PERSPECTIVE | FOCUS
https://doi.org/10.1038/s41591-021-01278-w

nature medicine

Nowcasting epidemics of novel pathogens: lessons from COVID-19

Joseph T. Wu[1,2] ✉, Kathy Leung[1,2], Tommy T. Y. Lam[1,2,3,4], Michael Y. Ni[1,5,6], Carlos K. H. Wong[7,8], J. S. Malik Peiris[1,9] and Gabriel M. Leung[1,2]

Epidemic nowcasting broadly refers to assessing the current state by understanding key pathogenic, epidemiologic, clinical and socio-behavioral characteristics of an ongoing outbreak. Its primary objective is to provide situational awareness and inform decisions on control responses. In the event of large-scale sustained emergencies, such as the COVID-19 pandemic, scientists need to constantly update their aims and analytics with respect to the rapidly evolving emergence of new questions, data and findings in order to synthesize real-time evidence for policy decisions. In this Perspective, we share our views on the functional aims, rationale, data requirements and challenges of nowcasting at different stages of an epidemic, drawing on the ongoing COVID-19 experience. We highlight how recent advances in the computational and laboratory sciences could be harnessed to complement traditional approaches to enhance the scope, timeliness, reliability and utility of epidemic nowcasting.

Epidemic nowcasting broadly refers to assessing the current state by understanding key pathogenic, epidemiologic, clinical and socio-behavioral characteristics of an ongoing outbreak (Fig. 1). Here we share our views on the functional aims, rationale, data requirements and challenges of nowcasting at different stages of an epidemic, drawing on the ongoing COVID-19 experience. Given the broad scope and highly multidisciplinary nature of epidemic research generally, and the overwhelming volume of work on COVID-19 in particular, we focus on the most directly relevant and salient aspects in epidemic nowcasting, illustrated by representative examples.

What is the causative pathogen, and where did it emerge?

When an emerging outbreak is first detected, the overriding nowcast priority is to identify the etiologic pathogen and its origin (which is often zoonotic for novel pathogens; for example, severe acute respiratory syndrome coronavirus (SARS-CoV)[1,2], Middle East respiratory syndrome coronavirus (MERS-CoV)[3,4] and SARS-CoV-2 (refs. [5,6])). To this end, public health practitioners and researchers conduct meticulous outbreak investigations that would ideally yield detailed line lists containing essential demographic (for example, age and occupation), epidemiologic (for example, exposure, travel history and contact lists), laboratory (for example, viral and serologic test results) and clinical (for example, medical records of symptoms, disease course, treatment outcomes and long-term prognosis) data of not only infected individuals but also those without disease who have been investigated[7], as well as their contacts. The most fundamental step in these initial investigations is the determination of case definitions, because all downstream operations, clinical management and data analyses are based on them. As the epidemic unfolds, health officials might revise these definitions to improve disease surveillance and control as they progressively learn more about the disease. For example, during the first wave of COVID-19 in China, the case definition was progressively broadened several times to allow more infections to be detected[8]. Each time the case definition was updated, the proportion of confirmed infections increased three- to sevenfold. As such, when interpreting case counts from different time periods or disparate health jurisdictions, caution should be taken to account for any temporal or contextual differences in case definitions.

Identifying the causative pathogen, its host population and the place of first emergence is critical for outbreak control and re-emergence prevention. Caution should be taken to distinguish the epicenter where cases were first reported from the place of first emergence; the two are not necessarily the same[9]. For instance, SARS-CoV was first detected and identified to be the causative agent of SARS in Hong Kong[10], but retrospective studies later traced its origin to Guangdong province in mainland China[11,12]. Advances in genomic epidemiology have rendered phylogenetics an indispensable tool for origin tracing. As such, data-sharing platforms that support real-time open access to pathogen sequences (for example, GISAID; https://www.gisaid.org/) are instrumental for epidemic nowcasting. Phylogenetics infers virus evolutionary history to elucidate[13]: (1) the time of emergence, as reflected by the depth of the tree root; and (2) whether the outbreak was caused by continual zoonoses or sustained human-to-human transmission. Although COVID-19 cases were first detected in Wuhan, the origin of SARS-CoV-2 remains elusive to date, mostly because animal and environmental samples from the early days of the outbreak have not been available for viral sequencing. In contrast, there is some evidence that SARS-CoV-2 might have emerged and circulated elsewhere unnoticeably before that[14,15]. Origin tracing thus requires global perspectives and multidisciplinary efforts, as demonstrated by the ongoing World Health Organization (WHO) mission that aims to resolve the origin of SARS-CoV-2 (ref. [16]).

Does the agent spread efficiently between humans?

A key question following identification of the etiological pathogen is whether transmission is limited to zoonosis or taking place efficiently between humans. Clusters of cases within families or transmission within health care settings are early clues indicating the latter.

[1]WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [2]Laboratory of Data Discovery for Health (D24H), Hong Kong, China. [3]State Key Laboratory of Emerging Infectious Diseases, School of Public Health, The University of Hong Kong, Hong Kong, China. [4]Joint Institute of Virology (Shantou University and The University of Hong Kong), Guangdong-Hongkong Joint Laboratory of Emerging Infectious Diseases, Shantou University, Shantou, China. [5]The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China. [6]Healthy High Density Cities Lab, HKUrbanLab, The University of Hong Kong, Hong Kong, China. [7]Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [8]Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [9]HKU-Pasteur Research Pole, The University of Hong Kong, Hong Kong, China. ✉e-mail: joewu@hku.hk

