



Social Media Big Data: The Good, The Bad, and the Ugly (Un)truths

Alton M. K. Chew^{1,2} and Dinesh Visva Gunasekeran^{1*}

¹Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore, Singapore, ²UCL Medical School, University College London (UCL), London, United Kingdom

Keywords: coronavirus—COVID-19, public health, big data, social media, health promotion

INTRODUCTION

Social media has been a defining component of life in the 21st century, monetising peer-to-peer sharing of information. This has led to the formation of powerful platforms leveraging artificial intelligence (AI) to effectively commoditise individual attention, with the average person spending over 2 h a day on social media (Statista, 2020). The ubiquity of these platforms underscores the need to better understand social media in the context of public health, in particular its potential risks and benefits. This will facilitate comprehensive assessment of its impact on healthcare services, development of mitigating strategies for its drawbacks, and identification of potential opportunities to leverage its strengths for the good of population health.

These sentiments were recently echoed by the World Health Organisation (WHO) in the context of the coronavirus disease 2019 (COVID-19) pandemic, calling upon member states to address misinformation through risk communication and timely dissemination of accurate information as a key pillar of national public health responses (World Health Organisation, 2020a). However, our collective experience during this pandemic highlights the limitations of our existing knowledge and approaches, as many countries continue to experience the spiralling social impact of misinformation related to COVID-19. This manuscript provides an overview of both the negative and positive public health impact of social media that has come to light in the course of COVID-19 with an emphasis on rampant disinformation during COVID-19, and concludes with potential future directions for research in this emerging area of public health.

Viral Infodemic in Social Media outpaces the Pandemic in the Streets—the (Un)truths

The COVID-19 outbreak that originated from Wuhan city, China, has already claimed over 1 million lives and infected over 40 million individuals since it was first reported in December 2019. With the initial spread of COVID-19 progressing to pandemic status, the WHO also recognized the parallel problem of widespread anxiety and emotional sharing of information through digital platforms. This imposed an additional strain on humanity's coordinated attempts to eradicate COVID-19, hindering public health risk communication (The Lancet Infectious Diseases, 2020). Experts have studied big data in online search and social media with many reports highlighting surges in online misinformation in each region even before the uptick in confirmed cases (Hong et al., 2020). Poor quality online sources have drowned out official advisories with fake and potentially harmful information (Cuan-Baltazar et al., 2020). This presents major health concerns given the public's difficulty in differentiating reliable from unreliable sources of medical information, even among individuals with good baseline health literacy (Huhta et al., 2018).

Further aggravating this situation, the excessive amounts of misinformation interspersed with official sources of information shared across social media platforms has placed a heavy burden on the public to discern true from false information. The WHO has since identified this phenomenon as an “infodemic”, making calls for novel solutions to moderate the flow of accessible information and stem medical disinformation (World Health Organisation, 2020b). A recent survey of social media users in Kurdistan, Iraq further emphasised the negative impact of this problem, whereby the

OPEN ACCESS

Edited by:

Kok-Leong Ong,
La Trobe University, Australia

Reviewed by:

Yee Ling Boo,
RMIT University, Australia
Yanchang Zhao,
Commonwealth Scientific and
Industrial Research Organisation
(CSIRO), Australia

*Correspondence:

Dinesh Visva Gunasekeran
mdcdvg@nus.edu.sg

Specialty section:

This article was submitted to
Medicine and Public Health,
a section of the journal
Frontiers in Big Data

Received: 30 October 2020

Accepted: 25 January 2021

Published: 01 June 2021

Citation:

Chew AMK and Gunasekeran DV
(2021) Social Media Big Data: The
Good, The Bad, and the Ugly
(Un)truths.
Front. Big Data 4:623794.
doi: 10.3389/fdata.2021.623794

researchers have reported how health-related content featured prominently in social media during COVID-19 with fake news accounting for over a quarter of “fear-inducing” content (Ahmad and Murad, 2020). This study was not appropriately designed to evaluate the relationship between social media content and fear, which was also inflated by selection bias. However, it does highlight the pervasiveness of this problem particularly considering the prominence of social media in modern life. This begets the need for solutions to adjudicate content in social media, and provide a check and balance to the spread of misinformation via these viral platforms.

