Original Research Article

Identifying how COVID-19-related misinformation reacts to the announcement of the UK national lockdown: An interrupted time-series study

Mark Green[1] , Elena Musi[2], Francisco Rowe[1],
Darren Charles[3], Frances Darlington Pollock[1] ,
Chris Kypridemos[4], Andrew Morse[1], Patricia Rossini[2],
John Tulloch[5,6], Andrew Davies[1], Emily Dearden[6,7],
Henrdramoorthy Maheswaran[8], Alex Singleton[1],
Roberto Vivancos[5,6] and Sally Sheard[4]

Abstract

COVID-19 is unique in that it is the first global pandemic occurring amidst a crowded information environment that has facilitated the proliferation of misinformation on social media. Dangerous misleading narratives have the potential to disrupt 'official' information sharing at major government announcements. Using an interrupted time-series design, we test the impact of the announcement of the first UK lockdown (8–8.30 p.m. 23 March 2020) on short-term trends of misinformation on Twitter. We utilise a novel dataset of all COVID-19-related social media posts on Twitter from the UK 48 hours before and 48 hours after the announcement (n = 2,531,888). We find that while the number of tweets increased immediately post announcement, there was no evidence of an increase in misinformation-related tweets. We found an increase in COVID-19-related bot activity post-announcement. Topic modelling of misinformation tweets revealed four distinct clusters: 'government and policy', 'symptoms', 'pushing back against misinformation' and 'cures and treatments'.

This article is a part of special theme on Studying the COVID-19 Infodemic at Scale. To see a full list of all articles in this special theme, please click here: https://journals.sagepub.com/page/bds/collections/studyinginfodemicatscale

[1]Department of Geography & Planning, University of Liverpool, Liverpool, UK
[2]Department of Communication & Media, University of Liverpool, Liverpool, UK
[3]Institute of Population Health Sciences, University of Liverpool, Liverpool, UK
[4]Department of Public Health & Policy, University of Liverpool, UK

[5]NIHR Heath Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, Liverpool, UK
[6]Public Health England, London, UK
[7]Greater Manchester Combined Authority, Manchester, UK
[8]Institute of Global Health Innovation, Imperial College London, London, UK

Corresponding author:
Mark Green, University of Liverpool, Liverpool L69 3BX, UK.
Email: mark.green@liverpool.ac.uk

Introduction

The 'infodemic' pandemic

Public health crises throughout human history have frequently seen the diffusion of misleading or inaccurate information (Mawdsley, 2020). The COVID-19 pandemic has introduced new dimensions to this old problem through being the first global pandemic occurring an information age which provides an environment of online platforms facilitating fast information sharing opportunities across large populations. Misinformation – defined here as any news, rumour, source or information that is misleading, inaccurate or not true and shared without clear intent to cause harm (where intent to harm was intended, this is often defined as disinformation) – has previously been identified as a global public health threat (Larson, 2018). Evidence has demonstrated that misinformation can spread both quicker and further along social networks than truthful information (Vosoughi et al., 2018). The potential for 'viral' misinformation, especially with few effective strategies to act against the constantly evolving nature of content, places considerable burdens on governments in dealing with the disease or may encourage individuals to act against their best interests (Pulido et al., 2020).

The close connectivity of humans through technical networks and a plethora of available information sources has enabled both official and unofficial COVID-19 information to thrive together in the same virtual environments, both competing for the attention of the audience. There are vast volumes of information and data available on the internet, most of which is unregulated (Swire-Thompson and Lazer, 2020). Medical and health information is particularly prevalent and often of low quality (Daraz et al., 2019). Social media networks contribute to these processes as users can share unvetted content due to the lack of a gatekeeping process, with organisations often slow to respond to controlling. With some population groups increasingly using the internet as their go-to source for health-related information (Daraz et al., 2019; Ofcom, 2020), closer scrutiny and analysis of how national government announcements are received online, both proactively and reactively, is therefore appropriate.

The uncertain, fast-moving, dynamic and unprecedented nature of the pandemic has increased the need for immediate information over the progress of COVID-19, its drivers and potential solutions (Pulido et al., 2020). This has been compounded with people spending greater times at home during social distancing measures (Yang et al., 2020). This diffusion of information is difficult to control especially when people want to stay informed (Cinelli et al., 2020). We define these processes of digitally enabled information contagion as the 'infodemic'.

How misinformation influences public health and health behaviours is not well understood (Swire-Thompson and Lazer, 2020). Misinformation is powerful because it plays on emotions, often through diverse means (Pennycook et al., 2020). Deceit can be most effective when viewed as plausible or paired with poor or pseudo-scientific material (Pelizza, 2020). Individuals often seek out information that support their views (i.e., confirmation bias), which can result in polarisation of attitudes and toxicifcation of online spaces (Cinelli et al., 2020; Schwarz et al., 2016). Social media plays an important role in how people react and respond to the pandemic, which can be compounded by algorithms promoting echo chambers of information sharing. 'Superspreader' moments can occur especially when items are shared or amplified by social influencers or accounts with large followings (e.g., celebrities or politicians) (Gruzd and Mai, 2020).

Emerging evidence about the scale of misinformation related to COVID-19

Worries about the role of misinformation influencing the course of COVID-19 have seen considerable effort towards monitoring the scale, reach and influence of misinformation. Studies estimating the scale of misinformation across internet technologies suggest encouraging signs that truthful information is far more prevalent than misinformation (Cinelli et al., 2020; Evanega et al., 2020; Ferrara, 2020; Pulido et al., 2020; Yang et al., 2020). This has not necessarily translated to low reach, with one UK survey estimating that half of people having come across some form of misinformation (Ofcom, 2020).

The most common types of misinformation tend to focus on cures or treatment (Evanega et al., 2020), including growing anti-vaccine movements (Burki, 2020), likely influenced by broader fears and anxiety of a newly discovered disease. Conspiracy theories have gained prominence, including one hoax centred around the hashtag #FilmYourHospital gaining 99,000 posts on Twitter (Gruzd and Mai, 2020). Other frequent content includes themes around the causes of COVID-19, mode of transmission and general uncertainty caused by unclear government messages on policies or actions (Kyriakidou et al., 2020). These combined forms of misinformation can be dangerous through disrupting preventative efforts to contain the spread of the disease. Bots – defined here as co-ordinated internet agents that simulate human activity to share content – have also played an important role, with greater likelihood of sharing misinformation (Ferrara, 2020; Yang et al., 2020).