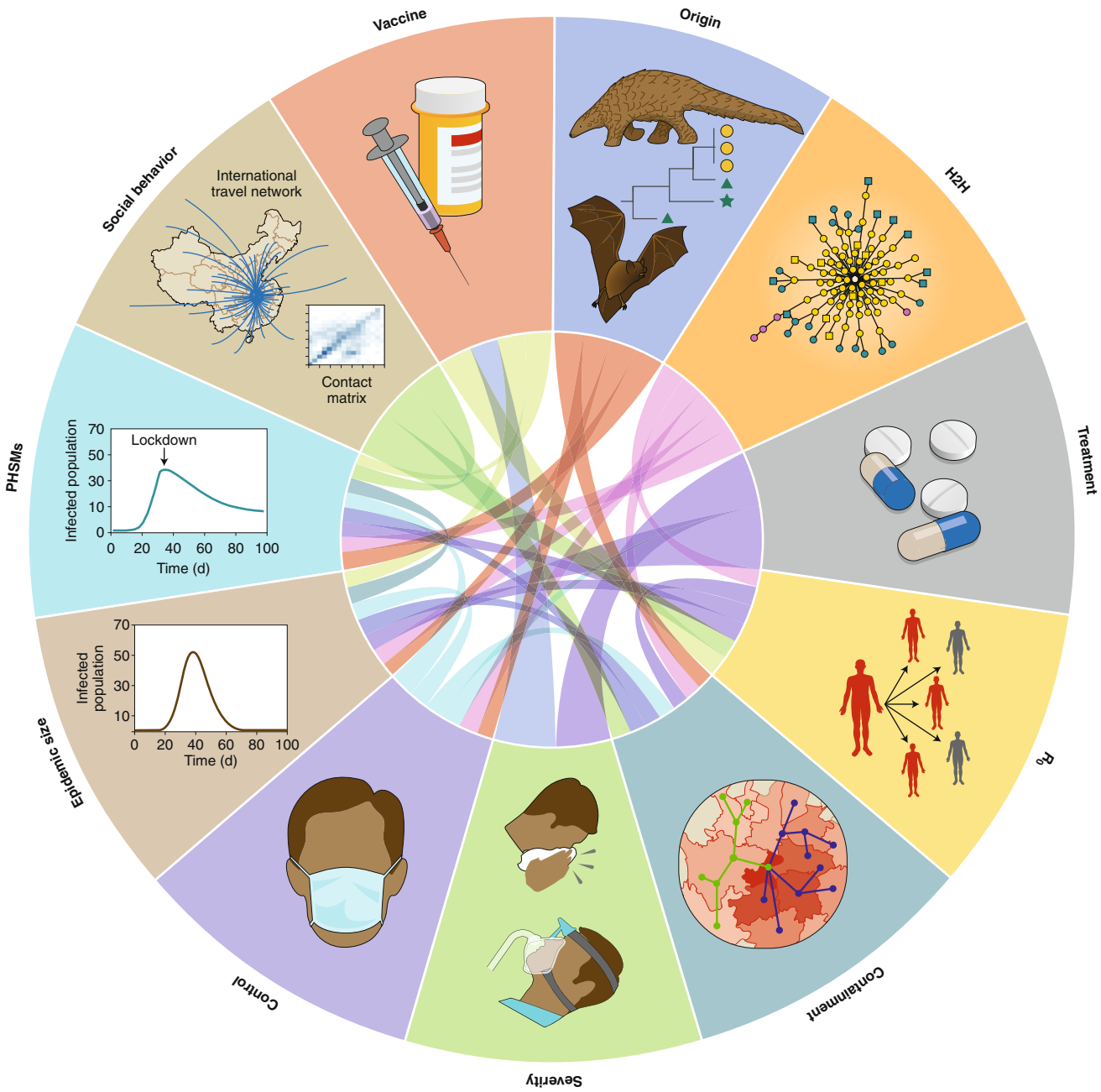

Fig. 1 | Main targets for epidemic nowcasting. The main targets of epidemic nowcasting are shown, highlighting the interconnectivity between all of the elements. H2H, human to human.

However, the detection of clusters does not prove human-to-human transmission, because exposure to a common source or multiple zoonotic sources may also explain such observations.

Detailed epidemiological investigations are crucial. Conclusive evidence of efficient human-to-human transmission for COVID-19 was provided through the investigation of a patient in Shenzhen (Guangdong province, China) who had no travel history to Wuhan and acquired infection through close contact with family members who had recently returned from Wuhan in early January 2020. Some of the other family members returning from Wuhan were also confirmed to have infection[17]. Since this family had no exposure to animals, game animal restaurants or the first recognized epicenter, Huanan Seafood Market in Wuhan, and they probably acquired infection from their relatives or others in Wuhan, this incident confirmed at least two generations of human-to-human infection.

In contrast, there are examples of very limited human-to-human transmission of zoonotic viruses, such as H5N1 influenza, without evidence of such a virus achieving sustained human-to-human transmission[18]. With MERS-CoV, there have been repeated explosive clusters of human-to-human transmission, particularly in nosocomial settings, both within the zoonotic epicenter (the Arabian Peninsula) and beyond (for example, the Republic of Korea in 2015), with some involving >100 infections and four or more generations of human-to human spread[19]. However, it has not become a pandemic, possibly because of early detection and implementation of aggressive containment measures and/or perhaps due to its inconsistent ability to spread efficiently from human to human in non-superspreading settings. The transition between exclusive zoonotic transmission and efficient human-to-human spread can be prolonged, throughout which, the strength of

epidemiologic evidence would change in a crescendo pattern such that cross-sectional assessment at any single point in time could be inconclusive.

Is there effective clinical treatment?

For the first few dozen, or even 100, individuals infected with a novel pathogen, clinical management other than supportive treatment is usually based on past experience with similar microbes and previous animal or human experimental studies if available. Clinical therapeutics against COVID-19 took reference from SARS and MERS, as well as more generally, with consequences of immune dysregulation (for example, cytokine storm) seen with infections of different influenza strains.

Early randomized trials from Beijing and Hong Kong showed that lopinavir–ritonavir[20] and remdesivir, respectively, as single agents[21] did not have significant clinical benefit for hospitalized patients with severe disease, but an antiviral plus immunomodulatory triple combination might be effective for hospitalized patients with mild-to-moderate disease[22]. Nevertheless, open-label design, insufficient power or short follow-up periods posed methodological challenges to these findings.

