

Modelling COVID-19

Alessandro Vespignani, Huaiyu Tian, Christopher Dye, James O. Lloyd-Smith, Rosalind M. Eggo, Munik Shrestha, Samuel V. Scarpino, Bernardo Gutierrez, Moritz U. G. Kraemer, Joseph Wu, Kathy Leung and Gabriel M. Leung

Abstract | As the COVID-19 pandemic continues, mathematical epidemiologists share their views on what models reveal about how the disease has spread, the current state of play and what work still needs to be done.

The challenges of modelling in a ‘war’

Alessandro Vespignani. Although there are inherent limitations to predictions in complex socio-technical systems, in recent years mathematical and computational models have successfully forecasted the size of epidemics and have been used to communicate the risks of uncurbed infectious disease outbreaks. Mathematical and computational models are ideal as forecasting tools and can also provide situational awareness when we lack good data, or define counterfactual scenarios that help disentangle the impact of pharmaceutical interventions and public health policies.

Peculiar to the field of computational epidemiology is the distinction between two different kinds of work: ‘peace time’ research when there are no health emergencies or threats, and what we call ‘war time’, when there are emergencies like the COVID-19 epidemic. During war time, we have to work with limited data, a constantly changing landscape and a lot of assumptions. The work must often be tactical, and what has been produced the day before often must be completely revised the day after because a new piece of information has arrived. At the same time, the challenges faced during infectious disease threats set the questions and problems for the rigorous and foundational research that allows the field to advance after the emergency is gone.

Early containment measures in China

Huaiyu Tian and Christopher Dye.

As the COVID-19 epidemic spread across China from Wuhan city in early 2020, it was vital to find out how to slow or stop it. We could not investigate the effectiveness of control measures in a controlled experiment or a clinical trial, and instead

had to rely on statistical and mathematical modelling. However, precise evaluation of particular interventions requires substantial data or assumptions: not only accurate characterization of the epidemic process itself but also government actions and even human behaviours, such as the three billion trips taken over the Chinese New Year holiday. We therefore constructed models in conjunction with a growing geocoded database on coronavirus epidemiology, human movement and public health interventions.

We took two approaches to the analysis. The first exploited natural variation in the distribution of COVID-19 cases, and in the type and timing of interventions. On the basis of statistical tests of association carried out with general linear models, we found that the unprecedented Wuhan city travel ban (affecting 11 million people) slowed the dispersal of infection to other cities by 3 days¹, delaying epidemic growth elsewhere in China. We found, too, that Chinese cities that pre-emptively implemented control measures — such as suspending intra-city public transport, closing entertainment venues and banning public gatherings — reported in the first week of their outbreaks one-third fewer cases than cities that started control later. Our second approach to analysis built these findings into a dynamic mathematical model, from which we calculated that China’s national emergency response prevented hundreds of thousands of cases that we otherwise expected to see during the first 50 days of the epidemic.

Causes and effects of superspreading

James O. Lloyd-Smith. As with severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) before it,

the epidemiology of COVID-19 has been punctuated by conspicuous superspreading events, in which an infected person transmits the virus to many more people than average. The average transmissibility of a pathogen is quantified by its basic reproduction number, R_0 , which is a bedrock concept in infectious disease dynamics. Yet biological, social and environmental factors — aided by a good dose of happenstance — give rise to significant individual variation around this average. This holds true for all pathogens to varying degrees, but evidence suggests that the emerging coronaviruses causing SARS, MERS and COVID-19 are systematically prone to superspreading^{2,3}.

Why do we care? Mathematically, for a given R_0 , a pathogen with more superspreaders must also have more infected individuals who do not contribute to onward spread. Such individual variation makes transmission chains more likely to die out, and outbreaks rarer but more explosive, than if every case was an average transmitter². This variation matters most when case numbers are small (early in the pandemic, or after successful outbreak suppression if the population remains susceptible), as countries try to prevent establishment of community transmission. Health officials must guard against complacency, recognizing that many importations will fade out by chance but a minority will ignite outbreaks that expand with shocking speed.

