Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020)

The Effects of an Informational Intervention on Attention to Anti-Vaccination Content on YouTube

Sangyeon Kim, Omer F. Yalcin, Samuel E. Bestvater,
Kevin Munger, Burt L. Monroe, Bruce A. Desmarais
Pennsylvania State University, 203 Pond Lab
University Park, Pennsylvania 16802
{szk922, oxy4, seb654, kmm7999, burtmonroe, bdesmarais}@psu.edu

Abstract

The spread of misinformation related to health, especially vaccination, is a potential contributor to myriad public health problems. This misinformation is frequently spread through social media. Recently, social media companies have intervened in the dissemination of misinformation regarding vaccinations. In the current study we focus on YouTube. Recognizing the extent of the problem, YouTube implemented an informational modification that affected many videos related to vaccination beginning in February 2019. We collect original data and analyze the effects of this intervention on video viewership. We find that this informational intervention reduced traffic to the affected videos, both overall, and in comparison to a carefully-matched set of control videos that did not receive the informational modification.

Introduction

In September 2018, a massive measles outbreak was reported in New York City—a city that had been regarded as measles-free for decades (Jacqueline 2019b). A similar outbreak affected the state of Oregon in early 2019 (Jacqueline 2019a). These events cannot be viewed as simple anomalies. The World Health Organization has reported that measles cases are growing, and they suggest that this trend is deeply related to shortcomings in vaccination coverage (WHO 2018). Individuals' decisions to opt out of vaccinations have been caused, in part, by exposure to misinformation on social media (Dunn et al. 2015). There are many versions of misinformation about vaccination: a link with autism (Motta, Callaghan, and Sylvester 2018), dangers of injections for infants (Callaghan et al. 2019), a conspiracy to support pharmaceutical companies (Attwell et al. 2018), and so forth. These false characterizations are not new, but social media can serve to disseminate them more widely and quickly than has been done in the past.

Research across several disciplines has addressed the topic of misinformation related to vaccination (Covolo et al. 2017; Pandey et al. 2010; Wang et al. 2019). Much of this work has focused on the differences in content between pro-vaccination and anti-vaccination videos (Covolo et al. 2017; Keelan et al. 2007), while others investigate how attitudes towards vaccination are affected by watching anti-vaccination videos (Robichaud et al. 2012). However, little research addresses the ways in which the dissemination of misinformation can be effectively reduced. Social media companies, such as YouTube, Facebook, Reddit, and Twitter, have developed tactics to deal with misinformation (Shen and Rose 2019; Pham et al. 2018). We conduct a case study of a recent intervention conducted by YouTube (O'Donovan 2019). In February 2019, YouTube added a link to, and a preview of, the Wikipedia page on "Vaccine Hesitancy" below videos addressing vaccines. We analyze the effects of YouTube's intervention on engagement with affected videos. Specifically, we ask whether the placement of the link affected the viewership rate of affected videos. In the absence of data on historical viewership, we develop a novel approach to using time-stamped comments as a proxy for viewership. Our central finding is that the rate of viewership was substantially decreased through the placement of the link.

An Intervention by YouTube

Misinformation about vaccination from YouTube videos is problematic for several reasons. Above all, existing studies find that videos with misinformation about vaccination tend to be more viral than videos about vaccination in general (Song and Gruzd 2017). Viewing anti-vaccine videos tends to induce further recommendations to anti-vaccine videos (Song and Gruzd 2017). Lastly, anti-vaccine videos receive more likes and shares by users than vaccination related videos in general (Covolo et al. 2017).

YouTube rarely removes videos (Jordan and Shorland 2019). However, in February 2019, YouTube implemented some interventions regarding the hosting of videos with anti-vaccination content. YouTube "demonetized" some heavily anti-vaccination channels, such as VAXXED TV, by deleting advertisements from their videos. To other videos, they added a Wikipedia link and page-preview to the page about "Vaccine Hesitancy" (O'Donovan 2019). Figure 1 shows an example of how the Wikipedia link appears on these videos. The link and the preview appear directly beneath the video.

