

Lessons from COVID-19 for rescalable data collection

Sangeeta Bhatia*, Natsuko Imai*, Oliver J Watson*, Auss Abbood, Philip Abdelmalik, Thijs Cornelissen, Stéphane Ghozzi, Britta Lassmann, Radhika Nagesh, Manon L Ragonnet-Cronin, Johannes Christof Schnitzler, Moritz UG Kraemer, Simon Cauchemez, Pierre Nouvellet, Anne Cori



Novel data and analyses have had an important role in informing the public health response to the COVID-19 pandemic. Existing surveillance systems were scaled up, and in some instances new systems were developed to meet the challenges posed by the magnitude of the pandemic. We describe the routine and novel data that were used to address urgent public health questions during the pandemic, underscore the challenges in sustainability and equity in data generation, and highlight key lessons learnt for designing scalable data collection systems to support decision making during a public health crisis. As countries emerge from the acute phase of the pandemic, COVID-19 surveillance systems are being scaled down. However, SARS-CoV-2 resurgence remains a threat to global health security; therefore, a minimal cost-effective system needs to remain active that can be rapidly scaled up if necessary. We propose that a retrospective evaluation to identify the cost-benefit profile of the various data streams collected during the pandemic should be on the scientific research agenda.

Introduction

The COVID-19 pandemic and the world's response have been unprecedented in scale and impact. In some countries, substantial resources were made available to support response operations and policy making, including the collection and dissemination of vast amounts of pandemic-related data. These data streams have been used to inform public health responses, reinforcing the value of timely and transparent data collection and analysis for decision making. The availability of novel data has catalysed the creation of innovative tools and methodologies.¹ However, the pandemic has also highlighted new challenges concerning the scalability, sustainability, and equity in the generation, access, and analysis of these data.²⁻⁴

Drawing on the European experience during the pandemic, we build on previous work identifying key data to support decision making during outbreaks.⁵ We explore the challenges and opportunities for future data collection efforts against lessons learnt during the pandemic. A key concept we draw attention to is the rescalability of data collection and processing systems—ie, the ability of these systems to adapt quickly and efficiently to changing priorities as a crisis evolves (eg, by changing data resolution or sampling as needed).

Key policy questions and evidence to inform them

Throughout an infectious disease outbreak, from its emergence, to its spread, and then its flare-ups, different policy questions are generated that could be addressed with different sources of quantitative evidence (figure). In this Personal View, we focus on five policy questions that emerge early in an epidemic, which need to be continuously evaluated to inform epidemic response.

Is an outbreak occurring and is it caused by a novel pathogen or a new variant?

Timely detection of public health concerns requires continuous monitoring and global surveillance to identify, verify, assess, and investigate potential threats.⁶ The first

detailed report on Dec 30, 2019, described a cluster of pneumonia cases of unknown cause in Wuhan, China; the cluster was detected by event-based surveillance in which professional networks monitored unstructured and structured data from multiple outlets.⁷

The new causative agent, SARS-CoV-2, was rapidly identified by genomic sequencing. The full viral genome was made publicly available on Jan 10, 2020, just 10 days after the Wuhan Municipal Health Commission reported the pneumonia cluster,⁸ allowing the rapid development of diagnostics.⁹ As the pandemic progressed, detection of new variants of concern was facilitated by routine genomic surveillance.

How dangerous is the pathogen?

Early characterisation of the pathogen, particularly its symptoms and severity, is crucial for generating case definitions and assessing the potential disease burden. Early disease severity estimates were initially informed by repatriation flights. Although not representative of the wider population, the passengers represented the first pool of infected individuals to be tested, without selection based on symptoms.^{10,11} Risk factors for severe disease were investigated in more representative populations with clinical data compiled across many countries by the International Severe Acute Respiratory and Emerging Infection Consortium.

To collate further information on symptoms, innovative solutions were adopted. For example, a new mobile app, the ZOE COVID-19 symptom tracker (version 3.0.1) was developed by King's College London (London, UK) and rolled out to volunteers throughout the UK.¹² This citizen science approach, which was launched less than 2 months after the first case was reported, allowed for rapid characterisation of the spectrum of symptoms, including the loss of sense of smell or taste.