Physicists can help by working to learn the causes of superspreading events. Many involve alternative modes of spread, such as airborne transmission, but inferring such mechanisms from imperfect data in complex environments raises many technical challenges. There are also unsolved mathematical and statistical problems in untangling the influences of individual biological variation and dynamic social contact networks that govern transmission opportunities. By understanding these causes, we can better target interventions to decelerate the spread of COVID-19.

Contact tracing and isolation

Rosalind M. Eggo. SARS-CoV-2 is a new pathogen with some key characteristics that many mathematical modellers were concerned could emerge together: quite high mortality and efficient transmission

between people. As the virus has made its way around the world, transmission models are proving invaluable to the response effort. Some models aim to provide answers that public health decision-makers need in real time: how effective possible interventions are likely to be. Our group recently modelled a classical intervention against infectious diseases: contact tracing and isolation⁴, in which the contacts of known cases are found, and, if they show symptoms, are isolated quickly. Doing so therefore

decreases the average number of new infections that each infected person creates, a quantity modellers call the reproduction number. By developing a model for this system, we were able to determine what fraction of a case's contacts must be found and isolated to control a new outbreak, based on the best-possible information on transmission and disease available at that time. However, for some parameters, such as for how much transmission occurs before the onset of symptoms, there was

little information. For these parameters, we provided scenarios of a range of values, so that decision-makers could see the effect under different regimes, and so that if that information became available in the future, the model would still be informative.

The contributors

Alessandro Vespignani is an Italian-American physicist, currently Sternberg Family Distinguished University Professor of Physics, Computer Science and Health Sciences at Northeastern University in Boston, USA. He is the director of the Network Science Institute, and is best known for his work on complex networks and his contributions to computational epidemiology by developing specific tools for analysing the global spread of epidemics.

Huaiyu Tian is a Professor of Ecology and Evolution of Infectious Disease at Beijing Normal University, China, and Oxford Martin Visiting Fellow at the University of Oxford. His interdisciplinary research focuses on the mechanistic processes that link biological and ecological change to disease dynamics. His lab combines geospatial computing, field surveillance, molecular epidemiology and ecological modelling.

Christopher Dye is a Visiting Professor of Zoology and Oxford Martin Visiting Fellow at the University of Oxford. He is a former Director of Strategy at the World Health Organization, and is currently editor of coronavirusexplained.ukri.org/en/, a website that explains the science of coronavirus outbreaks, hosted by UK Research and Innovation (UKRI).

James O. Lloyd-Smith is a professor in the Department of Ecology and Evolutionary Biology at the University of California, Los Angeles. His research explores the ecological and evolutionary dynamics of infectious disease in animal and human populations, with emphasis on the emergence of novel pathogens. His group combines mathematical models, statistical analysis, and laboratory, clinical and field studies to study diseases such as monkeypox, leptospirosis, influenza and now COVID-19.

Rosalind M. Eggo is a mathematical modeller focusing on directly transmitted viral pathogens and the severe outcomes resulting from infection. She works at the London School of Hygiene and Tropical Medicine.

Munik Shrestha is a postdoctoral researcher at Northeastern University in the Network Science Institute. He has contributed fundamental theory on message-passing algorithms for estimating the time-varying statistics of epidemics. He received his Ph.D. in physics from the University of New Mexico and was a Santa Fe Institute graduate fellow.

Samuel V. Scarpino is an assistant professor in the Network Science Institute at Northeastern University, with appointments in marine and environmental sciences, physics, and health sciences. He is also an ISI Foundation complexity fellow. Scarpino earned a Ph.D. in evolution, ecology and behaviour from the University of Texas at Austin and was a Santa Fe Institute Omidyar postdoctoral fellow.