Subsequently, the WHO launched the global SOLIDARITY trial, which found that four repurposed drugs (that is, remdesivir, hydroxychloroquine, lopinavir and interferon beta-1a) had little or no effect on reducing the overall mortality risk, ventilation initiation or duration of hospital stay among hospitalized patients[23], although the patient profile and timing of administration differed between different study sites and from the earlier Chinese trials.

In parallel, the United Kingdom led the world in robustly evaluating therapeutics against SARS-CoV-2 infection with the RECOVERY trial. Dexamethasone was found to confer a lower 28-d mortality risk among patients hospitalized with respiratory support[24], but lopinavir–ritonavir[25], hydroxychloroquine[26] and azithromycin[27] failed to show clinical benefit.

The SOLIDARITY and RECOVERY trials have been defining exemplars of best practice during the exigencies of the COVID-19 pandemic, and set a new benchmark for robust clinical science. Corollary activities, including rapid institutional review of protocols, operational execution of multi-country and multi-site studies, pervasive technology support and resource mobilization, including financing and study personnel, would be co-requisites of success.

In addition, emergency use authorization was rapidly granted by the US Food and Drug Administration for two monoclonal antibody regimens found to be efficacious in mild-to-moderate (mostly early-stage) COVID-19 disease[28,29]. Trials evaluating their prophylactic use (for example, in rare instances of vaccine contraindication) are ongoing. The role of novel biologics that are precisely designed for particular microbial threats will probably expand in future significant epidemics.

Finally, there will be instances where randomized evidence cannot be acquired quickly or realistically. During the 2009 H1N1 pandemic, there was intense, mostly misguided, controversy over the effectiveness of oseltamivir, particularly in those suffering from moderate-to-severe illness. This instance illustrates how careful interpretation of observational data is critical to informing public health practice, where inappropriate skepticism can be as damaging as casual acceptance of study findings[30,31]. For COVID-19, despite the success of the SOLIDARITY and RECOVERY trials and the monoclonal antibody trials, knowledge gaps remain regarding the timing of administration of single agents and combination treatment. For instance, propensity score-adjusted observational data of consecutive Chinese patients of Anhui and Hong Kong found that early administration of interferon beta-1b alone or in combination with oral ribavirin might reduce mortality and serious complications, as well as hasten recovery[32].

How easily does the pathogen spread?

The epidemic potential of a pathogen is characterized by its reproductive number, R, which is defined as the average number of secondary cases generated by a typical primary case[33]. The basic reproductive number, R_0, is a special case of R, namely when the entire population is susceptible. If $R<1$, the outbreak will die out without causing widespread transmission. Otherwise, the outbreak could grow exponentially, the probability and speed of which increase with R. Early estimates of R are inevitably fraught with substantial uncertainties due to data paucity and limited understanding of the natural history and transmission dynamics of the disease. Conventional approaches estimate R from line list data or case incidence time series and require careful considerations for under-reporting or other ascertainment biases (for example, larger clusters might be more likely to be identified)[34]. R can be estimated from the line list if it contains sufficient data for reconstructing the transmission chains or compiling the cluster size distribution[35,36]. However, for most outbreak settings, line list data are not sufficiently detailed to support these approaches, in which case, mathematical models can be used to estimate R from case incidence time series[37,38]. These modeling approaches typically require knowledge about the generation time (time between infection in an infector–infectee pair) or serial interval (time between symptoms onsets in an infector–infectee pair). The serial interval is often used to approximate the generation time because the onset time is more readily identifiable than the time of infection[39]. Robust estimation of the generation time and serial interval requires reliable contact-tracing data and adjustments for right truncation, the effect of interventions (for example, isolation), susceptible depletion and other factors[40,41]. Uncertainty regarding real-time estimates of generation time and serial interval was one of the main sources of uncertainty in nowcasted R estimates of COVID-19 (ref. [42]). For example, the first published estimate of the serial interval for COVID-19 was based on only six infector–infectee pairs in the Wuhan line list (the original epicenter of COVID-19), hence the high uncertainty in the resulting R estimate (95% credible interval = 1.4–3.9)[43]. To accelerate and broaden data availability for epidemic nowcasting, the Open COVID-19 Data Curation Group was launched during the early months of the pandemic to create a daily-updated open-access repository of individual-level information on laboratory-confirmed cases around the globe[44].

R can also be estimated from the growth rates of effective infected population sizes inferred from coalescent models[45] or directly from transmission models for the virus phylogeny and sequence data[46,47]. The R estimates for SARS-CoV-2 obtained by these genomic-based approaches were comparable to those obtained using traditional case-based approaches[48,49]. Genomic-based approaches are particularly useful when case reports are unavailable or unreliable, and the sequence-inferred genealogy among the viral strains preserves information of transmission history (for example, when transmission had been taking place unnoticed before the first reported case). However, the robustness of genomic-based methods is limited by the phylogenetic signal and resolution presented by the viral sequences.

Can the pathogen be contained locally?

Pathogens with pandemic potential could spread from the epicenter to other locations before or soon after the outbreak is first detected. This was the case for both the 2009 influenza pandemic (which emerged in Mexico, but the first case was detected in the United States[50]) and the COVID-19 pandemic (which was first reported in China and then detected in Thailand and Japan within a few weeks[51]). Disease transmission models could be used to estimate R and epidemic sizes at the epicenter from international case exportation time series and air travel data[52]. If near-real-time, aggregated human mobility data are available at high spatio-temporal resolution,

spatial transmission models can be used to nowcast the geographical spread of the pathogen beyond its epicenter before those epidemics become apparent in their localities[53]. The digital footprints underlying our everyday activities on mobile devices can be used to compile such data, a resource that did not exist until the current age of mobile ubiquity. Shortly after Wuhan was locked down on 23 January 2020, nowcasts based on mobility data provided by Tencent revealed that: (1) dozens to hundreds of COVID-19 cases had already spread to multiple major cities in China; and (2) nationwide epidemics would be inevitable within weeks unless local prevalence could be suppressed to near-elimination levels[54]. These findings prompted national authorities to implement a stringent national and subnational network of cordon sanitaires, with intensive active-case finding and contact tracing in all major cities[55].

