Journal of the American Medical Informatics Association, 27(7), 2020, 1139–1141
doi: 10.1093/jamia/ocaa063
Advance Access Publication Date: 29 May 2020
Perspective

Perspective

Following data as it crosses borders during the COVID-19 pandemic

Joseph M. Plasek,[1,†] Chunlei Tang,[1,†] Yangyong Zhu,[2] Yajun Huang,[3] and David W. Bates[1]

[1]Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA, [2]School of Computer Science, Fudan University, Shanghai, China, and [3]School of Economics, Fudan University, Shanghai, China

[†]JMP and CT contributed equally.

Corresponding Author: Chunlei Tang, PhD, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02120, USA; ctang5@partners.org

Received 3 April 2020; Revised 14 April 2020; Editorial Decision 14 April 2020; Accepted 14 April 2020

ABSTRACT

Data change the game in terms of how we respond to pandemics. Global data on disease trajectories and the effectiveness and economic impact of different social distancing measures are essential to facilitate effective local responses to pandemics. COVID-19 data flowing across geographic borders are extremely useful to public health professionals for many purposes such as accelerating the pharmaceutical development pipeline, and for making vital decisions about intensive care unit rooms, where to build temporary hospitals, or where to boost supplies of personal protection equipment, ventilators, or diagnostic tests. Sharing data enables quicker dissemination and validation of pharmaceutical innovations, as well as improved knowledge of what prevention and mitigation measures work. Even if physical borders around the globe are closed, it is crucial that data continues to transparently flow across borders to enable a data economy to thrive, which will promote global public health through global cooperation and solidarity.

Key words: health information interoperability, global health, medical economics, data science

Tracing the origins of new diseases through their growth into global pandemics, such as the 2019 RNA virus strain from the *Coronaviridae* family known as coronavirus disease 2019 (COVID-19), necessitates following the flow of relevant data. Two weeks after the first COVID-19 hospitalization, virologists conducted metagenomic RNA sequencing on a patient and published its molecular blueprint (a dizzying string of more than 34 000 letters) about a month later.[1,2] News reports and other biosurveillance-related data pointed to a cluster of pneumonia cases that an artificial intelligence–driven algorithm called BlueDot identified as being an outbreak on December 31, 2019, a week before global public health officials notified the public.[3] Outbreaks, such as on the Diamond Princess cruise ship, provided valuable information about how the disease is spread and its incubation period.[4,5] Electronic health record systems have augmented their self-reported travel screening questionnaires to help identify patients who have recently visited areas where community spread is present.[6]

Transportation data have been used to simulate the spread of a disease and estimate the effect of local and intercontinental travel restrictions.[7] Air, sea, and land transport networks continue to expand in reach, speed of travel, and volume of passengers carried, providing a vector for infectious disease spread. Simulations suggested cancelling the Spring Festival in China—a period known for crowded buses, trains, planes, and ferries culminating in an estimated 3 billion trips. Prescriptive analytics on outbreak data through algorithms or models can simulate possible outcomes and help answer the question of what we should do when the outbreak constitutes a public health emergency of local or international concern. Decision making about travel advisories and quarantines is done locally, and each locale has its own level of preparedness for an outbreak. Some areas have used innovative approaches; for example, Taiwan integrated its health insurance database with biometric entry and exit data to generate real-time alerts based on travel history and clinical symptoms to aid in case identification and has used these data to decide whom to quarantine and track at the border.[8] The Global Health Security Index encompasses disease prevention, detection, reporting, and response capabilities for each country. Countries with a higher Global Health Security Index score such as Singapore can identify undetected cases through increased epidemiological surveillance and contact tracing, leading to improved accuracy regarding disease prevalence.[9]

There are many potential international data sources for disease surveillance systems[10] to utilize. For example, ProMED Mail, a program of the International Society for Infectious Diseases, is useful for monitoring emerging diseases. Aggregators of local media reports and news feeds, such as DXY, Google News, Baidu News, SOS Info, and Moreover, are useful in identifying new cases early on. Monitoring social media feeds such as Facebook and Twitter as well as trends in Google search terms can be useful to have an idea about what is present in a community or what people are worried about. Tracking animal, agriculture, and environmental health data for potential sources of human disease is possible via the Wildlife Data Integration Network, the Food and Agriculture Organization of the United Nations, and the World Organization for Animal Health.

