Vaccine 37 (2019) 2057–2064

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

# Expressions of pro- and anti-vaccine sentiment on YouTube

Nikolaos Yiannakoulias [a,*], Catherine E. Slavik [a], Monika Chase [b]

[a] School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario, Canada
[b] School of Interdisciplinary Science, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:
Received 14 November 2018
Received in revised form 1 March 2019
Accepted 4 March 2019
Available online 9 March 2019

Keywords:
YouTube
Vaccine hesitancy
Participatory Internet
Text analysis

ABSTRACT

Billions of hours of YouTube content are viewed every day. Much of this content is aimed at entertainment, some of it is educational, and a considerable quantity is meant to influence or reinforce public opinion on a variety of matters, including health. Most of the content on YouTube is not created by professionals, public institutions or the traditional media, and instead is authored by private individual content creators. Given the potential impact of this medium for communicating health information, it is important for researchers and public health practitioners to understand the nature of health information as it is shared on YouTube. The primary objective of this research is to describe expressions of vaccine hesitancy content on YouTube, and specifically, compare the expression of pro- and anti-immunization sentiments. We do this by not only analyzing a systematic sample of influenza and measles immunization videos in terms of viewer analytics, but also by choice of language. We find that pro- and anti-immunization videos are common, but that videos with anti-immunization sentiment tend to be more 'liked'. We also find that a small number of words can be effectively used to identify anti-immunization content, an observation that could be useful for identifying trends in anti-immunization sentiment on social media. Our results suggest that public health experts may need to increase the profile of their content on YouTube, and that there may be some useful strategies for improving the public's disposition towards pro-immunization messaging.



## 1. Background

Despite the success of vaccines in reducing the global burden of infectious disease, public concerns about safety and efficacy persist. Vaccine hesitancy—an expression of concerns towards immunization that can lead to a delay or rejection of available and recommended scheduled immunization—may be responsible for lower immunization coverage in some populations [1,2]. Individuals that are vaccine-hesitant are characterised as having refused or delayed some vaccines while agreeing to others [3]. Some common misconceptions that have been linked to vaccine hesitancy include the belief that vaccines can overwhelm the immune system of a child [4], that there is no need for immunization due to the decline of many vaccine-preventable diseases [5] and that vaccines contain ingredients perceived to pose a risk to health [6].

The controversy surrounding immunization is not new and has existed since the development of vaccines themselves [7]. Recently, there has been a growth in research aimed at developing a deeper understanding of the causes and nature of vaccine hesitancy. Many factors could explain why people choose to delay or avoid immunization, with perhaps the most important being the sense that it is of low personal utility [8]. Other research suggests that a lack of trust in health professionals, government and/or the pharmaceutical industry as well as concerns over the expansion of the recommended immunization schedule for children are also key determinants of vaccine hesitancy [5,9]. A divide in opinion over vaccine safety and efficacy between some members of the public and scientists and medical experts may also be a driving force in fueling vaccine hesitancy and uncertainty by stoking vaccine controversy [10].

Any modern discussion of the causes of vaccine hesitancy must consider the role of new forms of information sharing found on the Internet. The Internet has been used to disseminate important information by government agencies, medical professionals and academia for decades. In recent years, the emergence of the participatory Internet (sometimes referred to as Web 2.0) has increased the availability of information generated by a broader range of content creators [11]. This has resulted in a shift away from the dominance of institutionally created information (published by academia, government and industry) to information created by non-institutional and independent users. This shift is marked by an increase in both the volume of participatory Internet content, as well as platforms that increase the reach of this content by

* Corresponding author.
E-mail address: yiannan@mcmaster.ca (N. Yiannakoulias).

https://doi.org/10.1016/j.vaccine.2019.03.001
0264-410X/

avoiding the filtering and vetting common to traditional media platforms. This change comes with some concern however; recently published research is revealing some worrying trends about health content on participatory Internet platforms like Twitter, Facebook, Instagram and YouTube. Specifically, the growth of the participatory Internet is accompanied by an increased sharing of misinformation that could be harmful to public health [12–14].

YouTube is a particularly important platform in the current participatory Internet landscape. YouTube reports over one billion users, and billions of hours of videos viewed each day [15]. Nevertheless, recent research suggests that some of the health-related content on the platform is problematic, due to misinformation, promoting of conspiracy theories or reinforcing negative stigma [16–23]. To date, there has been little research surveying the content of immunization information on YouTube. One recent exception suggests that YouTube is more populated by anti-immunization messaging than other online sources [13]. On the one hand, this suggests the platform is meeting a demand for alternative perspectives, but on the other hand, it may lead to a more misinformed public. There are a number of suggestions for solving the problem of misinformation on the participatory Internet, including increasing editorial control within the platform [13]. However, it is unclear if such strategies can successfully challenge the presence or influence of highly motivated content creators who have embraced a postmodern skepticism towards science and medical expertise [24]. In an analysis of website content on vaccines, Moran et al. (2016) found that anti-vaccination websites used a variety of strategies for preserving the anti-vaccine agenda in the face of public criticism, including sharing ideas about identity and individualism, mistrust in experts and other health and lifestyle decisions [25]. Recent research reveals the importance of understanding the language used to communicate on user-generated social media platforms, and suggests a pattern of positive and negative sentiment that could inform more effective content generation in the future [26].