The low prevalence of misinformation has not necessarily translated to few harms. While measuring the specific impacts of misinformation can be difficult, there has been observable impacts on populations. Panic buying of food and household items were commonplace prior to initial lockdowns in many nations (Mao, 2020). In Liverpool (UK), 5 G masts were set on fire following fears they were transmitting COVID-19 following conspiracy theories and disinformation (BBC, 2020). Rioting occurred in Ukraine following social media posts about citizens being evacuated from China (Miller, 2020). These large-scale events co-incide with smaller and incremental erosions of trust that undermine responses to COVID-19. Misinformation has also been used to discriminate or justify exclusion of people. For example, discussions in the US around how different racial groups may be immune to COVID-19 due to melanin levels (Pelizza, 2020).

Misinformation influencing policy narratives

When faced with an uncertain, unknown and unprecedented pandemic, the need for clear public health communication is imperative. Many national governments, including the UK, instigated daily briefings in the early stages of the pandemic. The immediate (e.g., 24–48 hours) response to messages is paramount to inform a population and avoid confusion.

Misinformation during the early stages of public health communication can disrupt or harm the process for sharing clear messages. Human responses to viewing or sharing misinformation may interject uncertainty among the population, especially if information gathering due to broader fears or anxiety demands immediate answers. Misinformation may lead to populations resisting policy decisions or public health advice, which could disrupt policy or reduce trust in official announcements. Misinformation can also appeal to public officials who may integrate content into the decision making process or amplify the misinformation so that it becomes prominent (Pelizza, 2020)

Tackling the spread of misinformation is therefore a key government priority. The role of misinformation in reacting to policy announcements is not well understood. There is very little understanding of its association with policy announcements, including whether coordinated and uncoordinated actions respond to critical moments in policy decision making (Swire-Thompson and Lazer, 2020). We need to identify whether there are critical periods to target interventions co-inciding with major policy announcements to reduce any potential impact of misinformation in disrupting the policy process.

'Infoveillance' – defined here as systematically utilising structured and unstructured online health information to monitor human responses to public health events (Eysenbach, 2011) – is a core approach to support public health policy decision making. Through combining Data Science, Communication Studies and Epidemiological techniques, we can harness novel data structures (including social media records) to understand how to effectively disrupt the spread of misinformation. However, there have been few applications of successful infosurveillance with COVID-19. For example, most of the initial studies responding to COVID-19 using Twitter data were descriptive simply reporting what people were tweeting (e.g., Boon-Itt, 2020; Ferrara, 2020; Kwon et al., 2020; Pulido et al., 2020; Stella et al., 2020; Stephens, 2020), although some more rigorous studies exist which tease out specific pathways about how misinformation operates (Aiello et al., 2020; De Santis et al., 2020; Gruzd and Mai, 2020). None of these examples are applied to evaluating a policy announcement or context. COVID-19 is not going to be the last pandemic. We need to get smarter at dealing with misinformation during public health crises to capitalise on it in future research and policy.

Aim

The aim of our project is to examine how trends in misinformation react to a major government policy announcement during a public health crisis. We utilise the case study of the UK lockdown announcement on 23 March 2020 (20:00–20:30). Specifically, we tackle the following research questions: (i) Was the number of Twitter posts identified as misinformation more or less common on Twitter up to 48 hours following the announcement (compared to 48 hours before it)? (ii) Were tweets from bots more or less frequent post-announcement? (iii) What misinformation topics were shared online during the study period and were they influenced by the announcement?

Methodology

Ethical approval

Ethical approval for the study was granted by the University of Liverpool's Research Ethics Committee (ref: 7654).

Twitter data

Our primary data source was Twitter. Twitter is a social media platform where users post short messages (called 'tweets') and interact with other users posts. Twitter is a useful surveillance resource for studying both the spread of misinformation (Swire-Thompson

and Lazer, 2020) and responses to infectious diseases (Tulloch et al., 2019) including COVID-19 (Aiello et al., 2020; Boon-Itt, 2020; De Santis et al., 2020; Ferrara, 2020; Gruzd and Mai, 2020; Kwon et al., 2020; Pulido et al., 2020; Stella et al., 2020; Stephens, 2020). We accessed UK tweets via Twitter's 'premium application programming interface (API)'. Most research utilising Twitter data for studying responses to COVID-19 uses the 'stream API' which provide real-time streams of tweets. The 'stream API' source does not capture all tweets and can introduce bias into dataset, with coverage varying between 1% and 40% of all tweets over time (Morstatter et al., 2013). Our approach allows us to interrogate trends among all tweets (including those of both original content and retweets) as well as allowing for a greater number of identification keywords to compile a more detailed COVID-19-related dataset.

We select all tweets 48 hours before (20.00 21 March to 20.00 23 March 2020) and 48 hours after the end of the announcement (20.30 23 March 2020 to 20:30 25 March 2020) as well as during the announcement. We selected the immediate time period following the announcement since it represents a critical period for ensuring consistent messages. Tweets from the UK were selected from using Twitter's assigned tweet or profile location 'Great Britain (GB)'. Keywords were selected through expert opinion, review of Twitter's own list of COVID-19-related trending terms, as well as following keywords selected in other Twitter analyses (see Online Appendix for full list). Our analytical dataset therefore consists only of COVID-19-related tweets.

Bots have been shown to be associated with sharing COVID-19 misinformation (Ferrara, 2020), and are prevalent on Twitter. We utilise the classifier 'tweetbo-tornot' for estimating whether a Twitter account is a bot or not (Kearney, 2018). The classifier is based on a trained machine learning model that has learnt to predict the likelihood of a user being a bot based on several features of their account including how often they tweet, number of followers, profile location and the content of their previous tweets (e.g., number of hashtags, mentions, capital letters). Previous evaluation of the resource found a 92% model accuracy (Kearney, 2018). We define bots as users with a predicted probability of ≥ 0.5.

Identifying misinformation on Twitter

Measuring misinformation is difficult due to the diverse and dynamic nature of content. We focus on identifying sources and tweets identified by 'fact checker' websites as viral misinformation. Our approach follows previous studies for measuring misinformation (Cinelli et al., 2020; Vosoughi et al., 2018). While we may capture tweets that post satirical content or ridicule misinformation, such tweets also maintain the circulation of misinformation that could be misinterpreted by others (e.g., satire may be taken seriously within a different cultural setting or if taken out of context).

