Evaluating the Efficacy of Facebook’s Vaccine Misinformation Content Removal Policies

**Authors:** David A. Broniatowski\textsuperscript{1,2*}, Jiayan Gu\textsuperscript{3}, Amelia M. Jamison\textsuperscript{4}, Lorien C. Abroms\textsuperscript{2,3}

**Affiliations:**

\textsuperscript{1}Department of Engineering Management and Systems Engineering, The George Washington University; Washington, DC, 20052, USA.
\textsuperscript{2}Institute for Data, Democracy, and Politics, The George Washington University; Washington, DC, 20052, USA.
\textsuperscript{3}Department of Prevention and Community Health, The George Washington University; Washington, DC, 20052, USA.
\textsuperscript{4}Department of Health, Behavior, and Society, Johns Hopkins University; Baltimore, MD, 21218, USA.

*Corresponding author. Email: broniatowski@gwu.edu*
Abstract: Social media platforms have attempted to remove misinformation about vaccines because it obstructs efforts to end the COVID-19 pandemic. We examined whether Facebook’s vaccine misinformation removal policies were effective. Posts and engagements in anti-vaccine pages were reduced to 29% and 23% of pre-policy levels, respectively, but recovered over the subsequent six months. Posts and engagements in pro-vaccine pages were also reduced – to 68% and 30% of pre-policy levels, respectively. Low-credibility content became more prevalent in anti-vaccine pages and groups, and high-credibility content became less prevalent in pro-vaccine pages. Links between anti-vaccine pages and coordinated inauthentic behavior were also reduced. Our results suggest that Facebook’s policies were only partially successful. Facebook’s attempts at self-regulation appear to have been resource intensive, and ineffective in the long term.

One-Sentence Summary: Facebook’s vaccine misinformation removal policies penalized pro-vaccine content and were only effective in the short term.
Widespread misinformation about vaccines contributes to vaccine hesitancy, stalling efforts to overcome the COVID-19 pandemic (1–6). In the United States alone, over 60 million eligible Americans remain unvaccinated (7) despite several safe and effective vaccines. Social media platforms have historically hesitated to remove misinformation, preferring, instead, to downrank or flag it. However, when faced with significant pressure from public health and government officials, several platforms have reversed this position.

We seek to understand whether policies that removed content were effective. Historically, self-regulation on social media platforms has not worked well. For example, Facebook’s prior attempts to remove hate speech have had partial success at best and have sometimes been counterproductive (8, 9). Additionally, illegal drug sales and posts have proliferated on Facebook, despite the company’s stated plans to remove them and repeated requests from the U.S. Food and Drug Administration (10, 11). More recently, reports have alleged that Facebook’s policies are inconsistently applied and promote online toxicity, human trafficking, child pornography, incitement to violence, organ selling, and vaccine refusal (12).

There is reason to doubt that platforms like Facebook can effectively self-regulate. Facebook’s attempts to combat vaccine misinformation using “soft” content moderation remedies met with limited success (13, 14). On March 7, 2019, Facebook reduced the ranking of anti-vaccine posts in Newsfeed and Search, banned anti-vaccine advertisements, and promoted authoritative vaccine information by placing banners in vaccine related groups and pages. Furthermore, policies that were intended to increase ad transparency appear to have disproportionately removed pro-vaccine ads (15). Despite these limitations, Facebook initially resisted implementing “hard” content moderation remedies (13, 14), such as removing objectionable content (16). However, in August 2020, Facebook began removing accounts associated with the QAnon conspiracy theory that explicitly promoted violence, expanding this ban to all QAnon accounts on October 6, 2020. QAnon embraced anti-vaccine conspiracies (17) leading Facebook to remove “Stop Mandatory Vaccination” – one of the largest anti-vaccine pages – on November 18, 2020, when its administrator came to endorse QAnon content. On December 3, 2020, Facebook announced its intention to remove false claims about COVID-19 (18). On February 8, 2021, Facebook extended this policy to vaccine misinformation in general (19), while promising to increase the credibility of information shared. Thus, this time period – between November 18th, 2020, and February 8th, 2021 – marks a significant strengthening of Facebook’s tactics in its fight against vaccine misinformation with the introduction of content removal policies.

The efficacy of these policies has recently been called into question. Preliminary reports argue that misinformation is still widespread on the platform (12, 20), although Facebook has questioned these conclusions (21). Our aim is to evaluate the efficacy of Facebook’s new content removal policies. If successful, these policies should have 1) removed existing anti-vaccine posts in the pages and groups that hosted them; 2) reduced engagement with these anti-vaccine posts, 3) reduced the growth of new anti-vaccine posts, pages, and groups, and 4) promoted links to credible health sources while demoting or removing links to misinformative sources. In contrast, pro-vaccine posts, pages, and groups should not have been affected. We therefore seek to answer two questions: (i) did Facebook’s new policies successfully reduce the number of anti-vaccine pages, groups, posts, and engagements on the platform? (ii) did Facebook successfully improve the credibility of vaccine-related content on the platform? Additionally, we seek to determine whether Facebook’s actions addressed some of the limitations of their prior policies. For example, anti-vaccine page administrators often post links to one another’s pages (22) to
circumvent the 2019 policy’s downranking of anti-vaccine content in Search and Newsfeeds. These links aid users to discover new anti-vaccine pages and encourage the formation of a common anti-vaccine narrative (23), leading us to ask: (iii) did Facebook reduce links between anti-vaccine Facebook pages? Finally, disinformation on social media platforms has been associated with artificial content amplification (e.g., “bots”) (24–26). Although Facebook prohibits “coordinated inauthentic behavior” (27), prior work has found significant amplification of misinformation about vaccines and COVID-19 (28, 29). This leads us to ask (iv) did Facebook successfully reduce coordinated efforts to spread anti-vaccine content?

Data and Definitions
We leveraged data from CrowdTangle (30) -- a public insights tool owned and operated by Facebook, which tracks interactions with public content from Facebook fan pages and public groups (“assets”). Fan pages (“pages”) are primarily for marketing and brand promotion with posts made by the page administrators. Groups serve as a forum for users to discuss topics of shared interest, with posts made by group members. For a given post, we defined engagements as the total number of interactions (shares, comments, likes and emotional interactions, such as sad, love, angry, etc.,) with that post.

The original sample
To answer our first research question, we used an interrupted time series design (see Materials and Methods) to compare the weekly number of Facebook posts and engagements in public anti- and pro-vaccine Facebook assets to what existed prior to policy implementation. To obtain these data, we searched CrowdTangle on November 15, 2020 -- three days prior to Facebook’s removal of “Stop Mandatory Vaccination” -- and identified a set of 216 (114 anti-vaccine and 102 pro-vaccine) English-language pages and 99 (91 anti-vaccine and 9 pro-vaccine) English-language groups with content pertaining to vaccination. On November 15, 2020, we collected all available public posts from this sample from November 15, 2019, through November 15, 2020 (the “pre-policy period”), and their associated engagements. We next collected all available public posts from this sample from February 8, 2021 -- when Facebook announced its anti-vaccine content ban -- through August 8, 2021 (the “post-policy period”), and their associated engagements. These data were collected on August 11, 2021.

The final sample
Throughout the duration of our study, assets may have been removed for violating Facebook’s policies. Additionally, new assets may have formed. Therefore, beyond following the assets identified in our initial search (the “original sample”) we ran a second search on July 21, 2021, to identify new assets (the “final sample”) using the same technique (see Materials and Methods). The final sample contained 79 (33 anti-vaccine and 46 pro-vaccine) pages and 139 (69 anti-vaccine and 70 pro-vaccine) groups. Of these, 27 (34%; 14, 42% anti-vaccine and 13, 28% pro-vaccine) pages and 114 groups (82%; 47, 68% anti-vaccine and 67, 96% pro-vaccine) were not present in the original sample. On August 12, 2021, we collected all available public posts from the final sample for the post-policy period.

Content credibility
To answer our second research question, we examined content credibility, following the approach used by several authors (31–34), who attribute the quality of a post to its publisher. Specifically, we used the “Iffy Index of Unreliable Sources” (https://iffy.news/index/) -- whose
ratings are strongly correlated with similar ratings from other sources (32) -- to identify a list of publishers that routinely post falsehoods and conspiracy theories. We considered all posts containing a link whose top-level domain was on this list to be low credibility. Similarly, following (32), we considered posts from “high quality health sources”, which we operationalized as academic and government sources, as high credibility. Of all posts containing a link pointing off the Facebook platform (“off-platform links”), we calculated proportions pointing to either low- or high-credibility sources.