Bernardo Gutierrez is a D.Phil. student in the Department of Zoology at the University of Oxford, and a guest researcher at Universidad San Francisco de Quito. He studies the evolution of emerging viruses and the integration of epidemiological and genomic data to track outbreaks. He began contributing to the Open COVID-19 Data Working Group during the start of the SARS-CoV-2 epidemic in January 2020.

Moritz U. G. Kraemer is a research fellow at the University of Oxford and founder of the Open COVID-19 Data Working Group. Moritz is an epidemiologist working on the integration of multiple data streams to better understand the dynamics of emerging infectious diseases.

Joseph Wu is a professor in the School of Public Health at the University of Hong Kong. He specializes in disease modelling and data science.

Kathy Leung is an infectious disease epidemiologist at the University of Hong Kong. She specializes in mathematical modelling of communicable and non-communicable diseases, including COVID-19, influenza, Middle East respiratory syndrome, human papillomavirus, colorectal cancer and breast cancer.

Gabriel M. Leung is an infectious disease epidemiologist, Dean of Medicine and Zimmern Professor of Population Health at the University of Hong Kong.

Realistic models require better data Munik Shrestha and Samuel V. Scarpino.

While firefighters in Australia were bringing the devastating 2019–2020 bushfires under control, a novel coronavirus was spreading rapidly through China, transitioning from a local outbreak into a global pandemic. Remarkably, the mathematical models scientists use to study the spreading of both fires and pathogens share many features. An important similarity is that in both cases, models and data show that the economic and health costs do not increase linearly with the number of people infected or hectares burned; instead, they explode, going from imperceptibly small to unimaginably large in what seems an instant. Herein lies the central challenge of studying such systems: both involve random, multiplicative processes characterized by exponential growth and discontinuous phase transitions, where the equilibrium number of cases can go from nearly zero to infinitely many in actually an instant. Accordingly, their mathematical properties depend on rare, exponentially costly and hard-to-study events.

However, there is at least one crucial difference between fires and epidemics: how readily the two systems can be monitored and measured. Whereas the effects of fires can often be seen with the naked eye, public health records contain information on only the largest outbreaks. As a result, although ecologists can study the past effects of numerous fires, both large and small, epidemiologists rely primarily on data from rare, population-spanning events. Why does this matter? At the phase transition, between local outbreak and global pandemic, the proportion infected can be scale free, meaning measuring only the large outbreaks leads to immeasurably large bias. Mathematical epidemiologists have developed models that can inform policy that saves lives, but the need remains for more refined instruments to gather empirical observations and test those models.

Open, detailed case data Bernardo Gutierrez and Moritz U. G.

Kraemer. Making accurate assessments about the spread of infectious diseases relies on the availability of robust

epidemiological data, a scarce commodity during growing epidemics and when resources are scarce. Official statistics are usually presented as aggregated data (for instance, newly reported confirmed cases by day) and tend to be shared on a limited basis, restricting access to the data by the scientific community at large. Although such information is important in tracking outbreaks, it does not include the tremendous detail of epidemiological information that is available across different platforms, such as news outlets, social media and official government health reports. Furthermore, epidemiological reporting standards vary considerably across these platforms and contain varying degrees of detail, making automated data collection challenging.

As an alternative approach, on the basis of experiences from past outbreaks, the Open COVID-19 Data Working Group opted for a crowdsourcing approach. An international team of volunteers curate data sources by hand and compile the data in a standardized format. This format presents epidemiological data at an individual case level, which allows the extraction of remarkably detailed information on case demographics, travel histories and high-resolution geographical distributions⁵. More importantly, the constant screening of various information sources in real time adds value to this approach and makes it an important tool for disease surveillance. A key aspect to guarantee its utility is the focus on data sharing through Google Sheets and the software development platform GitHub.

Data are openly accessible and [updated daily](#)⁶, and can be viewed [here](#).