How severe is the disease?

The clinical severity of infectious diseases is typically measured in terms of infection fatality risk (IFR) and symptomatic case fatality risk[56]. These severity estimates are used to project how health care demand and deaths would scale with epidemic size. Nowcasting severity and its risk factors is therefore crucial for helping health officials to strike the optimum balance between the socioeconomic costs of control measures and the health threat posed by an unmitigated epidemic[57]. However, real-time accurate estimation of severity is challenging because it requires reliable counts of severe cases and deaths, accurate estimation of epidemic size and careful adjustments for various sources of bias (for example, preferential ascertainment of severe cases and delayed reporting of death)[56]. For example, the Wuhan COVID-19 death toll was revised upward by 50% weeks after China lifted its nationwide lockdown because deaths from COVID-19 were difficult to count during the first wave in the absence of widespread testing[58]. Shortly after COVID-19 began to spread around the globe, its IFR in China was estimated to be 0.66%, with an increasing gradient with age[59]. Based on these age-specific IFRs and the nowcasted R estimates, it was projected that an unmitigated COVID-19 epidemic in the United Kingdom and United States would result in more than 510,000 and 2.2 million deaths, respectively[60]. These findings forewarned policymakers of the necessity for stringent public health and social measures (PHSMs) until safe and effective vaccines had become widely available, in order to prevent hospital overloads and mass deaths. By the same token, tens of billions of US dollars were invested to initiate national and cross-national programs (for example, Operation Warp Speed in the United States) that aim to facilitate and accelerate the development, manufacturing and distribution of COVID-19 vaccines, therapeutics and diagnostics[61].

Can pathogen spread be interrupted?

Health protection and the precautionary principle in public health mandate swift containment of emerging outbreaks regardless of their eventual epidemic potential. Containment measures invariably comprise isolation of cases and quarantine of exposed individuals who have not (yet) exhibited evidence of infection[62]. The main purposes of quarantine are to: (1) prevent potentially asymptomatic and presymptomatic individuals from unknowingly infecting other people; and (2) isolate and treat cases as soon as they are detected during quarantine. Health officials determine the quarantine duration based on the incubation period, which is the time from infection to symptom onset[63]. The 95th percentile of the incubation period distribution of COVID-19 was estimated to be around 12 d[64]. Accordingly, most jurisdictions require a quarantine duration of at least 14 d to ensure that nearly all quarantined individuals who are infected and would go on to show symptoms would be detected before release. If testing is co-implemented in quarantined persons, asymptomatically infected individuals would also be identified. The line list is typically the first source of data for estimating the incubation period and its dependence on age, disease severity and other factors. This highlights the importance of collating temporal information on exposure and symptoms during outbreak response—an essential task that is often overlooked or underappreciated when investigators or tracers are overwhelmed. Notably, only ten of the first 425 confirmed cases on the Wuhan line list provided the data necessary for estimating the incubation period[43].

The effectiveness of control measures initiated by symptoms onset (that is, including isolation and quarantine) can be substantially reduced if a large proportion of disease transmission occurs when the infectors have no symptoms (that is, when they are presymptomatic or asymptomatic)[65,66]. As such, nowcasting the contribution of presymptomatic and asymptomatic individuals to overall disease transmission is essential for gauging the speed and coverage of case finding and contact tracing needed to contain the epidemic. The proportion and infectiousness of asymptomatic COVID-19 cases remains poorly understood to date due to immense difficulties in detecting them and tracking their viral loads or infectees[67]. In contrast, it was estimated early on that symptomatic cases infected about 40% of their infectees before showing symptoms[68,69], thus implying that contact tracing would need to be very fast and efficient to be effective[65,66]. These findings prompted the rapid development and deployment of various platforms of digital contact tracing around the world. Although digital contact tracing is one of the pillars for successful COVID-19 control in countries such as China, Singapore and South Korea, the uptake of this intervention in some other jurisdictions—most notably, the United States and Europe—has been low due to privacy concerns and technical issues[70].

How many people have been or are currently infected?

As the epidemic unfolds, real-time estimation of cumulative incidence becomes the centerpiece of epidemic nowcasting, because knowledge about epidemic size is a prerequisite for inference of severity, population-level immunity and the time to and magnitude of the epidemic peak. Nowcasting the size of a growing epidemic is challenging because reported case counts are inevitably biased by the proportion of infections that are symptomatic, care-seeking behavior, the availability of tests and so on. Seroepidemiological studies, which use measurements of antibody response (for example, neutralization assays and enzyme-linked immunosorbent assay) to infer infective exposure to the pathogen, arguably provide the most direct and reliable data for estimating cumulative incidence[71]. Shortly after COVID-19 began its global spread, population-based seroepidemiological studies from multiple countries revealed that the cumulative incidence of COVID-19 was far below the potential herd immunity threshold, and yielded IFR estimates that were similar to previous model-based estimates[70,72], thus strengthening the evidence base for the necessity for sustained, stringent PHSMs. Even in places where the theoretical herd immunity threshold has been exceeded, it has been known to remain susceptible to further infection waves. The resurgence of COVID-19 in Manaus, Brazil is a case in point[73], from which lessons have been learned[74].

Real-time, large-scale serosurveillance has been mostly infeasible during the COVID-19 pandemic because the throughput, accuracy and reproducibility of commonly used serologic assays (especially the labor-intensive neutralization assays, which are regarded as the gold standard) are limited. Nevertheless, recent long-awaited advances in high-throughput multiplex serology (for example, VirScan and Sioma[75,76]) have set the stage for implementing serosurveillance as a standard arm of future epidemic surveillance.

