Research Article

Vaccine hesitancy in online spaces: A scoping review of the research literature, 2000-2020

We review 100 articles published from 2000 to early 2020 that research aspects of vaccine hesitancy in online communication spaces and identify several gaps in the literature prior to the COVID-19 pandemic. These gaps relate to five areas: disciplinary focus; specific vaccine, condition, or disease focus; stakeholders and implications; research methodology; and geographical coverage. Our findings show that we entered the global pandemic vaccination effort without a thorough understanding of how levels of confidence and hesitancy might differ across conditions and vaccines, geographical areas, and platforms, or how they might change over time. In addition, little was known about the role of platforms, platforms' politics, and specific sociotechnical affordances in the spread of vaccine hesitancy and the associated issue of misinformation online.

Authors: Timothy Neff (1), Jonas Kaiser (2), Irene Pasquetto (3), Dariusz Jemielniak (4), Dimitra Dimitrakopoulou (5), Siobhan Grayson (6), Natalie Gyenes (7), Paola Ricaurte (8), Javier Ruiz-Soler (9), Amy Zhang (10)
Affiliations: (1) Media, Inequality and Change Center, Annenberg School for Communication, University of Pennsylvania, USA; Berkman Klein Center for Internet & Society, Harvard University, USA, (2) Berkman Klein Center for Internet & Society, Harvard University, USA; Communication, Journalism, & Media, Suffolk University, USA, (3) School of Information, University of Michigan, USA; Shorenstein Center on Media, Politics and Public Policy, Harvard University, USA, (4) MINDS (Management in Networked and Digital Societies), Kozminski University, Poland, (5) Center for Constructive Communication, Massachusetts Institute of Technology, USA; Department of Communication and Media Research, University of Zurich, Switzerland, (6) School of Computer Science, University College Dublin, Ireland, (7) Digital Health Lab, Meedan, USA, (8) Tecnologico de Monterrey, Mexico; Berkman Klein Center for Internet & Society, Harvard University, USA, (9) Digital Democracies Institute, Simon Fraser University, Canada, (10) Allen School of Computer Science & Engineering, University of Washington, USA

Research questions

- RQ1: How was vaccine hesitancy—defined as "delay in acceptance or refusal of vaccination despite availability of vaccination services" (MacDonald & the SAGE Working Group on Vaccine Hesitancy, 2015, p. 4161)—in online spaces studied in the academic literature from 2000 to 2020?
- RQ2: What were the research gaps in the academic literature in relation to this area of investigation as of early 2020?

---

[1] *A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.*

Essay summary

- We searched the Web of Science database for articles researching aspects of vaccine hesitancy in online spaces. Of 236 articles selected for analysis, 100 were determined relevant to our research interest and content analyzed by human coders.

- We identified gaps in the academic literature pertaining to vaccine hesitancy and the associated issue of online misinformation prior to the pandemic in five areas: disciplinary focus; specific vaccine, condition, or disease focus; geographical focus; stakeholders and implications; research methodology.

- Although this literature has greatly expanded, as of early 2020 it had yet to fully grapple with key dimensions of how vaccine hesitancy may be expressed in online discourse, specifically how online vaccine hesitancy differs across diseases, digital spaces, and local contexts, and how it might change over time. In addition, as of early 2020 very little appeared in the literature about the role of social media platforms in preventing and addressing the spread of vaccine hesitancy and online misinformation.

- Most of the articles we analyzed were produced within the field of public health research. In general, more interdisciplinary research is needed.

Implications

Globally and with increasing intensity during the present COVID-19 pandemic, efforts have been taken to address adherence to vaccine implementation across a spectrum of vaccine confidence levels, where low levels can contribute to hesitancy to vaccinate, and, potentially, the spread of related misinformation. Our research examines the lower end of the vaccine confidence spectrum, focusing on vaccine hesitancy and related challenges.[2]

Though vaccine hesitancy, misinformation, and online environments offer different sets of issues, health officials and researchers acknowledge important overlaps, as achieving broad acceptance of vaccines requires understanding information ecosystems (Berman, 2020; Larson, 2020). Vaccine hesitancy also has associated challenges that are specific to different vaccines, including COVID-19 vaccines, yet the persistence of varying degrees of confidence in vaccines more broadly is an ongoing challenge, and online environments have long been gathering points where vaccine rumors and myths are shared (Burki, 2020).