Antibody tests are imminently needed

Joseph Wu, Kathy Leung and Gabriel M. Leung. Rapid and reliable assessment of the clinical severity of pandemic pathogens such as SARS-CoV-2 (the virus that causes COVID-19) is a top priority in pandemic response. In particular, characterization of the infection–fatality risk (the probability of dying among infected individuals) and symptomatic

case–fatality risk (the probability of dying among those who develop symptoms) relies on accurate estimates of the true epidemic size. Reported case counts are inevitably biased by the proportion of infections that are symptomatic, care-seeking behaviour and the availability of tests. Although these biases could be partially addressed by developing transmission models to analyse all available epidemiological and clinical data, a synergistic and complementary solution is to use seroepidemiological studies. These use measurements of antibody response across a population to infer infective exposure to (and, by extension, immunity against) the pandemic pathogen, and provide the most direct and reliable data for estimating true epidemic size. Integrating seroepidemiological data into transmission models can greatly reduce the uncertainty in the parameter estimates of clinical severity and transmission dynamics. At the time of writing, serological tests for COVID-19 are being developed and validated. Long-term longitudinal serological follow-up of recovered individuals is crucial for characterizing the underlying immunodynamics. For instance, it can reveal the strength and duration of protective immunity and thus the probability of reinfection. As such, large-scale serosurveys should be conducted regularly and prospectively planned over the course of the different waves of the pandemic and beyond to provide serological data from different age groups.

Alessandro Vespignani^{1,2}, Huaiyu Tian⁵, Christopher Dye⁶, James O. Lloyd-Smith⁵, Rosalind M. Eggo⁶, Munik Shrestha¹, Samuel V. Scarpino⁷, Bernardo Gutierrez^{4,7}, Moritz U. G. Kraemer^{4,8,9}, Joseph Wu¹⁰, Kathy Leung¹⁰ and Gabriel M. Leung¹⁰

¹Network Science Institute, Northeastern University, Boston, MA, USA.

²ISI Foundation, Turin, Italy.

³State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China.

⁴Department of Zoology, University of Oxford, Oxford, UK.

⁵Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA.

⁶Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.

⁷School of Biological and Environmental Sciences, Universidad San Francisco de Quito USFQ, Quito, Ecuador.

⁸Harvard Medical School, Harvard University, Boston, MA, USA.

⁹Boston Children's Hospital, Boston, MA, USA.

¹⁰WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

✉e-mail: a.vespignani@neu.edu; tianhuaiyu@gmail.com; christopher.dye@zoo.ox.ac.uk; jilloydsmith@ucla.edu; r.eggo@lshtm.ac.uk; m.shrestha@northeastern.edu; s.scarpino@northeastern.edu; bernardo.gutierrez@zoo.ox.ac.uk; joewu@hku.hk; moritz.kraemer@zoo.ox.ac.uk; ksmleung@hku.hk; deanmed@hku.hk

<https://doi.org/10.1038/s42254-020-0178-4>

Published online: 06 May 2020

1. Tian, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* <https://doi.org/10.1126/science.abb6105> (2020).
2. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
3. Kucharski, A. J. & Althaus, C. L. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Eurosurveillance* **20**, 14–18 (2015).
4. Hellewell, J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
5. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **21**, eabb4218 (2020).
6. Xu, B. & Kraemer, M. U. G. Open access epidemiological data from the COVID-19. *Lancet Infect. Dis.* **3099**, 30119 (2020).

Acknowledgements

J.O.L.-S. is supported by the US National Science Foundation (DEB-1557022), the Strategic Environmental Research and Development Program (SERDP, RC-2635) of the US Department of Defense, and DARPA PREEEMPT D18AC00031. The content of the information does not necessarily reflect the position or the policy of the US government, and no official endorsement should be inferred. M.U.G.K. acknowledges support from the Oxford Martin School. R.M.E. acknowledges funding from HDR UK (grant: MR/S003975/1).

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2020