Given the important role of participatory Internet in modern information sharing, and its potential for increasing levels of vaccine hesitancy, it is important to understand the nature of this information, what users think about it, and to develop strategies that improve access to higher quality health information while preserving the opportunities for individual expression that the participatory Internet affords. The primary objective of this analysis is to build on previous work describing immunization content on YouTube, and specifically, the expression of pro- and anti-immunization content. Our analysis involves the comparison of two different vaccines—measles and influenza—since evidence suggests that perceptions towards immunization appears to be vaccine specific [27–29]. Our specific approach studies both the content and viewer disposition toward YouTube videos related to immunization using a combination of video metadata (including views, likes and dislikes) and specific word use in the videos extracted from video transcripts. The analysis of video metadata is important since it can measure the attitudes of viewers towards the content contained in the videos. The analysis of specific words within the videos is important since it can help identify key words that signal disposition towards videos. Our comparison across vaccine types helps identify the degree to which content is disposed towards immunization generally, or specific to particular vaccines. The findings of this research may help inform future strategies to increase the profile of useful information on YouTube by increasing the relative appeal of more informative content.

## 2. Methods

To conduct this research, we first collected a sample of YouTube videos for analysis. A complete list of all immunization-related YouTube videos would be required to select a true random sample of videos. As far as the authors are aware, such a list has never been compiled, and would be very challenging to create. Further, this would provide a poor representation of watched video content, since many of the videos within the population of all immunization videos available are rarely viewed. Therefore, our strategy was to collect a sample of frequently viewed videos. We describe the process in detail in the following paragraphs.

Our approach was to select a sample using the Google Search function available in YouTube. The videos returned from this search were the sampling frame from which we then drew a sample for analysis. This required the selection of search phrases to use in the Google Search function available in YouTube. To select these search phrases, we first used Google Trends [30], to identify the relative frequency of potential search terms. Our intent was to find search terms that were widely used, and result in distinct sample frames for the influenza and measles immunization content. We tested a large number of search terms and phrases. According to Google Trends data, searches for "flu", "flu shot" and "flu immunization" were considerably more frequently used than searches for "measles", "measles vaccine", and "measles immunization". The use of the most general search terms alone ("flu" and "measles") returned a large number of videos unrelated to immunization. The search terms "flu shot" and "measles vaccine" (both without quotes) returned a large number of videos, with over one million for "flu shot" and over 45,000 for "measles vaccine". Further, there was less overlapping content using these search terms than using the "flu vaccine" and "measles vaccine" and "flu shot" and "measles shot". We therefore used the "flu shot" and "measles vaccine" search phrases (without quotes) to select the sampling frame of videos to view.

After typing these search terms into the YouTube search function, we selected videos in the order in which they appeared in the YouTube search results. Our original sample included the top 150 videos for "flu shot" and "measles vaccine" (henceforth referred to as flu and measles videos, respectively) based on the order returned in the YouTube searches. From this sample we gathered metadata for all of these videos using the Google YouTube API [31]. Metadata extracted included: views, 'likes', 'dislikes' and comment count. All metadata were extracted on May 9, 2018. We also extracted video transcripts, which record the text of words spoken in a video based on Google's automatic caption track [32]. Finally, we manually extracted duration of video, date of upload, and the number of uploader subscribers from each video. Subscribers are people with YouTube accounts who are notified when an uploader puts new content on their YouTube channel. Videos that were not in the English language, had been removed by YouTube at the time of viewing, or did not have complete metadata were excluded. This left samples of 141 flu videos and 134 measles videos.

Two researchers (NY and MC) viewed all videos in the sample that were 10 min or less in length and independently classified the content into (1) uploader type, (2) advocacy and (3) content categories, the last of which could have more than one type of content. When the researcher and research assistant disagreed on classification, we retained the more senior researcher's (NY) classification decision. For videos longer than 10 min NY was the lone classifier, leaving 220 videos classified by two researchers, and 55 videos classified by only one researcher.

Uploader type was classified as academic/government, traditional media/entertainment and independent. The independent category typically involved an uploader unaffiliated with any obvious organization, or in some cases, uploaders affiliated with strictly online media enterprises or businesses. This classification was done by observing the video, and visiting the uploader's channel page, and was done by both MC and NY. Of the 275 videos classified by both researchers, 5 videos differed in uploader classification.

**Table 1**
Target words.

adjuvant*, aerosol*, aids, allerg*, aluminum, ambulance*, aminoglycoside*, anaphyla*, antibod*, antigen*, antiviral, aspirin, autis*, autoimmune, bacteri*, blood*, booster, cellular, chemical*, chemotherap*, chickenpox, chiroprac*, chronic,clinic, congenital, contagio*, dentritic, diabetes, diagnos*, diarrhea, diphtheria, disease*, dna, doctor*, drug, dtap*, ebola, embryo, endocrine, epidemi*, fatigue, febrile, fetal, fetus, fever*, formaldehyde, fructose, gardasil, gastrointestinal, gelatin, gene*, gmo, hemagglutinin, hepatitis, hiv, homeopath*, hospital*, hospitalize, hygien*, illness*, immun*, inactivated, incubat*, infect*, inflamm*, influenza, injectable, inocul* inoculate, iodine, kidney, laboratory, lead, leucocyte*, lymph, lymphocyte*, marrow, measle*, medic*, membrane*, meningitis, merck, mercury, microphage*, monosodium, msg, mump* myelitis, nanoparticle, nasal, naturopath*, neomycin, neoplastic*, nephrotoxic, neurolog*, neurotoxin*, nurse*, oregano, organism*, outbreak, oxygen, pancreas, pandemic*, paraly*, pathogen*, pediatric, pertussis, pharma*, pharmaceutical*, phosphate, physician*, pill*, poison*, polio, polysorbate, pox, pregnan*, prescription*, protein*, quarantine, respiratory, rotavirus, rubella, sanita*, scien*, seizure*, serum, sick*, silver, smallpox, sorbitol, sterilize, stimulant*, strain*, subcutaneous, surgeon, susceptible, syndrome*, syringe, tetanus, titer, toxic*, toxoid*, transmi*, tuberculosis, typhoid, vaccin*, varicella, vax, virol*, viru*, virulen*, vomit*, whooping cough, yeast

