Electronic Journal of General Medicine
2021, 18(3), em284
e-ISSN: 2516-3507
https://www.ejgm.co.uk/

Review Article

MODESTUM
OPEN ACCESS

The Rise of the Big Data Doctor: Recent Advances and Warnings in Digital Public Health Technology

Victoria Dawe [1]*, John Hardie [2]

[1] Portsmouth Hospital University NHS Trust, UK
[2] Frimley Health NHS Foundation Trust, UK
***Corresponding Author:** victoria.dawe@nhs.net

Citation: Dawe V, Hardie J. The Rise of the Big Data Doctor: Recent Advances and Warnings in Digital Public Health Technology. Electron J Gen Med. 2021;18(3):em284. https://doi.org/10.29333/ejgm/9763

ARTICLE INFO

Received: 31 Oct. 2020

Accepted: 16 Jan. 2021

ABSTRACT

The continuing expansion of social media and big data presents positive opportunities for healthcare research, public health surveillance, and dissemination of health information. However, caution must be used as online data can be harvested and sold without permission from the public. This includes social manipulation, such as the targeting of voters during elections, or pre-determining the likelihood of an individual engaging in high-risk behaviours with associated social, employment, or insurance consequences.

The first independent COVID-19 contact tracing application (app) in the UK was unsuccessful. Subsequently, the UK Government has sought to develop apps with the help of large technology firms. These companies generate profit through surveillance capitalism - the mining of human digital activity for profit. Although large technology companies may have the resources available to assist with public health endeavours, their primary obligation to the shareholder should not be overlooked.

This narrative review explores recent positive advances in the use of data derived from social media. It also highlights concerns about the potential to abuse this technology for outcomes that are not in the best interests of our patients or general population.

Keywords: big data, social media, COVID-19

INTRODUCTION

Social interactions are not a new phenomenon; they have long been pivotal to human evolution and society. The move from physical to digital interactions became recorded as social media. Social Distancing during the COVID-19 pandemic has accelerated this process further, leading to increased digitalisation of social interactions and utilisation of digital resources to reduce spread of COVID-19.

Global technology companies have taken a keen interest in this digitalisation during the pandemic. It is estimated that global digital-health revenues, including from telemedicine, online pharmacies and wearable devices, will rise from $350bn in 2019 to $600bn in 2024 [1].

Healthcare is subject to heavy regulation and guidelines from government organisations, such as NHS England and the General Medical Council (GMC) in the United Kingdom (UK). Patient information is often siloed in systems from which data cannot be easily extracted, in order to preserve patient confidentiality. This is at odds with online data from social media posts and searches, which can be stored indefinitely and viewed externally. As the use of social media for healthcare purposes is increased, caution must be taken to ensure appropriate use of data and to maintain confidence and confidentiality.

This review discusses recent positive advances in the use of data derived from social media, and highlights concerns about the potential to abuse the technology for results that are not in the best interests of our patients or the public.

BIG DATA AND EPIDEMIOLOGY SURVEILLANCE

20% of the world population are active users of Facebook or Twitter, creating large volumes of data points, termed “big data” [2]. This data is stored indefinitely, resulting in an exponentially expanding, inexhaustible resource. Health data can be sourced from various user digital interactions; from individual social media posts declaring symptoms or signs, to geotagging a hospital as a location, or wearable technology data. Implicit health information may be mined from ‘exhaust’ processes, such as a change in grammar or punctuation used during internet searches, indicating a change in user mental state [3].

Big Data interpretation has a range of successes and areas for improvement across all specialities. Studies investigating the prevalence of Google searches for flu symptoms compared with the prevalence reported by the Centers for Disease Control and Prevention (CDC) showed 85% accuracy [4]. The benefits of using social media data over CDC collected data is that social media gives real time results while the CDC waits for laboratory test results to analyse a trend, before health officials mount an appropriate response**.** External events, however, can impact social data. One celebrity posting about flu caused an increase in searches that did not correlate with an increase in patients with flu according to the CDC [5]. The systems used to interpret Big Data need refining to be able to detect these anomalous increases in searches while more studies are needed to establish correlation between social media reports of flu and laboratory confirmed influenza. Poor correlation here could result in over treatment of those with milder cold symptoms or an inaccurate public health alert [5].