**Links between anti-vaccine pages**

To answer our third research question, we examined whether Facebook’s new policies reduced links between the anti-vaccine pages in our sample (we did not examine groups because Facebook’s prior policies focused on pages and links to groups were rare). Specifically, of all posts containing a link pointing to an asset on Facebook (“in-platform links”), we calculated the proportion pointing to the anti-vaccine pages in our samples.

**Coordinated link-sharing**

To answer our fourth research question, we examined whether Facebook’s policy reduced the prevalence of coordinated behavior. Following prior work (29, 35, 36), two pages or groups were considered “coordinated” if they routinely shared identical links “near simultaneously, which we operationalized as within 25 seconds of one another. Pairs of pages or groups were defined as “routinely coordinated” if they posted the same link “near simultaneously” more frequently than would be expected given link sharing behavior in unrelated contexts (see Materials and Methods).

**Results**

The original sample contained 119,091 posts in vaccine-related pages and 168,419 posts in vaccine-related groups for the pre-policy period, and 27,344 posts in pages and 36,601 posts in groups for the post-policy period (Table 1). Qualitatively, these posts focused on general vaccine content, with content adhering closely to existing typologies of pro- and anti-vaccine topics (37, 38) (Table S1). In the final sample, we identified 17,001 posts in pages and 248,971 posts in groups for the post-policy period. Qualitatively, content in the final sample was predominantly pro-vaccine, largely due to increased activity in groups. Compared to the original sample, posts in the final sample were more likely to focus specifically on COVID-19 vaccines, especially as these new vaccines were made available to the public. New pro-vaccine venues included regional “Vaccine Hunter” pages and groups focused on alerting individuals to specific locations they could access vaccines. The final sample also included new, more ambiguous pages that presented as forums for discussion of personal experiences of COVID-19 vaccine side effects, ostensibly for users to form their own judgments.

**Table 1. Number of posts in each sample.**

| Original Sample, Pre-Policy Period (November 15, 2019-November 15, 2020) | Anti-Vaccine (%) | Pro-Vaccine (%) | Total |
|---|---|---|---|
|  |  |  |  |
Effect on number of groups and pages

Although anti-vaccine posts still made up most of the posts in both pages and groups, the relative proportion of all anti-vaccine posts was significantly smaller in the post-policy period (RR = 0.72; 95% CI: 0.71-0.73 for pages and RR = 0.95; 95% CI: 0.95-0.96 for groups). This change was due, in part, to the removal of anti-vaccine groups and pages. We found that 28 (13%; 22, 19% anti-vaccine and 6, 6% pro-vaccine) pages and 24 (24%; 22, 24% anti-vaccine and 2, 22% pro-vaccine) groups in the original sample had been removed. Anti-vaccine pages (RR=3.28, 95% CI: 1.39-7.78), but not anti-vaccine groups (RR=1.09, 95% CI: 0.30-3.90) were significantly more likely to have been removed compared to their pro-vaccine counterparts. In addition, 5 (5%) anti-vaccine groups had changed their settings from public to private.

Effect on content volume and engagements

Figure 1 shows that, within the original sample, total posts and engagements per week decreased significantly for anti-vaccine, but also pro-vaccine assets (p<0.01 in all cases; Table S2) in the post-policy period, compared to the pre-policy trend. These results replicated for posts, but not for engagements, when restricting analysis to those anti-vaccine assets that had not been removed (p<0.02; Table S2). Despite this decrease, the weekly growth rate of posts and engagements in anti-vaccine pages (but not groups) increased (p<0.01) such that post and engagement counts are expected to have fully recovered after 121 weeks and 49 weeks, respectively. When comparing the final sample to posts in the pre-policy period, weekly anti-vaccine posts and engagements in pages all decreased significantly (p<0.001). In contrast, the growth rate of weekly engagements in anti-vaccine groups increased (p<0.001). There are now more engagements in anti-vaccine groups than would be expected prior to the policy’s
implementation. Additionally, the weekly growth rates of posts and engagements in anti-vaccine pages has significantly increased (p<0.001).

![Graph showing weekly posts and engagement counts for anti-vaccine and pro-vaccine assets.](image)

**Fig. 1. Weekly posts and engagement counts for anti-vaccine and pro-vaccine assets.** Dashed lines have been fit to data from the pre-policy period and are projected forward to the end of the post-policy period. Solid lines indicate average weekly post volumes for data in the post-policy period. All linear fits have been applied after applying a logarithmic transform.

**Effect on content credibility**

On average, 55% and 42% of posts in anti-vaccine pages and groups, respectively, contained links to external websites, with these proportions remaining roughly stable over time in both the original and final samples. Contrary to expectations, we found that the fraction of posts with links to low-credibility sources -- and engagement with these posts -- has increased for anti-vaccine groups in both samples (Fig. 2; p≤0.001 in all cases, Table S3). Among anti-vaccine pages, the fraction of posts with links to low-credibility sources also increased in the final sample, as did engagements with corresponding posts in the original sample (p≤0.001). Additionally, the proportion of links to high-credibility health sources in pro-vaccine pages has begun to decrease over time (p=0.01). We did not analyze low-credibility links in pro-vaccine assets because they were too infrequent.
Fig. 2. **Weekly proportions of posts, and engagements, containing off-platform links with links pointing to low-quality sources.** Dashed curves have been fit to data in the pre-policy period and are projected forward to the end of the post-policy period. Solid lines indicate average weekly link proportions for data in the post-policy period. All linear fits have been applied after applying a logistic transform to data.

**Effect on links between pages**

Forty-two percent of posts in Facebook pages contained an in-platform link, with roughly equal proportions across samples. Links between anti-vaccine pages were almost twice as common (RR=1.94, 95% CI: 1.54-2.45) in the last six months of the pre-policy period (results were qualitatively similar for the whole pre-policy period). In contrast, in the six-month-long post-policy period, links between pro-vaccine and anti-vaccine pages occurred at roughly equal rates (RR\textsubscript{original}=0.81, 95% CI: 0.57-1.15; RR\textsubscript{final} = 0.64, 95% CI: 0.36-1.13). The proportion of page-pairs that were linked also decreased (Fig. 3): compared to the last six months of the pre-policy period, anti-vaccine page network density in the original sample decreased significantly during the post-policy period (RR=0.43, 95% CI: 0.32-0.57). In contrast, pro-vaccine page network density in the original sample remained roughly the same (RR=1.03, 95% CI: 0.76-1.39). Comparing the final sample to data from pre-policy period, the pro-vaccine page network was significantly denser (RR=2.10, 95% CI: 1.50-2.95) and we did not detect a significant difference in density among corresponding anti-vaccine pages (RR=0.69, 95% CI: 0.41-1.16).
Fig. 3. **Network diagram depicting links between Facebook pages.** Red (green) nodes denote anti- (pro-) vaccine pages. Black nodes denote pages that were present in the pre-policy period but had been removed in the post-policy period. Hollow nodes denote pages that were present in the final sample, but not the original sample. Black edges denote links that were present both in the last six months of the pre-policy period, and in the post-policy period. Dashed grey edges denote links that were present in the last six months of the pre-policy period, but not in the in the post-policy period. Solid colored edges denote links that were not present in the last six months of the pre-policy period, but were present in the post-policy period: red edges link between two anti-vaccine pages, green edges link between two pro-vaccine pages, and blue edges link between an anti-vaccine and a pro-vaccine page. Node location and graph layout was determined using the Fruchterman-Reingold force-directed layout algorithm (39).

Figure 4 shows that, during the pre-policy period, the proportions of links between anti-vaccine pages, and engagements with those links, were increasing (p<0.01 in all cases, Table S4). Projecting this growth rate forward to the start of the post-policy period, we did not detect a reduction in the proportions of links or engagements between anti-vaccine pages (p>0.1 in all cases). However, following the policy’s implementation, the proportion of links between anti-vaccine pages (and corresponding engagements) started decreasing (p<0.001 in all cases). In
contrast, links between pro-vaccine pages (p<0.001), and corresponding engagements (p<0.05), seem to have decreased in prevalence at the start of the post-policy period.

