https://doi.org/10.1038/s41591-022-02054-0

Data-enabled responses to pandemics: policy lessons from COVID-19

Sangeeta Dhami, Deidre Thompson, Maha El Akoum, David W. Bates, Roberto Bertollini and Aziz Sheikh

Most health systems struggled to obtain and analyze real-time data during the COVID-19 pandemic, but places that succeeded can be studied to provide a model for data-enabled responses to future epidemics and pandemics.

The COVID-19 pandemic, which emerged from Wuhan, China, in December 2019, has resulted in at least 603 million cases and more than 6.4 million deaths worldwide (as of September 2022)[1]. There has been considerable additional disruption, morbidity and mortality resulting from the social, economic and health system consequences that ensued as several governments instituted a series of national and then more localized lockdowns[2]. The pandemic required a series of policy, public health and clinical decisions to be taken, with major consequences to societal functioning, economics and care provision. The taking of these decisions was always going to be complex, but for most places this was exacerbated by the lack of availability of relevant data[3]. By contrast, a handful of territories substantially developed their data capabilities over the course of the pandemic, generating important insights to guide their own national decisions and to inform international deliberations.

Key data sources should be available at various stages of a pandemic. Case studies of territories that have been positive outliers in their data capabilities allow potentially transferable lessons to be learned, in order to be better equipped to generate data-enabled responses to future epidemics and pandemics. As the COVID-19 pandemic is not yet over, the ideas contained in this paper should be seen as a work in progress.

Data requirements

All pandemics have distinctive dimensions that depend on the nature of the responsible infectious agent, the speed of national and international non-pharmacological responses, and the availability and deployment of vaccines and therapeutic agents. It is, however, possible to identify some core phases of pandemics and therefore consider the data sources that should ideally be available to support decision-making during these phases. The core phases of pandemics are summarized in the WHO (World Health Organization) Pandemic Phases Framework, which was originally developed for influenza[3].

Although most governments have, to some extent, developed their pandemic data response capabilities, a few have disproportionately contributed to the discovery of policy-relevant insights during COVID-19. Examples of such places include Iceland, Israel, Qatar, Scotland and Taiwan (some of which are discussed in Table 1).

Having relevant datasets available is fundamental, but insufficient, to ensure capacity for data-enabled policy responses to pandemics. Also needed are permissions to access data by different stakeholders, ideally coordinated and granted by a national scientific committee, and the ability to curate, link, analyze, visualize, interpret and communicate these data to government bodies, policy makers, health system leaders and other audiences, often across national boundaries. These are each time-consuming steps, but time is one luxury not available in the context of the exponential growth of infections seen in pandemics. It is therefore crucial that due attention is given to the data infrastructure and pipeline as part of national pandemic preparedness plans.

Data infrastructure

There is a need to access disparate data, including from electronic health records, travel and other health-related data, ideally on every person, in as close to real-time as possible. Key datasets can potentially be stored in a single central secure warehouse, as is the case for Qatar (Table 1). This requires adequate computational power, which can be substantial when dealing with millions of rows of data. Bringing together these disparate datasets can be done through deterministic or probabilistic approaches; where possible, this is most efficiently achieved using unique identifiers[4,5]. An alternative approach is to leave data in situ and deploy a service-orientated architecture (SOA) approach, which creates interfaces between disparate datasets through application programming interfaces (APIs). This requires upfront engineering costs, but offers the potential for periodic synchronized updates and accompanying substantial reductions in downstream resource demands.

Information governance

Access to health and other sensitive data needs to be carefully regulated[6] and requires a variety of processes to be in place, to ensure that data are not inappropriately used. These checks are typically extensive and time consuming. However, the risk balance in providing access to these data needs to be shifted in the context of global emergencies such as pandemics. It is therefore important that policies and plans are in place, which may require special legislation. For example, Taiwan passed legislation to allow access to mobile-phone data (Table 1). Similarly, a Control of Patient Information (COPI) notice was issued by the UK Government's then Secretary of State for Health and Social Care to allow sharing of confidential patient information among healthcare organizations and other relevant bodies in order to safeguard public health[7].