Advocacy was classified as anti-immunization, neutral and pro-immunization. A video was classified as anti-immunization when there was an explicit statement or recommendation against immunization, and pro-immunization when there was an explicit statement or recommendation in favour of immunization. For videos with mixed messages (for example, advocating immunization generally, but not with people with a history of allergies) then the overall sentiment of the video guided classification. Videos were classified as neutral when there were either both pro- and anti-immunization messages, when the content made no explicit recommendations, or when the overall sentiment towards immunization was otherwise unclear. Of the 220 videos classified by both researchers, 14 videos differed in advocacy classification.

We classified video content as informative, celebrity or personal. Videos could be flagged as having all three, two, one or none of these content types. *Informative* content discussed information related to immunization—such as vaccine components, vaccination schedule and the prevalence of vaccination in a population. We did not consider the accuracy of this information for the purposes of classification. *Celebrity* content included videos, stills or discussion of one or more celebrity figures in the video. *Personal* content included either a portrayal or description of personal experiences related to immunization—including the process of getting immunized, feelings before or after immunization, or experiences of illness associated or perceived to be associated with immunization. There was disagreement on the content classification for 6 of the 220 videos viewed by both researchers.

Of the total 275 videos, we extracted and analyzed 206 (116 flu, 90 measles) transcripts of video content. We used the R package cleanNLP [33] to find unique words in these transcripts. This included finding and identifying word stems to help ensure that contextually unimportant variations in word forms (such as differences between singular and plural) were coded as the same word. From this list of all the words in the video transcripts, we compiled a list of 155 immunization-related words—what we refer to as target words. All target words appeared in at least one flu and one measles video within our sample, and had to be relevant to the general topic of immunization, medical care or infectious illness (Table 1). This step was important to focus our analysis on the most topically relevant words. Word frequencies of the target words were calculated across all videos and by the classification categories. More sophisticated language processing—for example, using sentences, and the relationship between words in sentences—proved difficult due to the absence of punctuation in YouTube transcripts.

**Table 2**
Frequencies and percentages of video content categories for influenza and measles immunization videos on YouTube.

| | Flu shot (%) (n = 141) | Measles vaccine (%) (n = 134) | *p*-value (independent proportions z-test for difference) |
|---|---|---|---|
| Transcript available | 115 (81.56) | 88 (65.67) | 0.003 |
| Uploader | | | |
| Academic, Government and NGOs | 2 (1.42) | 12 (8.96) | 0.004 |
| Traditional media and entertainment | 14 (9.93) | 46 (34.33) | < 0.001 |
| Independent | 125 (88.65) | 76 (56.72) | < 0.001 |
| Advocacy | | | |
| Anti-immunization | 41 (29.08) | 26 (19.40) | 0.062 |
| Neutral | 77 (54.61) | 43 (32.09) | 0.018 |
| Pro-immunization | 23 (16.31) | 65 (48.51) | < 0.001 |
| Content (categories not mutually exclusive) | | | |
| Children | 47 (33.33) | 50 (37.31) | 0.490 |
| Information | 63 (44.68) | 97 (72.39) | < 0.001 |
| Celebrity | 12 (8.51) | 5 (3.73) | 0.010 |
| Personal story | 50 (35.46) | 22 (16.42) | < 0.001 |
| Upload year | | | |
| Pre 2013 | 13(9.22) | 14 (10.45) | 0.732 |
| 2013 | 12 (8.51) | 4 (2.99) | 0.050 |
| 2014 | 13 (9.22) | 13 (9.70) | 0.891 |
| 2015 | 15 (10.64) | 74 (55.22) | < 0.001 |
| 2016 | 31 (21.99) | 6 (4.48) | < 0.001 |
| 2017 | 48 (34.04) | 22 (16.42) | < 0.001 |
| 2018 (to March 3) | 9 (6.38) | 1 (0.75) | < 0.013 |

## 3. Results

Video frequencies and proportions across different content categories reveal a number of trends (Table 2). Most flu and to a lesser extent measles videos are uploaded to YouTube by independent content creators. A larger proportion of measles videos are uploaded by traditional media and entertainment compared to flu videos. A larger proportion of measles videos contain pro-immunization content when compared to flu videos, and a larger share of measles videos contained immunization information than flu videos. Flu videos contain more personal and celebrity content than measles videos. There is considerable year-to-year variation in uploaded content, with a notable spike in measles videos in 2015, and flu videos in 2017.

