Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter

On July 27, 2020, former US President Donald Trump retweeted to his then 84 million Twitter followers an online video, published by the Breitbart website, promoting misinformation about the anti-malarial drug, hydroxychloroquine.[1] The video, featuring Dr Stella Immanuel and a group called America's Frontline Doctors, falsely claimed that the combination of hydroxychloroquine, zinc, and azithromycin (an antibiotic) could cure COVID-19.[1] Twitter subsequently removed President Trump's retweet, temporarily suspended and removed tweets from his son, Donald Trump Jr, and removed and suspended other accounts sharing this false information as a violation of its COVID-19 misinformation policy.[1]

Concerns about hydroxychloroquine's safety and efficacy are well documented, including its association with cardiac adverse events, blood and lymph system disorders, liver problems, and kidney injury.[2] A study done in 34 US hospitals found that adults who were hospitalised with COVID-19 respiratory illness who were treated with hydroxychloroquine did not have improved clinical outcomes.[3] Safety concerns led to discontinuation of clinical trials by the US National Institutes of Health and WHO, but has not stopped the flow of misinformation and conspiracies espoused on social media platforms.

President Trump's misinformation tweet was not the first time he promoted hydroxychloroquine despite its unproven efficacy; on March 21, 2020, Trump tweeted that the drug could be a "game changer", on April 4 and April 15, he promoted its use in public events, and on May 18, he claimed that he was taking the drug along with a zinc supplement for prevention of COVID-19. In the same time period, studies found an increase in online search queries for purchasing hydroxychloroquine, a surge in hydroxychloroquine and chloroquine prescription fills and stockpiling, and evidence of illegal online availability of hydroxychloroquine from sources such as the dark web.[4–6]

To better understand the specific characteristics of hydroxychloroquine misinformation, we used big data approaches to analyse Twitter user conversations before and after President Trump's misinformation retweet on July 27. We chose this event because it was the first time Twitter removed a tweet from the President for being associated with COVID-19 misinformation.

Using the Python programming language, we first collected Twitter posts from the public streaming Twitter application programming interface from July 21–30, 2020, using "hydroxychloroquine" and "chloroquine" as keyword filters. We chose these dates as they directly preceded (hereafter referred to as the pre-period) and occurred after (hereafter referred to as post-period) President Trump's misinformation retweet from July 27. Because of the large volume of data collected, we used an unsupervised machine learning approach, called the biterm topic model (BTM), to extract themes from text of tweets as used in a previous COVID-19 study.[6]

The BTM grouped tweet texts that contained the same word related themes into clusters; the main themes of those clusters were an aggregation of topics of the text, which were then split into a bag of words in which a discrete probability distribution for all words for each theme was generated on the basis of a coherence score to determine an optimal value of clusters (hereafter referred to as k; appendix pp 2–5). Thematic groups outputted by the BTM were then reviewed for relevance to misinformation topics and tweets that were correlated to these clusters. Next, tweets correlated to clusters were retrieved and manually annotated.

See Online for appendix

Manual annotation was done to identify thematic misinformation topics on the basis of the unit of analysis, which was the text of each tweet reviewed, with categories adapted from previous studies on COVID-19 misinformation (appendix pp 8–10).[7,8] We also extracted the top 100 retweeted tweets in the pre-period and post-period to evaluate the types of misinformation that were highly propagated. Embedded hyperlinks to videos or external websites were reviewed to confirm that they matched the misinformation theme originally classified for the tweet text. Authors coded posts independently and achieved high intercoder reliability

	Percentage of coded dataset (%)*	Pre-period (%)*	Post-period (%)*
Rumour (treatment)	400 486 (32·4%)	109 012 (79·6%)	291 474 (32·1%)
Prevention and treatment (clinical treatment)	16 365 (1·3%)	3068 (2·2%)	13 297 (1·5%)
Prevention and treatment (folk medicine)	12 (<0·1%)	2 (<0·1%)	10 (<0·1%)
Authority action and policy (general)	97 995 (7·9%)	5355 (3·9%)	92 640 (10·2%)
Authority action and policy related to media management	236 491 (19·1%)	9308 (6·8%)	227 183 (25·0%)
Authority action and policy related to medical measures	167 616 (13·5%)	4691 (3·4%)	162 925 (18·0%)
Crisis situation updates	..	..	..
Conspiracy	125 830 (10·2%)	5527 (4·0%)	120 303 (13·3%)
Stigma	..	..	..

*Number and percentage of tweets with signal (ie, thematic misinformation topics) for misinformation.

Table 1: **Categories of hydroxychloroquine misinformation detected between July 21–30, 2020**

(κ=0·92) and met and conferred on correct classification for discrepancies by reconsulting misinformation categories.