We collected all posts displayed on five major fact checker websites in English (Snopes, Politifact, Health Feedback, Ferret and Fullfact) that identified posts as true, false or misleading (neither true nor false). Where a specific tweet was the source of misinformation, we identify tweets and retweets based on the tweet's unique ID. Where there was an alternative source, we identify whether URLs within tweets match the source URL. We then extract the websites that have been identified as sharing misinformation content (via News Guard Tech) and identify whether it came from a website associated with sharing misinformation. We also identify Twitter accounts associated with these websites and extract tweets sharing content from these accounts. Finally, we compiled a list of keywords from the World health Organization's (WHO) official guide to misinformation about COVID-19 and extract tweets containing these keywords (WHO, 2020).

Statistical analysis

Government announcements act as natural experiments allowing for novel and rigorous study designs that can approximate causal effects through interrupted time-series designs. We present results from three time-series modelling approaches to estimate whether misinformation was more or less prevalent following the announcement. For all analyses, we aggregated data to both hours and minutes and define the start of the television (TV) announcement (20:00 23 March 2020) as our point of intervention.

First, we model the temporal trend using a multilevel Negative Binomial regression model following previous recommendations (Bernal et al., 2017). Our outcome variable is the count of tweets classified as misinformation. We predict this using an offset term in the logged total number of tweets, time and a binary variable of whether an observation was pre- or post-announcement (20:00 23 March 2020). An offset term was included to account for varying number of tweets over time. Temporal trends were modelled in a flexible way using B-spline bases accounting for patterns of seasonality and temporal autocorrelation. A negative binomial model was fitted to account for overdispersion relaxing the restrictive assumption of the equality between the variance and the mean assumed by the standard Poisson model (Ver Hoef and Boveng, 2007). We report the β coefficient for the intervention

Table 1. Descriptive statistics of sample.

	Pre-announcement	Post-announcement	Total
All tweets	1,236,161	1,295,727	2,531,888
Tweets from bots	409,349	449,060	858,409
Tweets with misinformation	11,135	9037	20,172

estimated effect, which is equivalent to whether there is a 'level' change in the trend denoting whether misinformation was higher or lower overall following the announcement (Bernal et al., 2017). Incidence rate ratios are reported for the coefficient to aid interpretation, producing the 'relative' increase or decrease in misinformation. We did not hypothesise there was a change in the relative trend (i.e., slope change) post announcement.

Second, we utilise Bayesian Structural Time Series models (Brodersen et al., 2015). Here, we train our model to predict trends in our outcome (count of misinformation) pre-intervention to 'learn' the underlying trends. This is achieved through a state-space time-series model with flexible regression components to capture the linear trend, seasonality and associations to coefficients (both fixed and dynamic). A control variable was selected that was highly correlated to our outcome (total number of tweets) to help improve the training phase. We note the difficulty in selecting an appropriate control here (i.e., one that is correlated to the outcome), especially as it may also be affected by the intervention itself, but utilise it as the overall number of tweets is an important denominator that can explain the extent of possible misinformation. The method then estimates the counterfactual trend for our outcome post-intervention, based on observed trends in our control measures. We report the cumulative absolute (i.e., observed minus expected) and relative (i.e., ratio for observed over expected) effects for the post-intervention time period to infer whether there were an impact on trends.

Third, we use an 'Auto Regressive Integrated Moving Average' (ARIMA) regression. ARIMA models are commonly used for modelling time-series data since they explicitly account for autocorrelation in time-series observations. We utilise an automatic ARIMA modelling framework which tests different combinations of seasonality, lagged effects and temporal 'drift' measures to select a final model that performs best on model fit statistics (Hyndman and Khandakar, 2008). Models therefore have different specifications to best accommodate the observed temporal structures in our data. We report the β coefficient for the intervention estimated effect, which represents the absolute step or level change in the number of tweets post-intervention.

With no clear gold-standard approach for what is the 'best' methodology in using three methods we provide robust triangulation into evidence of any associations. Using multiple models helps to minimise any spurious effects or model specific findings that can lead to misleading results. Each method offers different approaches and strengths for estimating the impact of lockdown announcement. The Bayesian Structural Time Series model employs a counterfactual approach, whereas the other approaches focus on identifying changes in trends holistically.

We also classified misinformation tweets during the whole time period using a Latent Dirichlet Allocation model to identify 'topics' or 'types' of tweets based on the terms included with them (Blei et al., 2003). Raw tweets were cleaned through removing special characters and emojis, converting terms to lower case, removing 'stop words' (e.g., the, of, in etc.) and lemmatising terms (i.e., converting words to their root such as removing 'ing' or 'ed').

Analyses were conducted using R and Python software. All data and analytical code are openly available at: https://github.com/markagreen/misinformation_uk_lockdown_2020.

Results

Describing the extent of misinformation pre- and post-announcement

There were a total of 2,531,888 COVID-19-related tweets between 20:00 21 March and 20:30 25 March 2020 (Table 1); 51.2% of all COVID-19-related tweets were made following the announcement. We identified 20,172 tweets containing misinformation (0.8% of all COVID-19-related tweets) over the period, with fewer misinformation tweets post-announcement (45.8%). Finally, we estimate 858,409 tweets made potentially made by bots over the period, with slightly more tweets from bots post-announcement (52.3%). While COVID-19-related tweets from bots over the period were only slightly more likely to have contained misinformation (0.85%), the difference was statistically significant ($\chi^2 = 12.3$, p < 0.001).

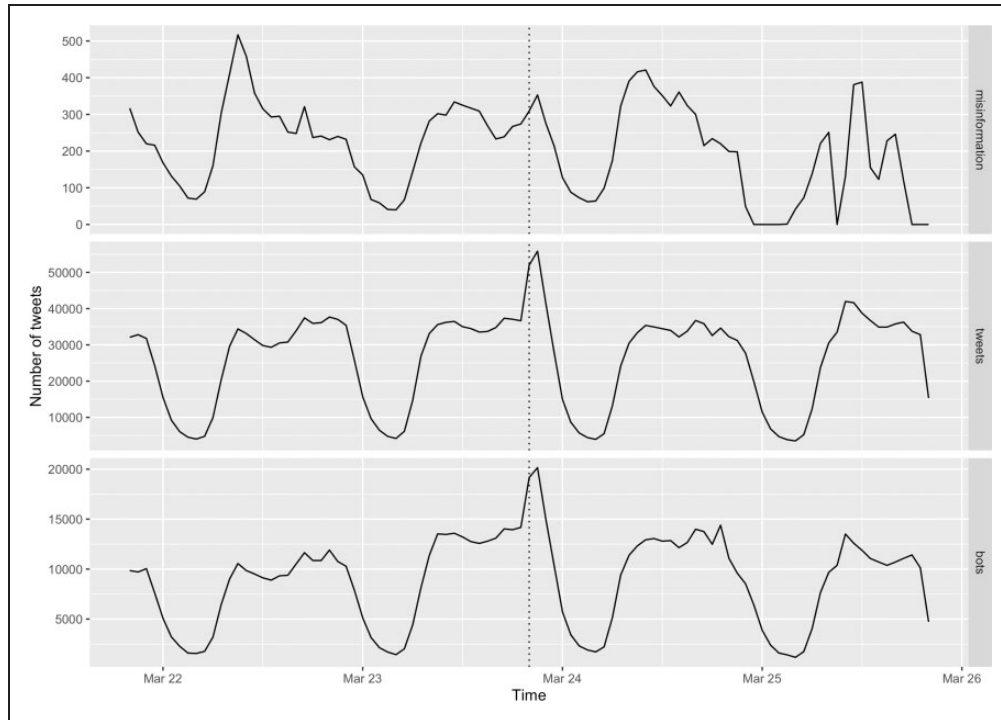Figure 1 plots how the frequency of our three tweet types varies across the time period. For all COVID-19-