How quickly is the pathogen spreading?

Characterising the transmissibility and natural history of the pathogen is crucial to assess potential health-care needs, how soon they will be needed, and how demand

Lancet Infect Dis 2023; 23: e383-88

Published Online

May 4, 2023

[https://doi.org/10.1016/S1473-3099\(23\)00121-4](https://doi.org/10.1016/S1473-3099(23)00121-4)

This online publication has been corrected. The corrected version first appeared at thelancet.com/infection on May 24, 2023

*Contributed equally

MRC Centre for Global Infectious Disease Analysis, Jameel Institute, School of Public Health, Imperial College London, London, UK
(S Bhatia PhD, N Imai PhD, O J Watson PhD, M L Ragonnet-Cronin PhD, Prof P Nouvellet PhD, A Cori PhD); Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK (O J Watson); Robert Koch Institute, Berlin, Germany (A Abbood MSc); WHO, Geneva, Switzerland (P Abdelmalik PhD, T Cornelissen PhD, J C Schnitzler PhD); Helmholtz Centre for Infection Research, Braunschweig, Germany (S Ghozzi PhD); ProMED-mail, International Society for Infectious Diseases, Brookline, MA, USA (B Lassmann MD); Blavatnik School of Government (R Nagesh PhD), Department of Biology (M U G Kraemer DPhil), and Pandemic Sciences Institute (M U G Kraemer), University of Oxford, Oxford, UK; Department of Ecology & Evolution, University of Chicago, Chicago, IL, USA (M L Ragonnet-Cronin); Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Université Paris Cité, CNRS UMR2000, Paris, France (S Cauchemez PhD); School of Life Sciences, University of Sussex, Brighton, UK (Prof P Nouvellet); NIHR Health Protection Research Unit in Modelling and Health Economics, Imperial College London, UK Health Security Agency, and London School of Hygiene & Tropical Medicine,

London, UK (A Cori, S Bhatia)
Modelling and Economics Unit,
UK Health Security Agency,
London, UK (S Bhatia)

Correspondence to:
Dr Anne Cori, MRC Centre for
Global Infectious Disease
Analysis, Jameel Institute, School
of Public Health, Imperial College
London, London W2 1PG, UK
a.cori@imperial.ac.uk

For more on the International
Severe Acute Respiratory and
Emerging Infection
Consortium see
<https://isaric4c.net/>

For more on the ZOE COVID-19
symptom tracker see
<https://covid.joinzoe.com/>

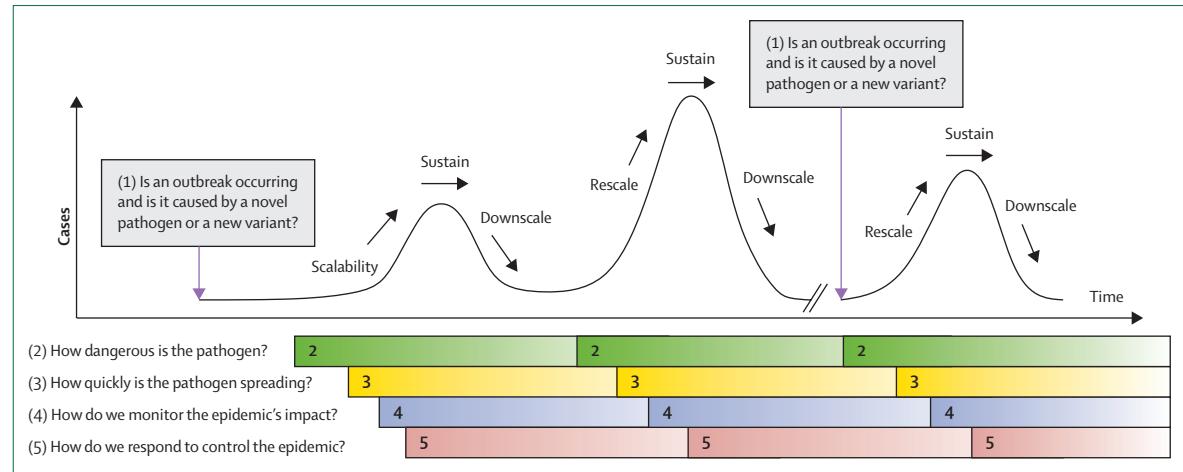


Figure: Policy questions and data needs across the different phases of an epidemic

