Harnessing social media in the modelling of pandemics – challenges and opportunities

Joanna Sooknanan[1] and Nicholas Mays[2]

[1] The University of The West Indies Open Campus, Bridgetown, Barbados.
ORCID 0000-0001-9618-1288

[2] Professor of Health Policy, Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London WC1H 9SH
ORCID 0000-0001-9808-8466

Corresponding Author

Joanna Sooknanan
The University of The West Indies Open Campus
Bridgetown, Barbados
Email: joannasooknanan@gmail.com

Abstract

As COVID-19 spreads throughout the world without a straightforward treatment or widespread vaccine coverage in the near future, mathematical models of disease spread and of the potential impact of mitigation measures have been thrust into the limelight. With their popularity and ability to disseminate information relatively freely and rapidly, information from social media platforms offers a user-generated, spontaneous insight into users' minds that may capture beliefs, opinions, attitudes, intentions and behaviour towards outbreaks of infectious disease not obtainable elsewhere. The interactive, immersive nature of social media may reveal emergent behaviour that does not occur in engagement with traditional mass media or conventional surveys. In recognition of the dramatic shift to life online during the COVID-19 pandemic to mitigate disease spread and the increasing threat of further pandemics, we examine the challenges and opportunities inherent in the use of social media data in infectious disease modelling with particular focus on their inclusion in compartmental models.

1 Introduction

The COVID-19 pandemic has led to an eruption of mathematical modeling efforts which have been used by many governments to guide their policy decisions for its containment. Though not unfamiliar to policy makers, this is the first time that infectious disease models have been used on such a global scale and with such urgency. In 2001, they were used most prominently in the UK to guide policy decisions related to an outbreak of foot and mouth disease (Woolhouse 2003). More recently, during the 2014 Ebola and the 2018 Zika epidemics, models were used to forecast

the course of the epidemic, allocate resources and evaluate the potential impact of interventions (Chretien et al. 2015; Keegan et al. 2017).

What is striking in the COVID-19 pandemic is the parallel increased usage of social media as an integral part of many people's lives. Use in July 2020 was 10.5% higher than in the same month in 2019 (DataReportal 2020a) as people in many countries were urged to "stay home, stay safe". Even after the initial easing of lockdowns and other social restrictions later in 2020, these acquired digital habits appeared to have become part of the "new normal" (DataReportal 2020b)

Hans Kluge (2020), WHO Regional Director for Europe, has argued that 'Behavioural insights are valuable to inform the planning of appropriate pandemic response measures.' In this regard, social media data offer a novel wealth of potentially useful information on public awareness, opinions, attitudes and beliefs, relevant to intentions and eventual behaviour. Given the major role of mathematical modelling in guiding health policy decisions in the COVID-19 pandemic at a national level, data from social media are thus, at least in principle, useful sources to help refine these models. Models, which though undoubtedly valuable, have been criticized for being too narrowly epidemiological and lacking input of social science data (Rhodes et al. 2020). Social media platforms provide one potential source of such social data.

COVID-19 is the sixth pandemic since the influenza pandemic of 1918. With more than five new diseases emerging in human populations every year – each with the potential to spread and become pandemic (IPBES 2020) - combined with our increasing use and dependence on technology, social media data offer an opportunity to enhance the capability of these models.

 The use of social media data in compartmental infectious disease models was introduced by Sooknanan and Comissiong (2020). We extend this analysis to discuss in greater detail the challenges and opportunities inherent in the use of social media data in infectious disease modelling.

2 Behavioural insights obtained from social media data

The most common forms of the compartmental models used to model infectious disease epidemics are variations of the basic susceptible-exposed-infected-removed (SEIR) or susceptible-infected-removed (SIR) models (Mohamadou et al. 2020). Their very simplicity - in line with Einstein's maxim that everything should be made as simple as possible, but not simpler - means that they may be easily modified as information about disease transmission is updated.

During an emerging disease outbreak, in the absence of straightforward treatments or widely available vaccines, governments are likely to be entirely dependent for long periods on the use of non-pharmaceutical interventions such as isolation, physical distancing, personal protective equipment such as face coverings and hand hygiene, to stem transmission. These practices all involve changes in behaviour - voluntary or mandated – which a sufficiently large proportion of the population must undertake consistently for them to be truly effective. Even with a vaccine, the

conspicuous presence of the anti-vaxxers movement and vaccine hesitancy online, has resulted in concerns about the likely levels of take-up of vaccination during an outbreak. These intentions and behaviours could potentially have a large impact on the future course of an outbreak and should be included in modeling efforts (Bae et al. 2021).