Fig. 4. Weekly proportions of posts containing in-platform links and corresponding engagements, pointing within the anti- and pro-vaccine page networks. Dashed curves have been fit to data from the pre-policy period and are projected forward to the end of the post-policy period. Solid lines indicate average weekly post volumes for the post-policy period. All linear fits have been applied to logistic transformed weekly proportion data. Links between pro- and anti-vaccine pages were rare (Fig. 3).

**Effect on coordinated link sharing**

Compared to the pre-policy period, anti-vaccine pages and groups in both samples had significantly smaller proportions of “near simultaneous” link shares (p<0.01, Table S5). Furthermore, in the last six months of the pre-policy period, links in anti-vaccine pages and groups were significantly more likely to be coordinated than corresponding pro-vaccine links (RR = 6.36, 95% CI: 3.37-11.95). In contrast, in the post-policy period, we did not detect a significant difference in coordination between anti- and pro-vaccine pages and groups (RR_{original} = 3.26, 95% CI: 0.98-10.85; RR_{final} = 1.26, 95% CI: 0.74-2.28).

Coordination was largely driven by near-simultaneous activity in groups, rather than pages, for both pro- and anti-vaccine content. 91% of the coordinated links in the post-policy period involve at least one new group with no significant difference between anti- and pro-vaccine content (RR_{original} = 1.15, 95% CI: 0.84-1.57; RR_{final} = 0.99, 95% CI: 0.90-1.09). Of the 25 new pro-vaccine groups and pages detected on July 21, 2021, none were connected to a previously identified group or page in our dataset. In contrast, 5 (33%) of the 15 new anti-vaccine groups that we detected coordinated directly with a group that had existed prior to the policy’s implementation, and all but 2 (13%) were reachable through a network of coordinated anti-
vaccine groups (see Fig. 5). As shown in Fig. 1, the popularity of these groups declined as vaccine supplies became more widespread, whereas activity in anti-vaccine groups continues to grow.

**Fig. 5. Network of coordinated links between Facebook groups and pages.** Square (circular) nodes denote groups (pages). Black nodes denote assets that had been removed or become private in the post-policy period. Hollow nodes denote assets that were present in the final, but not original, sample. Black edges denote links that were present both in the last six months of the pre-policy period and in the post-policy period. Dashed grey edges denote links that were present in last six months of the pre-policy period, but not in the post-policy period. Red edges link between two anti-vaccine groups or pages, and green edges link between two pro-vaccine groups or pages. Node location and graph layout was determined using the Fruchterman-Reingold force-directed layout algorithm (39).

**Discussion**

This study evaluated the efficacy of Facebook’s anti-vaccine content removal policies introduced between November 18, 2020 and February 8, 2021. The platform appears to have had success in curtailing some anti-vaccine content as indicated by significant reductions in publicly available anti-vaccine posts, engagements, links between pages, and coordinated behavior.

Facebook’s actions were not without costs: Pro-vaccine content was also penalized. This suggests that Facebook’s efforts to remove anti-vaccine content may have been imprecise, causing collateral damage and reducing the availability of information or discussions supporting vaccination.

Despite these costs, anti-vaccine content recovered during our period of observation. We found that anti-vaccine page post counts grew and, based on projections, would ultimately return to prior levels. This means that the gains made by these policies may be both costly and temporary.
Engagements with anti-vaccine content are also recovering and, in anti-vaccine groups, had already surpassed pre-policy projections.

Furthermore, it appears that Facebook’s efforts to remove anti-vaccine content may not have targeted the most egregious sources of information. Contrary to the intent of the policy, the proportion of links to high-credibility sources in pro-vaccine pages decreased, and proportions of low-credibility content in anti-vaccine Facebook assets increased. Thus, Facebook’s efforts to remove misinformative content may have had the unintended consequence of driving more attention to websites that routinely promote misinformation.

Facebook appears to have been particularly successful in preventing behaviors that circumvented the intent of prior policies. For example, the policy seems to have reversed the growth of links between anti-vaccine pages (although links between pro-vaccine pages also decreased), making some content harder to find outside of Search. Additionally, the policy seems to have reduced coordinated link sharing behavior in anti-vaccine assets.

Limitations to this study include that only public Facebook pages and groups were studied. We cannot make claims about information in comment threads and private spaces. This leaves lacunae in external evaluators’ ability to study Facebook’s policies and raises the prospect that anti-vaccine posts and engagements in pro-vaccine spaces may have been mislabeled. For example, pro-vaccine “vaccine hunter” pages and groups appear to have garnered significant engagement as demand for COVID vaccines increased in early 2021; nevertheless, we observed that these venues were sometimes flooded with sensationalized stories of vaccine harm and may have been targeted by anti-vaccine activists attempting to circumvent content policies, as has been reported elsewhere (12). Nevertheless, the data available through tools such as CrowdTangle constitute a critical window into the largest, and most public, venues on the platform, which are most likely to achieve high numbers of exposures.

Based on data available, Facebook’s policy appears to have been only partially effective. Initial gains are in danger of being reversed and Facebook’s algorithms seem unable to reliably distinguish between harmful and helpful content. Furthermore, deciding what types of content to allow may require significant in-house scientific expertise, especially given the swiftly-changing nature of our scientific understanding of the COVID-19 pandemic. Apparently, self-regulation requires continuous oversight, enforcement, adaptation, and the application of significant internal expertise, human judgment, and accountability principles. External, transparent evaluations are therefore crucial to ensure success.

References
1. World Health Organization, Ten health issues WHO will tackle this year (2019), (available at https://www.who.int/emergencies/ten-threats-to-global-health-in-2019).

2. CDC, How to Address COVID-19 Vaccine Misinformation. Centers for Disease Control and Prevention (2021), (available at https://www.cdc.gov/vaccines/covid-19/health-departments/addressing-vaccine-misinformation.html).

3. M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, A. Scala, The COVID-19 social media infodemic. Sci Rep. 10, 16598 (2020).
4. S. Loomba, A. de Figueiredo, S. J. Piatek, K. de Graaf, H. J. Larson, Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat Hum Behav* (2021), doi:10.1038/s41562-021-01056-1.

5. K. Kricorian, R. Civen, O. Equils, COVID-19 vaccine hesitancy: misinformation and perceptions of vaccine safety. *Human Vaccines & Immunotherapeutics*. 0, 1–8 (2021).

6. V. Carrieri, L. Madio, F. Principe, Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health Economics*. 28, 1377–1382 (2019).

7. CDC, COVID Data Tracker. *Centers for Disease Control and Prevention* (2020), (available at https://covid.cdc.gov/covid-data-tracker).

8. J. Guynn, Facebook apologizes to black activist who was censored for calling out racism. *USA TODAY* (2017), (available at https://www.usatoday.com/story/tech/2017/08/03/facebook-ijeoma-oluoh-hate-speech/537682001/).

9. J. Angwin, H. Grassegger, “Facebook’s Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children” (ProPublica, 2017), (available at https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms).

10. Analysis | The Technology 202: Researchers say it’s easy to find drugs on Facebook, Instagram and YouTube. *Washington Post*, (available at https://www.washingtonpost.com/politics/2021/03/15/technology-202-researchers-say-it-easy-find-drugs-facebook-instagram-youtube/).

11. Instagram has a drug problem. Its algorithms make it worse. *Washington Post*, (available at https://www.washingtonpost.com/business/economy/instagram-has-a-drug-problem-its-algorithms-make-it-worse/2018/09/25/c45bf730-bdbf-11e8-b7d2-0773aa1e33da_story.html).

12. The Facebook Files. *Wall Street Journal* (2021), (available at https://www.wsj.com/articles/the-facebook-files-11631713039).

13. E. Goldman, “Content Moderation Remedies” (SSRN Scholarly Paper ID 3810580, Social Science Research Network, Rochester, NY, 2021), doi:10.2139/ssrn.3810580.

14. J. Grimmelmann, The Virtues of Moderation. *Yale J.L. & Tech.* 17, 42–109 (2015).

15. A. M. Jamison, D. A. Broniatowski, M. Dredze, Z. Wood-Doughty, D. Khan, S. C. Quinn, Vaccine-related advertising in the Facebook Ad Archive. *Vaccine*. 38, 512–520 (2020).