Social media big data and public health during COVID-19—the Good and the Bad

The extensive usage of social media naturally includes large amounts of information regarding users’ unadulterated feelings and thoughts documented in a publicly visible platform. The collection of this information provides ‘naturally-occurring’ and publicly-visible big data, that can potentially be applied to improve Public Health responses during Natural disasters and emergencies such as Pandemics. Several such applications of big data have been described during the COVID-19 pandemic, including regular *COVID-19 Snapshot Monitoring (COSMO)* initiated in Germany to improve surveillance of misinformation and inform the development of policies and communication messages (Betsch et al., 2020).

Moreover, researchers from China demonstrated the use of big data from social media platform WeChat to identify trends in communication and searches for key words related to these topics. Using these “infodemiology” techniques, which analyse online user generated content (UGC) to inform public health applications, researchers could correlate digital big data to the progression of the pandemic unfolding in real-time (Lu and Zhang, 2020). By harnessing this readily available online information, researches have further demonstrated “infodemiology” techniques for applications such as planning of pandemic responses, optimising the flow of resources, and identifying growing themes of misinformation and/or public concerns real-time to develop targeted public health strategies and communications (Wong et al., 2020).

Emerging Research in Social Media Big Data for Public Health Interventions

The reports described in the previous section have illustrated “infodemiology” techniques that leverage big data from social media for timely insights that could inform the development of critical public health responses. Other researchers have further described the amalgamation of multimodal data from social media complemented by other sources such as traditional news media and online behaviour/market research agency platforms to inform the development of evidence-based public health interventions. These capabilities have been demonstrated for the evaluation of the public’s compliance to public health measures as well as the evaluation of national responses to help control the pandemic in China, using data from social media such as Weibo and Tik Tok, the People’s Daily major Chinese newspaper, and online market research agency platforms such as Mob-Tech research institute (Hua and Shaw, 2020).

Researchers have further demonstrated the potential utility of new techniques such as Online Ecological Recognition (OER) that combine big data with other emerging technology domains like artificial intelligence (AI) to develop predictive models (Li S. et al., 2020). This facilitated additional applications beyond the surveillance of information, to evaluating the mental health impact of the pandemic itself. The study demonstrated that negative emotions and sensitivity to social risks increased while positive emotions and life satisfaction plummeted (Li S. et al., 2020). The results from this particular study corroborated with the Behavioural Immune System (BIS) theory that people tend towards negative emotions when threatened by disease, whereby the spike in negative emotion was heightened during COVID-19 due to the infodemic.

However, another study of 17,865 Weibo users in China highlighted a silver lining regarding the impact of social media during this pandemic, whereby initial negative emotions (after COVID-19 was reported widely) were subsequently balanced by positive emotions as users leveraged social media platforms for peer-support, with trending topics such as “faith” and “blessing”. (Li S. et al., 2020). These terms reflect greater group cohesiveness given the threat to greater public, and these findings were further replicated in Lombardy based on data from Italy (Su et al., 2020). Notably, the increased group cohesion occurred in tandem with more monetary and supply donations to regions of need and key organisations including the Hubei Red cross (Li S. et al., 2020). It is thus evident that social media can be leveraged for positive impact, by helping to connect individuals during a crisis and improve individual alignment for common good. This has additional implications for other aspects of medication, including the use of these platforms for health promotion and raising awareness about critical health-related problems (Horrell et al., 2019). Through further investigation and refinement of these methods, public health organisations will be able to optimise response strategies in real-time by extrapolating trends in transmission, communication content, information flow, and population sentiment.

What Lies Ahead

Our article has highlighted the potential impact of social media big data to be a double-edged sword. Presently, the negative impact has gained much visibility and criticism, due to limited mechanisms for differentiating reliable information from misinformation, and mitigating the risks of the latter. Fortunately, increasing coordination between social media platform providers, non-governmental organisations, and governments have given rise to promising collaborations such as the “Share verified” initiative led by the United Nations (UN) to build a freely-accessible resource of reliable health content and front-end flags to redirect individuals to reliable sources in order to address misinformation. Ultimately, long-term solutions may require new legislation to govern the creation and dissemination of misinformation online. Regulations have been effectively applied for other public health challenges, such as tobacco advertising regulations to reduce population exposure to marketing and cues to smoke (Henriksen, 2012).