Models coupling behaviour–disease dynamics have been on the increase (Funk et al. 2010; Pananos et al. 2017). However, data are needed to "robustly estimate, and possibly predict, behavioural parameters" (Manfredi and D'Onofrio 2013). Behaviour is related to "attitudes, belief systems, opinions and awareness of a disease" (Funk et al. 2010). What modelers need access to continuously and in a timely manner is an indication of prevailing attitudes and perceptions that provide an indication of likely behaviour.

The participatory, informal and spontaneous nature of social media such as Twitter or Sina Weibo allows for almost real-time access to the uncensored feelings and opinions of users, their risk perceptions, intentions and their preventive behaviours - something that researchers would ordinarily have to collect painstakingly through surveys and/or focus groups.

Twitter, which is increasingly being recognised as an important source of such information, has been mined to study public opinion and sentiments regarding disease outbreaks such as influenza and COVID-19 (Signorini et al. 2011; (Boon-Itt and Skunkan 2020), as well as public opinion on vaccination (Tavoschi et al. 2020) and mask-wearing (He et al. 2021).

Data from social media indicative of trends in these behaviours or intentions may be used in models with suitable care. For example, researchers using Twitter found that 10% of all Tweets in the United States related to mask wearing were against (He et al. 2021). This information may then be used when modifying basic disease transmission models to reflect the prevailing sentiment (and related, inferred likely pattern of public behaviour) thereby increasing the accuracy of estimates of transmission dynamics.

However, some caution is advised when using these insights, since there may be a discrepancy between intentions and actual behaviour (Sheeran and Webb 2016), whereby social media sentiment which may be considered to reflect intent may differ from eventual behaviour offline (Smith et al. 2020; Social Media Research Group 2016). A comparison of the insights obtained from social media with subsequent behavioural data may then determine how these two sources relate to each other for eventual incorporation into the modelling process.

3. Incorporating social media data and its characteristic features into models

While social media use allows for the indirect observation of trends in population behaviour, these media also interact with their users to influence their behaviour, for example, by increasing people's awareness of non-pharmaceutical, treatment and preventive practices of others with whom they have a relationship and/or who are influential with them. Though behavioural change has been recognised as an important factor in reducing transmission, "the complex interplay of

changing epidemiology, media attention, pandemic control measures, risk perception, and public health behaviour" presents significant challenges to modelers (Betsch et al. 2020).

Traditionally, the effects of media have been included in compartmental models in two ways. One approach is to add compartments representing aware and unaware subpopulations. Transitions between unaware and aware compartments are assumed to take place at constant rates with aware people having an assumed lower risk of infection (Agaba et al. 2017). This "awareness" may be explicitly included within a media compartment $M(t)$ whose growth rate has generally been assumed to be proportional to the number of infected individuals (Greenhalgh et al. 2015; Misra et al. 2011 However, Kumar et al. (2021), while investigating the effects of social media on an influenza epidemic and the COVID-19 pandemic, considered a media compartment $M(t)$ consisting of the daily normalized number of Tweets over a specified period of time. Using their model for COVID-19 as an example, these data consisted of daily Tweets with health-related keywords such as "corona", "coronavirus", "covid", "quarantine" etc. from March 22, 2020 to July 20, 2020. Interaction of susceptibles with this compartment resulted in "aware" individuals who were influenced positively by the Tweets and thus changed their behaviour to reduce their social contacts and thus reduce their risk of infection.

The second way to include the effects of media use is as a reduction in transmission with a contact rate of the form $\beta S f(I)$ where $f(I)$ represents a decreasing function of the number of infected, such as e^{-mI} (Cui et al. 2008). In order to understand the relationship between the spread of H1N1 influenza and the number of Tweets, Huo et al. (2020) combined these two approaches by using a disease transmission rate reduced by $e^{-\alpha T}$ and an additional compartment $T(t)$ representing the number of Tweets about the H1N1 epidemic at time t. Unknown parameter values were estimated by data fitting to the percentage of Tweets that self-reported influenza and the officially recorded cases in England between May and December 2009.

