



Commentary

Fighting pandemics with digital epidemiology

Sasu Tarkoma^{a,*}, Suliman Alghnam^b, Michael D. Howell^c^a University of Helsinki, Pietari Kalmin katu 5, 00014, Finland^b King Abdullah International Medical Research Center (KAIMRC), King Saud Bin Abdulaziz University for Health Sciences (KSAU-HS), Al-Sheikh Jaber Al-Sabah St., 11426 Riyadh, Saudi Arabia^c Google Health, Palo Alto, CA, USA

ARTICLE INFO

Article History:

Received 31 July 2020

Revised 31 July 2020

Accepted 31 July 2020

Available online 25 August 2020

Digital epidemiologists conduct traditional epidemiological studies and health-related research using new data sources and digital methods from data collection to analysis [1,2]. According to Salathé, digital epidemiology is epidemiology building on digital data and tools, but a narrower definition defines it as epidemiology building on data generated and obtained with a primary goal other than conducting epidemiological studies [1]. Digital epidemiology provides insights into health and disease determinants in human populations by building on diverse digital data sources. Infectious diseases already account for > 50% of digital epidemiology studies [2], but in the current crisis the rapid understanding of disease spread, risk factors, and intervention impact at the population scale has never been more important to mitigate health and economic consequences.

Digital epidemiology can contribute to global health security through syndromic surveillance, public health surveillance, and early pandemic detection. There is, however, a need for open, accessible data and improved computing capability to bridge gaps in producing knowledge and making knowledge-based decisions. In addition to the availability and accessibility of data and digital tools in pandemic responses, it is important to improve health equity by promoting access to connectivity and digital health literacy skills and to consider the interaction between digital health technologies and society, culture, and the economy.

Digital systems have a potentially critical role in the early pandemic detection. The Program for Monitoring Emerging Diseases (ProMED), the Global Public Health Intelligence Network (GPHIN), and HealthMap have pioneered digital approaches to epidemic intelligence. Epidemic intelligence from governmental and public health agency sources is complemented with other data sources such as

from mobile phones, call centers, sensors, social media, and search engines. The WHO Alert and Response operations unit reported that > 60% of initial disease outbreak reports originate from informal sources [3]. Over a decade ago, Google Flu Trends pioneered the use of large-scale technological information for public health, first using search queries to track influenza-like illnesses [4]. However, meaningful limitations in predictive capability remain, maintaining model calibration over time is challenging, and some of the underlying data may not be publicly available [4]. More recently, other aggregated data for applications such as mobility pattern analysis have found use in public health [5]. Twitter has also become a frequently employed, publicly available data source [2].

Digital epidemiology and digital tools have had a profound role in understanding and mitigating the COVID-19 pandemic through analysis of diverse digital data sources such as smartphone, health register, and environmental monitoring data. Aggregate and anonymized smartphone data have been extensively used to study the pandemic and support decision-making; however, the pandemic's global scale requires coordinated regional, national, and global efforts in sharing, combining, and privacy-protecting data [5].

The WHO has published interim guidance for public health programs and governments regarding the ethical and appropriate use of digital proximity tracking for COVID-19 contact tracing. The COVID-19 Exposure Notifications API from Apple and Google shows how opt-in privacy-preserving digital contact tracing can be deployed ubiquitously, while the COVID Symptom Study shows how a smartphone application can be used to study COVID-19 symptoms and predict disease hotspots 5–7 days in advance [6]. MIT's open source Private Kit: Safe Paths is an example of an anonymous contact tracing platform that supports hotspot detection [7].

While the digital epidemiology toolkit is evolving rapidly, significant challenges must be addressed pertaining to data privacy, availability, and analysis. A lot of useful data are private and require data protection. State-of-the-art privacy solutions include data aggregation, data anonymization, differential privacy (DP), federated learning, and synthetic data generation techniques. Advances that might help in this regard include DP-based methods and decentralized data processing [7]. Privacy-preserving and decentralized machine learning (ML) have also become active research areas that can contribute to a paradigm shift in digital epidemiology. For example, federated learning is an ML technique that enables distributed model creation across multiple

* Corresponding author.