Reduction in the spread of communicable diseases and appropriate medical services could be directed to the correct locations earlier by using geotagged social media posts in combination with anonymised phone data. Bengtsson et al. used phone data to retrospectively map the movement of people following the January 2010 Haiti earthquake, showing that people left Port-au-Prince for three major areas, contrasting the governments real-time estimates [6]. Utilising this data in conjunction with phone data could allow accurate and timely deployment of medical personnel after a crisis.

Another study aimed to target healthcare resources to focal areas of cholera outbreaks. The investigation proposed that a cholera outbreak in Rwanda would result in a reduction in distance walked by unwell individuals, and therefore reduced phone movement [5]. There was a change to phone movement, however, this was two weeks preceding the outbreak of cholera (a waterborne disease), when unexpected floods occurred [5]. This emphasises the importance of not just analysing data, but also having an understanding of the disease and populations affected.

INDIVIDUALISED HEALTH SURVEILLANCE TOOLS

Many members of the public carry smart phones with health apps pre-installed by the manufacturer. Fitness activity trackers are commonly installed, recording distance walked, altitude climbed, and periods of low physical activity. Others allow manual input of data such as food and drink consumed. Many of these applications allow results to be published on social media [7]. These applications may have positive health benefits, such as those that allow diabetic patients to synchronise phones to glucose meters, reducing the frequency of hypoglycaemic episodes [8]. Healthcare resources could be targeted to those with a significant change to their physiological baseline, suggesting, for example, a diagnosis of COVID-19 [9]. Caution must be used, however, if identifiable data could influence health insurance premiums or stigmatise patients [7].

Health insurance companies in Switzerland, the UK, and South Africa, offer an incentive system reliant on automatic uploads from a fitness activity tracker [7]. The accuracy of the applications is still unknown. There is no clear evidence that an increase in activity recorded by a fitness tracker results in improved health outcomes [7]. Both automatic and inputted information can be intentionally and unintentionally manipulated. Those doing more activities could have other undisclosed habits, for example smoking, that are not monitored by the application, but have a more significant impact on health. Although financial incentives are currently offered on a reward system, it is possible that those unable or unwilling to have personal and behavioural data collected with activity monitoring in the future may be subject to higher upfront insurance premiums.

Public perception of allowing data to be accessed and analysed is one of concern and confidentiality is essential in a good doctor-patient relationship [9]. Care must be taken to ensure trust is not lost with the public over health matters, including those that the public post about on social media.

SOCIAL MEDIA FOR PUBLIC HEALTH INFORMATION

Public health campaigns have long used social media due to its cost-effective method of reaching large audiences not restricted by geography. The addition of interactive tagging features, such as web badges and hashtags, encourages user information engagement and sharing, whilst providing a means for public health workers to track message dissemination. A 2008 Worlds AIDS Day promotion using hashtags increased engagement via Blogs and Twitter [10]. The platform of social media has positively allowed users to destigmatise diseases through awareness and education. Two UK supermarkets sold stigmatising ‘mental patient’ and ‘psycho ward’ Halloween costumes in 2013 [11]. Twitter users responded with mass reposting of information, until the companies apologised and made voluntary donations to mental health charities [11].

Younger generations are increasingly diversifying from obtaining information from traditional online news platforms, such as official news websites [12]. At the start of the COVID-19 pandemic, it was suggested that China was controlling news channels, resulting in social media taking a dominant role in information sharing from China [13]. Social media remained a primary source of information in May 2020 for many in Kurdish Iraq where information from government websites is reported to be eroded as it is disseminated [14]. Social media was also found to have significant positive correlation with spreading panic, on all topics including health and diminishing shop supplies [14]. Furthermore, social media and it’s role in COVID-19 was demonstrated to have a negative psychological impact, especially in the 18-35 age group [14].