Analytical capability

Another key rate-limiting step in the ability to generate data-enabled insights is the lack of data processing and analytical capability. There is a need for trained staff who are ideally familiar with the datasets in question who can, at pace, check, clean, link, analyze and help to visualize data for policy audiences and others. This requires staff with a range of skills to work together[8]. Taking the time to develop, for example, a

Table 1 | Case studies of national data infrastructures used to support pandemic responses

Location	Context	Approach	Impact	Lessons
Qatar	Since 2011, Hamad Medical Corporation and the Primary Health Care Corporation have maintained a single electronic health record across 13 hospitals and 27 primary healthcare facilities. Qatar has used its advanced national, centralized eHealth system to execute evidence-based public health responses to the pandemic.	Hamad Medical Corporation compiled a centralized and standardized national SARS-CoV-2 reverse transcription polymerase chain reaction (RT–PCR) testing, hospitalization and immunization database, which includes: basic demographic information on all residents; vaccination records for the entire population; information on RT–PCR testing, including for those suspected to have a SARS-CoV-2 infection as well as traced contacts; and COVID-19 hospital admission data, with a WHO severity classification for each identified case. Scientific analyses of the surveillance and outbreak data are used to power infection transmission models that monitor and predict epidemiological trends, giving a real-time estimation of key indicators.	Key indicators were reviewed and validated by a national scientific committee and used to inform major public health policy decisions, such as predicting the earliest date for easing public health restrictions. The electronic health record database includes information for each of the 2.8 million residents in Qatar. More than 10.5 million RT–PCR tests were recorded.	Policy decisions should be guided by scientific knowledge Science-based, data-driven decision-making in Qatar during the COVID-19 pandemic helped to minimize economic losses. Robust data systems are essential to all health systems, as they allow the generation of knowledge related to the epidemiology of viruses and the efficacy of vaccines, which helps to guide effective policy responses.
Scotland	Scotland developed a national pandemic surveillance and evaluation platform, Early Assessment of Antiviral and Vaccine Effectiveness (EAVE), in response to the novel influenza A (H1N1) pandemic in 2009 (refs. [10,15]). The platform linked primary care, testing, vaccination, hospitalization and mortality data on about 250,000 people (5% of the population) managed through Public Health Scotland, and was put into hibernation after the end of the H1N1 pandemic.	Following the emergence of SARS-CoV-2, EAVE was repurposed to EAVE II and scaled up to cover nearly the entire Scottish population of 5.4 million people[16]. EAVE II brought together primary care, testing, sequencing, vaccination, hospitalization, intensive care unit, mortality and other datasets into Public Health Scotland. Datasets are securely linked using Scotland's unique identifier, the Community Health Index number.	EAVE II is one of the world's few national, end-to-end, near real-time COVID-19 data platforms. The EAVE II team were the first in the world to demonstrate the real-world effectiveness of first-dose COVID-19 vaccines in preventing hospitalizations[17]. They have produced many other major analyses, including demonstration of: increased severity of COVID-19 outcomes associated with Delta infection; reduced severity of associated with Omicron infection[18], and vaccine waning; and risk of serious outcomes after first, second and booster vaccine doses[19].	Repurposing EAVE to create EAVE II was a time-consuming process, as information governance procedures were not fit-for-purpose in the context of a pandemic. The lack of trained data processors and analysts with permissions to access these data increased the challenges. Data infrastructure and associated capabilities must be maintained and updated to allow quicker responses to future pandemics.
Taiwan	The Taiwan Communicable Disease Control Act (2007), passed four years after the outbreak of severe acute respiratory syndrome (SARS), waived data authorization and consent requirements in the context of an emerging infections disease[20]. Taiwan has an existing national health insurance database. The Taiwan Centers for Disease Control (CDC) Epidemic Intelligence Center has an existing automated early warning pandemic surveillance system (National Notifiable Disease Surveillance System), which includes automated analyses of a range of data sources[21].	On 20 January 2020, the Taiwan CDC activated the Central Epidemic Command Center. By 27 January 2020, the national health insurance database was linked with immigration and customs databases. Those at high risk of contracting SARS-CoV-2 infection were tracked through their mobile phones[22]. Patients' travel histories were made available to hospitals, clinics and pharmacies. Passive mobile phone geolocation data were used, among other sources, for contact tracing, for example on the *Diamond Princess* cruise ship.	Taiwan was able to rapidly reactivate its pandemic plans, including provision to use mobile phone data to support surveillance efforts, and which supported Taiwan's zero-COVID policy. For the first two years of the pandemic, this policy was effective in containing transmission, leading to a low number of cases, hospitalizations and deaths.	It is important to enact legislative changes that may prove helpful in the context of epidemics and pandemics as part of national pandemic plans. Safeguarding measures are needed to ensure that these data are not used outside of exceptional circumstances.

data dictionary and the sharing of source code can greatly increase efficiency of analysis and transparency of methods.