Several innovative strategies have also been trialed to mitigate the biases of reporting behavior, reporting delay, testing capacity and other factors that distort the convergence between surveillance data and true infection incidence. One promising strategy is to infer disease prevalence by quantitative monitoring of pathogen genomes in wastewater systems[77,78]. Digital participatory surveillance is

another potentially cost-effective substitute or complement for conventional syndromic and virologic surveillance, especially in resource-limited settings with weak health systems[79,80]. An integrative framework of next-generation epidemic nowcasting will probably employ advanced analytics to assimilate the myriad streams of conventional and novel surveillance data.

How effective are PHSMs?

Governments worldwide have implemented a wide range of PHSMs to suppress COVID-19, including travel restriction, physical distancing, rapid contact tracing and testing, school and workplace closure, curfews and targeted or nationwide lockdowns[81,82]. While it was clear a priori that these PHSMs would carry high economic and social costs, there remain large uncertainties about their effectiveness. As such, nowcasting the effectiveness of PHSMs is essential for informing countries how to optimally adjust their PHSM portfolios subject to their particular temporal and contextual feasibility constraints (for example, inertia to mask wearing, public fatigue to curfews and physical distancing, and social and political unrest in response to economic shutdown). For example, given the massive disruption to trade and commerce associated with sustained international travel restrictions, nowcasting of the optimal timing and duration of their implementation (which can be done by making integrative use of the above-mentioned data and analytics on global and local spread) is critical for guiding economic resumption[83].

The effectiveness of PHSMs in suppressing local transmissibility is typically inferred from temporal changes in R[84,85]. As such, these analyses are subject to all of the limitations and uncertainties that are inherent in the nowcasting of R. The effectiveness of PHSMs will not immediately manifest in case-based estimates of R because there is a delay of around 10 d between infection and reporting for COVID-19. One solution is to use physical contact mixing levels of the population as a proxy for disease transmissibility and monitor the former using conventional social contact surveys or digital human mobility data. Cross-sectional social contact surveys have been carried out in Wuhan and Shanghai, as well as in several European countries (for example, the CoMix study)[86,87]. These surveys aim to record the daily number of contacts of each participant before and after PHSMs are introduced or lifted. They are sometimes used to compare the geographical differences when localized interventions are implemented[88]. Aggregate data of human mobility trends have been made publicly available in near-real time by several technology enterprises, including Apple, Baidu, Google and Facebook. The impacts of individual PHSMs can be estimated more efficiently and accurately by considering mobility data from multiple sources in the reconstruction of transmission dynamics[88,89].

What are the socio-behavioral corollaries of epidemic control?

The success of PHSMs ultimately relies on the goodwill, and thus adherence, of the people. Therefore, it would be important to extend surveillance to include psychological and emotional wellbeing, as well as attitudes and beliefs[90]. Misalignment of individual and collective interests could hinder collective adherence, which could be further exacerbated over time by pandemic fatigue[90]. Behavioral studies have indeed shown that PHSMs that were effective in earlier waves may not be as effective in later waves[91], hence necessitating nowcasts of public adherence and socioeconomic impact to inform relaxation, reintroduction, sustainability and compensatory measures of PHSMs[92].

Metrics for nowcasting and epidemic control should include outcomes beyond mortality and morbidity, such as human flourishing and population wellbeing[93]. Psychobehavioral surveillance can detect unintended consequences of PHSMs and identify vulnerable groups and gaps. For example, stay-at-home orders and other measures have led to the largest enforced isolation period in human history[94]. This has contributed to reduced physical activity, social support and spiritual support, which in turn affects physical, mental and social wellbeing[95–97]. Schools were preemptively suspended early in the first waves of the COVID-19 pandemic by most countries. However, as the pandemic has lengthened, the long-term pedagogical, emotional and developmental downsides of school closure are increasingly recognized[98]. In addition, many older adults, who are at higher risk of death from COVID-19, are living alone or institutionalized and are failing to thrive due to policies that prohibit visitors or cross-household mixing[99].

To date, studies have predominantly employed online convenience samples, which are often unreliable for prevalence estimates of behaviors and disease burden[100]. Most studies have also been cross-sectional; thus, it is unclear whether mental health outcomes were due to the epidemic or preceded the outbreak. In contrast, population-representative panel studies can identify longitudinal determinants and improve causal inference, thus providing robust evidence to drive precision policy planning and evaluation[95,100,101]. However, prospective cohorts are resource intensive and not readily available, and respondent fatigue can preclude intensive follow-up[102]. Cohort studies can therefore be complemented by serial cross-sectional surveys that randomly sample the general population[92], and social media platforms (for example, Twitter and Weibo) might be harnessed for real-time monitoring of attitudes and wellbeing[102].

Are vaccines available?

Rapid development of COVID-19 vaccines had been particularly challenging because there was no previous licensed vaccine against any human coronavirus. The Ebola crisis in 2014 led to the formation of the Coalition for Epidemic Preparedness Innovations in 2017, with the mandate of developing vaccines—particularly generic platform technologies—to stop future epidemics. The Coalition for Epidemic Preparedness Innovations, in coordination with the WHO ACT Accelerator and national and commercial initiatives (for example, Operation Warp Speed), has facilitated unprecedented progress in developing multiple vaccines against SARS-CoV-2 through phase 3 clinical trials and emergency use authorization, and thus rollout for a handful of vaccines in several countries.

The duration of immunity remains unknown at present because the currently available data pertain to the first few months after vaccination. Serum neutralizing antibodies were found to persist for at least 1 year after severe or mild disease[103]. Since many of these vaccines produce neutralizing antibody titers at least comparable to that found after natural infection, it is reasonable to expect at least a similar duration of protection from vaccines. Natural infection appears to protect from both symptomatic infection and, to a lesser degree, asymptomatic infection, and both appear to last for at least 5 months[104]. It is unclear whether vaccines administered systemically via intramuscular injection would provide comparable protection from infection and disease.

Published data on vaccine efficacy pertain to protection from clinically symptomatic disease rather than evidence of infection, save for a subset of data from the AstraZeneca vaccine so far. Evidence of protection from clinical disease may not necessarily protect from transmission to a comparable degree[105]. Thus, vaccinated individuals, while protected from overt disease, may possibly be susceptible to asymptomatic or mildly symptomatic infection and may be able to transmit infection to others, with the corollary that PHSMs including travel restrictions may not be substantially relaxed.

