

Big Data and Digital Knowledge in Combating COVID-19

Caroline Paskarina^a, ^aUniversitas Padjadjaran, Indonesia, Email:
caroline.paskarina@unpad.ac.id

COVID-19 has made us aware that we need to actively participate by changing our behaviour to slow the ongoing spread of the global pandemic. We learned from many countries that the mode of governing the transmission of this virus is led by data-driven citizen awareness and timely government responses. Through a literature study of early publications discussing the efforts to cope with the outbreak, this article finds the rise of data-driven awareness as a basis to construct knowledge and further influence collective behaviour. Big data plays a strategic role in constructing knowledge of COVID-19, communicating awareness of personal hygiene, and building shared commitment to act in the form of self-isolation or social distancing to protect vulnerable others. It goes even further by encouraging governments to take strategic steps in handling the problem. These findings confirm the significance of big data as an instrument for social mobilisation and governing in the time of crisis while emphasising the importance of integrating data science as a basis for evidence-based decision making for further public policy.

Keywords: *Big data, Digital knowledge, Data-driven behaviour.*

Introduction

COVID-19 is a global pandemic that not only makes us alert about how vulnerable human beings are but also triggers our consciousness of humanity. Although there have been similar infectious disease outbreaks, such as Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), and other virus-related diseases, these outbreaks are not usually anticipated in a particular community, geographical region, or time (Huang, 2004; Oh et al., 2020). During the recent coronavirus or COVID-19 outbreaks, concern about data has become a strategic issue, both in terms of combating the outbreak and handling the social and economic impacts of the disease. From the previous studies, we learned how data about outbreaks is communicated by various forms of media, especially social media. Social media

platforms have functioned as immediate information channels from which the public can obtain disease-related information and exchange it with others in real-time. However, while the significant role of social media is acknowledged during infectious disease outbreaks, an explanation of exactly how social media affects the public's responses and behaviours has yet to be fully explored (Oh et al., 2020).

Previous studies have demonstrated the effect of social media during a pandemic situation. Some of the findings conclude that the media can substantially influence public perceptions of risk issues, such as in infectious disease outbreaks, through content that arouses fear of the disease's effects (Ali et al., 2019; Choi et al., 2017; Yoo et al., 2020; You et al., 2017). Another bulk of research also found how information about the outbreaks became viral and constructed public opinion to demand immediate government responses (Anderson et al., 2020; Scharmer, 2020; Wang et al., 2020). People tend to spread perceived risk in the form of messages via news media or their informal networks. Such messages are either amplified or weakened by various factors at each point of risk perception (Chong & Choy, 2018; Song et al., 2017).

Although social media is very influential as an instrument of communication, it's difficult to control the validity and accuracy of information spread in media. Misinformation, which includes false cure claims, conspiracy theories, and misleading information on the spread of the virus, hampers the efforts of governments to cope with the spread of COVID-19. During a pandemic situation, the public needs actual information to be able to respond appropriately to the uncertainties and crises they face. Misinformation that is not well managed by the government will have an impact through the public delegitimising true information and cause a decline in public trust in the government (Jang & Baek, 2019).

This article focuses on the production and transformation of data into behaviour. Currently, artificial intelligence and big data provide predictions through machine learning algorithms that subsequently analyse the data and provide visualisations for how the disease is growing and spreading. Big data is generally defined as large volumes of data that are produced routinely by organisations and are too complex for standard software packages to process (Mayer-Schönberger & Cukier, 2013). The utilisation of big data in various sectors is growing along with the increasing amount of data collected by each institution and the development of technology in storing and processing data into information.

What is interesting is to analyse what kind of knowledge is constructed through big data analytics and how this knowledge operates in combating infectious disease outbreaks, specifically COVID-19 or SARS-Cov-2 virus. As we know, data and information on COVID-19 is easy to gain, but the paradox is the bulk of this data is also advancing at lightning speed,

making it impossible to understand the complete picture. Big data analytics displayed in mathematical modelling and simulation (which are dispersed in social media) have shown the dynamics of measures against COVID-19. Throughout these days, the public could access curve on COVID-19—exponential, flattened, and otherwise—plotted in Excel and been reassured by them or scared by them or wondered whether we could trust them (Callaghan, 2020).