Figure 1. Number of COVID-19-related tweets over the study period by tweet type. Dotted line is 20:00 (23 March 2020).

related tweets, there is a cyclical pattern observed throughout the period, with numbers of tweets higher during daytime hours and falling during night time. There is a large increase in the number of COVID-19-related tweets following the announcement. For example, we observed a 42% increase in the number of tweets between 20:00–20:59 (n = 51,990) compared to 19:00-19:59 (n = 36,652) on the 23 March and a 38% increase compared to the 20:00–20:59 (n = 37,672) for the previous day. The large initial change in the trend is followed by a gradual return to the normal cycle pre-announcement.

Trends in the amount of misinformation follow a similar daily cycle to all COVID-19-related tweets, albeit slightly more erratic due to the smaller numbers involved. The graph does not appear to show any discernible increase following the announcement.

For the estimated number of COVID-19-related tweets from bots, we also find a similar daily cycle in the number of tweets. There was a clear increase in COVID-19-related tweets from bots following the announcement. The trend line also appears higher on the day of the announcement compared to other days. Figure A (Online Appendix) investigates this further, through showing an increase in the percentage of COVID-19-related tweets estimated from bots on the day of (23 March 2020) and day after the announcement (24 March 2020). For example, 30.8% of COVID-19-related tweets a day before the

announcement (22 March 2020) were from bots, compared to 36.4% on the day (23 March 2020) and 37.0% the following day (24 March 2020). The increase then declines back to 31.3% on 25 March 2020.

We compared mean engagement (e.g., replies, retweets, quotes, favourites) of these three types of tweets. While there were slightly more shares of COVID-19-related tweets pre-announcement (mean number of shares = 6.11) than compared to post-announcement (4.95), this difference was not statistically significant (Welch two sample t-test = 1.2, p = 0.2). Mean engagement of tweets estimated to be bots (2.88) was significantly lower than compared to those that were not bots (6.59) (t = 6.2, p < 0.001). Finally, tweets identified as misinformation had slightly higher numbers of tweets shared (7.04) than those not (5.50); however, this difference was not meaningful (t = 0.41, p = 0.7).

Estimating the impact of the announcement

We estimated whether the national lockdown announcement had an impact on trends in the number of tweets identified as misinformation. Table 2 presents the results across our three analytical models when using hourly data. Findings from the Negative Binomial regression model suggested there was a positive impact of the announcement on trends in misinformation. The estimated effect size was large

Table 2. Effect of the announcement on the hourly trend for the number of misinformation tweets about COVID-19 (48-hour pre- and post-announcement).

Model	Estimate	95% Lower	95% Upper	p
Negative binomial regression				
Relative effect (IRR)	1.913	1.014	3.610	0.045
Bayesian Structural Time Series				
Absolute effect	−2428	−3746	−1098	0.001
Relative effect (%)	−21%	−33%	−10%	0.001
ARIMA regression				
Absolute effect (beta)	−47	−79	−14	0.005

IRR: incidence rate ratio; ARIMA: Auto Regressive Integrated Moving Average.

Table 3. Effect of the announcement on the hourly trend for the number of bots tweeting about COVID-19 (48-hour pre- and post-announcement).

Model	Estimate	95% Lower	95% Upper	p
Negative binomial regression				
Relative effect (IRR)	1.131	1.049	1.220	0.001
Bayesian Structural Time Series				
Absolute effect	10,504	−26,715	50,308	0.308
Relative effect (%)	2.4%	−6.1%	11.0%	0.308
ARIMA regression				
Absolute effect (beta)	330	−36	595	0.077

IRR: incidence rate ratio; ARIMA: Auto Regressive Integrated Moving Average.

suggesting a 91% increase in misinformation tweets post-announcement, albeit highly uncertain in magnitude (95% confidence intervals (CIs) 1.014 to 3.610). The Bayesian structural Time Series model estimates a −21% (95% credible intervals −33% to −9.6%) change (i.e., decrease) in the number of misinformation tweets following the announcement (over the total period). This represents a total of 2428 fewer misinformation tweets 48 hours following the announcement than would have been expected. The ARIMA model also suggested a negative impact of the announcement on trends in misinformation on Twitter ($\beta = -46.469$, 95% CIs −79.286 to −14.101).

Results for when data is analysed by minute of the day demonstrate mixed associations (Online Appendix Table A). Each model gives a different result, with the negative binomial regression model suggesting an increase in the number of misinformation tweets following the announcement, the Bayesian Structural Time Series detecting no effect, and the ARIMA model suggesting a decrease. No conclusive relationships can therefore be derived from these analyses.

With trends in tweets from bots appearing to change in relating to the national lockdown announcement in Figure 1, we re-ran our analyses with the number of tweets from bots as our outcome variable (Table 3). The negative binomial regression model displayed a positive association, with a 13.1% increase (95% CIs 1.049 to 1.220) in the number of tweets from bots following the announcement. No statistically significant effect was detected for both the Bayesian Structural Time Series and ARIMA models. Similar findings were also displayed when running analyses at the minute level (Online Appendix Table B), with the only difference being a statistically significant positive effect now observed in the ARIMA model ($\beta = 5.972$, 95% CIs 4.246 to 7.697).