Transparency

As ever, it is important that analyses are undertaken in transparent ways[9] with provision for exploratory analyses. For example, it was unclear during the early stages of the pandemic which variables would be most useful to identify patients at greatest risk of poor COVID-19 outcomes, resulting in the need for several exploratory analyses. It is important that such exploratory analyses are transparently reported. Other recommendations for transparency include: reporting metadata; wherever possible, specifying statistical analysis plans (SAP) in advance and making these publicly available; making source code available through a repository such as GitHub; and, where possible,

BOX 1

Johns Hopkins SARS-CoV-2 testing dashboard

- A Chinese graduate student, Ensheng Dong, at Johns Hopkins University, USA, had studied epidemics and was concerned about the effect of COVID-19 on the people of China.
- After consultation with his advisor, Lauren Gardner, they created a dashboard to visualize SARS-CoV-2 cases using a geographic information system model that they had previously used to measure measles risk in the USA. The dashboard was interactive and provided near real-time data to track and visualize the location of cases of SARS-CoV-2, deaths and recovery, initially in China, but then including worldwide data.
- In February 2020, a team from Esri's ArcGIS Living Atlas of the World (https://livingatlas.arcgis.com) automated the task of data collection from China, and a team of volunteers was recruited from Johns Hopkins University to update and maintain the site.
- The dashboard (https://coronavirus.jhu.edu/map.html) has now evolved to a multi-layered resource that provides expert analyses and graphics for researchers, public health officials and the general public, and identified most newly infected countries ahead of the WHO[23].
- The dashboard reports cases at the province level in China; at the city level in the USA, Australia and Canada; and at the national level in other regions.
- The quick-thinking and prompt action of a graduate student led to the creation of an invaluable global data resource, which *TIME* magazine recognized as the 'go-to data source' for COVID-19 and named as one of the top 100 'best inventions of 2020'.
- A science-based team approach was essential to the rapid scaling-up of this effort.

BOX 2

Key policy recommendations

1. Development of the underlying data infrastructure, governance and analytic capacity to provide data for policymaking and respond to pandemics should be a core component of national pandemic preparedness plans.
2. Most territories have enhanced their data capabilities over the course of the COVID-19 pandemic, and these should be further developed and not allowed to regress. Repurposing COVID-19 data capabilities to help to respond to other major health concerns – such as influenza, pneumonia, cancer, cardiovascular diseases and mental health – will help to ensure that capabilities are maintained, enabling the rapid redeployment in any future epidemics or pandemics.
3. Identifying the range of data sources that can prove useful during various stages of an epidemic or pandemic will allow key data gaps to be identified and strategies to be prioritized and developed in order to plug these gaps as part of national data roadmaps.
4. Secure linkage of datasets greatly increases the range of questions that data can answer, especially using unique identifiers, which should be prioritized where they do not exist.
5. Developing capacity and capabilities should be a central component of national data and workforce strategies to bring together disparate datasets on entire populations, which is challenging in terms of computational ability, information security and governance, and the human capacity needed to process, link, analyze and interpret these data.
6. Countries should proactively review their legislative frameworks that govern the use of health data and should have provisions in place to expedite permissions for the use of health and health-related data in exceptional circumstances such as pandemics.
7. It is vital that public trust is maintained. A national commitment to transparency about access to and uses of data is crucial. The 'Five Safes Framework' of safe people, safe projects, safe settings, safe data and safe outputs is an example of a potentially effective approach.
8. Some territories have been able to make substantial progress with data-enabled responses to the COVID-19 pandemic, which other territories should study. Direct dialogue with policy teams in these territories will help identify potentially transferable lessons.
9. More attention should be given to citizen data science initiatives in the context of the pandemic using data from smartphones and other devices.
10. Mechanisms are needed for the efficient sharing and analysis of data and data-enabled insights between countries and regions, most likely through federated approaches to data analysis, where data remain within national jurisdictions.

making actual or synthetic data available to facilitate replication and validation studies and training of new analysts. While the immediate need is to provide insights to policy makers, there is considerable merit in also publishing analyses in preprints and peer-reviewed journals to allow independent verification of methods and to share insights with the global community

International co-operation

There are numerous instances where it is important to be able to run analyses across countries, regions or globally[10]. However, this is difficult because it is seldom possible to move sovereign datasets across national boundaries and so requires federated analyses to be undertaken with some form of data synthesis. The most prominent example has been the Johns Hopkins Coronavirus Resource Center COVID-19 Testing Dashboard (Box 1).