These two contrasting approaches to modeling take advantage of the assumed beneficial effects on behaviour of social media. At the same time, social media are awash with conspiracy theories, disinformation and fake news which are likely to have a detrimental effect on control strategies (Towers et al. 2015; Chandrasekaran et al. 2017; Kumar et al. 2020). An investigation into the extent to which misinformation or unverifiable information about the COVID-19 pandemic is spread on Twitter showed that misinformation accounted for 25% of Tweets (Kouzy et al. 2020). This "anti-information" may be incorporated into models by further refining Tweets (and hence the media compartments) into posts representing positive and negative information in terms of their likely relative effect on transmission dynamics (Huo and Zhang 2016) in order to modify the contact rate.

While these models highlight the importance of media, they neglect the unique features associated with social media use. Though similar, in that they both serve to impart information, the interactive, immersive nature of social media may precipitate and capture emergent behaviour (via intentions) that does not occur in engagement with traditional mass media such as television and

newspapers. Social media, with their AI-based recommendations and ease of use, encourage like-minded individuals to seek each other's virtual and, sometimes face-to-face company (Spohr 2017). Social learning theory (Bandura, 1978), suggests that a person's behaviour is affected by observing the behaviour of others. Consequently, users may reinforce and amplify each other's opinions and attitudes - both positive and negative - on any course of action to form "online echo chambers" (Burki 2019) unrelated to their consumption of conventional media messages.

This confirmation bias effect has only recently been incorporated in the modelling process in the shape of SIR-Opinion models by coupling disease dynamics with opinion dynamics (Tyson et al. 2020). Here, the susceptible compartment is divided into four subpopulations (with different infection rates) characterised by the strength of their attitudes either positive or negative towards for example, non-pharmaceutical interventions. The effects of echo chambers and amplification are captured by the interactions between these subpopulations via "influence functions" which can reinforce or weaken the strength of their opinions and attitudes, thus allowing for movement of individuals between these compartments as their opinions and attitudes change.

Though Tyson et al. (2020) acknowledged that "empirically measured influence functions" were not yet available to them, they used four different forms of these influence functions defined as linear, saturating, fixed-order saturating, and reverse-order saturating functions (all dependent on the number of infected people) to regulate these interactions. These influence functions were chosen so that as the disease prevalence increases, the influence of those with positive attitudes towards non-pharmaceutical interventions increases and that of those with negative attitudes decreases. While this model included salient features of social media, i.e. the effects of echo chambers and amplification, future models may also need to take account of the activities of "super-spreaders" of opinion. These are the small number of social media influencers with a wide reach who have the potential to alter the behaviour of large numbers of other people, either positively or negatively in terms of the impact on pandemic control. Though usually associated with commercial marketing and/or celebrity culture, their influence has been recognized by public health officials who have attempted to co-opt them to reinforce official public health messages (Bolat 2020; Archer et al. 2020). Modelers may need to do the same, especially when trying to estimate behavioural responses in specific population sub-groups.

4 Social media-based surveillance - early detection and monitoring

Researchers have recognised the potential of social media data as an alternative data source for tracking public health trends for a number of years. With each post or conversation, users leave a digital trace which may be mined to identify attitudes and opinions (sentiment analysis), as well as health status indirectly, or comments about their health. Not only can posts be accessed for content, but they may be linked to the demographic data and geographic location of the account holder through geo-tagging.

Typically, researchers access posts or "likes" (Mackey et al. 2020; Gittelman et al. 2015) made over a period of time and then identify which of these contain health information. A combination of human and computational approaches - including keyword filtering, crowdsourcing, computational algorithms and machine learning – may then be used to process and filter the data. For example, in a recent study to characterize the self-reporting of symptoms, experiences with testing, and mentions of recovery related to COVID-19 via social media, 4,492,954 Tweets were captured using the public streaming Twitter application programming interface (API) over the period March 3-20, 2020 (Mackey et al. 2020). These included terms in the English language such as "covid19," "corona," and "coronavirus". After further processing to identify relevant topic clusters with keywords such as "diagnosed," "pneumonia," "fever," "test," "testing kit," "sharing," "symptoms," "isolating," "cough," "ER" (emergency room), and "emergency room, and to remove duplicate Tweets, this was refined to 3465 Tweets. This led to the observation that though many posters reported symptoms they thought related to COVID-19, they were unable to get tested to confirm their concerns.