In this article, we conduct a scoping review[3] of the existing academic research on online vaccine hesitancy and suggest directions for future research. Though we acknowledge that vaccine hesitancy and the related issue of vaccine misinformation are not exclusively or most importantly online problems, our review provides a valuable snapshot of the state of research on this growing area of concern (RQ1), including potential research gaps (RQ2), prior to the COVID-19 pandemic. Given the impact of the pandemic on global dialogues about vaccination adherence, including effective communication to promote vaccination among diverse communities, it is useful to create a baseline for future research that can address challenges to vaccine uptake related to the present pandemic and future diseases. Our review

---

[2] Researchers have used both the term "vaccine hesitancy" and the term "vaccine confidence" to refer to sentiments about vaccines, "hesitancy" drawing attention to negative sentiments and "confidence" inclusive of positive sentiments. We use the term "vaccine hesitancy," which we operationalize as search terms used in our scoping review, such as "refusal," "skepticism," and "critical" (see Methods), referring to sentiments expressing the lower end of what has been studied as a spectrum of confidence levels (de Figueiredo et al., 2020; Larson, 2020; Orenstein et al., 2015).

[3] Unlike systematic reviews, scoping reviews seek to identify concepts and characteristics across extant literature rather than seeking to answer a specific question about a topic related to that literature (Munn et al., 2018).

also can serve as a point of comparison for understanding how the present pandemic tests prior research findings and influences future research.

Our review covers two decades in which research focusing on vaccine hesitancy online emerged and steadily evolved. In the early 2000s, researchers studied vaccine information on websites and even at this earliest stage of online research noted the prevalence of "misleading or inaccurate information" (article ID 19), the spread of anecdotal accounts of vaccine dangers, and the misrepresentation of the science behind vaccines (ID 6). Though we did not use the term "misinformation" in our database search string (see Methods), search results show that the relationship between online vaccine hesitancy and the quality of online information is a persistent concern in the research literature, and our review frequently encountered articles referencing "misinformation" (e.g., IDs 21, 60, 98, 104, 106, 138, 195) or expressing concerns for the accuracy of information (IDs 15, 23, 157).

With the emergence of participatory "Web 2.0" websites and technologies enabling Internet users to generate content and interact with each other via social media, researchers recognized the potential for the public spread of private concerns about vaccines (ID 10), which challenges the information-gatekeeping power of health professionals (ID 3). Research in the 2010s increasingly examined the sources of information about vaccines and network dynamics that help these sources spread information, finding that vaccine hesitancy is prevalent online (IDs 21, 73, 131) but often circulates in small but active and cohesive subgroups of Internet users (IDs 60, 92, 109, 244, 225). Researchers also have increasingly recognized nuances among different forms of vaccine hesitancy (IDs 34, 98) and have explored the power of storytelling vs. the power of facts and statistics prevalent on official health sites (IDs 23, 45, 138, 165, 187). Conclusions increasingly have urged monitoring and moderating social media platforms to stop the spread of misinformation (IDs 178, 243).

Our review identifies gaps in this literature in five areas:

- Disciplinary focus
- Disease and vaccine focus
- Geographical focus
- Stakeholders and implications
- Research methodology

In terms of disciplinary focus, a majority of the articles we analyzed were published in journals in the field of public health and medicine and focus on understanding how people share information about vaccines online. Articles identified online discourse that may contribute to vaccine hesitancy (ID 127); analyzed which subpopulations of users are likely to amplify vaccine misinformation (ID 15); investigated whether search engines return quality information sources (ID 83). These analyses aim to develop *tactics and tools* to help public health officials *surveil* in real-time public opinions and attitudes toward immunization and leverage online media to better communicate their messages.

Only 35 of the 100 articles we looked at examined vaccine hesitancy in relation to a specific disease and vaccine, most prominently measles, mumps, rubella (MMR), and the human papillomavirus (HPV), while 65 analyzed generic vaccine hesitancy or hesitancy across multiple diseases. The predominance of general studies is concerning because interventions should take account of the unique characteristics of specific vaccines and diseases.

Most articles we reviewed were conducted by Western research institutions and focused on vaccine hesitancy within Western contexts. We removed non-English articles prior to content analysis, but even considering this limitation of our review, the predominately Western focus in the extant research is clear. One consequence of this Western focus in the available research may be that the field becomes primarily aimed at developing vaccine-related digital communication strategies while overlooking issues of access to vaccines. Also, such strategies may be inappropriate in regions grappling with diseases not prominent in the West, such as polio.

In part due to the prominence of the field of public health in the extant research, health authorities and professionals are most often identified as top stakeholders in addressing vaccine hesitancy and online misinformation. Proposed actions for this category of stakeholders include getting involved in online groups that spread misinformation (ID 60); including stories alongside scientific facts in communication efforts (ID 209); and adopting "vaccine ambassador" programs, in which community members or health professionals share reliable information with select audiences to promote vaccine confidence and adherence (ID 23).

Though health professionals intervene at the nexus of vaccines, information, and patients, the dynamics of how information spreads involve a broader set of issues, such as online platforms' politics and sociotechnical affordances and coordinated efforts by actors external to the doctor-patient relationship. These dynamics also involve a broader set of stakeholders, such as platforms, news media organizations, and policymakers who can intervene at a structural level. These stakeholders are prominent in media studies, sociology, human-computer interaction, political science, and other fields positioned to make valuable contributions to research on vaccine hesitancy and misinformation online.