A total of 2 771 730 tweets were collected between July 21–30. We observed a sharp increase in the volume of hydroxychloroquine-related tweets immediately after President Trump's misinformation retweet, resulting in 2 523 766 tweets (91·1% of the entire dataset) collected in the post-period (appendix p 1). After using the BTM and extracting the top 100 retweeted tweets we were able to review 174 600 tweets and retweets from the pre-period and 1 063 399 tweets and retweets from the post-period, constituting 44·7% (n=1 237 999) of the entire dataset.

After analysing the content, 1 044 794 (84·4%) of the 1 237 999 reviewed tweets were identified as including hydroxychloroquine misinformation topics. Overall, there was a lower proportion of misinformation tweets observed in the pre-period (136 963 [78·4%] of 174 600 tweets) versus the post-period (907 831 [85·4%] of 1 063 339 tweets). In total, these misinformation tweets and retweets were liked (via a heart symbol) 2 193 021 times by 283 055 unique Twitter users, indicating social sharing and support of misinformation across hundreds of thousands of Twitter users. A metadata analysis of the twitter users who had posted misinformation found that users had an average of 2294 followers, were following an average of 1644 users, and had an average account creation date of 5·5 years, which is more indicative of non-bot accounts.

Hydroxychloroquine misinformation were grouped into four categories: (1) rumours (a sub-category of treatment), (2) prevention and treatment (subcategories of clinical treatment and folk medicine (ie, the belief that alternative medicines can help prevent and treat COVID-19), (3) authority action and policy (subcategories of medical measures, media management, and general), and (4) conspiracy (table). Common themes included users reposting the Breibart video with supportive comments and demanding access to hydroxychloroquine, such as circulating an online petition to make it over-the-counter (eg, an example of a paraphrased tweet: "Make #hydroxychloroquine OTC! It can prevent and is cheap!"). Other common themes included users making claims of undue censorship of positive hydroxychloroquine information by the media and government officials, unfounded conspiracies (eg, #FauciTheFraud, #HCQWorksFacuiKnewIn2005, and #BillGatesBioTerrorist, and a tweet claiming association with the conspiracy theory group QAnon), and promotion of personal use of hydroxychloroquine or being cured by hydroxychloroquine from both first-hand accounts and second-hand accounts (eg, an example of a paraphrased tweet: "I am with dozens of physicians not social distancing or wearing a mask because everyone is taking hydroxychloroquine for prevention, including me!"). Tweets also included links to misinformation hosted on other social media platforms, including YouTube, Facebook, Instagram, and LinkedIn. We also observed Twitter users interacting with misinformation sources inquiring about how to get access to, or buy, hydroxychloroquine.

By contrast, only 56 879 (5·4%; from 49 082 unique users) of 1 044 794 tweets that were identified as including misinformation-related topics expressed concern about hydroxychloroquine misinformation or introduced information about the potential risks associated with the drug. Most of these tweets

questioned the credibility of President Trump and Dr Immanuel, highlighted possible hydroxychloroquine side-effects, or provided factual statements or links to information about the scarce evidence of the drug's efficacy.

To conclude, our study detected over 1 million tweets, including various hydroxychloroquine misinformation topics, following President Trump's promotion of a misinformation video that was subsequently removed by Twitter. Most (84·4%) of the 1 237 999 tweets reviewed contained misinformation and occurred after the President's retweet, while a much smaller volume (5·4%) expressed concern, questioned the credibility of misinformation sources, or provided information about the risks associated with hydroxychloroquine for treatment of COVID-19.

Although misinformation topics existed before the President's retweet, this event appears to have generated a large volume and diversity of misinformation topics with high levels of user interaction. As of December, 2020, different versions of the misinformation video that catalysed this event are still actively circulating on Twitter and other social media platforms, leading to continued support for the drug's unsubstantiated use and calls for increased access. Despite Twitter's removal and suspension of particular tweets and accounts, this misinformation went viral, with media outlets reporting multiple versions of the video that was shared and viewed by millions of users, including those associated with the anti-vaccination movement and banned conspiracy group, QAnon.[9] Although Twitter permanently suspended President Trump's account on January 8, 2021, due to concerns about him inciting violence, the misinformation he generated about COVID-19 will undoubtedly persist.

Other studies have analysed misinformation from other social media platforms, websites, TV, newspapers, Google searches, online purchasing behaviour, and smaller volumes of Twitter data to characterise different trends and risks associated with the COVID-19 infodemic (defined by WHO as an overabundance of information, including deliberate attempts to disseminate wrong information to undermine public health response).[7,8,10] Our study adds additional evidence to this growing number of studies by characterising how a single retweet from a powerful public figure can propagate various misinformation topics in an online community. However, this study has limitations (appendix pp 6–7), including that it was done on a single platform and did not specifically assess whether users were bots or commercial accounts.