Video analytics metadata are presented in Table 3. Uploaders of flu videos have roughly twice the average number of subscribers compared to measles video uploaders. The mean number of video views was more than 10 times higher for flu videos than measles videos. Flu videos have more ‘likes’ and ‘dislikes’ than measles videos, as well as a higher likeability score, which is the ratio of

**Table 3**
Video analytics metadata.

| | Flu shot (n = 141) | | | | | Measles vaccine (n = 134) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | Standard deviation | Minimum | Maximum | Mean | Median | Standard deviation | Minimum | Maximum |
| Subscribers (1000 s) | 848.87 | 57.75 | 2590.97 | 0 | 21892.06 | 449.57 | 16.02 | 1372.20 | 0 | 12426.82 |
| Views (1000 s) | 1051.22 | 67.49 | 4749.47 | 131 | 38,019 | 83.62 | 0.76 | 618.55 | 9 | 6383 |
| Likes | 4060.00 | 716 | 13380.28 | 2 | 126,532 | 1545.80 | 4.5 | 11643.68 | 0 | 120,663 |
| Dislikes | 519.00 | 42 | 2096.04 | 0 | 17,610 | 61.01 | 1 | 501.25 | 0 | 5673 |
| Likeability (Likes -Dislikes)/ Views * 100 | 1.46 | 0.79 | 1.73 | −0.14 | 10.30 | 1.02 | 0.39 | 1.82 | −1.16 | 11.11 |
| Duration (minutes) | 9.16 | 4.82 | 12.82 | 0.20 | 113.85 | 6.63 | 2.48 | 25.00 | 0.37 | 280.45 |

the difference between ‘likes’ and ‘dislikes’ to the total number of views. Flu videos are on average longer in duration than measles videos. Assuming that the average viewer watches 10% of a video, the mean view time for flu videos in our sample is 16,064 h, and for measles videos, 924 h.

Table 4 shows summary statistics of word frequencies. Mean target word frequency is the frequency of target words averaged over the flu and measles videos; measles videos appear to contain slightly more target words than flu videos. When target word frequencies are calculated across the different advocacy classifications, it appears that anti-immunization videos contain more target words than neutral or pro-immunization videos. Numbers in parentheses are the ratio of average frequency of target words to the average frequency of all words multiplied by 100—effectively the percentage of the transcript text that is comprised of target words. Measles videos generally have a higher percentage of target words than flu videos; pro-immunization measles videos have almost twice the proportion of target words as pro-immunization flu videos. Pro- and anti-immunization videos tend to have a higher percentage of target words than neutral videos. There were no meaningful associations between word frequencies and likeability, or the number of subscribers, but there was a weak negative correlation ($R = -0.15$) between duration of video and percentage of target words.

Fig. 1 shows the mean likeability by the advocacy classes. Anti-immunization videos have the highest likeability for both flu and measles videos. Flu videos appear to have higher likeability than measles videos, although this is most evident for neutral video content. Fig. 2 shows video likeability for videos over time. The spike in 2018 is based on a smaller number of videos than other years as it only includes two full months of data. The only other noteworthy spike is in 2015, for measles videos. This is also the year that had the most measles video uploads (Table 2). For this particular year, the vast majority of measles view content was classified as pro-immunization, and had an average likeability of 1.80, which was well above the average likeability for pro-immunization videos (which was below 1.00). The anti-immunization content during this year had an average likeability of 1.50, well below the average likeability of anti-immunization measles videos over the entire time period (which was above 2).

Fig. 3 is a plot of ranked frequencies of the top 100 target words in pro-, neutral and anti-immunization videos. The x and y axes show the ranks of word frequencies for pro- and anti-immunization videos, with higher rank values corresponding to lower frequency. The size of the word is proportional to the ranked frequency of the neutral videos, with larger words signifying more frequently used, and smaller words signifying less frequently used. Words along the diagonal reference line have similar frequency for pro- and anti-immunization videos and words farther from the