In terms of research methodology, few of the articles we reviewed made causal inferences. It is therefore difficult to confirm underlying factors in the online spread of vaccine hesitancy and related misinformation, or to recommend concrete strategies. We also found few studies focused on changes in expressions of vaccine hesitancy over time, beyond the ebb and flow (volume) of online activity, a finding that is likely connected to platform restrictions on access to data. Yet, in order to identify effective interventions, it is imperative to understand the long-term growth of online communities, how they connect to other communities, and how issue framings evolve.

Finally, as of early 2020, little attention had been paid to social media platforms beyond Facebook and Twitter that are well-known for playing major roles in the distribution and amplification of online misinformation, such as YouTube and Reddit (Cinelli et al., 2020; Kaiser et al., 2021; Li et al., 2020). Research has begun to fill this gap, but more work needs to be done to understand persistent, platform-specific challenges, though recommendations for action abound: enhancing surveillance of misinformation; removing sources, as well as information; stepping up platform self-regulation; and engaging lawmakers, activists, and others in policy interventions (see, for example, Chou et al., 2020; Rutschman, 2020).

The core takeaway of this review is that research on vaccine hesitancy in online spaces would benefit from the continued and strengthened participation of disciplines that can offer a range of research approaches to online communication dynamics, including but not limited to anthropology, media and communications, human-computer interaction, information science, sociology, STS, and political science. Such participation would harness the strengths of these disciplines in research on specific practices, sociotechnical affordances, and structural factors—such as media policymaking, platform and media economics, and socioeconomic variables—that play roles in addressing and adequately responding to vaccine hesitancy and related misinformation in online and offline spaces during the present pandemic and future health crises. After presenting our findings, we suggest and elaborate five directions for future research.

Findings

Finding 1: Research on vaccine hesitancy is a rapidly growing field.

Our search of the Web of Science database surfaced 252 articles researching aspects of vaccine hesitancy that referenced online spaces, the earliest of which was published in 2000. We note that the publication

rate begins to climb in 2010, rising to 69 articles in 2019.[4] The rise to prominence of social media platforms Facebook and Twitter (both launched to the general public in 2006) coincides with this increased research interest in online spaces.
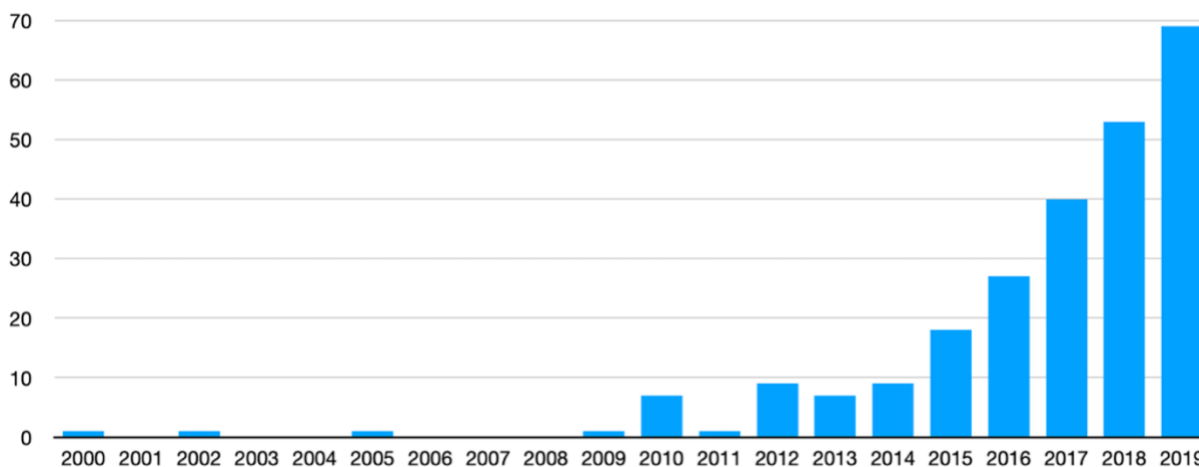

Figure 1. Number of publications, by year, found in search of Web of Science database, Feb. 12, 2020.

Of 236 articles selected for analysis (see Methods), 100 were determined relevant to our research interest as primarily focused on vaccine hesitancy in online spaces. Table 1 breaks down these 100 articles by digital platform analyzed.

Table 1. *Platforms analyzed (n = 100).[5]*

Social media platforms	Instagram	1%
	Twitter	14%
	Facebook	12%
	YouTube	7%
	Other social media	3%
	Multiple social media platforms	11%
	Other	2%
Other forms of digital media	Blog	2%
	News/online media	5%
	Wikipedia	1%
	Forums/communities	5%
	Websites	35%
	Other	13%

---

[4] A search using the same terms on April 17, 2021, showed that 182 articles were published between 2020 and early 2021, more than double the pace of publications in 2019.