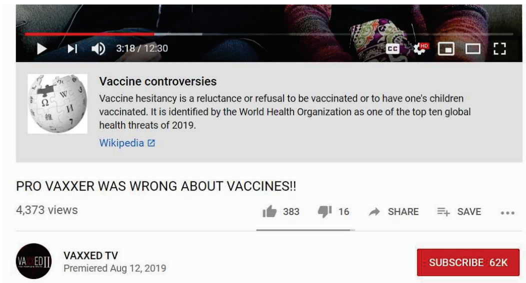

Figure 1: Screenshot of an affected video on YouTube

Research Objectives

Our goal is to determine whether YouTube's intervention had an effect on the viewership of affected videos. The intended effect was obviously to decrease the average user's engagement with a video, defined in part by YouTube as the total amount of time a video was viewed (Oliphant 2013). Total watch time would be reduced if viewers (1) navigated to the Wikipedia link, or (2) left the page as a result of the preview content, which defines Vaccine Hesitancy as a "global health threat." Since search rankings of videos are tied to total view time (Oliphant 2013; Venkatraman, Garg, and Kumar 2015), we suspect that, by leading users to shorten view times for the affected videos, the placement of the Wikipedia link would lead a video to rank lower in YouTube's search, in turn lowering viewership. We, unfortunately, cannot gather individual-user data on the effect of the Wikipedia link on view time, but we can analyze whether (and how) the aggregate trends in attention to the affected videos changed after the intervention, and how those changes compare to related videos that were not affected by the Wikipedia link.

Research Design

We first discuss how we gathered our sample of videos, through two methods: random walks through vaccine-related content; and manual collection through YouTube's basic search function. For the random walks, we built a list of fifty queries by adding "state name (e.g. Pennsylvania)" and "vaccination." We used state names to mimic a user searching for content regarding a state's vaccination policy, and also to induce some variation in the results at which each random walk began. Each walk included four steps: 1) search for the phrase; 2) click randomly on one of the first five results; 3) click randomly on one of the first five recommended videos; 4) repeat until: a) a video with the Wikipedia link for "Vaccine Hesitancy" is reached, b) the walk has taken ten steps, or c) neither the actual video title / description, nor any of the recommended video titles and descriptions contain the root "vaccin."

In addition to the random-walks, we searched directly for videos with the Wikipedia link by using a set of high-precision phrases in the YouTube search: "vaccine hesitancy," "vaccine denial," and "vaccine safety." In total, we found 105 videos that contained the Wikipedia link. Since we cannot accurately determine exactly when YouTube added the link, we use the date identified by BuzzFeed (O'Donovan 2019), February 22, 2019, as the intervention date.

Results

We are interested in understanding how the placement of the Wikipedia link affected viewership. One challenge is that the data available in the YouTube API (Sood 2019) does not provide time-stamped information about viewership. However, we can measure the current viewership count, along with the time-stamped comment count, using the YouTube API. We propose using the variation in the number of comments, including both original comments and replies, as as a proxy for the variation in the number of views, since all comments come with a time stamp. If comment and view counts are highly correlated, we will measure approximately the same behavior by focusing on comment counts. Below we present evidence that comment and view count are strongly correlated.