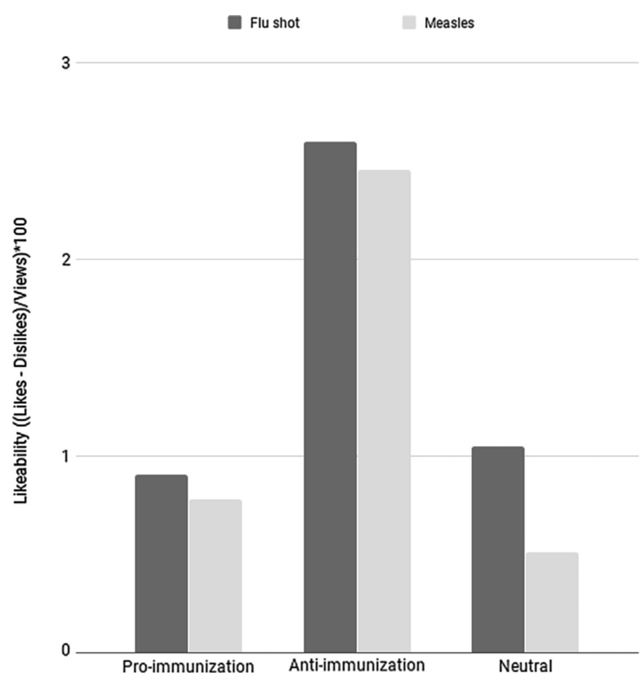


**Fig. 1.** Video likeability by immunization advocacy category.

**Table 4**
Target word frequencies for videos with transcripts.

| | Flu shot with transcript (n = 116) | | | | | Measles vaccine with transcript (n = 90) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Median | Standard deviation | Minimum | Maximum | Mean | Median | Standard deviation | Minimum | Maximum |
| Target words | 45.21 | 17.50 | 77.49 | 1 | 444 | 52.36 | 30 | 74.80 | 2 | 422 |
| Pro-immunization | 36.70 (3.76) | 19 | 40.161 | 1 | 141 | 34.02 (6.16) | 27 | 29.53 | 3 | 148 |
| Neutral | 14.45 (2.90) | 6 | 109.62 | 1 | 444 | 28.05 (3.18) | 18 | 123.97 | 3 | 422 |
| Anti-immunization | 83.91 (4.23) | 43 | 22.97 | 1 | 113 | 110.53 (4.48) | 61 | 22.52 | 2 | 96 |
| All words | 1328.54 | 689 | 2121.38 | 19 | 18,845 | 891.90 | 418 | 1378.98 | 34 | 7388 |
| Pro-immunization | 976.29 | 360 | 756.25 | 91 | 4849 | 551.82 | 476 | 1277.15 | 58 | 5707 |
| Neutral | 498.39 | 305 | 402.52 | 94 | 1413 | 911.44 | 749 | 697.56 | 62 | 2801 |
| Anti-immunization | 1983.61 | 1143 | 2191.54 | 260 | 7388 | 2464.56 | 1412 | 3255.41 | 120 | 18,845 |