[5] As some articles analyzed both social media and other forms of digital media, total percentage exceeds 100 percent.

Finding 2: Research is dominated by public health field, with a focus on how people talk about vaccines.

The majority of these articles appear in academic journals focusing on public health and medicine (76 percent). Few articles make causal inferences (9 percent). Table 2 shows the research focus[6] of the articles, which most often demonstrates interest in how people in online spaces talk about vaccines and vaccinations.

Table 2. Research focus (n = 100).

Identity (Who doubts or resists vaccines?)	3%
Communication (How do people talk about vaccines?)	64%
Information-seeking behavior (How do people search for information about vaccines?)	15%
Computational identification of vaccine sentiments	10%
Other	8%

Table 3 shows the main stakeholders (n = 139) in the articles. Researchers most often indicate that their studies have conclusions relevant to medical professionals and the public health sector. These conclusions often focus on steps medical practitioners should take to counter misinformation about vaccines.

Table 3. Stakeholders (n = 139).

Journalists/media	7.2%
Patients/relatives	5.8%
Policymakers/legislators	9.4%
Medical professionals/public health sector	46%
Technology companies/platforms	2.9%
Academic researchers	25.9%
Other	2.9%

The majority of stakeholders identified in articles published by public health journals are medical professionals (51 percent; n = 110), with academic researchers comprising 22 percent of stakeholders in these articles. In journals from other fields, academic researchers are identified as stakeholders 41 percent of the time, while medical professionals are identified as stakeholders only 28 percent of the time (n = 29).

Finding 3: Additional gaps include demographic groups, terminology, geography, and disease- or vaccine-specific research.

Very few articles include a focus on gender (11 percent), ethnicity (4 percent), or age (7 percent). When authors focus on gender, they usually study women's communication around vaccinations (e.g., in the context of Facebook groups). More research focused on these demographics and others would aid efforts to develop more relevant response approaches tailored to different social contexts.

We also note a wide variance in the terminology used for people who are doubtful of or otherwise resistant to vaccines. "Anti-vaccine" or "anti-vaccination" is the most common term found in these articles

---

[6] Intercoder reliability testing resulted in low coder agreement for "research focus" and "stakeholders" variables; see Methods.

(55 percent), followed by "vaccine hesitancy" (25 percent) and "vaccine critical" (7 percent). Authors occasionally use the terms interchangeably although they represent different concepts: "Vaccine hesitancy" reflects individuals or communities who may experience challenges related to confidence, complacency, or knowledge and awareness, while "anti-vaccination" reflects active opposition to vaccines. A preferable framework involves terminology broader than these categorizations but not often used in social media contexts: the spectrum of "vaccine confidence" (Larson, 2020).

Seventy-eight percent of the time, countries listed by Web of Science as the origin for the 252 articles originally returned by our search are located in North America or Western Europe (254 out of 325 instances). [7]

Finally, 35 percent of the articles relevant for our review (n = 100) focus on a specific vaccine or disease, with the most prominent diseases being the human papillomavirus (HPV) and the vaccine for mumps, measles, and rubella (MMR).

Directions for research

This scoping review suggests multiple directions for research to address online vaccine hesitancy and related vaccine misinformation. At a minimum, future reviews should examine how the present pandemic has changed the state of this research. In addition, we suggest the following directions for research.

1. *Broadening the scope*. One direction of research that would address multiple gaps in the present literature is interdisciplinary *comparative* research on online vaccine hesitancy across national, regional, local, and cultural contexts. Not only would such research help close gaps in our understanding of how vaccine hesitancy in online spaces differs from one vaccine and disease to another, but it also would provide opportunities for new research beyond Western contexts. Such research, well established in political science and media and cultural studies, also is capable of denaturalizing structural influences, such as economic, political, and media system dynamics, that encourage and amplify the online sharing of misinformation. Our review has noted that the extant research has over the years begun to engage more with the online sources of information rather than the quality of the information itself; interdisciplinary, comparative research is capable of bringing into view an additional set of structural sources or influences that can be addressed through policy interventions.

2. *Methods*. The existing research would be enhanced by qualitative, ethnographic fieldwork. Our review shows that information sharing *behaviors* rather than the *identities* of those who share information, including social milieus important to those identities, have been the most prominent focus of research. Researchers have used surveys to gather views on vaccines, but ethnographic fieldwork is capable of enhancing our understanding of how vaccine hesitancy relates, on a day-to-day basis, to a variety of social factors and how these factors relate to the online sharing of vaccine information. This fine-grained, nuanced fieldwork could inform intervention strategies that complement or even move beyond the prominent debate in the extant research over the effectiveness of facts vs. stories in encouraging vaccine uptake.