16. Y. Halon, Zuckerberg knocks Twitter for fact-checking Trump, says private companies shouldn’t be “the arbiter of truth.” *Fox News* (2020), (available at https://www.foxnews.com/media/facebook-mark-zuckerberg-twitter-fact-checking-trump).
17. A. M. Jamison, D. A. Broniatowski, M. Dredze, A. Sangraula, M. C. Smith, S. C. Quinn, Not just conspiracy theories: Vaccine opponents and proponents add to the COVID-19 ‘infodemic’ on Twitter. HKS Misinfo Review (2020), doi:10.37016/mr-2020-38.

18. Keeping People Safe and Informed About the Coronavirus. About Facebook (2020), (available at https://about.fb.com/news/2020/12/coronavirus/).

19. Reaching Billions of People With COVID-19 Vaccine Information. About Facebook (2021), (available at https://about.fb.com/news/2021/02/reaching-billions-of-people-with-covid-19-vaccine-information/).

20. The Disinformation Dozen | Center for Countering Digital Hate. CCDH, (available at https://www.counterhate.com/disinformationdozen).

21. How We’re Taking Action Against Vaccine Misinformation Superspreaders. About Facebook (2021), (available at https://about.fb.com/news/2021/08/taking-action-against-vaccine-misinformation-superspreaders/).

22. N. F. Johnson, N. Velásquez, N. J. Restrepo, R. Leahy, N. Gabriel, S. El Oud, M. Zheng, P. Manrique, S. Wuchty, Y. Lupu, The online competition between pro- and anti-vaccination views. Nature. 582, 230–233 (2020).

23. D. A. Broniatowski, A. M. Jamison, N. F. Johnson, N. Velasquez, R. Leahy, N. J. Restrepo, M. Dredze, S. C. Quinn, Facebook Pages, the “Disneyland” Measles Outbreak, and Promotion of Vaccine Refusal as a Civil Right, 2009–2019. Am J Public Health. 110, S312–S318 (2020).

24. J.-P. Allem, E. Ferrara, S. P. Uppu, T. B. Cruz, J. B. Unger, E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR public health and surveillance. 3, e98 (2017).

25. J. Uyheng, K. M. Carley, Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. J Comput Soc Sc. 3, 445–468 (2020).

26. V. S. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, The DARPA Twitter bot challenge. Computer. 49, 38–46 (2016).

27. Inauthentic Behavior | Transparency Center, (available at https://transparency.fb.com/policies/community-standards/inauthentic-behavior/).

28. D. A. Broniatowski, A. M. Jamison, S. Qi, L. AlKulaib, T. Chen, A. Benton, S. C. Quinn, M. Dredze, Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. American Journal of Public Health. 108, 1378–1384 (2018).

29. J. W. Ayers, B. Chu, Z. Zhu, E. C. Leas, D. M. Smith, M. Dredze, D. A. Broniatowski, Spread of Misinformation About Face Masks and COVID-19 by Automated Software on Facebook. JAMA Intern Med (2021), doi:10.1001/jamainternmed.2021.2498.
30. CrowdTangle Team, *CrowdTangle* (Menlo Park, California, United States, 2021; List IDs: 1475046, 1475047, 1584315, 1584316).

31. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 U.S. presidential election. *Science*. 363, 374–378 (2019).

32. L. Singh, L. Bode, C. Budak, K. Kawintiranon, C. Padden, E. Vraga, Understanding high- and low-quality URL Sharing on COVID-19 Twitter streams. *Journal of Computational Social Science*, 1–24 (2020).

33. C. Shao, G. L. Ciampaglia, A. Flammini, F. Menczer, in Proceedings of the 25th International Conference Companion on World Wide Web (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2016; https://doi.org/10.1145/2872518.2890098), *WWW ’16 Companion*, pp. 745–750.

34. A. Bovet, H. A. Makse, Influence of fake news in Twitter during the 2016 US presidential election. *Nat Commun*. 10, 7 (2019).

35. F. Giglietto, N. Righetti, L. Rossi, G. Marino, It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society*. 23, 867–891 (2020).

36. D. A. Broniatowski, “Towards Statistical Foundations For Detecting Coordinated Inauthentic Behavior On Facebook” (IDDP Report, The George Washington University Institute for Data, Democracy, and Politics, Washington, D. C., 2021), (available at https://iddp.gwu.edu/sites/g/files/zaxdzs3576/f/downloads/Coordinated%20Behavior%20On%20Facebook%20Broniatowski.pdf).

37. A. Kata, A postmodern Pandora’s box: anti-vaccination misinformation on the Internet. *Vaccine*. 28, 1709–1716 (2010).

38. A. Jamison, D. A. Broniatowski, M. C. Smith, K. S. Parikh, A. Malik, M. Dredze, S. C. Quinn, Adapting and Extending a Typology to Identify Vaccine Misinformation on Twitter. *Am J Public Health*. 110, S331–S339 (2020).

39. T. M. J. Fruchterman, E. M. Reingold, Graph drawing by force-directed placement. *Software: Practice and Experience*. 21, 1129–1164 (1991).

40. J. L. Bernal, S. Cummins, A. Gasparri, Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 46, 348–355 (2017).

41. J. Kurkowski, *john-kurkowski/tldextract* (2019; https://github.com/john-kurkowski/tldextract).

42. M. Nagode, B. Panic, J. Klemenc, S. Oman, *rebmix: Finite Mixture Modeling, Clustering & Classification* (2021; https://CRAN.R-project.org/package=rebmix).
**Funding:** The John S. and James L. Knight Foundation through the GW Institute for Data, Democracy, and Politics (DAB, LCA)

**Author contributions:**

- Conceptualization: DAB, JG, LCA
- Methodology: DAB, JG, AMJ, LCA
- Investigation: DAB, JG, AMJ
- Visualization: DAB
- Funding acquisition: DAB, LCA
- Project administration: DAB
- Supervision: DAB
- Writing – original draft: DAB
- Writing – review & editing: DAB, JG, AMJ, LCA

**Competing interests:** DAB has received consulting fees from Merck & Co. for participating in the 2021 Merck Global Vaccine Confidence Expert Input Forum, and has received a speaking honorarium from the United Nations Shot@Life Foundation. JG, AMJ, and LCA declare no competing interests.

**Data and materials availability:** Data in this study were obtained from CrowdTangle for Academics and Researchers, a third-party data provider owned and operated by Facebook. CrowdTangle list IDs are provided in the references. Anyone with a CrowdTangle account may access these lists and the corresponding raw data. Researchers may request CrowdTangle access at [https://help.crowdtangle.com/en/articles/4302208-crowdtangle-for-academics-and-researchers](https://help.crowdtangle.com/en/articles/4302208-crowdtangle-for-academics-and-researchers). CrowdTangle’s terms of service prohibit providing raw data to anyone outside of a CrowdTangle user’s account. The user can share the findings, but not the data. If a journal asks for data to verify findings, the CrowdTangle user may send a .csv, but it cannot be posted publicly, and the journal must delete it after verification. CSV files containing the raw data used in this study are therefore available upon request subject to these terms.
Supplementary Materials

Materials and Methods
We obtained all our data from CrowdTangle, a Facebook-owned tool that tracks interactions on public content from Facebook pages and groups, verified profiles, Instagram accounts, and subreddits. It does not include paid ads unless those ads began as organic, non-paid posts that were subsequently “boosted” using Facebook’s advertising tools. It also does not include activity on private accounts, or posts made visible only to specific groups of followers.

Procedure for identifying vaccine-related pages and groups in the original sample
We followed a three-step procedure to identify vaccine-related pages and groups.

1) We first identified a large set of pages and groups that mentioned vaccines at least once. To do so, we searched CrowdTangle on November 15, 2020, identifying and downloaded posts containing at least one keyword from the following list: “vaccine, vaxx, vaccines, vaccination, vaccinated, vax, vaxxed, vaccinations, jab”. Several of these posts contained content pertaining to guns, and to pet and other animal vaccines. Thus, we ran a second search excluding posts containing the following keywords: “gun, dog, cat, foster, adopt, shelter, vet, kennel, pet, chicken, livestock, kitten, puppy, paw, cow”. We performed separate searches for posts from pages and groups. At the time these searches were conducted, we retrieved the 299,981 most recent page posts and 299,904 group posts meeting search criteria before hitting CrowdTangle’s download limit. The earliest post using this procedure was timestamped September 7, 2020, for pages, and July 1, 2020, for groups.