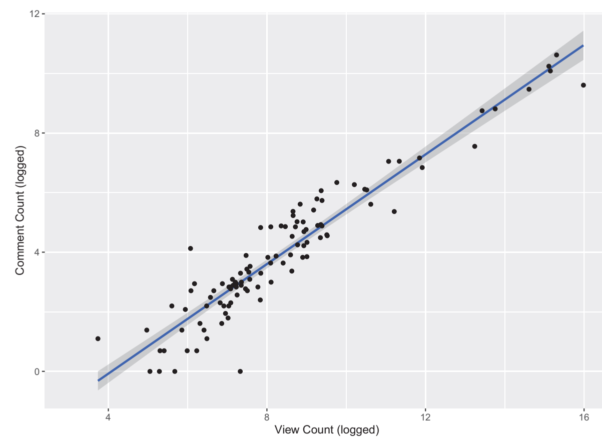

Figure 2: Correlation plot of view and comment counts

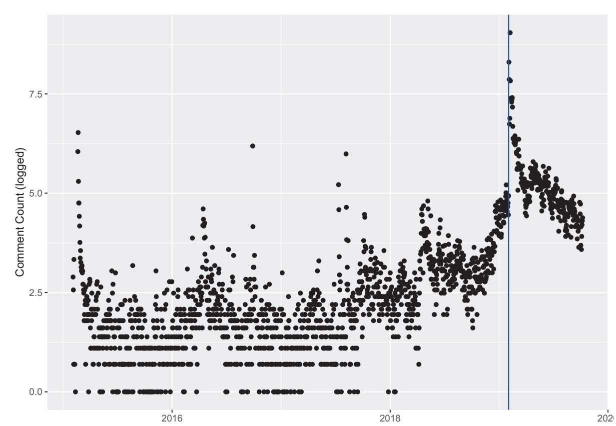

Figure 3: Effect of YouTube's intervention on videos from 2015 to 2019

Figure 2 is a scatter plot plot of the logged view counts and logged comment counts of our 105 videos. It shows a clear positive Pearson's correlation of 0.95 with 95% confidence interval of (0.9210387, 0.9628768). Since our main argument is about the longitudinal variation in the viewership of individual videos, not across-video comparisons, we also check longitudinal co-variation of view counts and comment counts within videos. Historical viewership data on videos is not available from YouTube, but the Wayback Machine, provided by the Internet Archive (Andersen 2013), offers historical coverage of some of the videos in our sample. There are 11 videos on our list that were archived by the Wayback Machine at least three times since January 1, 2018. For these videos, the average Pearson's correlation between view and comment count is approximately 0.99. We are confident in using comment counts as a measure of viewership.

Now we turn to the placement of the Wikipedia link and its effect. We begin with a series of visual analytics, and then move to a more precise estimation of the effect of the link placement using regression analysis. To evaluate the effects of YouTube's initiative, we selected the 102 videos that had at least one comment on September 1, 2018 or before and plot their comment traffic day by day between January 1, 2015 and October 1, 2019, in Figure 3. We go back to 2015 to ensure that the patterns we observe are not merely cyclical. The vertical date line shows February 22, 2019, the date known to be the time of YouTube's intervention. There is a clear downward trend after the intervention. Even though there are consistent positive outliers in January-March, the only period in which we observe a sustained downward trend is in the period following the intervention. The number of comments on videos clearly declines after the date of YouTube's action. We subject this result to a number of robustness checks—in terms of both analysis with regression, and visualization of more precise and limited subsets of data—to ensure that the pattern is not an artifact of some other broader trend.

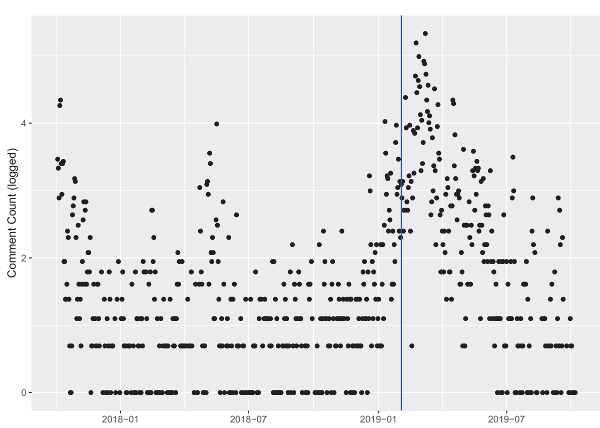

Figure 4: Effect of YouTube's intervention on videos from demonetized channels

The other potential concern regarding our finding is that the intervention's effect is driven by a few influential observations—in particular those videos associated with demonetized channels. To assess this concern, we divided the population by videos from demonetized channels and non-demonetized channels, and conducted the same analysis. The demonetized channels, which were reported to have been demonetized as part of YouTube's intervention (O'Donovan 2019), included in our sample are "Vaxxed TV," "LarryCook333," and "iHealthTube." The Wikipedia link appears on most pages of the videos related to vaccination in these channels. Since YouTube's demonetization decision affects the channel's incentives to disseminate videos, and to produce new content, the pattern shown in Figure 3 might be driven by videos from these channels. In total 74 videos in our sample came from demonetized channels while 31 are from non-demonetized channels.