3. *Different communities*. More research is needed on the different communities that exist within every society and that differ with regard to their vulnerability to vaccine misinformation, as well

---

[7] Web of Science uses addresses of author institutions to populate a "countries/regions" field used in our assessment.

as where and how they encounter it. Investigations into the role of gender, race, religious/spiritual beliefs and political ideology could highlight different online information pathways. These communities are not necessarily clearly delineated from one another but overlap; understanding these intersections is important when it comes to effective communication and outreach.

4. *Longitudinal research*. Most studies in our review that analyzed social media focused on a certain point in time or excluded time from their analyses. Longitudinal research is needed in order to highlight changes over time, identify patterns and discursive moments, and assess the impact of platform actions such as de-platforming bad actors.

5. *Access and data.* The gaps that our review finds in research on certain types of social media, research over time, and research on platform specific-interventions point to the broader, persistent issue of constraints on researcher access to online spaces and the challenges posed by frequently changing data formats and algorithms. These issues – so ubiquitous that they risk becoming invisible and uncritically accepted – often are left unsaid in research; by cataloging how researchers have encountered and addressed them, future literature reviews could contribute to the development of better research strategies and methods.

Methods

For our scoping review of the academic literature, we follow procedures outlined by Moher et al. (2009): identify relevant articles in a database; check these articles against other sources; screen out duplicates; assess articles for eligibility; and include the final list in the meta-assessment. Figure 2 shows our process as a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart (Page et al., 2021).
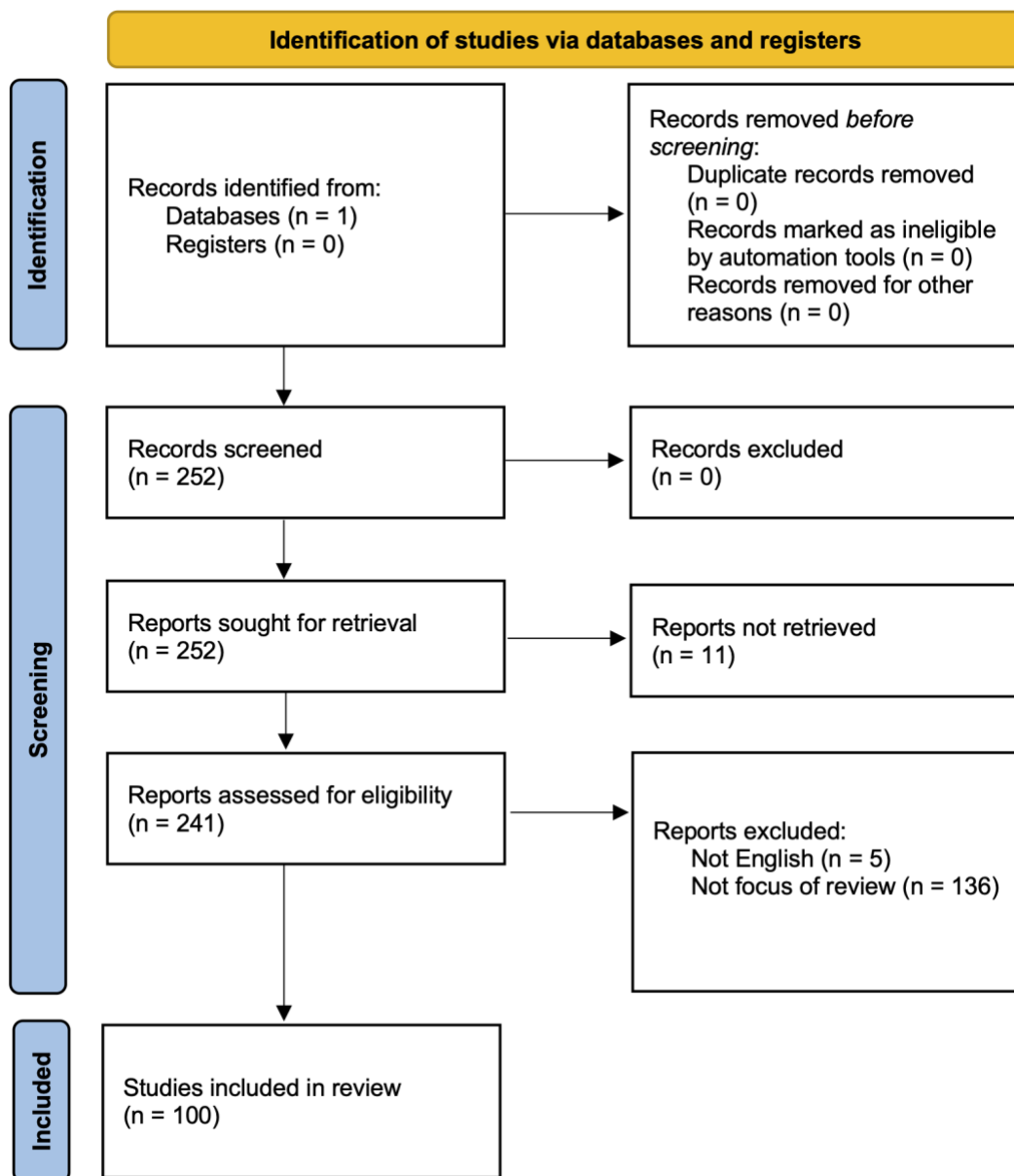
Figure 2. PRISMA flow chart for review.