Hence, this article seeks to advance the understanding of data-driven participation in the case of combating COVID-19 by identifying the process of knowledge creation, transfer, and transformation from big data to collective behaviour. The remainder of the article is organised as follows: The research methodology is described after this section, followed by the main themes of the literature review about the governance of COVID-19. The discussion will be outlined in the section after the results. Some conclusions close the article.

Methodology

This article used a literature study, operationalised with a systematic review method. A systematic review was focused on gathering and synthesising evidence from previously published research. The first phase was represented by the definition of the scope of the study in compliance with the objectives. The acronym PICO (Population, Intervention, Comparison, and Outcomes) was used to identify keywords and formulate search strings from research questions. The population in the acronym may refer to specific literature containing the discussion of COVID-19 and big data. Boolean search operators for scoping articles were “coronavirus” OR “COVID-19” AND “big data”, typed in the Advanced Search feature. Data was collected through journal articles published in the Scopus databased from January until March 2020.

Intervention refers to the method of governance applied in response to COVID-19. In this article, the author identified the method of governance through the technique of coding data into two themes, namely usage of big data and knowledge produced from big data. Then, themes were analysed by comparing various methods of governance and interpreting the results to explore how this process constructs collective behaviour from the public to participate in combating COVID-19.

Results

The results of the searching process with the keyword “COVID-19” in <http://scopus.com> (searched in March, 23rd, 2020), showed that 675 works of literature are discussing “coronavirus” OR COVID-19 throughout 2020, mostly in the subject area of medicine (487),

immunology and microbiology (174), and biochemistry, genetics, and molecular biology areas (83). When the keyword was changed to “coronavirus” OR “COVID-19” AND “big data”, there were only five articles. This means that most of the academic publications are still discussing COVID-19 per se, mainly as a virus that endangers the human health system. This article will analyse five publications to identify the usage of big data and knowledge produced that drives citizen participation and collective behaviour in response to COVID-19.

The literature review mapping shows that publications in the early days of the spread of COVID-19 discussed a lot of the efforts made by the governments of Taiwan and China in handling COVID-19. Both of these countries have experienced coronavirus outbreaks before, so they have institutions and health protocols to respond to the initial symptoms of the presence of this virus. More detailed findings are contained in the following data:

Table 1: The Governance of COVID-19

Study	Study Theme	Objective(s)	Usage of Big Data	Forms of Knowledge Produced by Big Data
Wang et al. (2020)	Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing	This article aims to inform about the actions that were implemented quickly in Taiwan and assess the effectiveness of these actions in preventing a large-scale pandemic.	Taiwan created big data for analysis by integrating a national health insurance database with an immigration and customs database, supported by new technology, including QR code scanning, online reporting of travel history and health symptoms, SMS (short message service), and a toll-free number served as a hotline	<ul style="list-style-type: none"> • Disease symptoms. • Disease diffusion. • Individuals with high risk. • How to prevent contagion. • How to respond.

Study	Study Theme	Objective(s)	Usage of Big Data	Forms of Knowledge Produced by Big Data
			for citizens reports.	
Oh, Lee, & Han (2020)	The Effects of Social Media Use on Preventive Behaviours during Infectious Disease Outbreaks: The Mediating Role of Self-relevant Emotions and Public Risk Perception	The study investigates three issues: (1) how social media use relates to fear and anger; (2) the extent to which self-relevant emotions predict public risk perception; and (3) how social media use affects preventive behaviours through psychological and cognitive mechanisms.	<ul style="list-style-type: none"> Information from social media is used to identify fear or anger toward infectious disease outbreaks. These emotions then construct risk prevention behaviour. 	Infectious disease-related risk information.
Chen, Yang, Yang, Wang, & Bärnighausen (2020)	COVID-19 control in China during the mass population movements at New Year	This article analyses the impact of Chinese Government policy to extend the Lunar New Year Holiday so that the	Chinese Government database regarding routine screening, contact tracing, and early detection and medical care of COVID-19 patients.	Protocol to control people movement and behaviour to protect people's health and minimise spreading of the virus.