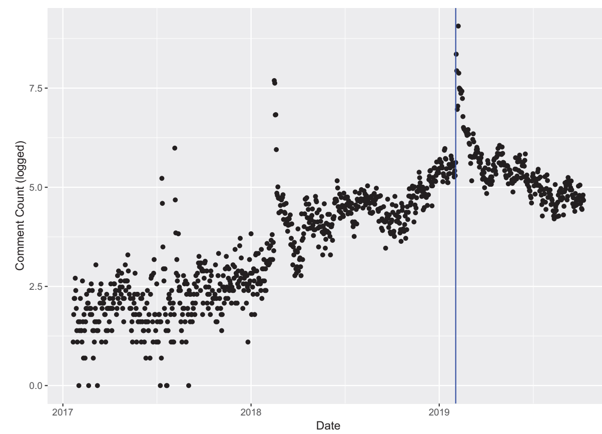

Figure 5: Effect of YouTube's intervention on videos from non-demonetized channels

The results depicted in Figures 4 and 5 demonstrate that the pattern shown in Figure 3 is not a product of outliers driven by demonetized channels.[1] In fact, the decreasing pattern after the intervention is much stronger for videos from non-demonetized channels.

Finally, we considered the possibility that the pattern identified in our analysis emerges from a broader trend affecting related videos. To evaluate this concern we collected "control videos" which do not include the Wikipedia link but are from the same channels as videos with the Wikipedia link—meaning they covered similar topics, and were uploaded on a similar date. We collected control videos with three criteria: 1) from the same channel as one of the original videos; 2) upload date similar or identical to the original videos; 3) view counts similar to original videos. We collected 52 control videos which satisfied these criteria. Figure 6 shows the trend for the control videos. There is a brief spike on the control videos in early 2019, but there is no significant downward trend.

To this point we have presented our results in terms of trend plots. Due to the dramatic nature of the trend, these are compelling, but these are not concisely quantified. In our final analysis, we present a precise estimate of the effect of the

[1] These plots are drawn starting in 2017 since there are no comments on the demonetized videos prior to 2017.

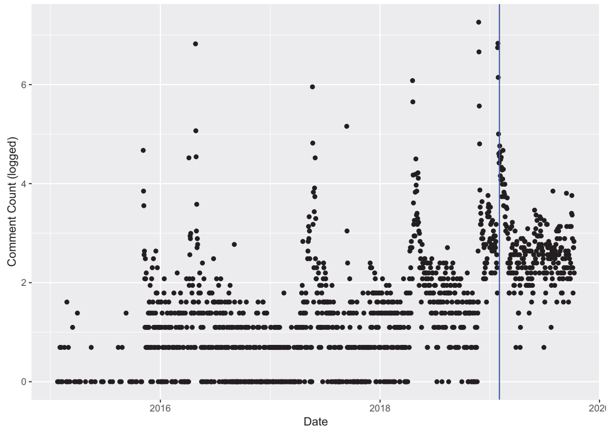

Figure 6: Comment counts of control videos

Post-intervention	-44.06^{**}
	(14.77)
Observations	3,264
Weeks	243
Week-by-week variance	2,247.85
Overall variance	59,547.63
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$	

Table 1: Regression estimate of the intervention effect. Standard error in parentheses.

intervention. In this analysis we define our outcome variable as the difference between an affected video, and its corresponding control video, in terms of the number of comments per week, over 243 weeks (01/28/2015–10/01/2019). This is analogous to a matching approach, in which the outcome is the difference between a treated video and its selected control video. Each affected video is included in the data from the first week that it was uploaded. Our predictor variable of interest is an indicator (1/0) of whether the respective week is post-intervention.We also include an intercept (i.e., fixed effect) for each video (unreported for the sake of brevity), to account for constant differences between the affected video and its control video (e.g., topic, quality, length). The fixed effects will account for any video-specific confounding variables that we did not measure (Plümper and Troeger 2007). We account for un-modeled temporal variation that affects all videos (e.g, seasonality) using weekly random effects. We estimate the model using the `lme4` package (Bates et al. 2015) in R.

The results of our regression analysis are presented in Table 1. We find that the intervention decreased the average difference in weekly comments between affected videos and their control videos by 44.06—a result that is statistically significant at the 0.05 level. This is a substantively significant result, as the average difference in weekly comments between affected and control videos, in our sample, is 37.05. The intervention effectively eliminated the difference in attention between the affected and control videos.

Conclusion

The spread of misinformation about medicinal practice, especially vaccination, can severely threaten public health. One common, and rapidly growing, conduit of misinformation is social media. It is important to understand what companies can do to limit the spread of misinformation. We studied an informational intervention implemented by YouTube. We found that the simple placement of an informational link below a video decreased engagement with affected videos. This intervention did not eliminate attention to these videos, but it did reverse an upward trend. The placement of informational links does not represent a complete method for limiting the spread of misinformation on YouTube, but, based on our results, it does represent a modestly effective tool for decreasing attention to misinformation.