Through discussion, experimentation, and comparison with different libraries of articles, we developed a string of search terms that surfaced a list of articles roughly commensurate with articles found in other databases. This search string operationalizes the concept of vaccine hesitancy by including terms related to research on the lower end of the spectrum of vaccine confidence, such as "refusal," "denial," and "skepticism:"

((“anti-vaccination” OR “anti-vaccine” OR “anti-vax” OR “vaccine hesitancy” OR “vaccine reluctance” OR “vaccine refusal” OR “anti-vaxxer” OR “anti-vaxx” OR “vaccine denial” OR “vaccine skepticism” OR “vaccine critical”) AND (“internet” OR “online” OR “social network analysis” OR “social network sites” OR “social media” OR “social networking” OR “Web 2.0” OR “websites”))

We chose the Web of Science Core Collection database for our scoping review due to its widely recognized quality, providing some assurance that the articles included in our review are well-researched and impactful.[8] However, the Scopus database also is prominent in literature reviews. As there are differences in coverage (Martín-Martín et al., 2021; Visser et al., 2021), we compared search results between the two databases and found 20 additional items indexed by Scopus from 2000 to the end of 2019 (roughly corresponding to our Web of Science search's time period) relevant to our study, three of which could not be accessed. A content analysis of these items found nothing that would significantly alter our findings, though we note a higher prevalence of articles from the field of computer science – most published in conference and workshop proceedings – among the additional Scopus articles (7 of 17 items) than among our relevant Web of Science articles (5 of 100 items).

Using our search string in Web of Science's title, abstract, and keywords field surfaced 252 articles on Feb. 12, 2020, approximately one month prior to the beginning of widespread COVID-19 precautions in the United States. After removing from this list inaccessible articles and articles not written in English, our final list for analysis included 236 articles. The full text of each of these articles was screened for inclusion in our findings.

Our codebook includes 24 variables, 11 of which require manual coding. Intercoder reliability testing involved 10 coders. Due to the difficulty of achieving high levels of agreement with such a large number of coders, a subset of three coders developed a shared codesheet for a 10 percent subsample of the articles. The other seven coders coded each of these articles separately, without access to the shared codesheet, and checked their coding against it, making changes when they agreed with the shared codesheet.

Final intercoder agreement (Holsti's) exceeded 0.75 for most variables, except for "stakeholders" (0.51 and 0.64 for two categorical variables sharing the same set of categories, as multiple stakeholders were allowed, coded in order of appearance in each article); and "research focus" (0.6). Coding for multiple stakeholders was complicated by requiring coders to enter codes in the order in which stakeholders appeared in each article; this complication does not affect the coders' ability to categorize stakeholders, and so we believe reliability for the stakeholders variable is higher than indicated. With regard to research focus, although articles focusing on how people talk about vaccines (communication) clearly dominated our corpus of research articles, at times coders found that some of these articles aimed to shed light on who these people are (identity) or primarily were about computational methods used to study them. We include findings for these two variables in our article, with the caution that intercoder reliability fell short of high confidence for them.

After intercoder reliability testing concluded, we randomly distributed the remaining 90 percent of the articles to our coders for final coding. We used SPSS software to generate descriptive statistics from the results of this coding.

Bibliography

Berman, J. M. (2020). *Anti-vaxxers: How to challenge a misinformed movement*. MIT Press.

Burki, T. (2020). The online anti-vaccine movement in the age of COVID-19. *The Lancet Digital Health, 2*(10), e504-e505. https://doi.org/10.1016/S2589-7500(20)30227-2

Chou, W. S., & Gaysynsky, A., & Cappella, J. N. (2020). Where we go from here: Health misinformation on social media. *American Journal of Public Health, 110*(S3), S273–S275. https://doi.org/10.2105/AJPH.2020.305905

---

[8] Web of Science developed the "impact factor" commonly used to assess the academic influence of journals (Garfield, 1994).