The relative frequency of mention of keywords associated with a disease has been found to be strongly correlated with the subsequent number of doctor visits and later reports of the number of people infected (Jordan et al. 2018; Aramaki et al. 2011; Marques-Toledo et al. 2017; Huo 2020). Recent studies (Gharavi et al. 2020; Li et al. 2020) have found a comparable pattern during the COVID-19 pandemic. In China, internet searches and social media data (keywords and sentiments) have been shown to be strongly correlated with daily incidence and exhibit an online peak 10 to 14 days before the peak of daily incidence from official data (Li et al. 2020). Similar research using Twitter showed a lag of 5 to 19 days between social media reports and official COVID-19 statistics in the United States (Gharavi et al. 2020).

Thus these social media data may be used alongside traditional surveillance data in models. For example, by using the percentage of Tweets which included phrases like "have flu", "have the flu", "have swine flu", and "have the swine flu" during the 2009 H1N1 outbreak to parameterise their SEI compartmental model, Pawelek et al. (2014) were able to reproduce the peaks of both the percentage of Tweets and of surveillance data showing the number of infections.

Therefore, social media data may provide not only the basis of an early warning system (ahead of official statistics) at the start of an outbreak but may be used subsequently in models to allow for an assessment of the likely progression of the disease during the outbreak. At the very least, trends in the data, and correspondingly in the outbreak, may be detected. From a practical point of view, this sort of lead time gives public health authorities vital opportunities to make policy decisions, inform the public and put in place the necessary arrangements, for example, to test, trace and isolate and/or treat an influx of cases. From a modelling point of view, it is also possible that such data may be used as a proxy for conventional data in countries where testing is limited or where there is a significant delay in reporting numbers due to test processing time and consequently reporting time. For example, the effect of a change in NPIs on incidence may be gauged by an analysis of social media data before and after the change. If these data are used in a model to

estimate parameters or to determine the form of a function, the model may then be used to forecast the effects of further changes related to those NPIs – at least with respect to social media sentiment.

5 Social media and internet-based data – the good, the bad and the ugly

Though the world has faced disease outbreaks before, one of the major differences with current and future outbreaks is the widespread use of social media. On the surface, the almost instantaneous, open access availability of some social media data, whereby modelers can bypass or eliminate the formal structures that were used previously to restrict access or share data, seems like a real boon to infectious disease modelers, particularly in countries where other sources of information are restricted and/or censored.

However, due to their volume and complexity, these data are generally difficult to process using traditional applications and tools. The analytic techniques needed differ from traditional statistical methods which generally cannot be used for analysis of social media data such as audio, images, video, and unstructured text. Suggestions for ways of incorporating multiple data sources in models such as by weighting data sources may be found in recent papers by De Angelis et al. (2015) and Gandomi and Haider (2015). The use of these "novel data streams" (Althouse et al. 2015) comes with concerns about the data themselves, particularly in terms of their representativeness, privacy concerns and the filtering of information by algorithms (De Angelis et al. 2015; Althouse et al. 2015; Lee et al. 2016).

In terms of sample validity, though social media data may not be limited to a particular geographic location, they are ultimately bound by the popularity of each platform in the region under consideration. In addition, different demographic groups use social media in different ways and to different degrees – traditionally, social media platforms have been the preserve of younger people. Thus the data cannot be assumed to be fully representative of the general population (Mellon and Prosser 2016). However, they may be more representative of important population sub-groups such as younger people whose behaviour may be particularly important to learn about since they are likely to have the most daily social contacts and are thus among the groups most likely to spread infection during a pandemic. This will have a bearing on how the data can be interpreted and used in public health terms.

Although there is a correlation between the occurrence of a disease and the likelihood that an individual posts about it or searches the internet for related information, it is tempting to make inferences regarding disease trends based solely on this online information. Researchers using the public health tool Google Flu Trends made such a misstep when they sought to provide real-time monitoring of influenza-like illnesses (ILI) activity exclusively through Google searches for influenza-related information. However, they overestimated the prevalence of influenza in the 2012-2013 and 2011-2012 seasons by more than 50% when compared to surveillance reports from the Centers for Disease Control and Prevention (CDC) (Lazer et al. 2014).