Identification of threats to public health requires continuous monitoring and surveillance. The purple arrows represent the detection of a novel pathogen or new variant. The first question is to assess (1) whether an outbreak is occurring and whether it was caused by a novel pathogen or a new variant. Once a causative agent has been identified, the second question (2) is to characterise morbidity and mortality (ie, how dangerous is the pathogen?) and the third question (3) is to understand transmissibility and its mode of transmission (ie, how quickly is the pathogen spreading?). While answering these questions, surveillance infrastructure should be rapidly scaled up to ensure that (4) the effect of the epidemic can be monitored. The final policy question that emerges simultaneously is (5) understanding how the ensuing epidemic can be controlled. Throughout the epidemic, surveillance systems must be sustained to monitor changes in pathogen transmission or severity. After an infection wave has subsided, surveillance can be downscaled as focus moves towards longer term disease surveillance and evaluating the need for changes in policy response. Flare-ups caused by a new pathogen or a new variant of a previously circulating pathogen generate similar policy questions.

could change over time. The likelihood and extent of sustained SARS-CoV-2 human-to-human transmission were first evaluated by the use of media reports on early confirmed COVID-19 cases outside mainland China and international air traffic data.^{13,14} Once transmission was established and diagnostic tests were available, the transmissibility of SARS-CoV-2 was continuously monitored with daily reported cases and deaths,¹⁵ PCR cycle threshold data,¹⁶ and analyses of genomic sequences.¹⁷ To provide a global overview of transmission trends, WHO regularly collated and published officially reported surveillance data on the WHO COVID-19 dashboard. Natural history characteristics of SARS-CoV-2 and heterogeneities in transmission as new variants of concern emerged were quantified by case investigation,¹⁸ contact tracing,¹⁹ and analysis of viral genome sequencing data.²⁰

How do we monitor the epidemic's impact?

Continuously tracking the epidemic's trajectory is important to quantify the burden of disease within populations and its variations over time and across regions. Initially, countries focused on estimating the number of imported COVID-19 cases. However, as case numbers increased, efforts shifted to capturing the total number of locally diagnosed cases. Variable testing access and capacity issues meant this number was not a consistent indicator of transmission levels. Thus, numbers of COVID-19 hospitalisations and deaths, which were more consistently reported, were used as proxies for the level of transmission. However, hospitalisations and deaths only represent a proportion of all infections and

are lagged indicators that reflect transmission levels in the previous 2-week or 3-week period due to the time between infection, developing severe symptoms, and death.

To measure unbiased, real-time transmission, infection prevalence must be quantified in representative samples of the population. In the UK, the Real-Time Assessment of Community Transmission (REACT) study and the Office of National Statistics COVID-19 Infection Survey, were instrumental in estimating trends in transmission and identifying risk factors for infection, thus helping to inform potential targeted control measures.²¹ Alternative methods were also leveraged, such as the use of wastewater surveillance to measure viral activity and monitor the changing dynamics of SARS-CoV-2 lineages in real time.²²

Beyond real-time estimation of transmission levels, understanding SARS-CoV-2 immunity is important to quantify the potential future burden and transmission. Cohort studies, such as the SARS-CoV-2 Immunity and Reinfection Evaluation (SIREN) study in UK health-care workers, were designed to assess the risk of reinfection, and therefore the duration of protection, through linking frequent serology and virus testing in a non-representative but high exposure and high retention cohort.²³

How do we respond to control the epidemic?

Interventions are often crucial to mitigate the public health impact of a novel pathogen. The collection of interventions considered will depend on the availability and effectiveness of individual interventions, the pathogen's characteristics, and the disease burden.

For more on the REACT study
see <https://www.reactstudy.org/>

For more on the WHO
COVID-19 dashboard see
<https://covid19.who.int/>

Furthermore, interventions will typically be continuously assessed to adjust response efforts as needed.