Acknowledgments

This work was supported in part by the Center for Social Data Analytics at Pennsylvania State University, NSF grants SES-1558661, and SES-1637089. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect those of the sponsors.

References

Andersen, H. 2013. A website owner's practice guide to the wayback machine. *J. on Telecomm. & High Tech. L.* 11:251.

Attwell, K.; Ward, P. R.; Meyer, S. B.; Rokkas, P. J.; and Leask, J. 2018. "do-it-yourself": Vaccine rejection and complementary and alternative medicine (cam). *Social Science & Medicine* 196:106–114.

Bates, D.; Mächler, M.; Bolker, B.; and Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48.

Callaghan, T.; Motta, M.; Sylvester, S.; Trujillo, K. L.; and Blackburn, C. C. 2019. Parent psychology and the decision to delay childhood vaccination. *Social Science & Medicine* 238:112407.

Covolo, L.; Ceretti, E.; Passeri, C.; Boletti, M.; and Gelatti, U. 2017. What arguments on vaccinations run through youtube videos in italy? a content analysis. *Human vaccines and immunotherapeutics* 13(7):1693–1699.

Dunn, A. G.; Leask, J.; Zhou, X.; Mandl, K. D.; and Coiera, E. 2015. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. *Journal of medical Internet research* 17(6):e144.

Jacqueline, H. 2019a. As measles outbreak flares, vaccination rates soar and some come off the fence. *New York Times*.

Jacqueline, H. 2019b. New york tackles 'largest measles outbreak' in state's recent history as cases spike globally. *CNN*.

Jordan, G., and Shorland, A. 2019. What is youtube pushing you to watch next?

Keelan, J.; Pavri-Garcia, V.; Tomlinson, G.; and Wilson, K. 2007. Youtube as a source of information on immunization: a content analysis. *jama* 298(21):2482–2484.

Motta, M.; Callaghan, T.; and Sylvester, S. 2018. Knowing less but presuming more: Dunning-kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science & Medicine* 211:274–281.

O'Donovan, C. 2019. Youtube just demonetized anti-vax channels. *BuzzFeedNews*.

Oliphant, T. 2013. User engagement with mental health videos on youtube. *Journal of the Canadian Health Libraries Association/Journal de l'Association des bibliothèques de la santé du Canada* 34(3):153–158.

Pandey, A.; Patni, N.; Singh, M. abd Sood, A.; and Singh, G. 2010. Youtube as a source of information on the h1n1 influenza pandemic. *American journal of preventive medicine* 38(3):e1–e3.

Pham, C. V.; Thai, M. T.; Duong, H. V.; Bui, B. Q.; and Hoang, H. X. 2018. Maximizing misinformation restriction within time and budget constraints. *Journal of Combinatorial Optimization* 35(4):1202–1240.

Plümper, T., and Troeger, V. E. 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis* 15(2):124–139.

Robichaud, P.; Hawken, S.; Beard, L.; Morra, D.; Tomlinson, G.; Wilson, K.; and Keelan, J. 2012. Vaccine-critical videos on youtube and their impact on medical students' attitudes about seasonal influenza immunization: a pre and post study. *Vaccine* 30(25):3763–3770.

Shen, Q., and Rose, C. 2019. The discourse of online content moderation: Investigating polarized user responses to changes in reddit's quarantine policy. In *Proceedings of the Third Workshop on Abusive Language Online*, 58–69.

Song, M. Y. J., and Gruzd, A. 2017. Examining sentiments and popularity of pro-and anti-vaccination videos on youtube. *Proceedings of the 8th International Conference on Social Media and Society* 17.

Sood, G. 2019. *tuber: Access YouTube from R*. R package version 0.9.8.

Venkatraman, A.; Garg, N.; and Kumar, N. 2015. Greater freedom of speech on web 2.0 correlates with dominance of views linking vaccines to autism. *Vaccine* 33(12):1422–1425.

Wang, Y.; McKee, M.; Torbica, A.; and Stuckler, D. 2019. Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine* 112552.

WHO. 2018. Measles cases spike globally due to gaps in vaccination coverage.