Study	Study Theme	Objective(s)	Usage of Big Data	Forms of Knowledge Produced by Big Data
		duration of the holiday would be sufficiently long to fully cover the suspected incubation period of COVID-19.		
Zhou et al. (2020)	Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV	This article applies the susceptible-exposed-infected-removed (SEIR) compartment model to estimate the basic reproduction number of the Wuhan novel coronavirus (2019-nCoV).	Data of infected individuals is simulated in mathematical modelling to predict the reproduction number of the virus and design of control measures.	Recommendation to control the extension of COVID-19 by identifying transmission processes.
Anderson et al. (2020)	How will country-based mitigation measures influence the course of the COVID-19 pandemic?	To identify uncertainty factors in coping with the virus spreading and mapping the mitigation strategies	Data of infected individuals are simulated in mathematical modelling to set priority decision in four phases of action entitled contain, delay,	Model-based predictions to help policymakers make the right decisions in a timely way, even with the uncertainties about COVID-19, specifically in

Study	Study Theme	Objective(s)	Usage of Big Data	Forms of Knowledge Produced by Big Data
		based on the phases of spreading.	research, and mitigate.	advising on how best to prevent transmission.

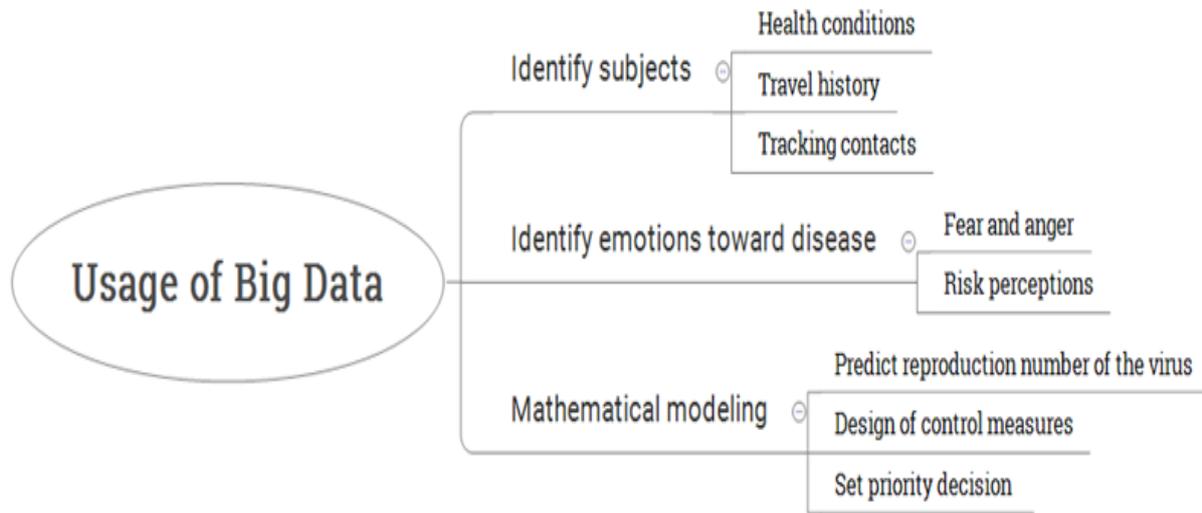
Source: research results, 2020

Discussion

From the results, we find that for some countries that had previous experience with infectious disease outbreaks, each government has established institutions or at least protocols to deal with the disease spreading. Also, past pandemics provide a database that can be used as research material and create knowledge that can be used when faced with similar situations. This condition makes several countries, such as China, Taiwan, and Singapore, more ready to respond to the spread of COVID-19 than other countries. Even so, the results of the literature mapping also show that social media plays an important role in channelling information about the latest developments in handling COVID-19 in various countries. Free access has even been allowed to several international scientific publishers to enable the public to learn from the handling models carried out by other countries and research that has been done. This shows the important role of big data in producing and reproducing knowledge about COVID-19.