Rapid sharing of viral genome sequences helped to identify viral targets (eg, the spike protein) and accelerated the development of vaccines and therapeutics. To evaluate the efficacy of such interventions, the gold standard is randomised control trials. Clinical trials of COVID-19 treatments (eg, RECOVERY²⁴ and DisCoVeRy²⁵) and vaccines were implemented at a rapid pace.

However, clinical efficacy does not translate directly into real-world effectiveness. The rapid roll-out of COVID-19 vaccines in multiple countries was followed by studies of real-world effectiveness, requiring information on infection prevalence and vaccine coverage. This information relies on robust population denominators, which were challenging to quantify as representative census data were often out of date.²⁶ However, in high-income settings such as Europe and Israel, the numerous, robust, vaccine effectiveness studies were invaluable to estimate the real-life effect of pharmaceutical interventions in the context of continuously emerging new variants. These studies helped inform the implementation of vaccination strategies, such as age-targeted booster programmes.

Although some randomised trials were done to assess specific non-pharmaceutical interventions, such as wearing masks, evaluation typically depended on indirect evidence or observational data, particularly interventions relying on physical distancing measures. The impact of changes in behaviours and physical distancing on transmissibility was estimated with aggregated mobility reports that were based on data routinely collected by Google and Apple Maps.²⁷ To further investigate how physical mixing changed with non-pharmaceutical interventions, and thus the effectiveness of policies, regular social contact surveys such as CoMix were informative.²⁸

These data generated insights into the effectiveness of non-pharmaceutical interventions. Assessment of the impact of specific interventions on transmission over time and across regions required detailed data on their implementation and uptake.²⁹ Such data had not been routinely or systematically collected before. Novel efforts, such as the Oxford Government Response Tracker, collected data on interventions implemented worldwide.³⁰ Other similar efforts were done in parallel; for example, by non-government organisations such as the ACAPS Government Measures Dataset.³¹ The European Centre for Disease Prevention and Control³² and WHO³³ eventually developed central repositories collating data from the multiple platforms. YouGov polls on knowledge, attitudes, and practices towards COVID-19 were done at unprecedented scale. These efforts generated open-source data, which showed heterogeneous and declining adherence to control measures over time.³⁴

Data challenges and opportunities

Scalability—initial ability to increase data collection and processing

Early in the pandemic, data collated manually from informal sources were crucial for addressing public health questions.³⁵ However, these manual, volunteer-led efforts became unsustainable as cases grew exponentially. The inherent limitations of tools used to collate case data collaboratively motivated the creation of new systems, such as Global.health, which standardised and automated data ingestion and validation. Global.health has collated over 100 million individual records of patients with COVID-19 from more than 100 countries, and is focused on curating detailed de-identified case records during the early phase of new emerging infectious disease outbreaks, most recently during the 2022 mpox (formerly known as monkeypox) outbreak.³⁶ The development of such systems was made possible by drawing on expertise from multiple disciplines including software engineering, public health, and data governance. Similarly, event-based surveillance of COVID-19 indicators has increasingly been streamlined over the course of the pandemic. Several international agencies have automated the retrieval of relevant data from official websites and social media channels worldwide.³⁷ Conversely, the Oxford COVID-19 Government Response Tracker continues to rely on volunteer-led collection of data on interventions globally. Although the initiative was scaled up successfully and sustainably through multiple COVID-19 waves, geographical coverage of the collated data has varied substantially. As countries relax COVID-19 measures, the waning public interest in the pandemic has made retention and recruitment of volunteers challenging, requiring more concerted efforts towards attracting and engaging volunteers, and a reassessment of the original model of contribution.

The development of new tools and adaptation of existing technology helped to scale the response. Digital contact tracing was a technical innovation in response to the rapid spread of SARS-CoV-2 that outstripped manual tracing efforts.³⁸ Despite some initial scepticism, most European countries adopted digital contact tracing, which vastly increased the volume of data that could be collected. Mobile health approaches, such as the UK ZOE COVID-19 symptom tracker, to collect self-reported symptoms were rolled out very rapidly after the detection of the first case. As an existing collaborative project between academia and a health science start-up, ZOE's ability to scale up rapidly and efficiently can therefore be partly attributed to leveraging existing technical competencies and infrastructures.