2) We next narrowed down our initial list of pages and groups to those that routinely mentioned vaccination, as follows: Given all pages and groups that had posted at least once about vaccines, we next retrieved available posts from these assets (“2020 initial vaccine assets”). As before, we retrieved as many posts as possible before hitting CrowdTangle’s download limit: the 299,994 most recent page posts and 299,969 group posts meeting search criteria. The earliest post from these 2020 initial vaccine assets was timestamped November 8, 2020, for pages, and November 13, 2020, for groups. We identified the subset of all 2020 initial vaccine assets for which at least 20% of posts retrieved contained “vacc” or “vax”, constituting the “2020 final vaccine assets”. We selected this 20% threshold by inspection and results were insensitive to this threshold (relaxing the threshold yielded more groups and pages characterized as “other”, which were not included in our analysis, and all groups and pages were checked for relevance by two authors – AMJ and JG). On November 15, 2020, we retrieved all posts from these 2020 final vaccine assets for a 12-month period starting on November 15, 2019 forward. We did not exceed the download limit for the 2020 final vaccine assets. The original sample constituted all 2020 final vaccine assets that were labeled as pro- or anti-vaccine. On November 15, 2020, we collected all posts in the original sample from November 15, 2019 to November 15, 2020. Additionally, on August 12, 2021, we collected available posts from February 8, 2021 through August 8, 2021.

3) We manually annotated the 2020 final vaccine assets as either pro-vaccine, anti-vaccine, or “other” (see Table S1). Specifically, two independent annotators (JG and AMJ) manually assessed each group and page following a two-tiered coding scheme. Entities were first labeled as either pro-vaccine, anti-vaccine, or other (Cohen’s $\kappa = 0.88$, 95% CI: 0.85-0.92). Next, a sub-category was assigned based on specific concerns, adapted from a
previously published coding scheme (38) (Cohen’s κ = 0.75, 95% CI: 0.71-0.79; Table S1). Categorization was based on content shared in the “about” section of each entity. When this section was left blank, annotators considered the entity’s title, any imagery used, and recent posts to decide. All entities were double-coded, with disagreements discursively reconciled. We retained all assets that were labeled as pro-vaccine or anti-vaccine.

Procedure for identifying vaccine-related pages and groups in the final sample
Since new pages and groups might have formed after November 15, 2020, we repeated this procedure on July 28, 2021. Using the same vaccine keywords (and excluding posts about pets and guns), we retrieved the 100,000 most recent page posts and 100,000 group posts meeting search criteria before hitting CrowdTangle’s download limit (this limit had been revised downward by CrowdTangle from ~300,000 posts to exactly 100,000 posts). The earliest post using this procedure was timestamped July 24, 2021, for pages, and July 17, 2021, for groups. We next retrieved posts from pages and groups that had posted at least once in our search. As before, we retrieved as many posts as possible before hitting CrowdTangle’s download limit: the 100,000 most recent page posts and 100,000 group posts meeting search criteria (“2021 initial vaccine assets”). The earliest post from these 2021 initial vaccine assets was timestamped July 26, 2021, for both pages and groups. Finally, as above, we identified the subset of all these assets for which at least 20% of posts retrieved contained “vacc” or “vax” (“2021 final vaccine assets”) and again annotated any assets that did not appear in the 2020 final vaccine assets as either pro-vaccine, anti-vaccine, or “other” (see Table S1). Additionally, we identified a new category of groups that we labeled as “Unclear” because they primarily contained discussion about side effects from COVID-19 vaccines and were neither explicitly pro- nor anti-vaccines. We next downloaded all available posts from these assets. In so doing, we hit the download limit for the 2021 final vaccine groups. CrowdTangle allows multiple historical queries, and each query is restricted in size to the 100,000 most recent posts. Thus, we conducted several queries with overlapping date ranges to ensure that we collected all available posts in the 2021 final vaccine assets from February 8, 2021, to August 8, 2021. Since some dates overlapped, we removed duplicate posts (i.e., posts in the same asset with the exact same content and timestamp) during these dates.

Calculating post volume and engagements
For each dataset, we calculated the weekly number of posts in anti- and pro-vaccine pages and groups. We also calculated the weekly number of engagements with these posts as the sum of all Likes, Shares, Comments, and other emotional engagements (Love, Wow, Haha, Sad, Angry and Care) reported by CrowdTangle. We applied a logarithmic transform to post volumes and engagements to control for data skew.

Interrupted Time Series Analysis
We used a quasi-experimental interrupted time series regression design (40) to analyze the effects of Facebook’s policies. The specific regression model that we used for count data (posts and engagements) was

$$\log (Y_t) = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 TX_t$$
where $\beta_0$ is the log-transformed number of posts or engagements at time $T = 0$ (November 15, 2019) with time units measured in weeks, $\beta_1$ is the weekly change in log-transformed post volume or engagements, (representing the underlying pre-intervention trend), $\beta_2$ is a step change following the intervention (where $X_t$ is a dummy variable that is 0 before February 8, 2021 and 1 otherwise) and $\beta_3$ indicates a change in the slope change following the intervention (using the interaction between time and intervention: $TX_t$ ). Similarly, for proportion data, the regression model that we used was

$$
\text{logit} \left( Y_t \right) = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 TX_t
$$

where

$$
\text{logit} \left( p \right) = \log \left( \frac{p}{1 - p} \right)
$$

**Identifying off-platform link credibility**

For each dataset, we extracted all URLs in the “Link” field returned by CrowdTangle. On occasion, these links used URL shortening services (e.g., bit.ly). Thus, if the “Final Link” field was non-empty, we used the “Final Link” field instead of the “Link” field. Next, using the TLDExtract Python module (41), we extracted the top-level domain (TLD) and suffix for each URL (for example, the top-level domain of www.example.com/this_is_an_example.html is example.com).

**Low-Credibility TLDs:** After removing links pointing to other Facebook content (i.e., with the facebook.com TLD), we calculated the weekly proportion of all remaining TLDs in anti-vaccine pages and groups listed on iffy.news (https://iffy.news/index/) on August 18, 2021. We conducted interrupted time series analyses on these weekly proportions after applying a logit transform to these proportions to control for floor and ceiling effects.

**High-Credibility TLDs:** All TLDs ending in .gov, .gc.ca, .mil, .nhs.uk, starting with gov., mygov., government., containing .govt. or .gov., or matching who.int, paho.org, un.org, canada.ca, ontario.ca, toronto.ca, or alberta.ca were coded as “government”. Similarly, all TLDs ending in .edu, containing .edu., .ac., thelancet.com, sciencedirect.com., medrxiv.org, pnas.org, apa.org, nature.com, sciencemag.org, nejm.org, bmj.com, mayoclinic.org, aaas.org, healthdata.org, researchgate.net, or rand.org were coded as “academic”. Similar to (32), we considered government and academic TLDs to be “high credibility”. After removing links pointing to other Facebook content (i.e., with the facebook.com TLD), we calculated the weekly proportion of all remaining government and academic TLDs. We conducted interrupted time series analyses on these weekly proportions after applying a logit transform to these proportions to control for floor and ceiling effects.

**Identifying in-platform links**

Each Facebook page and group has a unique numerical Facebook ID. In addition, Facebook pages have unique usernames. Using CrowdTangle, we identified and extracted the Facebook ID and username for each pro- and anti-vaccine page in each dataset. Next, we identified all links in each dataset pointing to content on these pages -- i.e., links beginning with www.facebook.com/<Facebook ID>/ or www.facebook.com/<username>/
For each dataset, we next calculated the weekly proportion of all links from anti-vaccine pages to other anti-vaccine pages within the same dataset, conducting the same analysis for links from pro-vaccine pages to other pro-vaccine pages. We conducted a third set of interrupted time series analyses on these weekly proportions after applying a logit transform to these proportions to control for floor and ceiling effects.

**Constructing page networks**

We constructed networks from anti-vaccine and pro-vaccine Facebook fan pages both before and after Facebook’s policies were implemented. The pre-policy period was twice as long as the post-policy period. Thus, to ensure that pre- and post-policy networks could be compared, we compared the last six months of the pre-policy period to the entirety of the six-month-long post-policy period. Specifically, a pair of pages were linked in this network if there was at least one URL pointing from one page to another in the 181-day-long (six-month) time period from May 18, 2020 through November 15, 2020 (before the removal of SMV) or the 181-day-long time period from February 8, 2021 through August 8, 2021 (after the policy).