Big data includes a broad and multidimensional understanding, which is characterised by its volume, velocity, variety, and veracity (Lazer et al., 2009; Mergel et al., 2016). Big data, referred to in the literature review mapping results, includes data on individual health conditions, travel records that have been made in a certain period, and data on the number of cases reported. By comparing the use of big data in various countries in handling COVID-19, it can be seen that there are three categories of big data usage: identifying subjects, identifying public emotions about the virus, and further analysing data through mathematical and statistical modelling (Figure 1).

Figure 1. Usage of Big Data

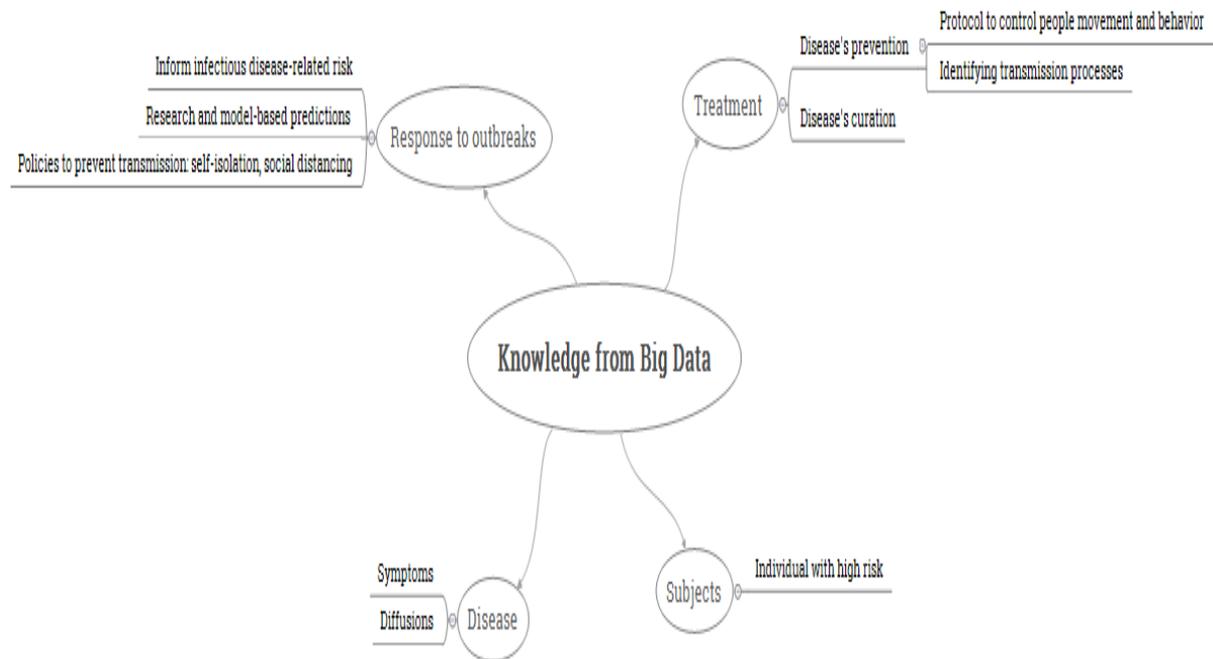


Source: Research results, 2020.