Although the magnitude of the pandemic sparked creative approaches to the collection and application of new data streams, it has also highlighted several challenges. Tools often used for outbreak analytics were challenged by the larger volume of data available during

For more on Global.health see
<https://global.health/>

the pandemic.³⁹ In analysis pipelines, the download or upload of datasets became a rate-limiting step. Fitting complex models, initially designed for real-time analysis, to data accrued over the past 2 years can now take several hours to days.

A critical appraisal of the successes and shortcomings of how systems scaled their response reveals important opportunities for future preparedness. Early in an outbreak, news and social media are likely to remain important sources of information.⁴⁰ Subject matter expertise will therefore continue to have an important role in detecting a signal from the large volume of data gathered during emerging outbreaks. Stronger partnerships between media and academia, the standardisation of data reporting across sources, and the automation of extraction, ingestion, validation, and analysis pipelines will improve the scalability of early response operations.^{41,42} The Epidemic Intelligence from Open Sources initiative led by WHO is a powerful example of a multisector approach.⁴³ Development of tools that can integrate systematic collection of data on interventions into routine surveillance systems will reduce the need for post-hoc data scraping and collation.

Rescalability—adapt data collection and processing to changing priorities

The challenges in scaling up data collection in early 2020 were repeated in subsequent waves (eg, mid-2021 in Europe) as surveillance was relaxed, and then had to be stepped up when new variants emerged. The changing priorities over the course of the pandemic have shown the value of developing surveillance systems that can be rapidly rescaled when needed. Agile systems could facilitate faster activation of comprehensive data collection when a novel pathogen emerges and in response to flare-ups of ongoing epidemics.

As many countries move towards a chronic phase of the pandemic, data collection efforts are being scaled down or shut down as dedicated funding finishes. In the UK, the REACT cross-sectional infection prevalence and sero-prevalence surveys, and free universal PCR testing ended in April, 2022,⁴⁴ and therefore genomic surveillance in the general population has ceased.⁴⁵ Reducing data collection efforts as case numbers decline is logical from an economic perspective. However, the downscaling of surveillance should be orchestrated to achieve a minimal cost-effective system that can still rapidly detect epidemic resurgence to quickly reactivate more comprehensive surveillance mechanisms.

Optimising the downscaling of surveillance requires evaluating the extent that each data stream was used to support epidemic management. Retracing the pathway from data to decision making is challenging but could be achieved through expert panel consultations and by analysing, when available, publicly released summaries of the scientific evidence considered by policy makers. This process will ensure the most informative and actionable

data sources are prioritised going forward. Similarly, assessing whether necessary scientific evidence could have been obtained with less data can highlight data streams that are important but could be scaled down. In addition to the volume of data, the quality of data and the synergy between the data streams is a key consideration when designing data collection infrastructure.

An important element of agility and resilience in systems is integrating evidence across multiple data streams, as over-reliance on individual sources can make the systems fragile. Multiplicity of data streams can also reduce inherent biases in different data streams. For instance, uneven smartphone ownership across age groups or geographical regions can bias the mobility patterns derived from mobile phone use.⁴⁶ Hence, these data should be augmented with other sources, such as surveys. Quantifying the monetary and non-monetary cost of collecting each data stream, which can vary across epidemic phases, is also crucial. Economies of scale could be made by identifying functional redundancies, and expanding and capitalising on existing surveillance systems.⁴⁷ The UK Health Security Agency has expanded its seasonal respiratory pathogen surveillance to include COVID-19 through cross-disease studies and multiplex testing.⁴⁸ Identifying where and how existing health systems could be strengthened—eg, laboratories or computational infrastructures—will also ensure that scalable systems are durable.

The cost-benefit of data collection should be evaluated over the long term, so that the pay-off of seemingly large upfront costs are accurately accounted for. For example, large investment in genomic surveillance in the UK, Denmark, and South Africa before SARS-CoV-2 emergence paid dividends in the capability to rapidly detect and characterise new variants. Similar investments in collecting, and regularly updating, baseline demographic, contact, and mobility data are also essential. These data are crucial to address a range of public health questions including the estimation of infection prevalence, severity, vaccine effectiveness, and vaccine stockpile needed, and would provide benefits well beyond a particular public health crisis.