Social media has increased the propagation of incorrect information, or even malicious misinformation [5]. The CDC (as an organisation with high credibility) was able to counter misperceptions about the spread of the Zika virus, but 39% of the tweets from the CDC about the Zika virus were replies to users rather than detecting misinformation and providing accurate data [15]. A single user correcting misperceptions about the Zika virus was found to cause no change to misperceptions, while a social source agreeing to a correction from the CDC was shown to backfire and reaffirm misperceptions [15]. Google acknowledged the threat of misinformation during the COVID-19 pandemic and explained that both Google searches and the YouTube homepage would ‘direct users to the WHO or other locally relevant authoritative organizations’ [16]. This highlights the difficult balance for healthcare professionals who not publicise their profession and credentials on social media. While the medical community is unable to monitor incorrect information on a large scale, part of public health promotion and education should centre on dissemination of reliable websites for medical information and engagement on social media.

EXAMPLES OF TECHNOLOGY USE FOR COVID-19 APPS

With the emergence of COVID-19 many authorities are looking for methods to monitor and trace large populations, with some utilising apps. While populations may have concerns about privacy and cyber security, most declare that they are willing to download apps voluntarily to their personal devices in order to contain the pandemic [17]. In China and Singapore, there has been notable success. In contrast, apps have been problematic in Europe and the United States, due to poor interoperability between mobile phone systems, and poor compliance of individuals to isolate when traced [18].

The UK was an early adopter of a centralised government 'NHS Track and Trace' app, with low energy Bluetooth mobile technology being used as the main technology to detect the proximity of other nearby mobile phones [9]. If a user has been in contact with another who subsequently tests positive for COVID-19 the app will alert that user that they have been in a 'high risk' encounter and require isolation [9]. The risk score is calculated by looking at a combination of time and proximity to someone who could have been asymptomatic for COVID-19 at the time. Early trials conducted on the Isle of Wight showed poor effectiveness of the Bluetooth proximity technology [19].

Singapore was the first country to release an app, 'TraceTogether', then further developed the app to be used as an electronic token worn as a lanyard or carried [20]. This appeals to the older generation as it does not require a smart phone to run, and users with privacy concerns do not need to be concerned about other applications [20].

The UK's second app was developed by international technology companies Google and Apple. The has user inputs as well as the automated Bluetooth function [9]. Users directly input information to the app by scanning QR Codes on arrival to specific venues as a method of 'checking-in' to each venue visited, so that if an outbreak is linked to a venue the owners and staff can be informed so more members of the public can be contacted by the venues own tracing system [21]. The app does not require GPS or the input of personal contact or identification information, but it does require Bluetooth and camera access [9].

While the UK and Singapore have adopted national approaches to apps using Bluetooth, the app is dependant on state in the USA, with some states not yet having one available [22]. This has created cross-state incompatibility, with the consequence that anyone traveling across state boundaries is at risk. The Association of Public Health Laboratories have released a national server so the states that use the tool will still have effective apps if the phone is in another state that is also utilising the tool [22].

China, Singapore and South Korea had early successes using surveillance applications, regulated on a government level. In Europe, the wide General Data Protection Regulation (GDPR) allowed for the rapid rollout of surveillance technology in an emergency, compatible with basic rights [23]. In contrast, and counterintuitively, the multiple narrow regulations from the insurance and healthcare sectors in the US made the use of surveillance applications more challenging [23].

CONCLUSION

Social media's expansion and adaptation allows many opportunities for research, public health surveillance, and dissemination of health information. It is yet to be seen how the second Test and Trace app performs in the UK, but its success will be reliant on a large proportion of the public downloading and using the app. Social media and the use of big data must be approached with caution. More research is required to establish how a patient can confirm the source or reliability of health information, and ensure that confidentiality and trust with the public is maintained.

Author contributions: All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Declaration of interest: No conflict of interest is declared by authors.