A common way to disseminate data about infections such as COVID-19 is through data visualizations and simulated disease models. These data products enable the public, policymakers, and scientists to quickly understand the global spread of COVID-19 at the population level, enabling forecasting at the local level. HealthMap[11,12] provides an accurate, continuously updated, and usable visualization tracking the global spread of COVID-19 over time. The John Hopkins dashboard tracks cases in different geographic areas across the globe.[13] The Institute for Health Metrics and Evaluation has developed a real-time data visualization and forecasting tool based on geocoded epidemiological information that includes (when available) symptoms, key dates (date of onset, admission, and confirmation), and travel history.[14]

These examples of data and data product flow across geographic borders are extremely useful to public health professionals for many purposes such as accelerating the pharmaceutical development pipeline; triaging clinician resources to a locale; and making decisions about intensive care unit rooms, where to build temporary hospitals (eg, Boston Hope Medical Center),[15] or where to boost supplies of personal protection equipment, ventilators, or diagnostic tests.[16] Providing data analytics tools for organizations that cannot share data or have limited analytical resources can also be helpful to help with virus response, better-coordinated care, reporting, and organizational operations. The health sectors in advanced economies can help developing countries via cross-border data product sharing, as they did with the Congo in the 2018 Ebola outbreak. This included early screening (eg, outbreak detection), continuing disease surveillance, advice regarding travel advisories, and ex situ medical treatment (e.g., medical tourism) and helped result in improved quality of care and reductions in cost.

Global virtual hackathons focused on COVID-19 such as the Observational Health Data Sciences and Informatics (OHDSI)[17] international communities' study-a-thon (March 26-29, 2020) and the Massachusetts Institute of Technology hackathon (April 3-5, 2020) have also spurred the development of cross-border clinical research studies and cross-border entrepreneurship, respectively. The output of these efforts are open-source ideas and tools to solve a variety of problems arising from the COVID-19 pandemic. The OHDSI efforts are focusing on a global baseline characterization of COVID-19 patients as well as the safety of hydroxychloroquine for COVID-19 treatment, drawing on collaborators and data spread across the globe.

The flow of COVID-19 data across borders also has economic implications. In the field of biomedical informatics, we sometimes ignore the economic effects that the data and data products we create and consume may have on the global economy, but it is worth examining them in the context of a global pandemic. Certainly, COVID-19 has had a devastating effect on the global economy, and that could affect public health in a variety of ways.[18] From a health data economy perspective, the inflows and outflows of data and information across geopolitical boundaries have the potential to generate enormous economic value in a digitally connected global healthcare economy.[19,20] The capital value of global COVID-19 data can be maximized when analyzed using descriptive, predictive, or prescriptive analytics for the purposes of clinical research, public health purposes, and pharmaceutical development. These cross-border data flows have the potential to be a driver of global economic growth, though altruism has largely dictated the free flow of data and ideas in the current crisis. COVID-19 data enable significant new opportunities for innovation and disruption within the health data economy, especially for emerging infectious diseases,[20] and telemedicine. Governance of data and data product sharing can take the form of the OHDSI network with the free flow of data products for a collective research gain, a commercial data-sharing model such as between Google and HCA Healthcare,[21] or a self-governance model like DataBox[22] in which profit sharing from the data transfer can be realized by the data owners.

The goal of flattening the curve is to reduce the reproduction ratio. The reproduction ratio is how many people that a person in one disease episode passes the disease along to (eg, if the reproduction ratio is 4, then that infected patient transmitted COVID-19 to 4 more people).[23] However, the number of reported cases may not be a very useful indicator unless you know something about how the COVID-19 testing is being conducted and how the data are being gathered.[23] When there are major differences between COVID-19 testing strategies, as there have been in this pandemic, it is difficult to make direct comparisons accurately, as the testing strategies can skew case counts.[23] Accurate estimation of the reproduction ratio depends on having comprehensive, diverse, and heterogeneous datasets to overcome the limitations of individual localized data sources. For COVID-19, countries that conducted comparatively high numbers of tests had lower mortality rates even though they reported high case counts that alarmed the public in the short run.[23] Tracking the viral mutations of COVID-19 cases in New York suggests that most cases were traced to travelers returning from Europe, and not Asia, as originally expected.[24] Missing this hidden spread due to insufficient testing and screenings at the borders meant that the suspension of air travel and mandatory quarantines for travelers from Europe occurred too late.

Global data on disease trajectories and the effectiveness and economic impact of different social distancing measures are essential to facilitate effective local responses to pandemics. Policymakers have used these data to inform their decisions regarding travel bans, quarantines, and economic stimulus. To facilitate the dissemination of knowledge regarding COVID-19 during the outbreak, publishers are prioritizing review of and offering free, open access to relevant research findings.[25] Sharing COVID-19 data freely and globally boosts the data economy, enabling quicker dissemination and validation of pharmaceutical innovations, as well as improving knowledge of what prevention and mitigation measures work. Even if physical borders around the globe are closed, it is crucial that data related to COVID-19 continue to transparently flow across borders to enable a data economy to thrive, which will promote global public health through global cooperation and solidarity.

AUTHOR CONTRIBUTIONS

CT, YZ, YH, and DWB built on and extended the initial idea. CT drafted the manuscript. All authors provided substantial contribution to paper edits. JMP, especially, filled up this manuscript with great content to increase its size. All the authors are accountable for the integrity of the work.

