Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Crowdsourcing data to mitigate epidemics

Published Online
February 20, 2020
https://doi.org/10.1016/S2589-7500(20)30055-8

See Articles page e201

See Online for appendix

Coronavirus disease 2019 (COVID-19) has spread with unprecedented speed and scale since the first zoonotic event that introduced the causative virus—severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—into humans, probably during November, 2019, according to phylogenetic analyses suggesting the most recent common ancestor of the sequenced genomes emerged between Oct 23, and Dec 16, 2019.[1] The reported cumulative number of confirmed patients worldwide already exceeds 70 000 in almost 30 countries and territories as of Feb 19, 2020, although that the actual number of infections is likely to far outnumber this case count.[2,3]

During any novel emerging epidemic, let alone one with such magnitude and speed of global spread, a first task is to put together a line list of suspected, probable, and confirmed individuals on the basis of working criteria of the respective case definitions. This line list would allow for quick preliminary assessment of epidemic growth and potential for spread, evidence-based determination of the period of quarantine and isolation, and monitoring of efficiency of detection of potential cases. Frequent refreshing of the line list would further enable real-time updates as more clinical, epidemiological, and virological (including genetic) knowledge become available as the outbreak progresses.

Therefore, from a public health viewpoint, a line list is indispensable. Hence, Kaiyuan Sun and colleagues' work,[6] published in *The Lancet Digital Health*, is very valuable at this key timepoint in the COVID-19 outbreak. Sun and colleagues harnessed Chinese social media, specifically a social network used by health-care professionals, to compile individual-level data on patients with COVID-19 and daily province-level case counts during January, 2020. They distilled this information into a crowdsourced line list, which, when analysed appropriately, aligns closely with that derived from official versions, such as the report published by the Chinese CDC on Jan 29, 2020.[7] For instance, the various delay intervals converge between the first 425 cases in Wuhan in the earlier report[7] and the 507 cases sourced from both China and overseas in the present Article. Although Sun and colleagues' work provides a valuable picture of the outbreak in real time, the geographical coverage is heterogeneous with only a small proportion of cases from the epicenter of Wuhan and Hubei province.

We surveyed different and varied sources of possible line lists for COVID-19 (appendix pp 1–4). A bottleneck remains in carefully collating as much relevant data as possible, sifting through and verifying these data, extracting intelligence to forecast and inform outbreak strategies, and thereafter repeating this process in iterative cycles to monitor and evaluate progress. A possible methodological breakthrough would be to develop and validate algorithms for automated bots to search through cyberspaces of all sorts, by text mining and natural language processing (in languages not limited to English) to expedite these processes.

In this era of smartphone and their accompanying applications, the authorities are required to combat not only the epidemic per se, but perhaps an even more sinister outbreak of fake news and false rumours, a so-called infodemic. The most obvious consequences of an infodemic are, at best, a noisy cacophony that confuses and can provoke irrational fear, even mass panic, and ultimately imposes a destabilising effect on society when precisely the opposite is required. The images of empty supermarket shelves in the most open free-trade economies of Singapore and Hong Kong, where fewer than 100 cases have been reported to date, provide a salutary reminder of the potential impact of such infodemics. Another example is the worldwide shortage of and some national export bans on face masks. Creating a resource such as Sun and colleagues have compiled in their work would allow scientists and lay observers alike to quickly fill knowledge vacuums that would otherwise fuel infodemics.

Related to the infodemic is the so-called geodemic of geopolitical considerations and nationalistic populism apparently being placed ahead of the science of outbreak control. A case in point relates to national border policies that have been suggested to contravene International Health Regulations.[4] Anecdotes of xenophobic treatment of people from different places or fellow natives who look different are doing substantial harm to building the extra solidarity necessary in such times.

Finally, the epidemic, infodemic, and geodemic all have economic costs.[5] During the severe acute

respiratory syndrome (SARS) outbreak in 2003, China accounted for 4% of global economic output compared with 16% today.[5] Despite the ongoing trade tensions since 2019, China's supply chains and production lines remain closely enmeshed with much of the world's trading markets. These economic uncertainties, of course, have not taken into account how the outbreak might affect the rest of the world, when cases have now been reported on most continents, including Africa.

Notwithstanding the above motivations, during the exigency of an outbreak, especially one with a doubling time of 1 week[2] in the world's most populated country, expecting a ready line list for analytical prosecution covering all domestic geographies within a few short weeks would be astounding. Even during the SARS outbreak in 2003, we worked through over 30 versions of the case-contact questionnaire before settling on the final version well over a month after the first case had been confirmed. Notably, China's health protection function is decentralised to provincial and local levels (with over 300 prefecture-level Centers for Disease Control and Prevention branches) and it remains a developing country with differing levels of epidemic preparedness along socioeconomic development gradients across a large geography.[8]

Crowdsourced data could be compiled and analysed as timely as, or perhaps even quicker than, officially released data. However, such future developments do not negate the overriding importance of the timely release and updating of official line lists with as much detail as ethics and confidentiality allow. However, such sourcing would go a long way to address and mitigate the epidemics, infodemics, and geodemics that the world will face in the years to come.

We declare no competing interests.

*Gabriel M Leung, Kathy Leung
gmleung@hku.hk

WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, Special Admistrative Region, China

1 Andersen K. Clock and TMRCA based on 27 genomes. Jan 25, 2020. http://virological.org/t/clock-and-tmrca-based-on-27-genomes/347 (accessed Feb 17, 2020).

2 Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020; published online Jan 31. https://doi.org/10.1016/S0140-6736(20)30260-9.

3 Dorigatti I, Okell L, Cori A, et al. Report 4: severity of 2019-novel coronavirus (nCoV). London: Imperial College London, Feb 10, 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-2019-nCoV-severity-10-02-2020.pdf (accessed Feb 14, 2020).

4 Habibi R, Burci GL, de Campos TC, et al. Do not violate the International Health Regulations during the COVID-19 outbreak. *Lancet* 2020; published online Feb 13. https://doi.org/10.1016/S0140-6736(20)30373-1.

5 How China's coronavirus epidemic could hurt the world economy. The Economist, Feb 13, 2020. https://www.economist.com/leaders/2020/02/13/how-chinas-coronavirus-epidemic-could-hurt-the-world-economy (accessed Feb 17, 2020).

6 Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digital Health* 2020; pubished online Feb 20. https://doi.org/10.1016/S2589-7500(20)30026-1.

7 Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020; published online Jan 29. DOI:10.1056/NEJMoa2001316.

8 Vlieg WL, Fanoy EB, van Asten L, et al. Comparing national infectious disease surveillance systems: China and the Netherlands. *BMC Public Health* 2017; **17:** 415.