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Nature Scientific Reports*, *10*, 1–10. https://doi.org/10.1038/s41598-020-73510-5

de Figueiredo, A., Simas, C., Karafillakis, E., Paterson, P., & Larson, H. J. (2020). Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: A large-scale retrospective temporal modelling study. *Lancet, 396*(10255), 898–908. https://doi.org/10.1016/S0140-6736(20)31558-0

Garfield, E. (1994). The Clarivate Analytics impact factor. Clarivate. https://clarivate.com/webofsciencegroup/essays/impact-factor/

Kaiser, J., Rauchfleisch, A., & Córdova, Y. (2021). Fighting Zika with honey: An analysis of YouTube's video recommendations on Brazilian YouTube. *International Journal of Communication, 15*, 1244–1262. https://ijoc.org/index.php/ijoc/article/view/14802

Larson, H. J. (2020). *Stuck: How vaccine rumours start – and why they don't go away.* Oxford University Press.

Li, H. O., Bailey, A., Huynh, D., & Chan, J. (2020). YouTube as a source of information on COVID-19: A pandemic of misinformation? *BMJ Global Health*, *5*(5), 1–6. https://doi.org/10.1136/bmjgh-2020-002604

MacDonald, N. E., & the SAGE Working Group on Vaccine Hesitancy. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine, 33*(34), 4161–4164. https://doi.org/10.1016/j.vaccine.2015.04.036

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, *126*, 871–906. https://doi.org/10.1007/s11192-020-03690-4

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6*(7), 1–6. https://doi.org/10.1371/journal.pmed.1000097

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*(143), 1–7. https://doi.org/10.1186/s12874-018-0611-x

Orenstein, W. A., Gellin, B. G., Beigi, R. H., Despres, S., Lynfield, R., Maldonado, Y., Mouton, C., Rawlins, W., Rothholz, M. C., Smith, N., Thompson, K., Torres, C., Kasisomayajula, V., & Hosbach, P. (2015). Assessing the state of vaccine confidence in the United States: Recommendations from the National Vaccine Advisory Committee. *Public Health Reports, 150*(6), 573–595. https://doi.org/10.1177/003335491513000606

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffman, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(71), 1–9. http://dx.doi.org/10.1136/bmj.n71

Rutschman, A. S. (2020). *Facebook's latest attempt to address vaccine misinformation – and why it's not enough* (Legal Studies Research Paper No. 2020-35). Health Affairs Blog, Saint Louis University. https://scholarship.law.slu.edu/faculty/544/

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies, 2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Competing interests
The authors report no conflicts of interest.

Ethics
The authors have no ethics approvals to declare that are relevant to the content of this article.

Data availability
All materials needed to replicate this study are available via the Harvard Dataverse:
https://doi.org/10.7910/DVN/MACUIS

Appendix: Codebook

Variables should receive one code only (not multiple codes), indicating assessment of the primary characteristic relevant for each variable. If an article passes the filter variable (V11), every other cell in that article's row in the codesheet must include a single numeral as coding (in other words, do not leave any cell in that article's row empty). The only exception is that article's cell under V17, which is a text variable and can be left empty.

V1. Article ID
(already filled out)

V2. Article title
(already filled out)[9]

V3. Coder
(fill out with name)

V4. Journal name
(already filled out)

V5. Academic field
(already filled out)

V6. Publication year
(already filled out)

V7. Author names
(already filled out)

V8. Author affiliations
(already filled out)

V9. Keywords
(already filled out)

V10. [DELETED]

V11. Relevant for our research interest?
This variable asks for the relevance of the article within the scope of our research interest. Relevance means these conditions are met: 1) online, vaccination-related communication (internet, social media, blogs, etc.) is referenced in the article's abstract; 2) one paragraph in the introduction, one paragraph in the methods section, and one paragraph in the conclusion mentions online vaccination-related communication; 3) the article must directly research anti-vaccination (hesitancy, opposition, etc.) communication expressed online. If the article only conducts online surveys, it should be coded (0) no.

---

[9] Based on the output from the Web of Science search, metadata was already filled out.

Important: Anti-vaccination needs to be addressed multiple times in the paper, i.e., papers that just mention the issue once are coded as 0.
0 = No (if "no," stop coding)
1 = Yes

V12. [DELETED]

V13. Open access
Is this article openly accessible?
0 = No
1 = Yes

V14. Social media platforms (code one)
Are social media platforms the focus of direct analysis (not simply mentioned)? If so, which ones?
(Note: Analysis of websites or blogs is not meant here; code 0)
0 = No
1 = Instagram
2 = Twitter
3 = Facebook
4 = YouTube
5 = Reddit
6 = Other social media platform (e.g., TikTok, VKontakte, etc.)
7 = Multiple platforms (multiple, specific social media platforms specified as included in article's analysis)
8 = Other (Only in case a different social media platform is analyzed)

V15. Other digital media (code one)
If not, what other digital platforms are the focus of direct analysis (not simply mentioned)?
0 = No (if purely social media, code 0)
1 = Blog
2 = News/Online media
3 = Wikipedia
4 = Forums or other online communities
5 = Websites
6 = Other (e.g., 4chan)
7 = Multiple platforms (multiple, specific platforms specified as included in article's analysis)

V16. Country of focus given by Web of Science
(already filled out)

V17. Country of focus different from Web of Science classification
If the country or countries of focus are different than the Web of Science classification, type in full country name(s), separated by comma; if there are four or more countries type "International;" if not leave the field empty.