7

Though this tool was eventually discontinued, the lessons learnt still resonate today. An oft cited reason for the inability of Google Flu Trends to accurately project the prevalence of influenza cases lay in the method of deriving the search terms. Algorithmic dynamics allowed for the recommendation of searches (autosuggest feature) to users after entering a term like fever or cough. Since people are more likely to interact with content suggested to them, this had the effect of skewing the terms people searched for.

Another possible contributor to this inaccuracy was attributed to a type of "echo chamber" effect. Online search behaviour is not restricted to the people suffering from a disease. Increased media reports predicting an active influenza season resulted in a spike in the number of searches for influenza related information in anticipation of becoming ill and a mistaken inference about the extent of seasonal influenza (Harris, 2014).

Despite this, the use of data gleaned from online sources has been steadily increasing in models. A major challenge is that these data "are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis" (Lazer et al. 2014). However, hybrid systems are increasingly being utilised where Google searches (e.g. Google Trends) are used in conjunction with social media data - whose content is carefully analyzed for context (Panuganti et al. 2020) - alongside frequently updated, traditional, surveillance sources. For example, an increase in searches for a particular illness could signal an outbreak, while concomitant to this, social media may be scrutinized for references to this illness and then conventional surveillance sources mobilized to provide additional data.

The challenge lies in consolidating all this diverse information to discover and quantify causal relationships between phenomena such as public opinion, attitudes and reported behaviour, as well as in identifying relationships between indicators of the extent and severity of the outbreak and patterns of Twitter and other social media activity. This calls for collaboration between data scientists, epidemiologists, social scientists and modelers to unravel these relationships so that the data can be used in models with a clearer understanding of what they mean and how they should be interpreted.

6 Conclusion

With more than half of the world's population currently using social media, now is an ideal time to rethink how we use data from these powerful platforms. Social media simultaneously reflect, forecast and shape behaviour. This spontaneous, "nontraditional" source of data may be mined to capture the public's attitudes, beliefs, opinions, awareness, intentions and reported behaviour towards an infectious disease, which can then be incorporated with caution into disease models. However, social media platforms are not simply passive observers and reporters of trends, but they also interact with their users, for example, by increasing social awareness about non-pharmaceutical interventions and treatment practices. During these uncertain times when

information changes quickly, it is imperative that modelers use the most up to date information sources to refine their models - social media data may be just the extra source needed to help accomplish this.

References

Agaba G, Kyrychko Y, Blyuss K (2017) Mathematical model for the impact of awareness on the dynamics of infectious diseases. Mathematical Biosciences 286:22-30.

Althouse B, Scarpino S, Meyers L et al. (2015) Enhancing disease surveillance with novel data streams: challenges and opportunities. EPJ Data Science. https://doi.org/10.1140/epjds/s13688-015-0054-0

Aramaki E, Maskawa S, Morita M (2011) Twitter catches the flu: detecting influenza epidemics using Twitter.In: Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics

Archer C, Wolf K, & Nalloor J. (2020) Capitalising on chaos–exploring the impact and future of social media influencer engagement during the early stages of a global pandemic. Media International Australia, 1329878X20958157.

Bae S, Sung E(C, Kwon O (2021) Accounting for social media effects to improve the accuracy of infection models: combatting the COVID-19 pandemic and infodemic. European Journal of Information Systems 1–14. doi: 10.1080/0960085x.2021.1890530

Bandura, A. (1977) Social learning theory. New York: General Learning Press.

Betsch C, Wieler L, Habersaat K (2020) Monitoring behavioural insights related to COVID-19. The Lancet 395:1255-1256. https://doi.org/10.1016/s0140-6736(20)30729-7

Bolat, E (2020) Why the UK government is paying social media influencers to post about coronavirus. The Conversation, (9 September 2020).

Boon-Itt S, Skunkan Y (2020) Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. JMIR Public Health and Surveillance 6:e21978. doi: 10.2196/21978

Burki T (2019) Vaccine misinformation and social media. The Lancet Digital Health 1:e258-e259. https://doi.org/10.1016/s2589-7500(19)30136-0

Chandrasekaran N, Gressick K, Singh V, Kwal J, Cap N, Koru-Sengul T, Curry CL (2017) The utility of social media in providing information on zika virus. Cureus 9(10):e1792

Chretien J, Riley S, George D (2015) Mathematical modeling of the West Africa Ebola epidemic. eLife. https://doi.org/10.7554/elife.09186