Infodemic events raise questions about how social media platforms manage health-related content, particularly in the context of health emergencies. Although social media is now ubiquitous, consensus on whether it balances, promotes, or jeopardises public health remains a polarising issue. Recognising the acute threat of the infodemic and its potential to undermine hard fought gains against COVID-19, several international organizations (eg, WHO, the UN, UNICEF, UN Development Programme, UNAIDS, The United Nations Educational, Scientific and Cultural Organization, and the International Telecommunication Union) issued a joint statement calling on member states, platforms, researchers, and civil society to collaborate on strategies and tools to combat misinformation, while also recognising the need to respect freedom of expression.

To address the digital challenge of the COVID-19 pandemic, multistakeholder cooperation—coupled with a whole-of-society approach, including directly engaging with affected social media communities—is needed. Inclusive and effective infodemic management needs to be approached from a health communication, public policy, and pandemic preparedness standpoint to address COVID-19, as well as future public health emergencies that are likely to emerge in what is now a post-digital era.

TKM, JL, and MCN are employees of the startup company S-3 Research. S-3 Research is a startup funded and supported by the National Institutes of Health–National Institute on Drug Abuse through a small business innovation and research contract for opioid-related social media research and technology commercialisation. We report no other competing interests and have not been asked by any organisation to be named on or to submit this manuscript. All information collected from this study was from the public domain and the study did not involve any interaction with users. User indefinable information was removed from the study results. This study conforms to the Twitter Terms of Service. The raw Twitter data cannot be directly shared as it would infringe the Twitter Developer Terms. However, a public dataset of only the tweet IDs of the data we collected is available (https://github.com/Mathison/Twitter_hydrochloroquine_July21_July30) for anyone to download the tweets used for this study from Twitter using Twitter's application programming interface. The classification of posts as "misinformation" tweets is a matter of independent assessment and opinion, rather than fact, informed by other peer-reviewed articles that similarly classified misinformation COVID-19 topics.

Tim K Mackey, Vidya Purushothaman, Michael Haupt, Matthew C Nali, Jiawei Li
tmackey@ucsd.edu

Department of Anesthesiology and Division of Infectious Disease and Global Public Health (TKM, MCN), Department of Healthcare Research and Policy (TKM, JL), Department of Family Medicine and Public Health (VP), and Department of Cognitive Science (MH), University of California, San Diego, San Diego, CA 92037, USA; Global Health Policy Institute, San Diego, CA, USA (TKM, VP, MCN, JL); S-3 Research, San Diego, CA, USA (TKM, MCN, JL)

1 Morrison S, Heilweil R. How Trump and his son helped make a Covid-19 conspiracy theorist go viral in a matter of hours. 2020. https://www.vox.com/recode/2020/7/28/21345008/donald-trump-jr-twitter-suspension-stella-immanuel-hydroxychloroquine-misinformation (accessed Aug 15, 2020).

2 US Food and Drug Association. FDA cautions use of hydroxychloroquine/chloroquine for COVID-19 outside of the hospital setting or a clinical trial due to risk of heart rhythm problems. 2020. https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or (accessed Aug 24, 2020).

3 Self WH, Semler MW, Leither LM, et al. Effect of hydroxychloroquine on clinical status at 14 days in hospitalized patients with COVID-19: a randomized clinical trial. *JAMA* 2020; **324:** 2165–76.

4 Vaduganathan M, van Meijgaard J, Mehra MR, Joseph J, O'Donnell CJ, Warraich HJ. Prescription fill patterns for commonly used drugs during the COVID-19 pandemic in the United States. *JAMA* 2020; **323:** 2524–26.

5 Liu M, Caputi TL, Dredze M, Kesselheim AS, Ayers JW. Internet searches for unproven COVID-19 therapies in the United States. *JAMA Intern Med* 2020; **180:** 1116–18.

6 Mackey TK, Li J, Purushothaman V, et al. Big data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: infoveillance study on Twitter and Instagram. *JMIR Public Health Surveill* 2020; **6:** e20794.

7 Lu J. Themes and evolution of misinformation during the early phases of the COVID-19 outbreak in China—an application of the crisis and emergency risk communication model. *Front Commun (Lausanne)* 2020; **5:** e0118093.

8 Islam MS, Sarkar T, Khan SH, et al. COVID-19-related infodemic and its impact on public health: a global social media analysis. *Am J Trop Med Hyg* 2020; **103:** 1621–29.

9 Frenkel S, Alba D. Misleading hydroxychloroquine video, pushed by the trumps, spreads online. 2020. https://www.nytimes.com/2020/07/28/technology/virus-video-trump.html (accessed Aug 24, 2020).

10 Niburski K, Niburski O. Impact of Trump's promotion of unproven COVID-19 treatments and subsequent internet trends: observational study. *J Med Internet Res* 2020; **22:** e20044.