In the absence of head-to-head comparison of different vaccines, it is difficult to compare vaccine efficacy data from different clinical trials, given that such trials have been carried out in regions with different forces of infection. While there are data on older adults for

some vaccines, suggesting that vaccine efficacy is not compromised by age, this remains to be established for others. There are also no clinical trial data as yet concerning pregnant women, children, the immunocompromised and other high-risk populations.

The emergence of different variant strains with enhanced transmissibility has been detected separately in the United Kingdom, South Africa and Brazil, suggesting convergent evolution. While it is unlikely that such antigenic change will completely evade vaccine-induced protection from disease, some reduction of efficacy is possible, and needs to be monitored[106]. Equitable vaccine allocation and delivery between and within countries have remained huge challenges. This may well become the Achilles heel of global human security in the coming year.

Future directions

The global scientific community has collectively risen to the challenge of COVID-19 by innovating and deploying in real time a wide range of collaborative tools and platforms to facilitate epidemic nowcasting, including the COVID-19 dashboards by Johns Hopkins University, the Coronavirus Government Response Tracker by the Oxford Martin School, the SARS-CoV-2 sequence repository at GISAID and its partners, and human mobility trend reports by Apple, Baidu and Google. In contrast, ineffective leadership, lack of coordination and inconsistent risk communication have seriously undermined epidemic nowcasting (and, more broadly, pandemic response), especially at the national level in some high-income countries, as attested by the almost inverse correlation between pre-COVID-19 assessed preparedness and actual performance in COVID-19 control across countries. As such, perhaps the most important lesson learned from our experience with COVID-19 is that the disconnect among policymakers, practitioners and scientists must be remedied in order to align evidence synthesis with public health operations and policies, with political implementation for the ongoing COVID-19 pandemic as well as future emergencies.

Going forward, epidemic nowcasting will require scientists to distill informative or actionable insights from an increasingly diverse range of data. Misinterpretation, misrepresentation or otherwise misuse of these nowcasts will fuel infodemics, as we have learned to our detriment during the ongoing pandemic. Such infodemics can be mitigated by developing sophisticated information systems that are professionally designed to robustly collate and curate epidemic nowcasts for policymakers and public health professionals, much like how the Bloomberg terminals provide real-time financial and economic nowcasts for professionals in the world of finance. Such systems will probably reduce the friction and latency in cross-disciplinary communication and consumption of epidemic intelligence. Once installed and made routine, these systems can strengthen our response not only to pandemics but also to less catastrophic but more frequent epidemics (such as influenza and Ebola), as well as non-infectious disease crises (such as climate change and environmental pollution).

However, the initial and continuous development of such systems will be challenging because it competes for the same pool of technical talents (most notably, in data science and software engineering) that are highly sought after by enterprises and startups in technology, finance and other industries that can support much more generous remunerative packages. In view of the massive economic damage imposed by pandemics, which was estimated to be US$16 trillion for COVID-19 (ref. [107]), world leaders should devise innovative incentives to support the development of such systems and other pandemic products, instead of relying on ad-hoc uncoordinated efforts of the research and development community for delivery. The US National Academy of Medicine-led Global Health Risk Framework, which was set up to create an effective global architecture for mitigating the threat of pandemics, made the same recommendation in 2016, alongside dozens of other recommendations that would have better prepared us for COVID-19 had these lessons been adopted and implemented more widely[108].

Received: 20 January 2021; Accepted: 3 February 2021;
Published online: 15 March 2021

References

1. Guan, Y. et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278 (2003).
2. Li, W. et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
3. Cotten, M. et al. Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. *Emerg. Infect. Dis.* **19**, 736–742 (2013).
4. Sabir, J. S. et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* **351**, 81–84 (2016).
5. Lam, T. T. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).
6. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
7. *Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction to Applied Epidemiology and Biostatistics* (Centers for Disease Control and Prevention, 2011); https://www.cdc.gov/csels/dsepd/ss1978/index.html
8. Tsang, T. K. et al. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study. *Lancet Public Health* **5**, e289–e296 (2020).
9. Lemey, P., Suchard, M. & Rambaut, A. Reconstructing the initial global spread of a human influenza pandemic: a Bayesian spatial-temporal model for the global spread of H1N1pdm. *PLoS Curr.* **1**, RRN1031 (2009).
10. Peiris, J. S. et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325 (2003).
11. Xu, R. H. et al. Epidemiologic clues to SARS origin in China. *Emerg. Infect. Dis.* **10**, 1030–1037 (2004).
12. Zhong, N. S. et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353–1358 (2003).
13. Lam, T. T., Zhu, H., Guan, Y. & Holmes, E. C. Genomic analysis of the emergence, evolution, and spread of human respiratory RNA viruses. *Annu. Rev. Genomics Hum. Genet.* **17**, 193–218 (2016).
14. Apolone, G. et al. Unexpected detection of SARS-CoV-2 antibodies in the prepandemic period in Italy. *Tumori* https://doi.org/10.1177/0300891620974755 (2020).
15. Gianotti, R., Barberis, M., Fellegara, G., Galvan-Casas, C. & Gianotti, E. COVID-19 related dermatosis in November 2019. Could this case be Italy's patient zero? *Br. J. Dermatol.* https://doi.org/10.1111/bjd.19804 (2021).
16. *WHO-Convened Global Study of the Origins of SARS-CoV-2* (World Health Organization, 2020); https://www.who.int/publications/m/item/who-convened-global-study-of-the-origins-of-sars-cov-2
17. Chan, J. F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514–523 (2020).
18. Kandun, I. N. et al. Three Indonesian clusters of H5N1 virus infection in 2005. *N. Engl. J. Med.* **355**, 2186–2194 (2006).
19. Alkhamis, M. A. et al. Temporal dynamics of Middle East respiratory syndrome coronavirus in the Arabian Peninsula, 2012–2017. *Epidemiol. Infect.* **147**, 1–10 (2018).
20. Cao, B. et al. A trial of lopinavir–ritonavir in adults hospitalized with severe COVID-19. *N. Engl. J. Med.* **382**, 1787–1799 (2020).
21. Wang, Y. et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* **395**, 1569–1578 (2020).
22. Hung, I. F.-N. et al. Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. *Lancet* **395**, 1695–1704 (2020).
23. WHO SOLIDARITY Trial Consortium.Repurposed antiviral drugs for COVID-19—interim WHO Solidarity trial results. *N. Engl. J. Med.* **384**, 497–511 (2021).
24. The RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with COVID-19—preliminary report. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2021436 (2020).
25. Horby, P. W. et al. Lopinavir–ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet* **396**, 1345–1352 (2020).
26. The RECOVERY Collaborative Group. Effect of hydroxychloroquine in hospitalized patients with COVID-19. *N. Engl. J. Med.* **383**, 2030–2040 (2020).
27. RECOVERY Collaborative Group et al. Azithromycin in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet* **397**, 605–612 (2021).