The knowledge produced by big data includes identification of subjects, namely people who were potentially exposed to COVID-19. The data used is sourced from travel history, health conditions, and other diseases that may affect the subject, as well as contact history, at least within the previous 14 days. Identification of other diseases that infect the subject forms knowledge about the vulnerability of the subject's health conditions, which in turn determines what the subject will be treated with. After the subject is identified, big data produces knowledge about how the subject will be organised based on categorisation. There are 3 categories created: positive, subject under surveillance, and patient under supervision. The method of handling is determined by the categorisation. The legitimacy of the method of handling the subject is strengthened by the regime of truth through the simultaneous dissemination of information about how other countries deal with the spread of viruses, best practices that can be followed, the distribution of relevant research results, mathematical modelling as evidence of knowledge claims about who the subject is, why it is governed in a certain way, and who has the authority to govern it.

The use of big data analytics generates knowledge about the essential things of COVID-19. This knowledge includes the characteristics of the virus itself, subjects who have a high risk of contracting the virus, handling the spread of the virus, and the government's response to the pandemic situation (Figure 2). These various dimensions of knowledge show that not only is the virus the target to be controlled, but it also affects humans and institutions at a systemic level. Data becomes an instrument to conduct surveillance of the physical condition and behaviour of individuals in their social relations, which then becomes an way for the government to formulate relevant and effective policies to deal with the spread of the virus (Callaghan, 2020).

Figure 2. Knowledge Produced from Big Data Analytics



Source: Research results, 2020.

This mechanism of knowledge shows that the big data that is spread online and through the results of previous research is used to:

First, inform about the workings of COVID-19, methods of prevention and treatment, as well as the latest situation developments related to handling COVID-19 in various countries. Social media and various personal communication platforms, such as WhatsApp and short message service, are widely used for this purpose.

Second, visualise the enemy. Several sites, such as worldometers.info/coronavirus and similar platforms at the local level, such as <https://pikobar.jabarprov.go.id/#/> (in West Java Province, Indonesia), <https://corona.jakarta.go.id/en/map> (in Jakarta), <https://kawalcovid19.id/> (non-government initiated platform to update the situation on COVID-19) present data visualisation with graphs and maps of the spread that makes it easy for readers to understand the danger of this virus spreading. Data uploaded by these sites is also used by academics to make simulations through mathematical modelling. Data science is needed to not only develop the models but also to determine in which ways they're wrong and which ways they're useful. This both predicts the development of the situation and analyses the effectiveness of various possible interventions.

Third, big data analytics becomes a source of information for individuals, communities, and governments to determine the actions to be taken. Visualised big data constructs the reality of COVID-19, generates knowledge about that reality, and directs the actions of individuals, communities, and governments in responding to that reality.

How do the production and dissemination of knowledge then construct collective behaviour in response to COVID-19? Previous studies of risk prevention behaviour show that behaviour can change when individuals assess the risk they face that cannot be controlled. In the case of the spread of COVID-19, the mapping of literature shows that big data is also a disciplinary instrument used by individuals, communities, and governments to reshape behaviour. This disciplinary mechanism can be seen in the recommendations produced by various research results and practical experiences of other countries, which are amplified through the dissemination of information on social media. Social distancing, self-isolation, quarantine, handwashing regularly, and the use of masks are recommendations conveyed repeatedly to discipline individual bodies.

Unlike previous research, which emphasises fear as a factor that drives behavioural change in the face of uncertainty, the response to COVID-19 shows that the public tends to change its behaviour according to its perception of the risks that may arise due to exposure to COVID-19. This perception arises through the acceptance of various sources of information and data presented in various forms of media. The COVID-19 pandemic situation presents similar experiences in various countries, and this is the basis for the public to test the validity of the data presented. Comparison with other countries' experiences is part of learning that helps the public to adapt to the desired changes of behaviour to fight the coronavirus.

