakyriakopoulos, O., Serrano, J.C.M., Hegelich, S.: Political communication on social media: A tale of hyperactive users and bias in recommender systems. Online Social Networks and Media 15, 100058 (2020)
37. Mickoleit, A.: Social media use by governments: A policy primer to discuss trends, identify policy opportunities and guide decision makers (2014)
38. Saha, K., Torous, J., Ernala, S.K., Rizuto, C., Stafford, A., De Choudhury, M.: A computational study of mental health awareness campaigns on social media. Translational behavioral medicine 9(6), 1197–1207 (2019)
39. Baumgartner, J.: Pushshift api (version 1.0). API Documentation, Pushshift. https://pushshift.io/api-parameters (2018)
40. Boe, B., Pedersen, A., Mellor, T.: Python Reddit API Wrapper (2016)
41. Jiang, Z.P., Levitan, S.I., Zomick, J., Hirschberg, J.: Detection of mental health from reddit via deep contextualized representations. In: Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis, pp. 147–156 (2020)
42. LaViolette, J., Hogan, B.: Using platform signals for distinguishing discourses: The case of men’s rights and men’s liberation on reddit. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 13, pp. 323–334 (2019)
43. Mohr, J.W., Bogdanov, P.: Introduction—topic models: What they are and why they matter. Poetics 41(6), 545–569 (2013)
44. Blei, D.M.: Probabilistic topic models. Commun ACM 55(4), 77–84 (2012)
45. Roberts, M.E., Stewart, B.M., Airoldi, E.M.: A model of text for experimentation in the social sciences. Journal of the American Statistical Association 111(515), 988–1003 (2016)
47. E, R.M., Stewart, B.M., Tingley, D.: Stm: R Package for Structural Topic Models. J Stat Softw, (2014). J Stat Softw
48. CSSE, J.: Covid-19 data repository by the center for systems science and engineering (csse) at johns hopkins university (2020)
49. Organization, W.H., et al.: WHO Director-General’s opening remarks at the media briefing on COVID-19-11 March 2020. Geneva, Switzerland (2020)
50. Grimmer, J., Stewart, B.M.: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis 21(3), 267–297 (2013)
51. Buntain, C., Golbeck, J.: Identifying social roles in reddit using network structure. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 615–620 (2014)
52. Seale, H., Heywood, A.E., Leask, J., Sheel, M., Durheim, D.N., Bolsewicz, K., Kaur, R.: Examining australian public perceptions and behaviors towards a future covid-19 vaccine. BMC Infectious Diseases 21(1), 1–9 (2021)
53. Grimmer, J., Stewart, B.M.: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis 21(3), 267–297 (2013)
54. Buntain, C., Golbeck, J.: Identifying social roles in reddit using network structure. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 615–620 (2014)
55. Organization, W.H., et al.: WHO Director-General’s opening remarks at the media briefing on COVID-19-11 March 2020. Geneva, Switzerland (2020)
56. Buttenheim, A.M., Joyce, C.M., Ibara, J., Agas, J., Feenster, K., Handy, L.K., Amin, A.B., Omer, S.B.: Vaccine exemption requirements and parental vaccine attitudes: an online experiment. Vaccine 38(11), 2620–2625 (2020)
57. Pennycook, G., Rand, D.G.: Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences 116(7), 2521–2526 (2019)
58. Barnett, P.A., Hoskins, C.E., Alhoti, J.A., Carpenter, L.J.: Reducing public misinformation about organ donation: An experimental intervention. Journal of Social Distress and the Homeless 18(1-2), 57–73 (2009)
59. Wesson, D.E., Lucey, C.R., Cooper, L.A.: Building trust in health systems to eliminate health disparities. Jama 322(2), 111–112 (2019)
60. Nyhan, B., Reifler, J., Richey, S., Freed, G.L.: Effective messages in vaccine promotion: a randomized trial. Pediatrics 133(4), 835–842 (2014)
61. Goldenberg, M.J.: Antivaccination movement exploits public’s distrust in scientific authority. BMJ 367 (2019)
62. Evanega, S., Lynas, M., Adams, J., Smolenyak, K., Insights, C.G.: Coronavirus misinformation: quantifying sources and themes in the covid-19 ‘infodemic’. JMIR Preprints (2020)
63. Shah, P.D., Calo, W.A., Gilkey, M.B., Boynton, M.H., Dailey, S.A., Todd, K.G., Robichaud, M.O., Margolis, M.A., Brewer, N.T.: Questions and concerns about hpv vaccine: a communication experiment. Pediatrics 143(2) (2019)
64. Buttenheim, A.M., Joyce, C.M., Ibara, J., Agas, J., Feenster, K., Handy, L.K., Amin, A.B., Omer, S.B.: Vaccine exemption requirements and parental vaccine attitudes: an online experiment. Vaccine 38(11), 2620–2625 (2020)
65. Pennycook, G., Rand, D.G.: Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences 116(7), 2521–2526 (2019)
66. Barnett, P.A., Hoskins, C.E., Alhoti, J.A., Carpenter, L.J.: Reducing public misinformation about organ donation: An experimental intervention. Journal of Social Distress and the Homeless 18(1-2), 57–73 (2009)
67. Wesson, D.E., Lucey, C.R., Cooper, L.A.: Building trust in health systems to eliminate health disparities. Jama 322(2), 111–112 (2019)
68. Amaya, A., Bach, R., Keusch, F., Kreuter, F.: New data sources in social science research: Things to know before working with reddit data. Social Science Computer Review, 0894439319895305 (2019)
69. Duggan, M., Smith, A.: Cell internet use 2013. Washington, DC: Pew Research Center (2013)