28. Weinreich, D. M. et al. REGN-COV2, a neutralizing antibody cocktail, in outpatients with COVID-19. *N. Engl. J. Med.* **384**, 238–251 (2021).
29. Chen, P. et al. SARS-CoV-2 neutralizing antibody LY-CoV555 in outpatients with COVID-19. *N. Engl. J. Med.* **384**, 229–237 (2021).
30. Jefferson, T. et al. Oseltamivir for influenza in adults and children: systematic review of clinical study reports and summary of regulatory comments. *Br. Med. J.* **348**, g2545 (2014).
31. Dobson, J., Whitley, R. J., Pocock, S. & Monto, A. S. Oseltamivir treatment for influenza in adults: a meta-analysis of randomised controlled trials. *Lancet* **385**, 1729–1737 (2015).
32. Wong, C. K. H. et al. Clinical outcomes of different therapeutic options for COVID-19 in two Chinese case cohorts: a propensity-score analysis. *EClinicalMedicine* https://doi.org/10.1016/j.eclinm.2021.100743 (2021).
33. Heesterbeek, H. et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, aaa4339 (2015).
34. Cauchemez, S., Hoze, N., Cousien, A., Nikolay, B. & Ten Bosch, Q. How modelling can enhance the analysis of imperfect epidemic data. *Trends Parasitol.* **35**, 369–379 (2019).
35. Adam, D. C. et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020).
36. Faye, O. et al. Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect. Dis.* **15**, 320–326 (2015).
37. Gostic, K. M. et al. Practical considerations for measuring the effective reproductive number, R_t. *PLoS Comput. Biol.* **16**, e1008409 (2020).
38. King, A. A., Domenech de Celles, M., Magpantay, F. M. & Rohani, P. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proc. Biol. Sci.* **282**, 20150347 (2015).
39. Vink, M. A., Bootsma, M. C. & Wallinga, J. Serial intervals of respiratory infectious diseases: a systematic review and analysis. *Am. J. Epidemiol.* **180**, 865–875 (2014).
40. Park, S. W., Champredon, D. & Dushoff, J. Inferring generation-interval distributions from contact-tracing data. *J. R. Soc. Interface* **17**, 20190719 (2020).
41. Ali, S. T. et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* **369**, 1106–1109 (2020).
42. Park, S. W. et al. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. *J. R. Soc. Interface* **17**, 20200144 (2020).
43. Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
44. Xu, B. et al. Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci. Data* **7**, 106 (2020).
45. Pybus, O. G. et al. The epidemic behavior of the hepatitis C virus. *Science* **292**, 2323–2325 (2001).
46. Stadler, T. et al. Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.* **29**, 347–357 (2012).
47. Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J. & Frost, S. D. Phylodynamics of infectious disease epidemics. *Genetics* **183**, 1421–1430 (2009).
48. Lai, A., Bergna, A., Acciarri, C., Galli, M. & Zehender, G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *J. Med. Virol.* **92**, 675–679 (2020).
49. Volz, E. M. *SARS CoV 2 Phylodynamics* (MRC Centre for Global Infectious Disease Analysis, Imperial College London, 2020); http://sarscov2phylodynamics.org
50. Fraser, C. et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324**, 1557–1561 (2009).
51. Leung, G. M., Cowling, B. J. & Wu, J. T. From a sprint to a marathon in Hong Kong. *N. Engl. J. Med.* **382**, e45 (2020).
52. Imai, N. et al. *Report 3: Transmissibility of 2019-nCoV* (Imperial College London, 2020).
53. Buckee, C. O. et al. Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).
54. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).
55. Burki, T. China's successful control of COVID-19. *Lancet Infect. Dis.* **20**, 1240–1241 (2020).
56. Lipsitch, M. et al. Potential biases in estimating absolute and relative case-fatality risks during outbreaks. *PLoS Negl. Trop. Dis.* **9**, e0003846 (2015).
57. Li, Z. et al. Active case finding with case management: the key to tackling the COVID-19 pandemic. *Lancet* **396**, 63–70 (2020).
58. Graham, F. Daily briefing: why the coronavirus death toll in Wuhan, China, has just jumped by 50%. *Nature* (17 April 2020).
59. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677 (2020).
60. Ferguson, N. et al. *Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand* (Imperial College London, 2020).
61. Lurie, N., Saville, M., Hatchett, R. & Halton, J. Developing COVID-19 vaccines at pandemic speed. *N. Engl. J. Med.* **382**, 1969–1973 (2020).
62. *About Quarantine and Isolation* (Centers for Disease Control and Prevention, 2020); https://www.cdc.gov/quarantine/quarantineisolation.html
63. Lessler, J. et al. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect. Dis.* **9**, 291–300 (2009).
64. Lauer, S. A. et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020).
65. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151 (2004).
66. Peak, C. M., Childs, L. M., Grad, Y. H. & Buckee, C. O. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc. Natl Acad. Sci. USA* **114**, 4023–4028 (2017).
67. Nogrady, B. What the data say about asymptomatic COVID infections. *Nature* **587**, 534–535 (2020).
68. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
69. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
70. The Lancet Digital Health. Contact tracing: digital health on the frontline. *Lancet Digit Health* **2**, e561 (2020).
71. Metcalf, C. J. et al. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet* **388**, 728–730 (2016).
72. Gudbjartsson, D. F. et al. Humoral immune response to SARS-CoV-2 in Iceland. *N. Engl. J. Med.* **383**, 1724–1734 (2020).
73. Buss, L. F. et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **371**, 288–292 (2021).
74. Sabino, E. C. et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452–455 (2021).
75. Norman, M. et al. Ultrasensitive high-resolution profiling of early seroconversion in patients with COVID-19. *Nat. Biomed. Eng.* **4**, 1180–1187 (2020).
76. Shrock, E. et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **370**, eabd4250 (2020).
77. Wurtzer, S. et al. Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Euro Surveill.* **25**, 2000776 (2020).
78. Peccia, J. et al. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* **38**, 1164–1167 (2020).
79. Garg, S., Bhatnagar, N. & Gangadharan, N. A case for participatory disease surveillance of the COVID-19 pandemic in India. *JMIR Public Health Surveill.* **6**, e18795 (2020).
80. Pullano, G. et al. Underdetection of COVID-19 cases in France threatens epidemic control. *Nature* **590**, 134–139 (2020).
81. Hale, T., Petherick, A., Phillips, T. & Webster, S. *Variation in Government Responses to COVID-19* Blavatnik School of Government Working Paper **31** (Univ. Oxford, 2020).
82. Haug, N. et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).
83. Russell, T. W. et al. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *Lancet Public Health* **6**, e12–e20 (2021).
84. Li, Y. et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect. Dis.* **21**, 193–202 (2021).
85. Brauner, J. M. et al. Inferring the effectiveness of government interventions against COVID-19. *Science* https://doi.org/10.1126/science.abd9338 (2020).
86. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
87. Jarvis, C. I. et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med.* **18**, 124 (2020).
88. Davies, N. G. et al. Association of tiered restrictions and a second lockdown with COVID-19 deaths and hospital admissions in England: a modelling study. *Lancet Infect. Dis.* https://doi.org/10.1016/S1473-3099(20)30984-1 (2020).
89. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
90. Van Bavel, J. J. et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
91. Cowling, B. J. et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Health* **5**, E279–E288 (2020).