In some countries, the role of the government is more focused on handling victims who have been identified and reported as having the virus, while prevention is carried out independently by practicing social distancing and working from home. Big data, as a disciplinary instrument, is also assisted by technology for social surveillance. Examples include body temperature sensor devices, a mobile or wearable app that tracks contacts automatically by the proximity of devices, and quick test tools to identify affected subjects. With big data tools of this sort, all compromised individuals can be tracked or isolated. Big data can help the government effectively forecast the development of a given pandemic, and to do that, we need to integrate the collection of the data in surveillance.

Aside from being a disciplinary instrument, data-driven change of behaviour also requires ongoing reproduction of knowledge, especially concerning preparedness to face risks in pandemic situations and the formation of individual and community resilience. In the future, COVID-19 and various public health problems will still be challenges for mankind, so data-

driven efforts are needed to anticipate the situation. Knowledge reproduction is an important part of ensuring the availability of valid, integrated, and meaningful data as material for policymaking. In this context, conducting ongoing multidisciplinary research is a crucial effort that needs to be continued to develop knowledge. This is not only to produce vaccines but also to prepare the public for the social, economic, and psychological impacts arising from interventions in managing the pandemic.

A review of previous studies also shows that the government can adopt policies through various pathways. An example is South Korea's experience in dealing with the spread of outbreaks in the past. The government implements policies to conduct as many rapid tests as possible to identify the spread of the virus. China's experience in applying strict restrictions has also been adopted by governments of other countries, both in the form of lockdowns and wide-scale restrictions. However, it cannot be denied that these policies also contain uncertainties, especially related to how long the policies will be implemented. During an uncertain pandemic, the need for academic and technical expertise becomes very important to support government policymaking (Weible et al., 2020). Therefore, the use of big data needs to be directed at evidence-based policymaking so that every decision taken has a rational and objective basis.

Conclusion

The coronavirus crisis opens up an opportunity to improvise new ways of governing uncertainty. Learning from the experience of various countries, it turns out that data can be used to direct changes in individual and community behaviour to fight the spread of the virus. In a 'normal' situation, our behaviour is constructed by regulation imposed by the government or by rational choice thinking based on cost and benefit analysis. However, these mechanisms tend to be irrelevant in a crisis. Therefore, it is necessary to optimise a data-driven approach by using big data to formulate evidence-based policy in an attempt to combat this pandemic.

Governments often make difficult decisions in uncertainty and amidst time constraints. These decisions must be led by data to create a systemic impact. Information from the results of big data analysis need to be well communicated to the public to grow public trust in the government. For this reason, the government needs to routinely convey the latest situation developments in handling viruses, as well as other information needed by the public to maintain their health. Thus, the public can participate in handling the spread of the virus with the help of official, accountable information from the government. Meanwhile, individuals and communities can help reduce the spread by disciplining themselves in social distancing, personal hygiene, and participating in self-isolation. Although the use of big data opens up



greater opportunities for evidence-based policymaking, several limitations need to be addressed. Clarity of data sources and the overall data production process are crucial issues regarding the validity of the quality of the data and the knowledge produced. Such a combination approach is needed as the knowledge of the virus is still learned and developed.

Acknowledgment

This article is part of research funded by the Ministry of Research and Technology/National Research and Innovation Agency as well as the Republic of Indonesia through the Competitive Basic Higher Education Research Scheme (2019-2020). The author would like to thank the Ministry for providing the grant and the Centre for Decentralisation and Participatory Development Research, Universitas Padjadjaran for supporting this research. A preliminary draft of this article was already presented at the 16th Symposium Kebudayaan Indonesia-Malaysia at Universiti Kebangsaan Malaysia in May 2020. This article is a revised and enlarged version of the previous draft presented in that seminar.