Figures

Figure 1: Regression where the outcome variable was the proportion of each document dedicated to each topic. Values were generated from a regression where the outcome variable was the proportion of each document dedicated to each topic, given the selected STM model, and with various vaccine events as main explanatory variables. Topics on the right of the zero line were more likely to be brought up after the indicated event. Confidence intervals (95%) included both regression uncertainty and measurement uncertainty from the STM model.
Figure 2: **Graph of subreddits that have COVID-19 vaccine-related posts**
The largest subgraph above shows subreddits that have COVID-19 vaccine-related posts. Node sizes represent subreddit user sizes, node edges indicate shared COVID-19 vaccine posts between subreddits, and node labels indicate the subreddit name. Node color is based on centrality (red is higher centrality, yellow is lower centrality). Edge direction was based on whether a node had >50% of its posts prior to its adjacent connecting node e.g. A->B if >50% of A’s shared posts were made before B. We used larger labels for subreddit nodes with >5,000 COVID-19 vaccine posts.
Table 1: Structural topic model results from 266,840 documents, March 11 2020 - December 14 2020, including the topic proportion and the top 10 words associated with each topic.

| Expected Topic Proportions | Topics Title                                                      | Top 10 words                                                                 |
|-----------------------------|-------------------------------------------------------------------|-------------------------------------------------------------------------------|
| 0.1036                      | Severity of COVID-19                                              | die, risk, life, normal, live, sick, yes, stop, stay, serious                 |
| 0.0941                      | Hope for a swift end to the pandemic                              | long, term, hope, safe, next, available, wait, shot, pretty, rush             |
| 0.0861                      | Suspicion of science and mainstream media                         | believe, anti, tell, real, science, liter, bad, media, cure, trust            |
| 0.0796                      | Evidence based COVID-19 discussion                                | read, fact, understand, inform, clear, true, person, reason, post, evidence   |
| 0.0702                      | COVID-19 transmission rates and patterns                          | spread, population, herd, high, risk, rate, number, reduce, hospitalization, mortality |
| 0.0633                      | Effect of the virus on the human body                             | common, system, response, human, influenza, body, cold, mutate, strain, similar |
| 0.0577                      | Impact of pandemic on communities and hospitals                   | home, school, family, care, person, stay, live, learn, hospital, help          |
| 0.0523                      | Global effects of pandemic                                        | world, country, public, global, travel, economy, social, govern, open, state  |
| 0.0500                      | COVID-19 vaccine trials and developments                          | data, phase, trial, clinic, safety, studies, drug, efficacy, severe, receive  |
| 0.0478                      | COVID-19 statistics                                               | death, million, rate, dead, season, die, total, number, second, near           |
| 0.0381                      | Economy and markets                                               | market, companies, stock, product, industry, supply, price, sell, demand, billion |
| 0.0376                      | Federal policies                                                 | trump, president, nation, state, elect, administration, federal, house, unit, response |
| 0.0359                      | Government aid and expenses                                       | govern, money, free, pay, business, cost, spend, economy, system, support      |
| 0.0325                      | Long term effects of COVID-19 and related diseases                | medical, damage, cancer, doctor, heart, polio, measles, harm, medicine, child, blood |
| 0.0310                      | Black Lives Matter protests                                       | game, watch, video, police, fire, sport, street, post, show, red               |
| 0.0309                      | COVID-19 vaccine conspiracy theories and misinformation           | bill, world, power, human, war, russia, mark, control, chip, conspiracy        |
| 0.0294                      | COVID-19 research                                                 | research, link, studies, found, scientific, respiratory, science, paper, associate, article |
| 0.0291                      | Masks and social distancing                                      | wear, mask, social, protect, spread, face, public, hand, person, distance     |
| 0.0160                      | Racism on social media                                            | chance, please, remember, pass, remove, black, message, thank, stick, attend  |
| 0.0148                      | Trump’s leadership                                                | march, trump, perfect, control, disappear, march, anybody, fine, false, great  |

Note: The topic proportions indicated the proportion of the corpus that belongs to each topic.
Figures

Figure 1
Regression where the outcome variable was the proportion of each document dedicated to each topic. Values were generated from a regression where the outcome variable was the proportion of each document dedicated to each topic, given the selected STM model, and with various vaccine events as main explanatory variables. Topics on the right of the zero line were more likely to be brought up after the indicated event. Confidence intervals (95%) included both regression uncertainty and measurement uncertainty from the STM model.

Figure 2
Graph of subreddits that have COVID 19 vaccine-related posts. The largest subgraph above shows subreddits that have COVID-19 vaccine-related posts. Node sizes represent subreddit user sizes, node edges indicate shared COVID-19 vaccine posts between subreddits, and node labels indicate the subreddit name. Node color is based on centrality (red is higher centrality, yellow is lower centrality). Edge direction was based on whether a node had >50% of its posts prior to its adjacent connecting node e.g. A->B if >50% of A’s shared posts were made before B. We used larger labels for subreddit nodes with >5,000 COVID-19 vaccine posts.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.pdf
- SupplementaryTables.tex