Other examples include analyses of data across UK nations to investigate the effect of lockdown measures on health system functioning, investigation of rare vaccine safety signals, such as cerebral venous sinus thrombosis[11,12], the impact of variants of concern (Gamma in Brazil and Delta in Scotland) on disease severity and waning of vaccine effectiveness[13] and work undertaken across more than 40 countries through the International COVID-19 Data Alliance (ICODA; https://

icoda-research.org) to investigate the effect of lockdown measures on perinatal outcomes[14].

Conclusions

Ready access to high quality multi-dimensional data is fundamental to generating effective evidence and informed policy responses to pandemics, but most places have struggled with this. Many analyses need to extend across international boundaries, which is most likely to be achieved through federated analytical approaches, but will require coordination between governments. A few territories have excelled in health data science during the pandemic, which offers a framework that might be developed and deployed in future epidemics and pandemics (Box 2).

Sangeeta Dhami[1], Deidre Thompson[2], Maha El Akoum[2], David W. Bates[3], Roberto Bertollini[4] and Aziz Sheikh[5] ✉

[1]NHS Lothian, Edinburgh, UK. [2]World Innovation Summit for Health, Doha, Qatar. [3]Harvard Medical School, Boston, MA, USA. [4]Ministry of Public Health, Doha, Qatar. [5]Usher Institute, University of Edinburgh, Edinburgh, UK.

✉e-mail: aziz.sheikh@ed.ac.uk

Published online: 13 October 2022

References

1. WHO. *WHO Coronavirus (COVID-19) Dashboard*; https://covid19.who.int/ (accessed 6 September 2022).
2. UNDP. *COVID-19: Socio-Economic Impact*; https://www.undp.org/coronavirus/socio-economic-impact-covid-19 (accessed 6 September 2022).
3. WHO. *Pandemic Influenza Preparedness and Response*; https://www.ncbi.nlm.nih.gov/books/NBK143061/ (2015).
4. Lewis, S. *TechTarget* https://go.nature.com/3fuPPUp (accessed 8 September 2022).
5. DbBee. https://go.nature.com/3UOdIGU (accessed 8 September 2022).
6. NHS Research Authority. https://go.nature.com/3dTcQQE (2021).
7. NHS Digital. https://go.nature.com/3y258dQ (accessed 8 September 2022).
8. Allsopp R. et al. *OMDDAC* https://doi.org/10.13140/RG.2.2.30601.98405 (2021).
9. NHS. https://go.nature.com/3CklyQk (2021).
10. Yang, L. et al. *Arc. Med. Sci.* **17**, 829–837 (2021).
11. Shah, S. A. et al. *eClinicalMedicine* **49**, 101462 (2022).
12. Kerr, S. et al. *PLoS Med.* **19**, e1003927 (2022).
13. Katikireddi, S. V. et al. *Lancet* **399**, 25–35 (2022).
14. Stock, S. et al. *Wellcome Open Res.* **6**, 21 (2021).
15. Simpson, C. et al. *Health Technol. Assess.* https://doi.org/10.3310/hta19790 (2015).
16. Simpson, C. et al. *BMJ Open* **10**, e039097 (2020).
17. Vasileiou, E. et al. *Lancet* **397**, 1646–1657 (2021).
18. Sheikh, A. et al. *Lancet* **397**, 2461–2462 (2021).
19. Sheikh, A. et al. *Lancet Infect. Dis.* **22**, 959–966 (2022).
20. Chen, C. M. et al. *J. Med. Internet Res.* **22**, e19540 (2020).
21. Taiwan Centers for Disease Control. https://go.nature.com/3CjYCS6 (2020).
22. Wang, C. J. et al. *JAMA*. **323**, 1341–1342 (2020).
23. Dong, E. et al. *Lancet Infect. Dis.* **20**, 533–534 (2020).

Acknowledgements

We thank S. Afdhal and Z. Sheikh for their comments and contributions to earlier drafts of this paper. The original report, of which this is an abridged version, was funded by the Qatar Foundation.

Competing interests

A.S. has served on several UK and Scottish Government COVID-19 advisory bodies and was a consultant to the World Innovation Summit for Health (WISH) on its report on data-enabled responses to pandemics. D.W.B. reports grants and personal fees from EarlySense, personal fees from CDI Negev, equity from ValeraHealth, equity from Clew, equity from MDClone, personal fees and equity from AESOP, personal fees and equity from Feelbetter, equity from Guided Clinical Solutions, grants from IBM Watson Health, outside the submitted work, and a patent pending (PHC-028564 US PCT), on intraoperative clinical decision support. All other authors report no conflicts.