Sharing existing protocols for data collection and analysis,⁴⁹ with open source, ready-to-use data formats and tested analytic tools, will facilitate rapid reactivation of data collection and improve equity. Institutional memory of surveillance and analysis pipelines should be maintained through documentation and training materials, which will facilitate knowledge transfer and onboarding of new recruits. Sustained funding to retain skilled data collectors and analysts is crucial. Finally, building and maintaining trusted, interdisciplinary, collaborative networks among key stakeholders with dedicated funding and coordination between epidemics will be essential for the rapid and effective scale-up of surveillance efforts leading to timely and actionable analyses.

Conclusion

As we emerge from the acute phase of the COVID-19 pandemic, the successes and missed opportunities should be reflected on to better prepare for a resurgence of COVID-19 and for future epidemics and pandemics. Although predicting the emergence of the next novel pathogen is not possible, we can anticipate some of the questions that will need to be addressed, and the data needed to answer them. The COVID-19 pandemic, and the sequential emergence of novel variants, has underscored the importance of scalable, rescalable, and sustainable data collection systems, and associated analytical tools.

Looking ahead, investments between pandemics will be key to retain individual expertise, maintain networks between stakeholders globally and the broad scientific community, and sustain baseline surveillance. Such investments will ensure a rapid, structured scale-up of data collection and analyses required to mitigate the impact of the next emerging pathogen.

Contributors

SB, AC, and PN were involved in the conceptualisation of the study and the acquisition of funding. SB, NI, OJW, AC, and PN wrote the original draft, and prepared the data visualisation included in the work. All authors reviewed the manuscript and gave critical review, commentary, or revision at all stages of the manuscript preparation.

Declaration of interests

AC has received payment from Pfizer for teaching mathematical modelling of infectious diseases. All other authors declare no competing interests. SG was supported by the project SORMAS@DEMIS of the German Ministry of Health. SC acknowledges financial support from the EU's Horizon 2020 research and innovation programme under grants 874/35 (VEO) and 101003589 (RECOVER), the Investissement d'Avenir programme, the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases programme (grant ANR-10-LABX-62-IBEID), Santé Publique France, the INCEPTION project (PIA/ANR-16-CONV-0005), AXA, and Groupama. OJW was supported by a Schmidt Science Fellowship in partnership with the Rhodes Trust. NI is currently employed by the Wellcome Trust. The Wellcome Trust had no role in the preparation of the manuscript or the decision to publish. AC, SB, NI, OJW, MLR-C, and PN acknowledge funding from the Medical Research Centre (MRC) Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK MRC and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the EU. AC was supported by the Academy of Medical Sciences Springboard scheme, funded by the AMS, Wellcome Trust, UK Department for Business, Energy and Industrial Strategy, the British Heart Foundation, and Diabetes UK (reference SBF005\1044). AC acknowledges funding from the National Institute for Health and Care Research (NIHR) Health Protection Research Unit in Modelling and Health Economics, a partnership between the UK Health Security Agency, Imperial College London, and the London School of Hygiene & Tropical Medicine (grant code NIHR200908), and from the International Society for Infectious Diseases (Mapping the Risk of International Infectious Disease Spread II). AC and SB acknowledge funding from Imperial College London through the European Partners Fund. The funding was used to organise a workshop that brought together leading experts from the UK, Germany, and France, and public health experts from the European Centre for Disease Prevention, International Society for Infectious Diseases, and WHO. PA has an unpaid advisory role on the Advisory Council for Epivere.

Acknowledgments

The views expressed are those of the authors and not necessarily those of the NIHR, UK Health Security Agency, or the Department of Health and

Social Care. The funders had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