**Identifying “near simultaneous” links**

Facebook's community standards disallow “coordinated inauthentic behavior”, stating:

*In line with our commitment to authenticity, we don't allow people to misrepresent themselves on Facebook, use fake accounts, artificially boost the popularity of content, or engage in behaviors designed to enable other violations under our Community Standards. This policy is intended to protect the security of user accounts and our services, and create a space where people can trust the people and communities they interact with.* ([https://www.facebook.com/communitystandards/inauthentic\_behavior](https://www.facebook.com/communitystandards/inauthentic\_behavior))

Nevertheless, prior work (35) suggests that this type of behavior is widespread on the platform. Building on this work, we use a technique to identify coordinated activity based on “near-simultaneous link sharing”. The intuition is that if two assets routinely share the same URL at roughly the same time, these two entities are coordinated.

We operationalized “near-simultaneous” link sharing in a manner that was robust to the specific query being used. Specifically, we conducted several different keyword queries on CrowdTangle (including the keywords used in this study) to identify posts pertaining to several different health topics (36). For each post containing the same URL, we calculated the interarrival time -- i.e., the time difference -- in seconds, between that each successive share of that URL. Treating each second as an opportunity for a given URL to be shared by an entity (group or page), we expect interarrival times to follow a Poisson distribution. We therefore used the REBMIX (42) package in R to fit a mixture of Poisson distributions to each query (the number six was selected by visual inspection).

Averaging across these queries, we constructed a mixture of two Poisson distributions with $\mu_1=14.15$, $\mu_2=39.94$, and respective weights of 59.25% and 40.75%. Per this mixture model, we define “near simultaneous” link sharing behavior as any pair of URLs that are shared within at most 25 seconds of one another. This number is comparable to the threshold defined by (35), but appears to be consistent across several search queries.
Identifying Coordinated Entities

We consider a pair of pages or groups to be coordinated if they have more “near simultaneous” links in common than would be expected if link interarrival times were drawn from a Poisson distribution with a mean of $\mu^2=39.94$ seconds (i.e., the slower of the two distributions defined above). According to this distribution, the cumulative probability of a link interarrival time that is less or equal to 25 seconds is extremely unlikely -- 0.78%. We therefore used binomial tests to test the hypothesis that the number of “near simultaneous” link pairs for a given pair of pages or groups was significantly higher than 0.78%. Edges between entities were retained if they were significant at the $p<0.05$ level after controlling for multiple comparisons using the Holm-Bonferroni procedure. Note that this threshold is conservative -- defining link pairs with average interarrival times as 39.94 seconds as “not simultaneous” still results in a very small likelihood of “near simultaneous” links that are less than 25 seconds.
### Table S1. Asset, post, and engagement counts for each sub-category of asset in each sample.

| Position       | Category                        | Pre-Policy Period | Original Sample Post-Policy Period | Final Sample Post-Policy Period |
|----------------|---------------------------------|-------------------|------------------------------------|---------------------------------|
| Anti-Vaccine   | Alternative medicine            | 6                 | 3                                 | 0                               |
|                | Civil liberties                 | 43                | 34                                | 14                              |
|                | Conspiracy                     | 24                | 15                                | 4                               |
|                | Conspiracy & Civil liberties    | 0                 | 0                                 | 0                               |
|                | Limited info                   | 0                 | 0                                 | 0                               |
|                | Morality issues                | 2                 | 2                                 | 1                               |
|                | Other Safety concerns          | 7                 | 7                                 | 0                               |
|                |                                 | 32                | 24                                | 14                              |
| Pro-Vaccine    | Anti anti-vaccine arguments    | 24                | 17                                | 6                               |
|                | Other                          | 4                 | 3                                 | 4                               |
|                | Pro science                    | 4                 | 3                                 | 3                               |
|                | Pro-policy                     | 3                 | 3                                 | 1                               |
|                | Promotion                      | 56                | 53                                | 25                              |
|                | Safe & effective               | 11                | 9                                | 7                               |
|                |                                 |                   |                                    |                                 |
| Unclear        | Side effects                   | 0                 | 0                                 | 0                               |
| Other          | Foreign language               | 3                 | 3                                 | 3                               |
| General public health                      | News sharing                  | Non-relevant                  | Other          | Pharmacy | Research | Unrelated |
|-------------------------------------------|-------------------------------|------------------------------|----------------|----------|----------|-----------|
| 10 7,631 4,314,424                       | 9 5,405 428,921               | 7 13,054 1,389,605           |                |          |          |
| 12 46,347 1,049,350                      | 11 26,223 1,296,455           | 12 41,658 3,117,059          |                |          |          |
| 1 4,771 169,829                          | 1 2,120 180,011               | 0 0 0                        |                |          |          |
| 23 9,341 246,847                         | 17 3,159 116,082              | 4 317 11,175                 |                |          |          |
| 1 1,847 2,838                            | 1 824 142                     | 0 0 0                        |                |          |          |
| 17 2,578 24,806                          | 16 907 25,433                 | 11 381 11,326                |                |          |          |
| 29 5,721 109,829                         | 22 1,441 29,587               | 14 4,915 8                   |                |          |          |

| Anti-Vaccine | Groups |
|--------------|--------|
| Alternative medicine | 1 733 1,448 |
| Civil liberties | 23 37,505 570,405 |
| Conspiracy & Civil liberties | 40 84,896 936,471 |
| Limited info | 1 2,086 19,851 |
| Morality issues | 1 686 1,193 |
| Other Safety concerns | 2 836 2,543 |
|                | 7 16,008 660,879 |
|                | 17 20,345 146,787 |

| Pro-Vaccine | Groups |
|-------------|--------|
| Anti anti-vaccine arguments | 5 2,154 107,125 |
| Other | 1 44 250 |
| Pro science | 0 0 0 |
| Pro-policy | 0 0 0 |

2
| Category                | Count | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 | Value 6 | Value 7 | Value 8 | Value 9 |
|-------------------------|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Promotion               | 1     | 1,314   | 32,967  | 1       | 229     | 5,269   | 41      | 173,391 | 6,070,310|
| Safe & effective        | 1     | 1,812   | 70,087  | 1       | 2,278   | 159,339 | 3       | 8,124   | 397,397 |
| Unclear                 |       | 0       | 0       | 0       | 0       | 0       | 7       | 2,581   | 91,982  |
| Other                   |       |         |         |         |         |         |         |         |         |
| Foreign language        | 17    | 53,038  | 1,290,575| 13      | 48,515  | 979,479 | 9       | 91,593  | 851,470 |
| General public health   | 2     | 1,287   | 8,585   | 2       | 1,071   | 1,415   | 0       | 0       | 0       |
| News sharing            | 4     | 15,428  | 122,203 | 3       | 1,820   | 725     | 1       | 4,471   | 186,399 |
| Non-relevant            | 0     | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       |
| Other                   | 5     | 1,613   | 11,134  | 2       | 64      | 221     | 0       | 0       | 0       |
| Pharmacy                | 0     | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       |
| Research                | 1     | 312     | 879     | 1       | 104     | 372     | 0       | 0       | 0       |
| Unrelated               | 11    | 24,127  | 590,506 | 10      | 8,865   | 264,276 | 5       | 6,737   | 97,663  |
Table S2. Regression tables for interrupted time series analyses of post and engagements counts. A logarithmic transform has been applied to control for data skew.