However, in the case of online misinformation, the enforcement of such legislation will be significantly more complexed, given the scale of individuals as potential sources as opposed to corporations that are stakeholders within the tobacco industry. This will likely require methods such as COSMO for big data surveillance, with incorporation of AI analytics of the social media big data to scale up enforcement. This also begets consideration of developing alternatives and complements to social media as sources of reliable health information hosting and exchange, with several recently launched in response to COVID-19 misinformation. Online health communities (OHCs) have drawn increasing interest in the domain of virtual social networks due to their potential to amplify positive impact such as peer-support and quality data as a source of health evidence (Smith et al., 2017; Audrain-Pontevia et al., 2019), as well as mitigate against negative impact through policies against the promotion of inaccurate information, as well as configurations that involve medical practitioners in moderation and content generation (Eysenbach, 2000; AskDr, 2020).

However, even with these measures in place, studies have highlighted the potential for lapses to occur that can be difficult to detect (Huh et al., 2016). Therefore, there is a growing need for OHCs that leverage the strengths of social media platforms with additional embodiments that mitigate against its weaknesses. These may be configurations that empower verified medical experts with digital tools to moderate the content and flow of information. These applications of OHCs for patients with chronic pain and mental health disorders that are likely to progress and increase in prevalence during COVID-19 have

been described in earlier reviews led by relevant specialist (Chew et al., 2020; Li L. W. et al., 2020). These digital platforms represent potential areas for future research and cross-disciplinary collaborations between technology partners, clinicians and regulators to enhance public health responses.

CONCLUSION

Social media has likely been a significant contributor to the dissemination of misinformation and fear in this pandemic, particularly given the lack of information arbitration and controls of viral false information (Li L. W. et al., 2020). However, several applications of social media big data about health-related content for public health communication measures have been discussed in this study. These studies shed light on the potential positive impact and applications of social media in public health. Much is still unknown, and it would be impossible to weigh the benefits and drawbacks of social media in healthcare at this stage. The only certainty is that social media is likely to remain a prominent feature of modern life, and more research is needed to better understand this domain to amplify its positive impact and to mitigate against the negative.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