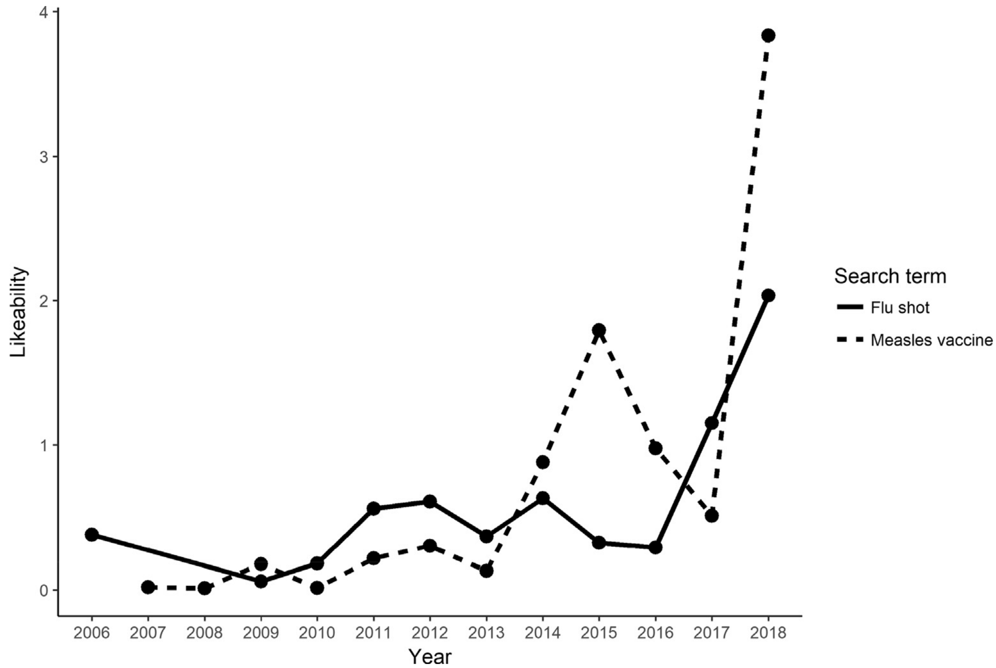


**Fig. 2.** Video likeability by year.

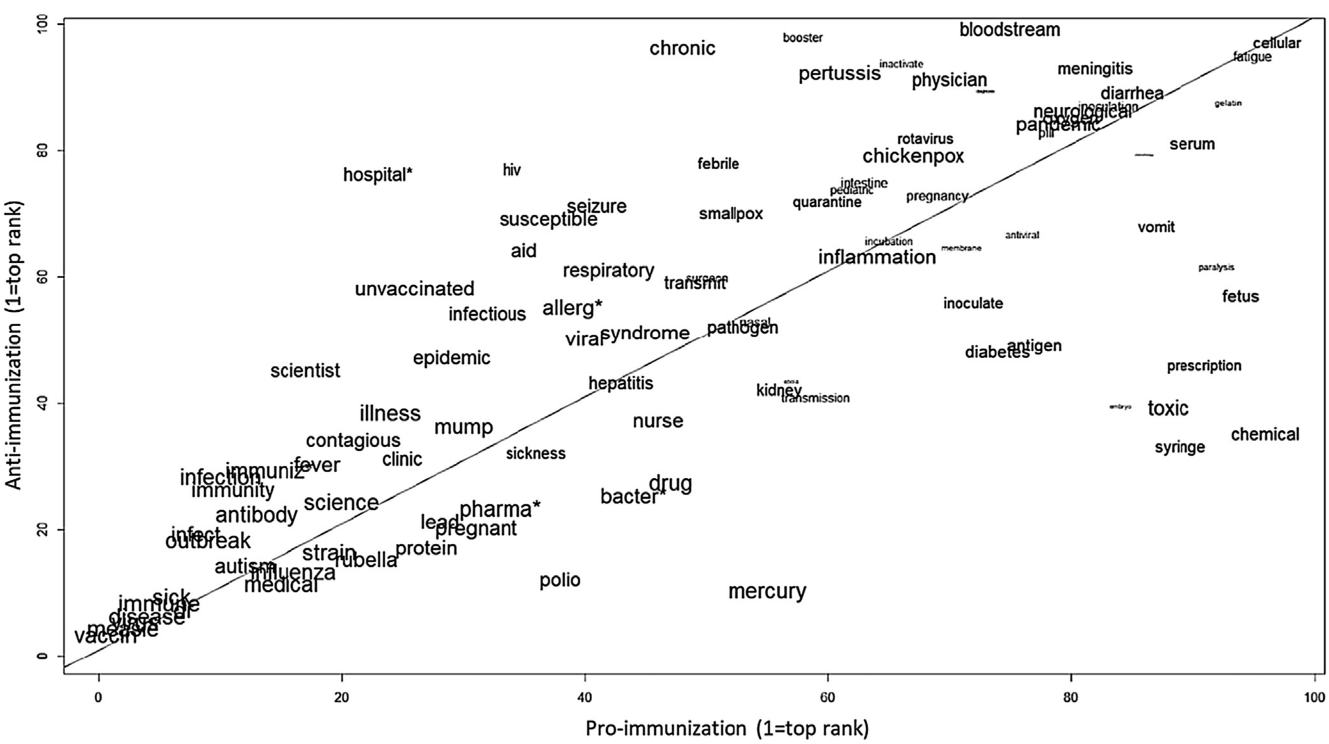


**Fig. 3.** Ranking of word frequencies by advocacy category.

reference line were used with differing frequencies. Words closer to the x-axis are relatively more common in anti-immunization videos and words closer to the y-axis are relatively more common in pro-immunization videos. Several differences are noteworthy. In the bottom left-hand corner are words that are common for pro-, anti- and neutral videos, including words like 'sick', 'vaccine' and 'disease'. Divergence in ranks is greater for less frequent words and, some specific target words appear to distinguish pro- and anti-immunization videos in particular. The word 'mercury' is used with considerably more relative frequency in anti-immunization videos, as were the words 'syringe', 'chemical' and 'toxic'. For pro-immunization videos, the words 'chronic', 'hospital*', 'unvaccinated' and 'HIV' are more common than for anti-immunization videos.

## 4. Discussion

### 4.1. Views and likeability

Similar to earlier research [34], our results show that YouTube is an important medium for expressing anti-immunization sentiment, and that overall, viewers express more positive sentiments towards anti-immunization videos than pro-immunization or neutral videos. Whether this is because anti-immunization viewers are more likely to express a negative sentiment than pro-immunization viewers, or there are just more anti-immunization viewers overall is unclear. What is clear is that YouTube is a particularly important forum for sharing anti-immunization feelings, which could be both an expression of existing vaccine hesitancy, as well as a mechanism for inspiring and reinforcing vaccine hesitancy in some populations of viewers. This could be particularly worrisome in populations that most frequently use YouTube—such as younger adults [35]—who may formulate important anchoring beliefs based on YouTube content that then influences them throughout their lives.

Flu videos in our sample had higher average likeability, more views and more subscribers. They also tended to be less informative, and were more likely to be classified as neutral or anti-immunization in tone than measles videos. Notably, nearly half of the measles videos were classified as pro-immunization, and less than 17% of flu videos were classified as pro-immunization. This could be expressed in the choice of language; measles videos had twice the proportion of target words when compared to flu videos. It could also be related to the content uploader as well; measles videos were more likely to be uploaded by academic, government and media sources than flu videos. A number of the flu videos have a very large number of views (over 20 million). These videos are personal videos of children getting influenza shots and acting in a way that is peculiar or amusing. Measles immunization videos were more likely to offer general information about immunization than influenza immunization videos, and less frequently showed children receiving shots.

The apparent connection between relatively higher likeability of flu videos that are simultaneously less informative, come from less authoritative sources and are more likely to contain anti-immunization content than measles videos suggests that immunization content on YouTube differs by vaccine. Compared to flu videos, measles videos are less common, and appear to be more focussed on delivering information rather than personal stories about immunization experiences. We can only speculate on why measles videos are systematically different from flu videos; it could be related to the target ages of immunization, the vaccination schedule, the perception of efficacy or media attention resulting from spikes in measles incidence. This latter explanation is consistent with our observation that the year with the largest number of measles video uploads in our sample is 2015, which coincides with a spike in reported measles cases in the California in late 2014 and early 2015 [36].

Interestingly, 2015 is also the year with highest likeability (1.79) for measles videos (with the exception of 2018, which is based on only two months of data), a value higher than the average for flu videos (1.46) and measles videos overall (1.02). Furthermore, pro-immunization measles videos in 2015 had a higher likeability than neutral or anti-immunization videos in that year. If this pattern is connected to an increased incidence of measles infections in some populations (and in particular, California), then it suggests that not only is there an increase in the number of uploads during periods of increased infection, but that viewers are more positively disposed to informative and pro-immunization measles content during times when risks of infection from measles is higher. This observation is consistent with the idea that attitudes towards immunization may respond to certain real world events; specifically, when confronted with tangible dangers of vaccine preventable illnesses, people may shift their disposition toward information about the benefits of immunization. If this is the case, it suggests that pro-immunization content on YouTube is most likely to have a useful impact when published during or immediately following a cluster of vaccine-preventable infection.

### 4.2. Word use of content authors

Anti-immunization content often includes the language of immunization science. Paradoxically, scientific language may be important for establishing the authority and expertise of the uploader to the viewing audience, even while the anti-immunization movement often appears to be part of a broader postmodern rejection of traditional science [24]. In spite of the similarity in language use overall, some key words seem to differentiate between the pro- and anti-immunization content—specifically, certain words are more frequently present in anti-immunization videos and also infrequently present in pro-immunization videos. This includes words like "mercury" and "toxic", which were more present in anti-immunization videos than pro-immunization videos. This observation is consistent with evidence of links between a fear of chemicals and health concerns in communities, including preferences for 'natural food' [37] and general concerns about the risk of 'chemicals' [38].