|                | Original Sample | Posts | Anti-Vaccine | Pro-Vaccine | Anti-Vaccine | Groups | Pro-Vaccine |
|----------------|-----------------|-------|--------------|-------------|--------------|--------|-------------|
|                |                 |       | β          | SE         | t            | β      | SE         | t            | β      | SE         | t            | β      | SE         | t            | β      | SE         | t            | β      | SE         | t            |
| β₀: Intercept at Nov. 15, 2019 | 7.41            | 0.03  | 216.46 *** | 6.35        | 0.03 196.17 *** | 7.59   | 0.06 129.28 *** | 3.86         | 0.09 42.93 *** |
| β₁: Weekly change | 0.00            | 0.00  | -0.11       | 0.00        | 0.00 2.26 * | 0.02   | 0.00 8.02 *** | 0.02         | 0.00 8.38 *** |
| β₂: Step change on Feb. 8, 2019 | -1.23           | 0.07  | -18.45 ***  | -0.38       | 0.06 -5.99 *** | -1.22  | 0.12 -10.63 *** | -1.16        | 0.18 -6.60 *** |
| β₃: Slope change on Feb. 8, 2019 | 0.01            | 0.00  | 3.02 **     | 0.00        | 0.00 0.90  | -0.03  | 0.01 -5.58 ** | 0.00         | 0.01 -0.30  |

|                | Original Sample | Engagements | Anti-Vaccine | Pro-Vaccine | Anti-Vaccine | Groups | Pro-Vaccine |
|----------------|-----------------|-------------|--------------|-------------|--------------|--------|-------------|
|                |                 | β          | SE         | t            | β          | SE         | t            | β          | SE         | t            | β          | SE         | t            | β          | SE         | t            |
| β₀: Intercept at Nov. 15, 2019 | 12.35           | 0.06       | 224.03 *** | 11.96       | 0.12 98.47 *** | 11.32   | 0.09 128.93 *** | 7.90        | 0.16 51.14 *** |
| β₁: Weekly change | -0.01           | 0.00       | -4.89 ***   | 0.01        | 0.00 1.91  | -0.03  | 0.00 -9.98 *** | 0.01        | 0.01 2.15 * |
| β₂: Step change on Feb. 8, 2019 | -1.46           | 0.11       | -13.55 ***  | -1.19       | 0.24 -5.00 *** | -0.58  | 0.17 -3.35 **  | -1.12       | 0.30 -3.70 *** |
| β₃: Slope change on Feb. 8, 2019 | 0.03            | 0.01       | 5.50 ***    | 0.02        | 0.01 1.96  | 0.01   | 0.01 1.12  | 0.04        | 0.02 2.68 ** |

Original Sample, Excluding Removed Assets

|                | Posts |
|----------------|-------|
|                | β      | SE         | t            | β      | SE         | t            | β      | SE         | t            |
| β₀: Intercept at Nov. 15, 2019 | 7.00  | 0.03       | 202.37 ***  | 6.35  | 0.03 182.91 *** | 3.86  | 0.09 34.77 *** |
| β₁: Weekly change | 0.00  | 0.00       | -0.11       | 0.00  | 0.00 2.26  | 0.02  | 0.00 8.02  *** | 0.02  | 0.00 8.38 *** |
| β₂: Step change on Feb. 8, 2019 | -1.23 | 0.07       | -18.45 ***  | -0.38 | 0.06 -5.99 *** | -1.22 | 0.12 -10.63 *** | -1.16 | 0.18 -6.60 *** |
| β₃: Slope change on Feb. 8, 2019 | 0.01  | 0.00       | 3.02 **     | 0.00  | 0.00 0.90  | -0.03 | 0.01 -5.58 ** | 0.00  | 0.01 -0.30  |
| $\beta_0$: Intercept at Nov. 15, 2019 | 6.72 | 0.10 | 66.17 | ** | 6.42 | 0.11 | 58.55 | ** | 7.47 | 0.16 | 46.104 | ** | 3.62 | 0.37 | 9.76 | ** |
| $\beta_1$: Weekly change | 0.00 | 0.00 | -0.77 | 0.00 | 0.00 | -0.50 | 0 | 0 | 0.886 | 0.03 | 0.01 | 2.80 | ** |
| $\beta_2$: Step change on Feb. 8, 2019 | -0.43 | 0.08 | -5.73 | ** | -0.20 | 0.08 | -2.50 | * | -0.34 | 0.12 | -2.795 | ** | -0.99 | 0.28 | -3.60 | ** |
| $\beta_3$: Slope change on Feb. 8, 2019 | 0.01 | 0.00 | 3.36 | ** | 0.01 | 0.00 | 1.72 | 0.01 | 0.01 | -3.585 | ** | 0.00 | 0.01 | -0.29 |

**Engagements**

| $\beta_0$: Intercept at Nov. 15, 2019 | 11.70 | 0.30 | 38.76 | ** | 11.39 | 0.52 | 22.14 | ** | 10.38 | 0.35 | 29.88 | ** | 7.47 | 0.61 | 12.32 | ** |
| $\beta_1$: Weekly change | -0.02 | 0.01 | -2.80 | ** | 0.02 | 0.01 | 1.91 | -0.02 | 0.01 | -2.77 | ** | 0.01 | 0.02 | 0.87 |
| $\beta_2$: Step change on Feb. 8, 2019 | -0.04 | 0.23 | -0.17 | -0.04 | 0.38 | -4.43 | ** | 0.06 | 0.26 | 0.22 | -0.83 | 0.45 | -1.84 |
| $\beta_3$: Slope change on Feb. 8, 2019 | 0.04 | 0.01 | 3.96 | ** | 0.01 | 0.02 | 0.36 | 0.00 | 0.01 | 0.40 | 0.04 | 0.02 | 1.81 |

**Final Sample**

| Posts | 7.41 | 0.04 | 194.56 | ** | 6.35 | 0.04 | 147.54 | ** | 7.59 | 0.06 | 129.96 | ** | 3.86 | 0.10 | 40.06 | ** |
| $\beta_1$: Weekly change | 0.00 | 0.00 | -0.10 | 0.00 | 0.00 | 1.70 | 0.02 | 0.00 | 8.06 | 0.02 | 0.00 | 7.82 | ** |
| $\beta_2$: Step change on Feb. 8, 2019 | -1.19 | 0.07 | -16.06 | ** | -0.11 | 0.08 | -1.25 | -1.14 | 0.11 | -10.01 | ** | 4.56 | 0.19 | 24.22 | ** |
| $\beta_3$: Slope change on Feb. 8, 2019 | 0.02 | 0.00 | 6.28 | ** | 0.04 | 0.00 | 9.87 | ** | 0.00 | 0.01 | 0.23 | -0.12 | 0.01 | -12.45 | ** |

**Engagements**
|              | β₀: Intercept at Nov. 15, 2019 | β₁: Weekly change | β₂: Step change on Feb. 8, 2019 | β₃: Slope change on Feb. 8, 2019 |
|--------------|-------------------------------|-------------------|-------------------------------|---------------------------------|
| β₀           | 12.35 0.06 206.92 ***         | -0.01 0.00 -4.52 *** | -1.88 0.12 -16.09 ***         | 0.07 0.01 11.35 ***            |
| β₁           | 11.96 0.12 98.97 ***          | 0.01 0.00 1.92    | -0.52 0.24 -2.19 *           | 0.02 0.01 2.08 *              |
| β₂           | 11.32 0.09 125.73 ***         | -0.03 0.00 -9.74 *** | -0.12 0.18 -0.66             | 0.06 0.01 6.61 ***            |
| β₃           | 7.90 0.11 70.17 ***           | 0.01 0.00 2.95 **  | 4.90 0.22 22.26 ***          | -0.09 0.01 -8.41 ***          |

Note. ***=p<0.001, **=p<0.01, *=p<0.05. β=regression coefficient. SE=Standard Error. t=Student’s t statistic.
Table S3. Regression tables for interrupted time series analyses of proportions of posts and engagements with off-platform URLs that contain links to low- and high-credibility sources. A logistic transform has been applied to proportion data to control for floor and ceiling effects.