REFERENCES

- Ali, K., Zain-ul-abdin, K., Li, C., Johns, L., Ali, A. A., & Carcioppolo, N. (2019). Viruses Going Viral: Impact of Fear-Arousing Sensationalist Social Media Messages on User Engagement. *Science Communication*, 41(3), 314–338. <https://doi.org/10.1177/1075547019846124>
- Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The Lancet*, 395(10228), 931-934. [https://doi.org/10.1016/S0140-6736\(20\)30567-5](https://doi.org/10.1016/S0140-6736(20)30567-5)
- Callaghan, S. (2020). COVID-19 Is a data science issue. *Patterns*, 1(2), 105-110. <https://doi.org/10.1016/j.patter.2020.100022>
- Chen, S., Yang, J., Yang, W., Wang, C., & Bärnighausen, T. (2020). COVID-19 control in China during mass population movements at New Year. *The Lancet*, 395(10226), 764–766. [https://doi.org/10.1016/S0140-6736\(20\)30421-9](https://doi.org/10.1016/S0140-6736(20)30421-9)
- Choi, D. H., Yoo, W., Noh, G. Y., & Park, K. (2017). The impact of social media on risk perceptions during the MERS outbreak in South Korea. *Computers in Human Behavior*, 72, 422–431. <https://doi.org/10.1016/j.chb.2017.03.004>
- Chong, M., & Choy, M. (2018). The social amplification of haze-related risks on the internet. *Health Communication*, 33(1), 14–21. <https://doi.org/10.1080/10410236.2016.1242031>
- Huang, Y. (2004). The sars epidemic and its aftermath in China: A political perspective. <https://www.ncbi.nlm.nih.gov/books/NBK92479/>
- Jang, K., & Baek, Y. M. (2019). When information from public health officials is untrustworthy: The use of online news, interpersonal networks, and social media during the MERS outbreak in South Korea. *Health Communication*, 34(9), 991–998. <https://doi.org/10.1080/10410236.2018.1449552>
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., & Jebara, T. (2009). Computational social science. *Science*, 323(5915), 721-723.
- Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton: Mifflin Harcourt.
- Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, 76(6), 928–937.
- Oh, S. H., Lee, S. Y., & Han, C. (2020). The effects of social media use on preventive behaviors during infectious disease outbreaks: The mediating role of self-relevant

- emotions and public risk perception. *Health Communication*, 00(00), 1–10. <https://doi.org/10.1080/10410236.2020.1724639>
- Scharmer, O. (2020). Eight emerging lessons: From coronavirus to climate action. <https://medium.com/presencing-institute-blog/eight-emerging-lessons-from-coronavirus-to-climate-action-683c39c10e8b>
- Song, J., Song, T. M., Seo, D. C., Jin, D. L., & Kim, J. S. (2017). Social big data analysis of information spread and perceived infection risk during the 2015 Middle East respiratory syndrome outbreak in South Korea. *Cyberpsychology, Behavior, and Social Networking*, 20(1), 22–29. <https://doi.org/10.1089/cyber.2016.0126>
- Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to COVID-19 in Taiwan. *Jama*, 1–2. <https://doi.org/10.1001/jama.2020.3151>
- Weible, C. M., Nohrstedt, D., Cairney, P., Carter, D. P., Crow, D. A., Durnová, A. P., Heikkilä, T., Ingold, K., McConnell, A., & Stone, D. (2020). COVID-19 and the policy sciences: initial reactions and perspectives. *Policy Sciences*, 53(2), 225–241. <https://doi.org/10.1007/s11077-020-09381-4>
- Yoo, W., Paek, H. J., & Hove, T. (2020). Differential effects of content-oriented versus user-oriented social media on risk perceptions and behavioral intentions. *Health Communication*, 35(1), 99–109. <https://doi.org/10.1080/10410236.2018.1545169>
- You, M., Joo, J., Park, E., Noh, G. Y., & Ju, Y. (2017). Emerging infectious disease content in newspaper editorials: Public health concern or leadership issue? *Science Communication*, 39(3), 313–337. <https://doi.org/10.1177/1075547017705392>
- Zhou, T., Liu, Q., Yang, Z., Liao, J., Yang, K., Bai, W., Lu, X., & Zhang, W. (2020). Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *Journal of Evidence-Based Medicine*, 13(1), 3–7. <https://doi.org/10.1111/jebm.12376>