It is unclear if the word choices of the (largely) independent anti-immunization community are specifically coordinated to use particular language, however the common use of some specific terms suggests at least a thematic similarity among members of the anti-immunization community. It is worth noting that the YouTube transcripts used in our analysis may reveal more unscripted and less edited narratives than written content on social media and the Internet more broadly. This is because the transcripts from YouTube are extracted from spoken words, and the videos themselves are often free-form narratives and conversational in style. In this way, YouTube may capture a more authentic choice of language than would be extracted from text that has been edited and curated on high profile anti-immunization websites.

These observations are noteworthy for at least two reasons. First, they suggest that authors of anti-immunization content on YouTube may seek the veneer of scientific credentials even while their underlying message challenges the scientific consensus on immunization. Whether the authors actually have scientific credentials cannot be determined from videos alone, but their choice of language does suggest that these content developers believe that their audiences value the language of scientific authority. This suggests that the pro- immunization content is disliked more for the message itself rather than the language that is communicated. This might suggest that pro- immunization content may be most effective when communicating accurate science without proselytizing. This strategy might not be effective at changing the hardened opinions of much of the anti- immunization community, however it may be more effective for viewers who have yet to formulate concrete negative opinions towards vaccination.

Second, in spite of the overall similarities, there appear to be sufficient differences in language use that simple machine learning systems could be effective at detecting and identifying anti-immunization content on the participatory Internet—for YouTube as well as other social media platforms. Such systems could be used to identify and even predict geographical or temporal shifts in vaccine hesitancy patterns on social media, and better prepare experts for declines in immunization uptake that may emerge over time and/or at specific locations. This is particularly true if these findings apply to social media platforms like Twitter, which can contain information about the geographic location of a Tweet; emergent clusters of users Tweeting the words "mercury", "toxic", "chemical" and the like may signal a need for more careful observation and/or intervention. Such a system would need to observe both the frequency of content uploaded, but also indicators such as 'shares', 'likes' and other forms of interaction between users if it is to be useful for measuring engagement with content.

### 4.3. Limitations

There are several important limitations to this research, many of which concern the data we collected and used in our analysis. First, it is difficult to make direct comparisons of measles and flu video metadata since the search terms were chosen based on their frequency and to reduce overlapping content. As a result, it is possible that different search terms would have resulted in different content, and in turn, different findings. Another important data limitation is that we did not rate the accuracy or comprehensiveness of content in our analysis. This is challenging since it requires careful viewing of videos to assess not only the language used, but also the visuals used in the videos. The scope of our data collection did not include video comments, so we are unable to comment on the information generated by video viewers. However, based on a casual scan of this content, we would suggest that there is very little information content in user comments.

Our approach for classifying content was not validated in terms of accuracy, but only in consistency. It is possible that the researchers introduced a bias into the results through a systematic coding error. For videos coded by two researchers, there was a high degree of agreement in coding, but long videos were only viewed by one researcher, and it is difficult to know if these data were coded with the same degree of precision. It is important to keep these limitations in mind when interpreting our results, and further research is justified to determine whether our findings are generally meaningful.

There are some important limitations to using YouTube as the sole source of data for understanding attitudes towards immunization on the participatory internet. For one, videos do not contain detailed geographic information, so it is difficult to know analyze the content in a way that would be useful for understanding regional or local patterns of immunization behaviour. In many cases, content authors can be identified by country, but this lacks precision that could be useful for understanding shifts in perspective at scales where it is most likely to occur. Similarly, much of the available video content on YouTube is in English, and therefore, is likely to greatly under represent the breadth of views that exist. Future research should consider combining multiple platforms of user-generated content in order to ensure a more comprehensive coverage of content. This could be especially useful if platforms with the potential for greater geographic detail (such as Twitter) were included in future work.

## 5. Conclusions

The participatory Internet allows people to connect and share experiences and information in ways that can be empowering, entertaining, and improve quality of life. User-generated content associated with these technologies can influence ideas, and bring together communities of people with a variety of perspectives. Public health agencies and other expert communities have very little profile on YouTube, and the content they author is generally viewed negatively by viewers. Searches for immunization content on YouTube frequently show material that is related to personal stories, such as clips of children wincing when receiving a shot, rather than content about how vaccines work or the individual and social benefits of immunization. One obvious recommendation is that health agencies should do more to increase the profile of their content on YouTube, however, the manner in which this is done may be important. Our results suggest that the timing of upload could be important. As such, uploading and marketing information on YouTube and other participatory Internet platforms may be most useful during and immediately following periods of general concern, such as an outbreak of vaccine-preventable disease. We also found evidence that anti-immunization content uploaders use terminology similar to pro-immunization content uploaders, but in spite of this, subtle differences in terminology exist that could be useful for detecting trends in anti-immunization sentiment communicated on social media.

## Conflict of interest

The authors declare that there are no conflict of interests involved in the conducting of the research or preparation of the manuscript.

## References


[1] Dubé E, Gagnon D, Ouakki M, Bettinger JA, Guay M, Halperin S, et al. Understanding vaccine hesitancy in Canada: results of a consultation study by the Canadian Immunization Research Network. PLoS ONE 2016;11(6):e0156118.

[2] Eskola J, Duclos P, Schuster M, MacDonald NE. How to deal with vaccine hesitancy? Vaccine 2015;33(34):4215–7.

[3] Larson HJ, Jarrett C, Eckersberger E, Smith DM, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. Vaccine 2014;32(19):2150–9.