V18. Study methodology (code one; based on whether the study uses these key terms)
Search article for specific terms. We do NOT code what we think the researchers are doing, we code what the scholars are describing their own study as. If they are not specifying their method, code unspecified.

If they both say qualitative and quantitative or quantitative and computational, etc. code mixed.
1 = Theoretical (e.g., literature reviews)
2 = Qualitative and descriptive statistics
3 = Quantitative and/or statistical significance of results assessed (look for p values)
4 = Computational (e.g., classifiers, machine learning, topic modeling, network analysis)
5 = Mixed
6 = Unspecified/other

V19. Causal inference (code one; based on whether the study uses these key terms)
This variable is aimed at understanding if the authors are claiming to have identified a statistically significant causal relationship between variables (e.g., misinformation -> anti-vaccination attitudes).
Signal words to look out for: causal, causality, direct relationship.
0 = No
1 = Yes

V20. Type of journal (code one)
1 = Public health/medicine
2 = Media and communications
3 = Anthropology
4 = Computer science
5 = Social sciences/humanities (e.g., economics, political science, linguistics, etc.)
6 = General (*Science*, *Nature*, *PLOS One*, *PNAS*)
7 = Other

V21. What is the main issue?
(automatically filled out)

V22. Title & abstract language used (code one)
1 = Anti-vaccine/anti-vaccination
2 = Vaccine hesitancy
3 = Vaccine skepticism
4 = Vaccine denial
5 = Vaccine critical
6 = Uses at least two terms interchangeably
7 = Other

V23. Stakeholders
The following variables assess which stakeholders are named by the authors in the article. Code one variable per stakeholder. The stakeholders have to be clearly identified by the authors. For example, if the article mentions that it has findings that could be helpful for "policymakers and health professionals," code V23a as 3, and then code V23b as 4. Note that this also follows the order in which stakeholders are mentioned (policymakers mentioned first, so coded in V23a; health professionals mentioned second, so coded in V23b). Please place a code in each variable, even if that code is 8 for the first three and 0 for the last, indicating no stakeholders are mentioned.

V23a. Are stakeholders named in the article, and which ones? (code one)
1 = Journalists/the media
2 = Patients and their relatives

3 = Policymakers and legislators
4 = Medical professionals and the public health sector
5 = Technology companies/platforms
6 = Academic researchers ("more research needed")
7 = Other
8 = None

V23b. Are stakeholders named in the article, and which ones? (code one)
1 = Journalists/the media
2 = Patients and their relatives
3 = Policymakers and legislators
4 = Medical professionals and the public health sector
5 = Technology companies/platforms
6 = Academic researchers ("more research needed")
7 = Other
8 = None

V23c. Are stakeholders named in the article, and which ones? (code one)
1 = Journalists/the media
2 = Patients and their relatives
3 = Policymakers and legislators
4 = Medical professionals and the public health sector
5 = Technology companies/platforms
6 = Academic researchers ("more research needed")
7 = Other
8 = None

V23d. More than 3 stakeholders named
0 = No
1 = Yes

V24. Demographic variables of research subjects
Focus of the article. This means: Is the article targeted at a specific demographic (e.g., mothers in Berlin, or in Boulder, Colorado, etc.). If there's no such focus or if the article only mentions genders, ethnicities, or ages, (e.g., 51% of survey respondents were male, 49% female; 15% were 18 or younger, 50% were 19 to 45, 35% were 46 or older), then code 0.

V24a. Gender
0 = No
1 = Yes

V24b. Ethnicity
0 = No
1 = Yes

V24c. Age
0 = No
1 = Yes

V25. Research focus (code one)
This variable asks for the unit of analysis as specified in the article (usually in the methods section).
1 = Identity: Who is vaccine hesitant? (e.g., surveys, polls, papers on collective identity and beliefs, etc.)
2 = Communication: How do people talk about vaccination? What do people say about vaccination? (e.g., analysis of anti-vaccination forums/subreddits, anti-vaccination hashtags; analysis of discourse on vaccinations as a whole)
3 = Information behavior: How do people search for information about vaccination? (information seeking and retrieval practices about vaccination)
4 = Conspiracy theories (e.g., vaccine hesitant as part of bigger conspiracy theory communities)
5 = Computational identification of anti-vaccination sentiments
6 = Other