Classifying misinformation throughout the period

Our final analysis classifies the types of tweets identified as misinformation. Following evaluation of model performance, a four topic solution was selected (see Online Appendix Figure B). The characteristics of the four topics are displayed in Figure 2. We classified each misinformation tweet as belonging to the topic with the highest gamma value (i.e., likelihood of belonging to each topic) to estimate their prevalence in our data. We describe each of our four clusters, including adding names to aid their interpretation:

1. *Government and policy (22.5%):* The common terms and themes followed debate, evaluation and anger towards government and policy decision making
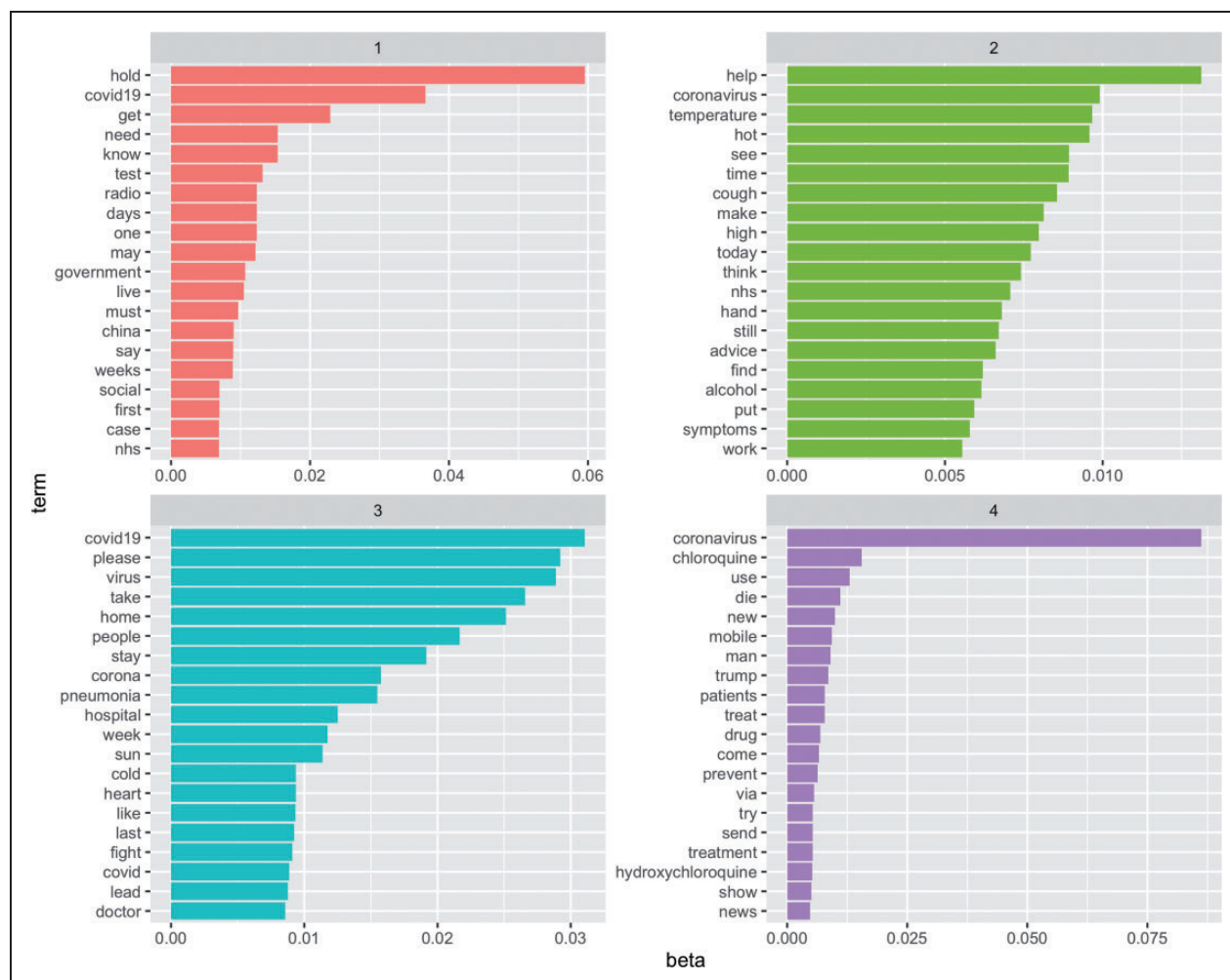
Figure 2. Top 20 most common terms found in four topics identified within COVID-19-related tweets. Numbers refer to the specific clusters, with 1 = government and policy, 2 = symptoms, 3 = pushing back against misinformation, and 4 = cures and treatments.

related to the pandemic. Discussion included a range of topics about the National Health Service (NHS) as well as the ability of the government to act and control the virus. Some tweets were also critical or blamed China for the pandemic as well.

2. *Symptoms (31.5%):* A range of topics were discussed within the cluster, with many centred on symptoms of COVID-19. Common symptoms included high 'temperature' and 'cough' as well as some discussion of ways to minimise risk or aid symptoms (e.g., 'alcohol' or 'hand'). Tweets were for both individuals discussing their own symptoms or sharing information on incorrect ways to diagnose the disease.

3. *Pushing back against misinformation (17.4%):* Tweets were often engaging with misinformation directly flagging and critiquing, for instance, misleading information. These included individuals

asking others to 'take the disease or situation seriously', 'stay at home' or socially distance and expressing disappointment at what was being shared. Many tweets stressed the importance of policy actions to help healthcare cope. While tweets did not always represent harmful misinformation, tweets debunking misinformation ironically can increase the likelihood that some false information will be believed since they make it more salient (e.g., by repeating it) or widening its circulation (Schwarz et al., 2016).