Cui J, Sun Y, Zhu H (2008) The Impact of Media on the Control of Infectious Diseases. Journal of Dynamics and Differential Equations 20:31-53. https://doi.org/10.1007/s10884-007-9075-0

DataReportal (2020a) Social Media Users — DataReportal – Global Digital Insights. In: DataReportal – Global Digital Insights. https://datareportal.com/social-media-users. Accessed 29 Nov 2020

DataReportal (2020b) Digital 2020: October Global Statshot. In: DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2020-october-global-statshot. Accessed 7 Dec 2020

De Angelis D, Presanis A, Birrell P et al. (2015) Four key challenges in infectious disease modelling using data from multiple sources. Epidemics 10:83-87. https://doi.org/10.1016/j.epidem.2014.09.004

Doogan C, Buntine W, Linger H, Brunt S (2020) Public Perceptions and Attitudes Towards COVID-19 Non-Pharmaceutical Interventions Across Six Countries: A Topic Modeling Analysis of Twitter Data (Preprint). Journal of Medical Internet Research. https://doi.org/10.2196/21419

Funk S, Salathé M, Jansen V (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of The Royal Society Interface 7:1247-1256. https://doi.org/10.1098/rsif.2010.0142

Gandomi A, Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35:137-144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gharavi, E., Nazemi, N., & Dadgostari, F. (2020) Early outbreak detection for proactive crisis management using Twitter data: Covid-19 a case study in the US. *arXiv preprint arXiv:2005.00475*.

Gittelman S, Lange V, Gotway Crawford CA, et al (2015) A New Source of Data for Public Health Surveillance: Facebook Likes. Journal of Medical Internet Research. doi: 10.2196/jmir.3970

Greenhalgh D, Rana S, Samanta S et al. (2015) Awareness programs control infectious disease – Multiple delay induced mathematical model. Applied Mathematics and Computation 251:539-563. https://doi.org/10.1016/j.amc.2014.11.091

Harris J (2014) Big data hubris. In: The Data Roundtable. https://blogs.sas.com/content/datamanagement/2014/03/26/big-data-hubris/. Accessed 18 Mar 2021

He L, He C, Reynolds TL, et al (2021) Why do people oppose mask wearing? A comprehensive analysis of US tweets during the COVID-19 pandemic. Journal of the American Medical Informatics Association. doi: 10.1093/jamia/ocab047

Huo HF, Zhang XM (2016) Modeling the influence of Twitter in reducing and increasing the spread of influenza epidemics. SpringerPlus 5(1):88

Huo H, Jing S, Wang X, Xiang H (2020) Modeling and analysis of a H1N1 model with relapse and effect of Twitter. Physica A: Statistical Mechanics and its Applications 560:125136. doi: 10.1016/j.physa.2020.125136

IPBES Workshop On Biodiversity And Pandemics (2020) IPBES Workshop On Biodiversity And Pandemics. The IPBES Bureau and Multidisciplinary Expert Panel

Jordan S, Hovet S, Fung I, et al (2018) Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. Data 4:6. doi: 10.3390/data4010006

Keegan L, Lessler J, Johansson M (2017) Quantifying Zika: Advancing the Epidemiology of Zika With Quantitative Models. The Journal of Infectious Diseases 216:S884-S890. https://doi.org/10.1093/infdis/jix437

Kluge, H. H. (2020) Behavioural insights are valuable to inform the planning of appropriate pandemic response measures. http://www.euro.who.int/en/mediacentre/sections/statements/2020/statement-behavioural-insights-are-valuable-toinform-the-planning-of-appropriate-pandemic-response-measures

Kouzy R, Abi Jaoude J, Kraitem A et al. (2020) Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus. doi: 10.7759/cureus.7255

Kumar S, Xu C, Ghildayal N et al. (2021) Social media effectiveness as a humanitarian response to mitigate influenza epidemic and COVID-19 pandemic. Annals of Operations Research. doi: 10.1007/s10479-021-03955-y

Kumar Chandan Srivastava, Deepti Shrivastava, Kumar Gaurav Chhabra et al. (2020) Facade of media and social media during COVID-19: A review. International Journal of Research in Pharmaceutical Sciences 11:142-149. doi: 10.26452/ijrps.v11ispl1.2288

Lazer D, Kennedy R, King G, Vespignani A (2014) The Parable of Google Flu: Traps in Big Data Analysis. Science 343:1203–1205. doi: 10.1126/science.1248506