92. Yang, B. et al. Changing disparities in COVID-19 burden in the ethnically homogeneous population of Hong Kong through pandemic waves: an observational study. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciab002 (2021).
93. VanderWeele, T. J., McNeely, E. & Koh, H. K. Reimagining health—flourishing. *J. Am. Med. Assoc.* **321**, 1667–1668 (2019).
94. Fancourt, D., Steptoe, A. & Bu, F.Trajectories of anxiety and depressive symptoms during enforced isolation due to COVID-19 in England: a longitudinal observational study. *Lancet Psychiatry* **8**, 141–149 (2021).
95. Ni, M. Y. et al. Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: cross-sectional survey. *JMIR Ment. Health* **7**, e19009 (2020).
96. Ni, M. Y. et al. Determinants of physical, mental and social well-being: a longitudinal environment-wide association study. *Int. J. Epidemiol.* **49**, 380–389 (2020).
97. Holmes, E. A. et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry* **7**, 547–560 (2020).
98. Viner, R. M. et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *Lancet Child Adolesc. Health* **4**, 397–404 (2020).
99. Verbeek, H. et al. Allowing visitors back in the nursing home during the COVID-19 crisis: a Dutch national study into first experiences and impact on well-being. *J. Am. Med. Dir. Assoc.* **21**, 900–904 (2020).
100. Pierce, M. et al. Says who? The significance of sampling in mental health surveys during COVID-19. *Lancet Psychiatry* **7**, 567–568 (2020).
101. Ni, M. Y. et al. Depression and post-traumatic stress during major social unrest in Hong Kong: a 10-year prospective cohort study. *Lancet* **395**, 273–284 (2020).
102. Ni, M. Y., Leung, C. M. & Leung, G. M. The epidemiology of population mental wellbeing in China. *Lancet Public health* **5**, e631–e632 (2020).
103. Lau, E. H. Y. et al. Neutralizing antibody titres in SARS-CoV-2 infections. *Nat. Commun.* **12**, 63 (2021).
104. Lumley, S. F. et al. Antibody status and incidence of SARS-CoV-2 infection in health care workers. *N. Engl. J. Med.* **384**, 533–540 (2021).
105. Peiris, M. & Leung, G. M. What can we expect from first-generation COVID-19 vaccines? *Lancet* **396**, 1467–1469 (2020).
106. *Statement on the Sixth Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Coronavirus Disease (COVID-19) Pandemic* (World Health Organization, 2021); https://www.who.int/news/item/15-01-2021-statement-on-the-sixth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic
107. Cutler, D. M. & Summers, L. H. The COVID-19 pandemic and the $16 trillion virus. *J. Am. Med. Assoc.* **324**, 1495–1496 (2020).
108. *Global Health Risk Framework: Resilient and Sustainable Health Systems to Respond to Global Infectious Disease Outbreaks* (National Academies of Sciences, Engineering and Medicine, 2016).

Acknowledgements

The authors are supported by the Health and Medical Research Fund (grant numbers CID-HKU2, COVID190126 and COVID190210), the General Research Fund (grant number 17110020), a special grant of the InnoHK program from the Government of the Hong Kong Special Administrative Region, and the National Natural Science Foundation of China (NSFC) Excellent Young Scientists Fund (Hong Kong and Macau) (grant number 31922087). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.T.W. and G.M.L. conceived the study. All authors contributed to drafting and finalizing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-021-01278-w.

Correspondence should be addressed to J.T.W.

Peer review information Joao Monteiro was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

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