4. *Cures and treatments (28.6%):* The final topic represented tweets discussing cures and treatments for COVID-19. These included medicines being pushed as effective remedies such as chloroquine as well as preventative strategies both linked to transmitting COVID-19 or the causes of the disease (e.g., sixth most common term was 'mobile' reflecting
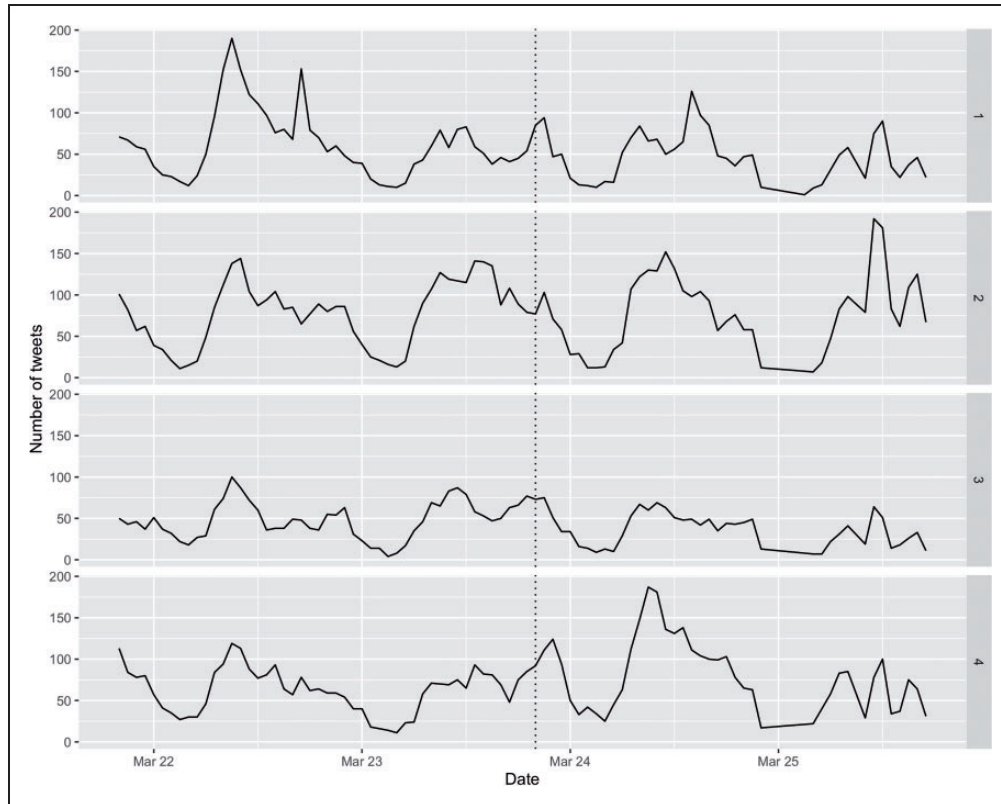
Figure 3. Hourly trends in the number of misinformation tweets by topic. Dotted line is 20:00 (23 March 2020). Numbers refer to the specific clusters, with 1 = government and policy, 2 = symptoms, 3 = pushing back against misinformation, and 4 = cures and treatments.

discussion of links to 5G mobile phone masts). Language was often emotive when describing these topics (e.g., fourth most common term was 'die'). The eighth most common term was 'trump', reflecting misinformation being shared by the US President Donald Trump at the time.

Figure 3 presents trends in each topics over the period. Topic 1 ('government and policy') was more common during 22 March 2020, then following a consistent trend otherwise. The topic displays a slight increase in tweets following the announcement, although in line with other short term increases over the period and not consistent over the longer term. Topic 2 ('symptoms') was fairly cyclical throughout the period. While there was a small increase following the announcement, the increase appears to follow similar increases over the period and was not sustained. Topic 3 ('pushing back against misinformation') was the least frequent topic throughout the period, with a consistent temporal pattern. Finally, topic 4 ('cures and treatments') shows an increase immediately following the announcement and this increase continues throughout the 24 March 2020. Indeed by 03:00 on the 24

March 2020, it represented just over half of all tweets identified as misinformation (although there were few tweets during this period).

We repeated our time-series analyses to estimate the effect of the announcement on trends in the topics (Table 4). Here, we utilise the same model specifications as used previously and, however, use total number of misinformation tweets as the offset or control rather than the total number of tweets. For topic 1 ('government and policy'), a negative association was detected in the Bayesian Structural Time Series model suggesting fewer tweets for the topic post announcement. The result contrasted with the other two models were no associations was detected. Topic 2 ('symptoms') resulted in no effects observed in the Bayesian Structural Time Series and ARIMA models, with a negative effect observed in the Negative Binomial regression model. Topic 3 ('pushing back against misinformation') detected negative associations in the Bayesian Structural Time Series and ARIMA models (no association in the Negative Binomial model), suggesting fewer tweets following the announcement. Finally, a positive association was detected for topic 4 ('cures and treatments') across all

Table 4. Effect of the announcement on the hourly trends for misinformation topics (48-hour pre- and post-announcement).

Topic	Name	Estimate	Lower	Upper	p
Bayesian Structural Time Series (absolute values)					
1	Government and policy	−369	−736	−31	0.018
2	Symptoms	21	−310	358	0.465
3	Pushing back against misinformation	−411	−614	−218	0.001
4	Cures and treatments	800	581	1025	0.001
Bayesian Structural Time Series (relative (%) values)					
1	Government and policy	−16%	−32%	−1.3%	0.018
2	Symptoms	0.7%	−10%	12%	0.465
3	Pushing back against misinformation	−22%	−32%	−12%	0.001
4	Cures and treatments	33%	24%	42%	0.001
Negative binomial regression (IRR values)					
1	Government and policy	1.077	0.879	1.320	0.474
2	Symptoms	0.581	0.510	0.661	<0.001
3	Pushing back against misinformation	0.838	0.700	1.004	0.0558
4	Cures and treatments	1.800	1.561	2.074	<0.001
ARIMA regression (beta values)					
1	Government and policy	−8.57	−21.29	4.14	0.186
2	Symptoms	−17.67	−40.21	4.86	0.124
3	Pushing back against misinformation	−9.02	−15.15	−2.88	0.004
4	Cures and treatments	16.29	3.94	28.64	0.010

IRR: incidence rate ratio; ARIMA: Auto Regressive Integrated Moving Average.

three models. The result suggests that the announcement saw a significant increase in the number of misinformation tweets related to 'cures and treatments', with a fairly large effect size reported across each model.

Discussion

Key results

Our study demonstrates a novel analysis of Big Data to evaluate whether the announcement of the UK national lockdown had an impact on trends in the sharing of misinformation on Twitter. While we found evidence that there were more COVID-19-related tweets following the policy announcement on national TV, we did not find consistent evidence suggesting it also led to an overall increase of misinformation being shared on Twitter. Analyses of hourly tweets across three analytical models suggested there were fewer tweets with misinformation post-announcement, although results were mixed when analysing trends at the minute level. In contrast, we found evidence of an increase in the number of tweets from bots post-announcement; however, tweets from bots were also less likely to be shared by users than normal tweets. We also present a novel classification of tweets identified as misinformation identifying four distinct topics: 'government and policy', 'symptoms', 'pushing back against misinformation' and 'cures and treatments'. While most types of

misinformation did not change or even decreased following the announcement, we found evidence for an increase in misinformation about 'cures and treatments' in the 48 hours following the announcement.

Interpretation

Our findings suggest that governments can be assured that moments of policy communications may not be immediately disrupted by misinformation overall. It follows findings from other countries where most population groups, including those with low trust of the government, clearly understood major policy messages (Barari et al., 2020; Ofcom, 2020). Where major national announcements offer certainty in policy responses and expectations, they may act specifically against the lure of misinformation. Regular and clear messaging from governments may therefore be encouraged throughout pandemics. Public engagement and openness in decision making can help to build trust (Pieczka and Escobar, 2013). Such implications should be considered against possible message saturation that may lead to disengagement (Baseman et al., 2013).