|                      | Low Credibility Anti-Vaccine Posts | Low Credibility Anti-Vaccine Engagements | Pro-Vaccine Posts |
|----------------------|-----------------------------------|------------------------------------------|------------------|
|                      | Original Sample | Final Sample | Original Sample | Final Sample | Original Sample | Final Sample | Original Sample | Final Sample |
| $\beta_0$: Intercept at Nov. 15, 2019 | $\beta$ | SE | t | $\beta$ | SE | t | $\beta$ | SE | t | $\beta$ | SE | t | $\beta$ | SE | t | $\beta$ | SE | t |
|                      | -1.62 | 0.06 | -28.41 | *** | -1.80 | 0.06 | -29.20 | *** | -1.62 | 0.05 | -30.00 | *** | -1.8 | 0.06 | -29.53 | *** |
| $\beta_1$: Weekly change | 0.00 | 0.00 | -0.31 | 0.00 | 0.00 | -1.24 | 0.00 | 0.00 | -0.33 | -0.0 | 0.0 | -1.253 | |
| $\beta_2$: Step change on Feb. 8, 2019 | 0.17 | 0.11 | 1.54 | 0.44 | 0.12 | 3.76 | *** | 0.80 | 0.10 | 7.84 | *** | 0.38 | 0.11 | 3.306 | ** |
| $\beta_3$: Slope change on Feb. 8, 2019 | 0.01 | 0.01 | 1.83 | 0.02 | 0.01 | 4.14 | *** | -0.01 | 0.01 | -1.28 | 0.02 | 0.01 | 4.22 | *** |
| $\beta_0$: Intercept at Nov. 15, 2019 | -1.58 | 0.10 | -16.08 | *** | -1.59 | 0.10 | -15.81 | *** | -1.58 | 0.09 | -17.32 | *** | -1.59 | 0.11 | -14.85 | *** |
| $\beta_1$: Weekly change | 0.00 | 0.00 | -0.13 | -0.01 | 0.00 | -3.14 | ** | 0.00 | 0.00 | -0.14 | -0.01 | 0.00 | -2.95 | ** | |
| $\beta_2$: Step change on Feb. 8, 2019 | 1.00 | 0.19 | 5.39 | *** | 0.69 | 0.19 | 3.64 | *** | -0.14 | 0.17 | -0.84 | 0.70 | 0.20 | 3.47 | ** |
| $\beta_3$: Slope change on Feb. 8, 2019 | -0.01 | 0.01 | -1.46 | 0.02 | 0.01 | 2.02 | * | 0.03 | 0.01 | 3.55 | ** | 0.01 | 0.01 | 1.34 | |
|                       | Anti-Vaccine Posts |                               | High Credibility Anti-Vaccine Posts |                               |                                 | Pro-Vaccine Posts |                               |
|-----------------------|--------------------|--------------------------------|-------------------------------------|-----------------------------------|------------------------------|-------------------|---------------------------------|
| **β₀**: Intercept at Nov. 15, 2019 | -3.15 0.11 -27.80 *** | -2.97 0.15 -19.65 *** | -3.69 0.23 -15.95 *** | -2.59 0.11 -24.8 *** | -2.59 0.09 -28.95 *** |
| **β₁**: Weekly change | 0.00 0.00 -0.39 | -0.02 0.01 -3.88 *** | 0.00 0.01 -0.59 | 0.002 0.0 0.542 | 0.00 0.00 0.63 |
| **β₂**: Step change on Feb. 8, 2019 | -0.36 0.21 -1.67 | 0.30 0.28 1.04 | 0.00 0.01 | 0.00 0.02 3.01 ** | 0.00 0.02 3.01 ** |
| **β₃**: Slope change on Feb. 8, 2019 | 0.00 0.01 -0.32 | 0.01 0.01 1.02 | 0.00 0.01 | 0.00 0.02 3.01 ** | 0.00 0.02 3.01 ** |
| \( \beta_0 \): Intercept at Nov. 15, 2019 | -4.10 | 0.36 | -11.28 | *** | . | . | . |
| \( \beta_1 \): Weekly change | . | . | . | . | 0.00 | 0.01 | -0.36 | . | . | . |
| \( \beta_2 \): Step change on Feb. 8, 2019 | -0.74 | 0.68 | -1.08 | . | . | . | . |
| \( \beta_3 \): Slope change on Feb. 8, 2019 | . | . | . | . | 0.03 | 0.03 | 0.93 | . | . | . |

**Note.** ***=p<0.001, **=p<0.01, *=p<0.05. \( \beta \)=regression coefficient. SE=Standard Error. t=Student’s t statistic, . = no regression coefficients could be calculated because at least one week in the sample had zero links.
Table S4. Regression tables for interrupted time series analyses of proportions of posts and engagements with in-platform URLs pointing from one anti-vaccine (pro-vaccine) page to another anti-vaccine (pro-vaccine) page in our sample. A logistic transform has been applied to proportion data to control for floor and ceiling effects.

|                     | Original Sample | Posts | Final Sample |                      |                      |
|---------------------|-----------------|-------|--------------|----------------------|----------------------|
|                     | Anti-Vaccine    |       |              |                      |                      |
| β₀: Intercept at Nov. 15, 2019 | -1.30 0.05 | -26.80 *** | -0.16 0.04 | -3.84 *** | -1.30 0.06 | -21.51 *** | -0.16 0.05 | -3.20 ** |
| β₁: Weekly change   | 0.01 0.00 | 4.15 *** | 0.00 0.00 | -0.07  | 0.01 0.00 | 3.34 **  | 0.00 0.00 | -0.06  |
| β₂: Step change on Feb. 8, 2019 | 0.05 0.09 | 0.54  | -0.31 0.08 | -3.95 *** | -0.19 0.11 | -1.67  | -0.41 0.09 | -4.46 ***|
| β₃: Slope change on Feb. 8, 2019 | -0.04 0.01 | -7.93 *** | 0.01 0.00 | 2.64 * | -0.03 0.01 | -5.71 *** | 0.01 0.01 | 1.41  |
|                     | Pro-Vaccine     |       |              |                      |                      |
| β₀: Intercept at Nov. 15, 2019 | -0.51 0.10 | -5.16 *** | 0.95 0.15 | 6.15 *** | -0.51 0.11 | -4.67 *** | 0.95 0.16 | 6.13 ***|
| β₁: Weekly change   | 0.02 0.00 | 6.86 *** | 0.00 0.01 | 0.47  | 0.02 0.00 | 6.20 *** | 0.00 0.01 | 0.47  |
| β₂: Step change on Feb. 8, 2019 | 0.07 0.18 | 0.41  | -0.73 0.29 | -2.53 * | 0.00 0.20 | 0.01  | -0.74 0.29 | 2.54 * |
| β₃: Slope change on Feb. 8, 2019 | -0.06 0.01 | -6.15 *** | 0.02 0.01 | 1.41  | -0.05 0.01 | -5.19 *** | 0.02 0.02 | 1.42  |

Engagements

|                     | Anti-Vaccine    |       |              |                      |                      |
| β₀: Intercept at Nov. 15, 2019 | -0.51 0.10 | -5.16 *** | 0.95 0.15 | 6.15 *** | -0.51 0.11 | -4.67 *** | 0.95 0.16 | 6.13 ***|
| β₁: Weekly change   | 0.02 0.00 | 6.86 *** | 0.00 0.01 | 0.47  | 0.02 0.00 | 6.20 *** | 0.00 0.01 | 0.47  |
| β₂: Step change on Feb. 8, 2019 | 0.07 0.18 | 0.41  | -0.73 0.29 | -2.53 * | 0.00 0.20 | 0.01  | -0.74 0.29 | 2.54 * |
| β₃: Slope change on Feb. 8, 2019 | -0.06 0.01 | -6.15 *** | 0.02 0.01 | 1.41  | -0.05 0.01 | -5.19 *** | 0.02 0.02 | 1.42  |

*Note.* ***=*p<0.001, **=*p<0.01, *=p<0.05. β=regression coefficient. SE=Standard Error. t=Student’s t statistic.
Table S5. Regression tables for interrupted time series analyses of proportions of successive URL pairs shared within 25 seconds of one another (i.e., “near simultaneous” link sharing). Counts for pages and groups have been combined. A logistic transform has been applied to proportion data to control for floor and ceiling effects.

| Coordinated Link Proportions | Original Sample | Final Sample |
|------------------------------|----------------|--------------|
|                               | Anti-Vaccine   | Pro-Vaccine  | Anti-Vaccine | Pro-Vaccine |
| $\beta_0$: Intercept at Nov. 15, 2019 | -1.86 0.14 | -13.18*** | -3.47 0.22 | -15.47*** | -1.86 0.13 | -13.88*** | -3.47 0.21 | -16.74*** |
| $\beta_1$: Weekly change      | 0.00 0.01 | 0.68 | 0.02 0.01 | 2.27* | 0.00 0.00 | 0.71 | 0.02 0.01 | 2.45* |
| $\beta_2$: Step change on Feb. 8, 2019 | -0.99 0.28 | -3.60** | -0.01 0.43 | -0.02 | -1.85 0.27 | -6.79*** | -0.79 0.38 | -2.08* |
| $\beta_3$: Slope change on Feb. 8, 2019 | -0.04 0.01 | -3.17** | -0.02 0.02 | -0.97 | 0.02 0.01 | 1.43 | 0.00 0.02 | -0.24 |

Note. ***=p<0.001, **=p<0.01, *=p<0.05. $\beta$=regression coefficient. SE=Standard Error. t=Student’s t statistic.