REFERENCES

1. Maida A. The dawn of Digital medicine. 2020. Available at: https://www.economist.com/business/2020/12/02/the-dawn-of-digital-medicine (Accessed: 28 December 2020).
2. Asch DA, Rader DJ, Merchant RM. Mining the social mediome. Trends in molecular medicine. 2015;21(9):528-9. https://doi.org/10.1016/j.molmed.2015.06.004 PMid:26341614 PMCid:PMC4662876
3. Zuboff S. The age of Surveillance Capitalism: The fight for a human future at the new frontier of power. London, UK: Profile Books, 2019.
4. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PloS one. 2013;8(12):e83672. https://doi.org/10.1371/journal.pone.0083672 PMid:24349542 PMCid:PMC3857320
5. Schmidt CW. Using social media to predict and track disease outbreaks. Environmental health perspectives. 2012;120(1). https://doi.org/10.1289/ehp.120-a30
6. Bengtsson L, Lu X, Thorson A, Garfield R, Von Schreeb J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med. 2011;8(8):e1001083. https://doi.org/10.1371/journal.pmed.1001083 PMid:21918643 PMCid:PMC3168873
7. Martani A, Shaw D, Elger BS. Stay fit or get bit – ethical issues in sharing health data with insurers' apps. Swiss Medical Weekly. 2019;149:w20089. https://doi.org/10.4414/smw.2019.20089 PMid:31256413
8. Feuerstein-Simon C, Bzdick S, Padmanabhuni A, Bains P, Roe C, Weinstock RS. Use of a Smartphone Application to Reduce Hypoglycemia in Type 1 Diabetes: A Pilot Study. Journal of Diabetes Science and Technology. 2018;12(6):1192-9. https://doi.org/10.1177/1932296817749859 PMid:29291641 PMCid:PMC6232731
9. Burgess M. Everything you need to know about the NHS test, track and trace app. 2020. Available at: https://www.wired.co.uk/article/nhs-covid-19-tracking-app-contact-tracing (Accessed: 28 December 2020).

10. Anderson J, Gomez M. From Flickr to a blogging call to action: User generated content lessons learned from AIDS.gov. 2009. Available at: https://www.slideshare.net/aidsgov/aidsgovs-usergenerated-content-presentation-for-cdcs-national-conference-on-health-communication-marketing-and-media-2009 (Accessed: 28 December 2020).
11. BBC News. Asda and Tesco withdraw Halloween patient outfits. 2013. Available at: https://www.bbc.co.uk/news/uk-24278768 (Accessed: 28 December 2020).
12. Kalogeropoulos A. How Younger Generations Consume News Differently. 2019. Available at: http://www.digitalnewsreport.org/survey/2019/how-younger-generations-consume-news-differently/ (Accessed: 28 December 2020).
13. BBC News. Coronavirus: What did China do about early outbreak? 2020. Available at: https://www.bbc.co.uk/news/world-52573137 (Accessed: 28 December 2020).
14. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. Journal of Medical Internet Research. 2020;22(5):e19556. https://doi.org/10.2196/19556 PMid:32369026 PMCid:PMC7238863
15. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. Science Communication. 2017;39(5):621-45. https://doi.org/10.1177/1075547017731776
16. Pichai S. Coronavirus: How we're helping. 2020. Available at: https://blog.google/inside-google/company-announcements/coronavirus-covid19-response/ (Accessed: 28 December 2020).
17. Altmann S, Milsom L, Zillessen H, et al.. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth. 2020;8(8):e19857. https://doi.org/10.2196/19857 PMid:32759102 PMCid:PMC7458659
18. Bachtiger P, Adamson A, Quint JK, Peters NS. Belief of having had unconfirmed Covid-19 infection reduces willingness to participate in app-based contact tracing. NPJ Digital Medicine. 2020;3:146. https://doi.org/10.1038/s41746-020-00357-5 PMid:33299071 PMCid:PMC7648058
19. Burgess M. Why the NHS COVID-19 contact app failed. Available at: https://www.wired.co.uk/article/nhs-tracing-app-scrapped-apple-google-uk (Accessed: 28 December 2020).
20. BBC News. Singapore distributes Covid contact-tracing tokens. 2020. Available at: https://www.bbc.co.uk/news/business-54143015 (Accessed: 28 December 2020).
21. Kelion L, Cellan-Jones R. NHS Covid-19 app: One million downloads of contact tracer for England and Wales. 2020. Available at: https://www.bbc.co.uk/news/technology-54270334 (Accessed: 28 December 2020).
22. Morrison S. Americans are one step closer to a national contact tracing app for Covid-19. 2020. Available from: https://www.vox.com/recode/2020/10/2/21497729/covid-coronavirus-contact-tracing-app-apple-google-exposure-notification (Accessed: 28 December 2020).
23. Bradford L, Aboy M, Liddell K. COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes. J Law Biosci. 2020;7(1):lsaa034. https://doi.org/10.1093/jlb/lsaa034 PMid:32728470 PMCid:PMC7313893