- Ahmad, A. R., and Murad, H. R. (2020). The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. *J. Med. Internet Res.* 22 (5), e19556. doi:10.2196/19556
- AskDr (2020). Making Reliable Health Information Accessible to All. Available at: <https://www.askdr.co/> (Accessed June 1, 2020).
- Audrain-Pontevia, A. F., Menvielle, L., and Ertz, M. (2019). Effects of Three Antecedents of Patient Compliance for Users of Peer-to-Peer Online Health Communities: Cross-Sectional Study. *J. Med. Internet Res.* 21 (11), e14006. doi:10.2196/14006
- Betsch, C., Wieler, L. H., Habersaat, K., and group, C. (2020). Monitoring Behavioural Insights Related to COVID-19. *Lancet* 395 (10232), 1255–1256. doi:10.1016/S0140-6736(20)30729-7
- Chew, A. M. K., Ong, R., Lei, H. H., Rajendram, M., K V, G., Verma, S. K., et al. (2020). Digital Health Solutions for Mental Health Disorders During COVID-19. *Front. Psychiatry* 11, 582007. doi:10.3389/fpsy.2020.582007
- Cuan-Baltazar, J. Y., Muñoz-Perez, M. J., Robledo-Vega, C., Pérez-Zepeda, M. F., and Soto-Vega, E. (2020). Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health Surveill.* 6 (2), e18444. doi:10.2196/18444
- Eysenbach, G. (2000). Towards Ethical Guidelines for E-Health: JMIR Theme Issue on eHealth Ethics. *J. Med. Internet Res.* 2 (1), E7. doi:10.2196/jmir.2.1.e7
- Henriksen, L. (2012). Comprehensive Tobacco Marketing Restrictions: Promotion, Packaging, Price and Place. *Tob. Control.* 21 (2), 147–153. doi:10.1136/tobaccocontrol-2011-050416
- Hong, Y. R., Lawrence, J., Williams, D., and Mainous, A. (2020). Population-Level Interest and Telehealth Capacity of US Hospitals in Response to COVID-19: Cross-Sectional Analysis of Google Search and National Hospital Survey Data. *JMIR Public Health Surveill.* 6 (2), e18961. doi:10.2196/18961
- Horrell, L. N., Lazard, A. J., Bhowmick, A., Hayes, S., Mees, S., and Valle, C. G. (2019). Attracting Users to Online Health Communities: Analysis of LungCancer.Net's Facebook Advertisement Campaign Data. *J. Med. Internet Res.* 21 (11), e14421. doi:10.2196/14421
- Hua, J., and Shaw, R. (2020). Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. *Int. J. Environ. Res. Public Health* 17 (7). doi:10.3390/ijerph17072309
- Huh, J., Marmor, R., and Jiang, X. (2016). Lessons Learned for Online Health Community Moderator Roles: A Mixed-Methods Study of Moderators Resigning from WebMD Communities. *J. Med. Internet Res.* 18 (9), e247. doi:10.2196/jmir.6331
- Huhta, A. M., Hirvonen, N., and Huotari, M. L. (2018). Health Literacy in Web-Based Health Information Environments: Systematic Review of Concepts, Definitions, and Operationalization for Measurement. *J. Med. Internet Res.* 20 (12), e10273. doi:10.2196/10273
- Li, L. W., Chew, A. M. K., and Gunasekeran, D. V. (2020). Digital Health for Patients with Chronic Pain during the COVID-19 Pandemic. *Br. J. Anaesth.* 125 (5), 657–660. doi:10.1016/j.bja.2020.08.003
- Li, S., Wang, Y., Xue, J., Zhao, N., and Zhu, T. (2020). The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. *Int. J. Environ. Res. Public Health* 17 (6), 2032. doi:10.3390/ijerph17062032
- Lu, Y., and Zhang, L. (2020). Social Media WeChat Infers the Development Trend of COVID-19. *J. Infect.* 81 (1), e82–e83. doi:10.1016/j.jinf.2020.03.050
- Smith, H., Bulbul, A., and Jones, C. J. (2017). Can Online Discussion Sites Generate Quality Data for Research Purposes?. *Front. Public Health* 5, 156. doi:10.3389/fpubh.2017.00156
- Statista (2020). Daily time spent on social networking by internet users worldwide from 2012 to 2019. Available at: <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/> (Accessed October 1, 2020).

- Su, Y., Xue, J., Liu, X., Wu, P., Chen, J., Chen, C., et al. (2020). Examining the Impact of COVID-19 Lockdown in Wuhan and Lombardy: A Psycholinguistic Analysis on Weibo and Twitter. *Int. J. Environ. Res. Public Health* 17 (12), 4552. doi:10.3390/ijerph17124552
- The Lancet Infectious Diseases (2020). The COVID-19 Infodemic. *Lancet Infect. Dis.* 20 (8), 875. doi:10.1016/S1473-3099(20)30565-X
- Wong, M. Y. Z., Gunasekaran, D. V., Nusinovi, S., Sabanayagam, C., Yeo, K. K., Cheng, C.-Y., et al. (2021). Telehealth Demand Trends During the COVID-19 Pandemic in the Top 50 Most Affected Countries: Infodemiological Evaluation. *JMIR Public Health Surveill* 7 (2), e24445.
- World Health Organisation (2020a). COVID-19 Strategic Preparedness and Response: Operational Planning Guidance to Support Country Preparedness and Response. Available at: <https://www.who.int/publications/i/item/draft-operational-planning-guidance-for-un-country-teams> (Accessed July 11, 2020).
- World Health Organisation (2020b). Managing the COVID-19 Infodemic: Promoting Healthy Behaviours And Mitigating the Harm from Misinformation And Disinformation. Available at: <https://www.who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation> (Accessed September 29, 2020).

Conflict of Interest: Author DG reports directly relevant equity investment in AskDr, an online health community (OHC). 22 Author DG reports other equity investments in Doctorbell (acquired by MaNaDr, Mobile Health), VISRE, and Shyfts that are not relevant to the content of this manuscript. Author DG also reports appointment as Physician Leader (Telemedicine) at Raffles Medical Group, and serving as an advisor to university-affiliated technology developers and start-up companies involved in the development of patient engagement systems in Southeast Asia.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chew and Gunasekeran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.