E-mail address: sasu.tarkoma@helsinki.fi (S. Tarkoma).

decentralized devices that store data locally [8]. Synthetic data generation is also a promising paradigm that can offer a high level of privacy; however, there is an inherent tradeoff between data privacy and utility [7]. Privacy-protected data disclosure remains an active research topic.

A vital requirement for any data-driven system is the validation of both the input data and the generated models. Purely data-driven approaches have limited predictive capability and their results may be difficult to interpret in different contexts [4, 9]. However, the combination of data-driven models and domain-specific knowledge would appear to be a promising research avenue, as is ensuring that digital epidemiology and ML promote algorithmic fairness and avoid worsening health disparities.

New legislative measures could also support efficient, ethical, and privacy-preserving combinations of data sets and sources. For example, the EU Member States have developed a technology and data use toolbox to combat and exit from the COVID-19 crisis [10]. The new health data rules in the US within the 21st Century Cures Act are another example of how third parties can work more easily with health data. Within the EU and Nordic Countries, Finland has pioneered a one-stop shop for secondary use of health and social data with a new Act and a new permit authority (Findata) for sharing health-related datasets.

Combining aggregate and privacy-protected diverse data sources such as mobility, health, environmental, and city data is expected to help understand and mitigate the consequences of pandemics. The digital epidemiology toolkit is likely to be supported by advances in ML, privacy-enhancing technologies, data/model validation and explainability, and national and transnational policy measures. Increasing data availability and access combined with advances in open source data processing and analysis pave the way for scalable digital epidemiology supporting world health security.

This article is published as part of G20 Riyadh Global Digital Health Summit (11–12 August 2020) activities. Saudi Arabia hosted this virtual summit to leverage the role of digital health in the fight against current and future pandemics.

Authors' contributions

All authors wrote the manuscript and reviewed and approved the final version of the paper.

Declaration of Interests

None.

References

- [1] Salathé M. Digital epidemiology: what is it, and where is it going. *Life Sci Soc Policy* 2018;14(1). doi: [10.1186/s40504-017-0065-7](https://doi.org/10.1186/s40504-017-0065-7).
- [2] Park HA, Jung H, On J, Park SK, Kang H. Digital epidemiology: use of digital data collected for non-epidemiological purposes in epidemiological studies. *Health Inform Res* 2018;24:253–62. doi: [10.4258/hir.2018.24.4.253](https://doi.org/10.4258/hir.2018.24.4.253).
- [3] World Health Organization. Epidemic intelligence - systematic event detection. Available at: <https://www.who.int/csr/alertresponse/epidemicintelligence/en/> (accessed 25 May 2020).
- [4] Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Global Health* 2017;7:185–9. doi: [10.1016/j.jegh.2017.06.001](https://doi.org/10.1016/j.jegh.2017.06.001).
- [5] Nuria O, Lepri B, Sterly B, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci Adv* 2020;6:eabc0764. doi: [10.1126/sciadv.abc0764](https://doi.org/10.1126/sciadv.abc0764).
- [6] Drew DA, Nguyen LH, Steves CJ, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020;368:1362–7. doi: [10.1126/science.abc0473](https://doi.org/10.1126/science.abc0473).
- [7] Raskar R, Schunemann I, Barbar R, et al. Apps Gone Rogue: maintaining personal privacy in an epidemic. Whitepaper 2020 Available at: <https://arxiv.org/pdf/2003.08567.pdf>. (accessed 25 May 2020).
- [8] McMahan BH, Moore E, Ramage D, Hampson S, Arcas BAY. Communication-efficient learning of deep networks from decentralized data. In: *Proceedings of the 20th international conference on artificial intelligence and statistics*, 54; 2017. p. 1273–82.
- [9] Eckmanns T, Füller H, Roberts SL. Digital epidemiology and global health security: an interdisciplinary conversation. *Life Sci Soc Policy* 2019;15(2).
- [10] European Commission. Recommendation on a common union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data. 2020. Available at: <https://ec.europa.eu/digital-single-market/en/news/coronavirus-recommendation-use-mobile-data-response-pandemic> (accessed 25 May 2020).