We find greater Twitter activity following the national lockdown announcement, representing populations reacting to the broadcast. While this increase in overall Twitter activity did not result in an increase in misinformation overall, we did find an increase in tweets from bots. This finding might be some cause for concern in the sharing of policy announcements, given

that bots are associated with sharing misinformation about COVID-19 (Ferrara, 2020; Yang et al., 2020). While tweets from bots were less likely to be shared by others within our study, which may indicate they were less problematic actors on social media, they were also common which suggest a high scale of activity that could be harmful. Controlling the prevalence and sharing abilities of bots may be an effective tool for governments in reducing misinformation.

Clear policy announcements might be important but are likely to only be one part of a suite of effective policy actions. The every-changing and unprecedent nature of COVID-19, much like any novel pandemic, complicates the ability to monitor misinformation as themes, issues and information may be changing rapidly. The dynamic nature of misinformation formats, types and channels for sharing, as well as the difficulty in controlling large quantities of information across the internet, means developing effective strategies for disrupting or inhibiting the spread of misinformation is challenging (Larson, 2018; Swire-Thompson and Lazer, 2020). We have little information and understanding of what we are dealing with, especially how to effectively deal with it on a global scale that accounts for multiple contexts. COVID-19 is not going to be the last pandemic; we need to use this time to catch up to the problem of misinformation in public health that we have largely ignored or poorly understood so that we can capitalise on it for the future (both for research and policy).

The study period we use is focused on the immediate time frame (i.e., 48 hours post-announcement) and future research should consider the longer term implications. Trends in cases of misinformation are time-relevant and often occur with short spikes over a few days (De Santis et al., 2020; Gruzd and Mai, 2020; Ofcom, 2020). The accumulation of posts, theories and ideas over the longer term might result in population responses that have public health implications (e.g., promoting conspiracy theories) or the gradual erosion of trust in policy efforts (Larson, 2018; Pulido et al., 2020). Greater research is needed on the longer term implications of misinformation on populations and public health as well as the timely evaluation of the communication of public health strategies.

We present evidence of four distinct types of misinformation that can help to characterise the themes and issues being discussed on social media. The topics found largely match findings from other research across different international contexts, displaying key misinformation themes relating to COVID-19 (Evanega et al., 2020; Ferrara, 2020; Gruzd and Mai, 2020; Yang et al., 2020). Our analyses suggested that there was an increase in tweets related to the topic 'cures and treatment' post-announcement. This may

represent individuals responding to the new national restrictions seeking to share 'opportunities' to end it. Given that many of the themes related to the topic can be dangerous, this represents a key area for policy makers to target. The 'pushing back against misinformation' topic was the least prevalent topic and declined following the announcement. This may represent a strategy governments could chose to focus their efforts on through larger scale and co-ordinated social media campaigns following key policy announcements. Such a recommendation should be made cautiously, given the mixed evidence on the effectiveness of educational campaigns across public health issues (Kyriakidou et al., 2020; Pennycook et al., 2020; Schwarz et al., 2016; Swire-Thompson and Lazer, 2020).

Our project provides a framework for how we can utilise Big Data for the real-time monitoring and evaluation of how misinformation is being shared online in relation to key policy moments. Capitalising on natural experiments, as in our study, can help to build rigorous evidence to guide the design and identification of effective interventions for tackling misinformation about COVID-19. Our approach can also be applied to different outcomes and public health issues both within and outside of the COVID-19 pandemic.

Limitations

The findings derived from our dataset might have limited wider generalisability. Twitter users are not representative of the UK population, being over-represented by younger age groups, males and individuals of higher socio-economic status (Sloan, 2017). We only consider one social media platform and extending our analysis to incorporate other platforms that each have different demographics and purposes will help to broaden the generalisability of our results. Evidence suggests that smaller platforms, such as Parler, might form 'echo chambers' for sharing misinformation which has been censored on other platforms (Cinelli et al., 2020). Such relationships may be ignored when considering more popular platforms, such as Twitter, limiting our understanding.

While we followed similar approaches implemented elsewhere for estimating whether tweets represented misinformation (Cinelli et al., 2020; Vosoughi et al., 2018), identifying misinformation in big datasets is difficult and our approach may be limited. For example, we may have underestimated the scale of misinformation through not being able to capture general misleading comments, rumours or those not reported by fact checking organisations (Swire-Thompson and Lazer, 2020). There is greater need for effective ways to capture the dynamically changing nature of misinformation that can help to support research and policy.

Similarly, we also estimate whether Twitter users are 'bots' which may introduce some bias into our data. While the approach we used has previously reported a 92% accuracy in classifying Twitter users as bots or not (Kearney, 2018), there may be some error introduced due to the classification model being trained on older data than what we use.

Although our analyses are rigorous and utilise novel time-series data, they can only tell us whether there was an increase in misinformation. They do not consider the mechanisms or networks through which misinformation may spread across social media. We do not consider if different population groups responded differently to lockdown announcement and were more or less likely to share misinformation following it. While these topics were beyond the scope of our study and not covered in our study aim, they are arguable more useful for identifying actionable strategies for minimising the spread and reach of misinformation. Future research should consider evaluating these specific pathways and how they are affected or disrupted by key policy announcements.

Conclusion

While the scale of misinformation being shared during the COVID-19 pandemic is of significant public health concern, our study presents evidence that key government policy announcements may help to minimise misinformation (at least in the immediate 48 hours following). Our results do not mean that governments should ignore the potential impact of misinformation circulating on social media platforms during or before key policy announcements. In particular, there appears to have been an increase in dangerous misinformation about 'cures and treatment'. Rather, an agreed approach is needed beyond the 48-hour timeline post announcement for official outlets on social media platforms (e.g., local authorities, government departments and public health bodies) to coordinate and synchronise their targeted messages in counteracting misinformation. It is a scenario which requires officials to provide a balance between freedom of expression and open public discourse, clear and consistent presentation of scientific evidence and the constant promotion of reliable sources for pandemic information. If public health messages are clear, consistent and part of an agreed approach, this has a higher potential pushing back against the infodemic.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Mark Green [ID] https://orcid.org/0000-0002-0942-6628
Frances Darlington Pollock [ID] https://orcid.org/0000-0001-5544-4459

Supplemental material

Supplemental material for this article is available online.