Lee E, Asher J, Goldlust S et al. (2016) Mind the Scales: Harnessing Spatial Big Data for Infectious Disease Surveillance and Inference. Journal of Infectious Diseases 214:S409-S413. https://doi.org/10.1093/infdis/jiw344

Li C, Chen L, Chen X et al. (2020) Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Eurosurveillance. https://doi.org/10.2807/1560-7917.es.2020.25.10.2000199

Mackey T, Purushothaman V, Li J, et al (2020) Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health and Surveillance. doi: 10.2196/19509

Manfredi P, D'Onofrio A (2013) Modeling the interplay between human behavior and the spread of infectious diseases. Springer, Berlin

Marques-Toledo C, Degener C, Vinhal L et al. (2017) Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level. PLOS Neglected Tropical Diseases 11:e0005729. https://doi.org/10.1371/journal.pntd.0005729

Mellon J, Prosser C (2016) Twitter and Facebook are Not Representative of the General Population: Political Attitudes and Demographics of Social Media Users. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2791625

Misra, A., A. Sharma and J. Shukla (2011) Modeling and analysis of effects of awareness programs by media on the spread of infectious diseases. *Mathematical and Computer Modelling*, 53, 5-6, 1221-1228

Mohamadou Y, Halidou A, Kapen P (2020) A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. Applied Intelligence 50:3913-3925. https://doi.org/10.1007/s10489-020-01770-9

Pawelek KA, Oeldorf-Hirsch A, Rong L (2014) Modeling the impact of twitter on influenza epidemics. Math Biosci Eng 11(6):1337–1356

Pananos AD, Bury TM, Wang C, et al (2017) Critical dynamics in population vaccinating behavior. Proceedings of the National Academy of Sciences 114:13762–13767. doi: 10.1073/pnas.1704093114

Panuganti BA, Jafari A, MacDonald B, DeConde AS (2020) Predicting COVID-19 Incidence Using Anosmia and Other COVID-19 Symptomatology: Preliminary Analysis Using Google and Twitter. Otolaryngology–Head and Neck Surgery 163:491–497. doi: 10.1177/0194599820932128

Rhodes T, Lancaster K, Lees S, Parker M (2020) Modelling the pandemic: attuning models to their contexts. BMJ Global Health. doi: 10.1136/bmjgh-2020-002914

Sheeran P, Webb T (2016) The Intention-Behavior Gap. Social and Personality Psychology Compass 10:503-518. https://doi.org/ 10.1111/spc3.12265

Signorini A, Segre A, Polgreen P (2011) The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS ONE 6:e19467. doi: 10.1371/journal.pone.0019467

Smith, L E, Potts, H W, Amlot, R, Fear, N T, Michie, S, Rubin, J. (2020) Adherence to the test, trace and isolate system: results from a time series of 21 nationally representative surveys in the UK (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] study). *medRxiv*.

Social Media Research Group (2016) Using social media for social research: An introduction. Social Media Research Group: United Kingdom.

Sooknanan J, Comissiong D (2020) Trending on Social Media Integrating Social Media into Infectious Disease Dynamics. Bulletin of Mathematical Biology. https://doi.org/ 10.1007/s11538-020-00757-4.

Spohr D (2017) Fake news and ideological polarization. Business Information Review 34:150-160. https://doi.org/10.1177/0266382117722446

Tavoschi L, Quattrone F, D'Andrea E et al. (2020) Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy. Human Vaccines & Immunotherapeutics 16:1062-1069. doi: 10.1080/21645515.2020.1714311

Towers S, Afzal S, Bernal G, Bliss N, Brown S, Espinoza B, Jackson J, Judson-Garcia J, Khan M, Lin M et al (2015) Mass media and the contagion of fear: the case of Ebola in America. PloS One 10(6):e0129179

Tyson, R C, Hamilton, S D, Lo, A S, Baumgaertner, B O, Krone, S M (2020).The Timing and Nature of Behavioural Responses Affect the Course of an Epidemic. Bulletin of Mathematical Biology, 82(1), 14. https://doi.org/10.1007/s11538-019-00684-z

Woolhouse M (2003) Foot-and-mouth disease in the UK: What should we do next time? Journal of Applied Microbiology 94:126-130. https://doi.org/10.1046/j.1365-2672.94.s1.15.x.