References

- 1 Kogan NE, Clemente L, Lautaud P, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Sci Adv* 2021; **7**: eabd6989.
- 2 Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med* 2020; **26**: 463–64.
- 3 Kraemer MUG, Scarpino SV, Marivate V, et al. Data curation during a pandemic and lessons learned from COVID-19. *Nat Comput Sci* 2021; **1**: 9–10.
- 4 Kucharski AJ, Hodcroft EB, Kraemer MUG. Sharing, synthesis and sustainability of data analysis for epidemic preparedness in Europe. *Lancet Reg Health Eur* 2021; **9**: 100215.
- 5 Cori A, Donnelly CA, Dorigatti I, et al. Key data for outbreak evaluation: building on the Ebola experience. *Philos Trans R Soc Lond B Biol Sci* 2017; **372**: 20160371.
- 6 Kaiser R, Coulombe D, Baldari M, Morgan D, Paquet C. What is epidemic intelligence, and how is it being improved in Europe? *Euro Surveill* 2006; **11**: E060202.4.
- 7 ProMED-mail. Subject: PRO/AH/EDR> Undiagnosed pneumonia—China (HU): RFI. Dec 19, 2019. <https://promedmail.org/promed-post/?id=6864153> (accessed April 7, 2022).
- 8 WHO. Pneumonia of unknown cause—China. Jan 5, 2020. <https://www.who.int/emergencies/diseases-outbreak-news/item/2020-DON229> (accessed May 6, 2022).
- 9 Virological. Novel 2019 coronavirus genome. Jan 11, 2020. <https://virological.org/t/novel-2019-coronavirus-genome/319> (accessed April 7, 2022).
- 10 Menkir TF, Chin T, Hay JA, et al. Estimating internationally imported cases during the early COVID-19 pandemic. *Nat Commun* 2021; **12**: 311.
- 11 Xu M, Du Z, Shan S, et al. RiskEstim: a software package to quantify COVID-19 importation risk. *Front Phys* 2022; **10**: 835992.
- 12 Smolinski MS, Crawley AW, Olsen JM, Jayaraman T, Libel M. Participatory disease surveillance: engaging communities directly in reporting, monitoring, and responding to health threats. *JMIR Public Health Surveill* 2017; **3**: e62.
- 13 Abbott S, Hellewell J, Munday J, Funk S. The transmissibility of novel coronavirus in the early stages of the 2019–20 outbreak in Wuhan: exploring initial point-source exposure sizes and durations using scenario analysis. *Wellcome Open Res* 2020; **5**: 17.
- 14 Imai N, Dorigatti I, Cori A, Riley S, Ferguson N. Report 1: estimating the potential total number of novel coronavirus cases in Wuhan City, China. Imperial College London, 2020. <https://doi.org/10.25561/77149> (accessed April 7, 2022).
- 15 Abbott S, Hellewell J, Thompson RN, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res* 2020; **5**: 112.
- 16 Hay JA, Kennedy-Shaffer L, Kanjilal S, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science* 2021; **373**: eabh0635.
- 17 Popa A, Genger J-W, Nicholson MD, et al. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci Transl Med* 2020; **12**: eabe2555.
- 18 Laxminarayan R, Wahl B, Dudala SR, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* 2020; **370**: 691–97.
- 19 Hu S, Wang W, Wang Y, et al. Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *Nat Commun* 2021; **12**: 1533.
- 20 Volz E, Mishra S, Chand M, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 2021; **593**: 266–69.
- 21 Dean N. Tracking COVID-19 infections: time for change. Nature Publishing Group UK, Feb 8, 2022. <https://doi.org/10.1038/d41586-022-00336-8> (accessed April 7, 2022).
- 22 Directorate-General for Environment. Coronavirus response: monitoring of wastewater contributes to tracking coronavirus and variants across all EU countries. March 17, 2022. https://ec.europa.eu/environment/news/coronavirus-response-monitoring-wastewater-contributes-tracking-coronavirus-and-variants-across-all-2022-03-17_en (accessed April 11, 2022).

23 Hall VJ, Foulkes S, Charlett A, et al. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). *Lancet* 2021; **397**: 1459–69.

24 RECOVERY Collaborative Group, Horby P, Lim WS, et al. Dexamethasone in hospitalized patients with Covid-19—preliminary report. *N Engl J Med* 2021; **384**: 693–704.

25 Ader F. Protocol for the DisCoVeRy trial: multicentre, adaptive, randomised trial of the safety and efficacy of treatments for COVID-19 in hospitalised adults. *BMJ Open* 2020; **10**: e041437.

