Open access epidemiological data from the COVID-19 outbreak

Coronavirus disease 2019 (COVID-19) is spreading rapidly across China, and as of Feb 16, 2020, had been reported in 26 countries globally. The availability of accurate and robust epidemiological, clinical, and laboratory data early in an epidemic is important to guide public health decision-making.[1] Consistent recording of epidemiological information is important to understand transmissibility, risk of geographic spread, routes of transmission, and risk factors for infection, and to provide the baseline for epidemiological modelling that can inform planning of response and containment efforts to reduce the burden of disease. Furthermore, detailed information provided in real time is crucial for deciding where to prioritise surveillance.

Line list data are rarely available openly in real time during outbreaks. However, they enable a multiplicity of analyses to be undertaken by different groups, using various models and assumptions, which can help build consensus on robust inference. Parallels exist between this and the open sharing of genomic data.[2]

We have built a centralised repository of individual-level information on patients with laboratory-confirmed COVID-19 (in China, confirmed by detection of virus nucleic acid at the City and Provincial Centers for Disease Control and Prevention), including their travel history, location (highest resolution available and corresponding latitude and longitude), symptoms, and reported onset dates, as well as confirmation dates and basic demographics. Information is collated from a variety of sources, including official reports from WHO, Ministries of Health, and Chinese local, provincial, and national health authorities. If additional data are available from reliable online reports, they are included. Data are available openly and are updated on a regular basis (around twice a day).

We hope these data continue to be used to build evidence for planning, modelling, and epidemiological studies to better inform the public, policy makers, and international organisations and funders as to where and how to improve surveillance, response efforts, and delivery of resources, which are crucial factors in containing the COVID-19 epidemic.

The epidemic is unfolding rapidly and reports are outdated quickly, so it will be necessary to build computational infrastructure that can handle the large expected increase in case reports. Data sharing will be vital to evaluate and maintain accurate reporting of cases during this outbreak.[3]

*Bo Xu, *Moritz U G Kraemer, on behalf of the Open COVID-19 Data Curation Group*

moritz.kraemer@zoo.ox.ac.uk

Department of Zoology, University of Oxford, Oxford OX1 3SZ, UK

1 Morgan O. How decision makers can use quantitative approaches to guide outbreak responses. *Philos Trans R Soc B Biol Sci* 2019; **374:** 20180365.

2 Yozwiak NL, Schaffner SF, Sabeti PC. Data sharing: make outbreak research open access. *Nature* 2015; **518:** 477–79.

3 Heymann DL. Data sharing and outbreaks: best practice exemplified. *Lancet* 2020; **395:** 469–70.