References

Aiello L, et al. (2020) How epidemic psychology works on social media: Evolution of responses to the COVID-19 pandemic, *arXiv*. Available at: https://arxiv.org/pdf/2007.13169.pdf (accessed 27 April 2021).

Barari S, et al. (2020) Evaluating COVID-19 public health messaging in Italy: Self-reported compliance and growing mental health concerns, *medRxiv*. Cold Spring Harbor Laboratory Press. DOI: 10.1101/2020.03.27.20042820. https://www.medrxiv.org/content/10.1101/2020.03.27.20042820v2

Baseman JG, et al. (2013) Public health communications and alert Fatigue. *BMC Health Services Research* 13(1): 295.

BBC (2020) *Liverpool 5G phone mast damaged in arson attack*. Available at: www.bbc.co.uk/news/uk-england-merseyside-52818328 (accessed 6 December 2020).

Bernal JL, Cummins S and Gasparrini A (2017) Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology* 46(1): 348–355.

Blei D, Ng A and Jordan M (2003) Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.

Boon-Itt S and Skunkan Y (2020) Public Perception of the COVID-19 pandemic on twitter: Sentiment Analysis and Topic Modeling Study. *JMIR Public Health and Surveillance* 6(4): e21978.

Brodersen K, et al. (2015) Inferring causal impact using Bayesian Structural Time-Series models. *Annals of Applied Statistics* 9: 247–274.

Burki T (2020) The online anti-vaccine movement in the age of COVID-19. *The Lancet Digital Health* 2(10): e504–e505.

Cinelli M, et al. (2020) The COVID-19 social media Infodemic. *Scientific Reports* 10: 16598.

Daraz L, et al. (2019) Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the internet. *Journal of General Internal Medicine* 34(9): 1884–1891.

De Santis E, Martino A and Rizzi A (2020) An infoveillance system for detecting and tracking relevant topics from Italian tweets during the COVID-19 event. *IEEE Access* 8: 132527–132538.

Evanega S, et al. (2020) Coronavirus misinformation: quantifying sources and themes in the COVID-19 'infodemic'. https://allianceforscience.cornell.edu/wp-content/uploads/2020/10/Evanega-et-al-Coronavirus-misinformation-submitted_07_23_20-1.pdf

Eysenbach G (2011) Infodemiology and infoveillance: Tracking online health information and cyberbehavior for public health. *American Journal of Preventive Medicine Elsevier* 40(5): S154–S158.

Ferrara E (2020) What types of COVID-19 conspiracies are populated by Twitter bots?, *arXiv*. https://arxiv.org/abs/2004.09531

Gruzd A and Mai P (2020) Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society* 7(2): 2053951720938405.

Hyndman RJ and Khandakar Y (2008) Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 27(3): 1–22.

Kearney M (2018) *tweetbotornot*. Available at: https://github.com/mkearney/tweetbotornot (accessed 27 April 2021).

Kwon J, et al. (2020) Defining facets of social distancing during the COVID-19 pandemic: Twitter Analysis. *Journal of Biomedical Informatics Elsevier Inc* 111: 103601.

Kyriakidou M, et al. (2020) *Government and media misinformation about COVID-19 is confusing the public*. Available at: https://blogs.lse.ac.uk/covid19/2020/05/07/government-and-media-misinformation-about-covid-19-is-confusing-the-public/ (accessed 9 December 2020).

Larson HJ (2018) The biggest pandemic risk? Viral misinformation. *Nature* 562: 309.

Mao F (2020) *Coronavirus panic: Why are people stockpiling toilet paper?* Available at: www.bbc.co.uk/news/world-australia-51731422 (accessed 6 December 2020).

Mawdsley H (2020) *Fake news and the flu*. Available at: https://wellcomecollection.org/articles/XXIeHhEAACYAIdKz (accessed 6 December 2020).

Miller C (2020) *A viral email about coronavirus had people smashing buses and blocking hospitals*. Available at: www.buzzfeednews.com/article/christopherm51/coronavirus-ukraine-china (accessed 6 December 2020).

Morstatter F, et al. (2013) Is the sample good enough? Comparing data from Twitter's streaming {API} with Twitter's firehose, *CoRR*, abs/1306.5. Available at: http://arxiv.org/abs/1306.5204 (accessed 27 April 2021).

Ofcom (2020) *Covid-19 news and information: consumption and attitudes*. Available at: www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/news-media/coronavirus-news-consumption-attitudes-behaviour (accessed 27 April 2021).

Pelizza A (2020) "No disease for the others": How COVID-19 data can enact new and old alterities. *Big Data & Society SAGE Publications Ltd* 7(2): 2053951720942542.

Pennycook G, et al. (2020) Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable Accuracy-Nudge intervention. *Psychological Science* 31(7): 770–780.

Pieczka M and Escobar O (2013) Dialogue and science: Innovation in policy-making and the discourse of public engagement in the UK. *Science and Public Policy* 40(1): 113–126.

Pulido CM, et al. (2020) COVID-19 infodemic: More retweets for science-based information on coronavirus than for false Information. *International Sociology* 35(4): 377–392.

Schwarz N, Newman E and Leach W (2016) Making the truth stick & the myths fade: lessons from cognitive psychology. *Behavioral Science & Policy* 2(1): 85–95.

Sloan L (2017) Who tweets in the United Kingdom? Profiling the Twitter population using the British social attitudes survey 2015. *Social Media + Society* 3(1): 2056305117698981.

Stella M, Restocchi V and De Deyne S (2020) #lockdown: Network-enhanced emotional profiling in the time of COVID-19. *Big Data and Cognitive Computing* 4(2): 14.

Stephens M (2020) A geospatial infodemic: Mapping Twitter conspiracy theories of COVID-19. *Dialogues in Human Geography*. 10(2): 276–281.

Swire-Thompson B and Lazer D (2020) Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health* 41(1): 433–451.

Tulloch JSP, et al. (2019) Mapping tweets to a known disease epidemiology; a case study of Lyme disease in the United Kingdom and Republic of Ireland. *Journal of Biomedical Informatics: X* 4: 100060.

Ver Hoef JM and Boveng PL (2007) Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology* 88(11): 2766–2772.

Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science (New York, N.Y.)* 359(6380): 1146–1151.

WHO (2020) *Coronavirus disease (COVID-19) advice for the public: Mythbusters*. Available at: www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters (accessed 27 April 2021).

Yang K-C, Torres-Lugo C and Menczer F (2020) Prevalence of low-credibility information on Twitter during the COVID-19 outbreak, *arXiv*. Available at: https://arxiv.org/abs/2004.14484 (accessed 27 April 2021).