[4] Prymula R. Controversies in vaccination. Eur Rev 2013;21(S1):S56–61.

[5] Colgrove J, Bayer R. Could it happen here? Vaccine risk controversies and the specter of derailment. Health Aff 2005;24(3):729–39.

[6] Song G. Understanding public perceptions of benefits and risks of childhood vaccinations in the United States. Risk Anal 2014;34(3):541–55.

[7] Baker JP. The pertussis vaccine controversy in Great Britain, 1974–1986. Vaccine 2003;21(25–26):4003–10.

[8] Meyer SB, Lum R. Explanations for not receiving the seasonal influenza vaccine: an Ontario Canada based survey. J Health Commun 2017;22(6):506–14.

[9] Yaqub O, Castle-Clarke S, Sevdalis N, Chataway J. Attitudes to vaccination: a critical review. Soc Sci Med 2014;112:1–11.

[10] Dixon G, Clarke C. The effect of falsely balanced reporting of the autism–vaccine controversy on vaccine safety perceptions and behavioral intentions. Health Educ Res 2013;28(2):352–9.

[11] Cormode G, Krishnamurthy B. Key differences between Web 1.0 and Web 2.0. First Monday, 13(6); 2008.

[12] Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic—analysis of Facebook as a social media health information platform. Am J Infect Control 2017;45(3):301–2.

[13] Venkatraman A, Garg N, Kumar N. Greater freedom of speech on Web 2.0 correlates with dominance of views linking vaccines to autism. Vaccine 2015;33(12):1422–5.

[14] Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? BMJ 2014;349:g6178.

[15] YouTube n.d. YouTube by the numbers. [Access July 18, 2018]. <https://www.youtube.com/yt/about/press/>.

[16] Yiannakoulias N, Tooby R, Sturrock SL. Celebrity over science? An analysis of Lyme disease video content on YouTube. Soc Sci Med 2017;191(10):57–60.

[17] Nerghes A, Kerkhof P, Hellsten I. Early public responses to the Zika-virus on YouTube: prevalence of and differences between conspiracy theory and informational videos. In: Proceedings of the 10th ACM Conference on Web Science. ACM; 2018. p. 127–34.

[18] Murugiah K, Vallakati A, Rajput K, Sood A, Challa NR. YouTube as a source of information on cardiopulmonary resuscitation. Resuscitation 2011;82(3):332–4.

[19] Butler DP, Perry F, Shah Z, Leon-Villapalos J. The quality of video information on burn first aid available on YouTube. Burns 2013;39(5):856–9.

[20] Gonzalez-Estrada A, Cuervo-Pardo L, Ghosh B, Smith M, Pazheri F, Zell K, et al. Popular on YouTube: a critical appraisal of the educational quality of information regarding asthma 6. In: Allergy & Asthma Proceedings. p. 121–6.

[21] Lewis SP, Heath NL, Sornberger MJ, Arbuthnott AE. Helpful or harmful? An examination of viewers' responses to nonsuicidal self-injury videos on YouTube. J Adolesc Health 2012;51(4):380–5.

[22] Hussin M, Frazier S, Thompson JK. Fat stigmatization on YouTube: a content analysis. Body Image 2011;8(1):90–2.

[23] Yoo JH, Kim J. Obesity in the new media: a content analysis of obesity videos on YouTube. Health Commun 2012;27(1):86–97.
[24] Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm–an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012;30(25):3778–89.
[25] Moran MB, Lucas M, Everhart K, Morgan A, Prickett E. What makes anti-vaccine websites persuasive? A content analysis of techniques used by anti-vaccine websites to engender anti-vaccine sentiment. J Commun Healthcare 2016;9(3):151–63.
[26] Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017;35(29):3621–38.
[27] Little KE, Goodridge S, Lewis H, Lingard SW, Din S, Tidley M, et al. Occupational vaccination of health care workers: uptake, attitudes and potential solutions. Public Health 2015;129(6):755–62.
[28] Hak E, Schönbeck Y, De Melker H, Van Essen GA, Sanders EAM. Negative attitude of highly educated parents and health care workers towards future vaccinations in the Dutch childhood vaccination program. Vaccine 2005;23(24):3103–7.
[29] Heininger U. An internet-based survey on parental attitudes towards immunization. Vaccine 2006;24(37–39):6351–5.
[30] Google Trends; 2018. <https://trends.google.com>. [Accessed, November 2017].
[31] Google Developers; 2018. <https://developers.google.com/youtube/> [Accessed, May 2018].
[32] Google Developers; 2018. <https://developers.google.com/youtube/v3/docs/captions>. [Accessed, May 2018].
[33] Arnold T. A tidy data model for natural language processing using cleanNLP. The R J 2017;9(2):1–20.
[34] Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007;298(21):2482–4.
[35] The latest YouTube Stats on audience demographics: who's tuning in <https://www.thinkwithgoogle.com/data-collections/youtube-viewer-behavior-online-video-audience/> [Accessed October 2018].
[36] Zipprich J, Winter K, Hacker J, Xia D, Watt J, Harriman K. Measles outbreak–California, December 2014-February 2015. MMWR Morb Mortal Wkly Rep 2015;64(6):153–4.
[37] Dickson-Spillmann M, Siegrist M, Keller C, Wormuth M. Phthalate exposure through food and consumers' risk perception of chemicals in food. Risk Anal 2009;29(8):1170–81.
[38] Slovic P. Perception of risk. Science 1987;236(4799):280–5.