26 Barnes O, Burn-Murdoch J. UK statistics tsar rebukes Health Security Agency over flawed jabs data. Financial Times, Oct 29, 2021. <https://www.ft.com/content/a51f85a2-e338-4057-81bc-75447b7660b3> (accessed May 17, 2022).

27 Nouvellet P, Bhatia S, Cori A, et al. Reduction in mobility and COVID-19 transmission. *Nat Commun* 2021; **12**: 1090.

28 Universiteit Hasselt. The CoMix study. <https://www.uhasselt.be/en/aparte-sites-partner-en/epipose/the-comix-study> (accessed April 13, 2022).

29 Sharma M, Mindermann S, Rogers-Smith C, et al. Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. *Nat Commun* 2021; **12**: 5820.

30 Hale T, Angrist N, Goldszmidt R, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* 2021; **5**: 529–38.

31 ACAPS. COVID-19 Government Measures Dataset. March 18, 2020. <https://www.acaps.org/covid-19-government-measures-dataset> (accessed April 20, 2022).

32 European Centre for Disease Prevention and Control, the Joint Research Centre of the European Commission. Response measures database. <https://covid-statistics.jrc.ec.europa.eu/RMeasures> (accessed April 20, 2022).

33 WHO. Tracking public health and social measures: a global database. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm> (accessed April 20, 2022).

34 Jones SP, Imperial College London Big Data Analytical Unit, YouGov. Imperial College London YouGov Covid Data Hub, v1.0. 2020. <https://github.com/YouGov-Data/covid-19-tracker> (accessed April 20, 2022).

35 Xu B, Gutierrez B, Mekaru S, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* 2020; **7**: 106.

36 Kraemer MUG, Tegally H, Pigott DM, et al. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. *Lancet Infect Dis* 2022; **22**: 941–42.

37 European Centre for Disease Prevention and Control. COVID-19 situation update for the EU/EEA, as of 26 May 2022. May 26, 2022. <https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea> (accessed May 27, 2022).

38 Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci USA* 2004; **101**: 6146–51.

39 Public Health England. PHE statement on delayed reporting of COVID-19 cases. GOV.UK. Oct 4, 2020. <https://www.gov.uk/government/news/phe-statement-on-delayed-reporting-of-covid-19-cases> (accessed April 11, 2022).

40 Sadilek A, Caty S, DiPrete L, et al. Machine-learned epidemiology: real-time detection of foodborne illness at scale. *NPJ Digit Med* 2018; **1**: 36.

41 Desai A, Nouvellet P, Bhatia S, Cori A, Lassmann B. Data journalism and the COVID-19 pandemic: opportunities and challenges. *Lancet Digit Health* 2021; **3**: e619–21.

42 Conlan AJK, Klepac P, Kucharski AJ, et al. Human mobility data from the BBC pandemic project; published online Feb 23. *medRxiv* 2021. <https://doi.org/10.1101/2021.02.19.21252079> (preprint).

43 WHO. Epidemic Intelligence from Open Sources (EIOS). <https://www.who.int/initiatives/eios> (accessed June 24, 2022).

44 Prime Minister's Office. Prime Minister sets out plan for living with COVID. GOV.UK. Feb 21, 2022. <https://www.gov.uk/government/news/prime-minister-sets-out-plan-for-living-with-covid> (accessed April 11, 2022).

45 Richardson S. It is vital that we retain some reliable covid-19 surveillance tools. *BMJ* 2022; **377**: o900.

46 Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. The impact of biases in mobile phone ownership on estimates of human mobility. *J R Soc Interface* 2013; **10**: 20120986.

47 Directorate-General for Health and Food Safety. COVID-19—sustaining EU preparedness and response: looking ahead. April 26, 2022. https://ec.europa.eu/health/publications/covid-19-sustaining-eu-preparedness-and-response_en (accessed May 27, 2022).

48 Nature Research Custom Media, Thermo Fisher Scientific. Multiplex testing helps provide critical insights as respiratory viruses surge. <https://www.nature.com/articles/d42473-021-00318-w> (accessed April 11, 2022).

49 WHO. Coronavirus disease (COVID-19) technical guidance: the unity studies: early investigation protocols. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations> (accessed April 12, 2022).

Copyright © 2023 Elsevier Ltd. All rights reserved.