ACKNOWLEDGMENTS

The authors thank Sheng Wang, PhD, for valuable comments and suggestions on early drafts. The content is solely the responsibility of the authors.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Wu F, Zhao S, Yu B, *et al*. A new coronavirus associated with human respiratory disease in China. *Nature* 2020; 579 (7798): 265–9.
2. Severe acute respiratory syndrome coronavirus 2 isolate in Wuhan-Hu-1, complete genome. GenBank: MN908947.3. https://www.ncbi.nlm.nih.gov/nuccore/MN908947 Accessed March 3, 2020.
3. Niiler E. An AI epidemiologist sent the first warnings of the Wuhan virus. https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings Accessed January 25, 2020.
4. Apuzzo M, Rich M, Yaffe-Bellany D. Failures on the Diamond Princess shadow another cruise ship outbreak. https://www.nytimes.com/2020/03/08/world/asia/coronavirus-cruise-ship.html Accessed March 16, 2020.
5. Zhu N, Zhang D, Wang W, *et al*. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; 382 (8): 727–33.
6. Miliard M. Epic pushes out software update to help spot coronavirus. https://www.healthcareitnews.com/news/epic-pushes-out-software-update-help-spot-coronavirus Accessed January 24, 2020.
7. Chinazzi M, Davis JT, Ajelli M, *et al*. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 2020; 368 (6489): 395–400.
8. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. *JAMA* 2020; 323 (14): 1341–2.
9. Niehus R, Salazar PMD, Taylor A, *et al*. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan. China that depend on reported cases in international travelers. medRxiv. https://www.medrxiv.org/content/10.1101/2020.02.13.20022707v2 Accessed March 3, 2020.
10. Mandl KD, Overhage JM, Wagner MM, *et al*. Implementing syndromic surveillance: a practical guide informed by the early experience. *J Am Med Inform Assoc* 2003; 11 (2): 141–50.
11. HealthMap. Boston Children's Hospital. https://www.healthmap.org/covid-19 Accessed February 18, 2020.
12. Kraemer M. I'm a researcher who's helped change how we tackle pandemics like coronavirus forever—this is what we've learned. https://www.independent.co.uk/voices/coronavirus-covid-19-pandemic-outbreak-data-research-cdc-who-a9406281.html Accessed March 17, 2020.
13. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; 20 (5): 533–4.
14. Xu B, Gutierrez B, Mekaru S, *et al*. Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* 2020; 7: 106. doi: 10.1038/s41597-020-0448-0.
15. Erickson JI. "Boston Hope" Medical Center opens at Boston Convention and Exhibition Center. https://www.massgeneral.org/news/coronavirus/boston-hope-medical-center-opens Accessed March 10, 2020.
16. Pagel C, Utley M, Ray S. Covid-19: how to triage effectively in a pandemic. *BMJ Opin*. https://blogs.bmj.com/bmj/2020/03/09/covid-19-triage-in-a-pandemic-is-even-thornier-than-you-might-think Accessed March 9, 2020.
17. Observational Health Data Sciences and Informatics. COVID-19 updates page. https://www.ohdsi.org/covid-19-updates Accessed April 7, 2020.
18. Frazee G. How the coronavirus' economic toll could also affect public health. https://www.pbs.org/newshour/economy/making-sense/how-the-coronavirus-economic-toll-could-also-affect-public-health Accessed March 30, 2020.
19. Dobbs R, Manyika J, Woetzel J. Digital globalization: the new era of global flows. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Globalization/Global%20flows%20in%20a%20digital%20age/Global_flows_in_a_digital_age_Full_report%20March_2015.ashx Accessed March 20, 2015.
20. Tang C, Plasek JM, Bates DW. Rethinking data sharing at the dawn of a health data economy: a viewpoint. *J Med Internet Res* 2018; 20 (11): e11519.
21. Kent JH. Google Cloud launch COVID-19 data sharing platform. https://healthitanalytics.com/news/hca-google-cloud-launch-covid-19-data-sharing-platform Accessed April 9, 2020.
22. Zhu Y, Xiong Y, Liao Z, *et al*. [Self-governing openness of data]. *Big Data Res* 2018; 4 (2): 3–14.
23. Silver N. Coronavirus case counts are meaningless. https://fivethirtyeight.com/features/coronavirus-case-counts-are-meaningless/?fbclid=IwAR1gpC1Zblt_rPRfBkMUZKBVNNKSTrIS3WcS_P6gwE1uGvCp98CspXRCzTE Accessed April 4, 2020.
24. Zimmer C. Most New York coronavirus cases came from Europe, genomes show. https://www.nytimes.com/2020/04/08/science/new-york-coronavirus-cases-europe-genomes.html Accessed April 8, 2020.
25. Calling all coronavirus researchers: keep sharing, stay open. *Nature* 2020; 578 (7793): 7.