

# Using open-source intelligence to identify early signals of COVID-19 in Indonesia

Yoser Thamtono,<sup>a</sup> Aye Moa<sup>a</sup> and Chandini Raina MacIntyre<sup>a</sup>

Correspondence to Yoser Thamtono (yosertamtono@protonmail.com)

**Objective:** Open-source data from online news reports and informal sources may provide information about outbreaks before official notification. This study aims to evaluate the use of open-source data from the epidemic observatory, *EpiWATCH*, to identify the early signals of pneumonia of unknown cause as a proxy for COVID-19 in Indonesia.

**Methods:** Using open-source data on pneumonia of unknown cause in Indonesia between 1 November 2019 and 31 March 2020 (extracted from *EpiWATCH*, an open-source epidemic observatory), a descriptive analysis was performed to identify the trend of pneumonia of unknown cause in Indonesia before official notification of COVID-19 cases.

**Results:** A rise in reports of pneumonia of unknown cause was identified in Indonesia, starting from late January 2020. There were 304 reported cases of pneumonia of unknown cause, 30 of which occurred before the identification of the first COVID-19 cases on 2 March 2020. The early signals of pneumonia of unknown cause in Indonesia may indicate possible unrecognized circulation of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) before official detection.

**Discussion:** Open-source data may provide rapid, unvalidated information for early detection of outbreaks. Although unvalidated, such information may be used to supplement or trigger investigation and testing. As *EpiWATCH* sources global information, this methodology can be repeated for other countries within the Western Pacific Region, and for other diseases.

On 31 December 2019, China reported an increased occurrence of pneumonia of unknown cause in Wuhan,<sup>1</sup> which was later confirmed as coronavirus disease 2019 (COVID-19), a respiratory illness caused by the newly discovered coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). On 13 January 2020, Thailand became the first country to identify imported cases of COVID-19. Other countries in South-East Asia, including Singapore and Malaysia, also reported their first imported cases in January 2020.<sup>2,3</sup> On 31 January 2020, the World Health Organization (WHO) declared COVID-19 a public health emergency of international concern,<sup>4</sup> and on 11 March 2020, it declared the outbreak a pandemic.<sup>5</sup> Indonesia reported its first two confirmed positive cases of COVID-19 on 2 March 2020.<sup>6</sup>

Early identification of an infectious disease outbreak is essential, to allow for immediate initiation of public health interventions and to mitigate global impacts of transnational spread of the disease. Traditional public

health surveillance, especially where there is low testing capability, is often not timely due to delays in reporting and validation by local health authorities.<sup>4</sup>

Open-source intelligence may serve as a valuable tool for rapid epidemic surveillance.<sup>5</sup> WHO reports that more than half of early epidemic information can be obtained through unofficial sources, including online news outlets and social media.<sup>6</sup> This study aims to evaluate the use of open-source data from the epidemic observatory, *EpiWATCH*, to identify the early signals of pneumonia of unknown cause as a proxy for SARS-CoV-2 circulation in Indonesia.

## METHODS

*EpiWATCH* is a semi-automated open-source epidemic observatory that collects and analyses outbreak data from publicly available sources, such as online news outlets and social media.<sup>7</sup> *EpiWATCH* has been used to collect outbreak data since 2016. The observatory col-

<sup>a</sup> Biosecurity Program, Kirby Institute, Faculty of Medicine, University of New South Wales, Sydney, Australia.

Published: 17 February 2021

doi: 10.5365/wpsar.2020.11.2.010

lects information on outbreaks of diseases and emerging infections, globally, to detect early signals and trends, which can then be used by researchers. The system was established and is managed by the Australian National Health and Medical Research Council (NHMRC) Centre for Research Excellence, Integrated Systems for Epidemic Response (ISER).<sup>7</sup>

*EpiWATCH* gathers open-source data from online news outlets and social media through an intelligent and modular system. To enable enhanced surveillance, searches are modified for specific languages and for specific infectious disease syndromes. Modifications include changing and adding keywords and languages for searching. The observatory system supports various intelligent data-gathering algorithms, including natural language processing algorithms, regular expression matching and supervised machine-learning algorithms. The data are stored in a PostgreSQL database.<sup>8</sup> Within *EpiWATCH*, reports are reviewed and entered manually by the team, to ensure collection of key data points identified by the automated data-gathering system and prevent duplication of reports.

In this study, enhanced surveillance was performed through *EpiWATCH* using keywords reflecting SARS or pneumonia, such as “pneumonia, lung infection, lung inflammation, severe acute respiratory infection (SARI), cough, fever, cough AND fever”, with Indonesia as a location, in Bahasa Indonesia (the country’s official language). The obtained reports were manually reviewed, and any reported cases of pneumonia of unknown cause in Indonesia dated between 1 November 2019 and 31 March 2020 were included in the *EpiWATCH* database. Reports matching the keywords were analysed and were excluded if they were not reporting cases from Indonesia or cases of pneumonia of unknown cause, or if they were reporting duplicate news or pneumonia cases with a known cause. Indonesia was chosen for this study as its first official COVID-19 notification was made later than that of other South-East Asian countries.

A descriptive epidemiological analysis was performed using Microsoft Excel 2016 to group cases by geolocation and by the date on which they were reported. Data reported from Indonesia’s Ministry of Health were also reviewed and compared with the news reports.

## Ethics and permissions

No ethics approval was required for this study because it uses publicly available open-source data from online media.

## RESULTS

Between November 2019 and March 2020, *EpiWATCH* documented 217 entries or reports related to pneumonia of unknown cause in Indonesia. Six duplicates were removed, leaving a total of 211 included reports. The highest number of reports (184 entries) were from March 2020, after the official identification of the first cases of COVID-19 in Indonesia. However, a steady increase in the reports of pneumonia of unknown cause was seen from late January 2020 (five entries) to the end of February 2020 (17 entries).

These 211 entries correspond to 304 reported cases of pneumonia of unknown cause in Indonesia, 30 (9.9%) of which occurred before the identification of the first COVID-19 cases on 2 March 2020. Among those 30 cases, 17 were suspected for COVID-19, including the case of a toddler from China, aged 18 months, who was treated for pneumonia in Lombok, West Nusa Tenggara, on 27 January 2020.<sup>9</sup> COVID-19 diagnosis was ruled out based on a blood laboratory count that was suggestive of bacterial infection; however, no further results of polymerase chain reaction (PCR) or bacterial culture were reported.<sup>9</sup>

**Fig. 1** shows the count of pneumonia of unknown cause from 1 November 2019 to 31 March 2020 as obtained from *EpiWATCH* reports. In November and December 2019, there were five *EpiWATCH*-reported cases of pneumonia of unknown cause. The number of reports increased slightly from late January (six identified cases) to February 2020 (19 identified cases), six weeks before the official identification of the first two COVID-19 cases in Indonesia. In March 2020, the count of *EpiWATCH*-reported cases of pneumonia of unknown cause increased significantly to 274 cases, which may represent an increase in COVID-19 cases or growing media attention and reporting following the official notification of COVID-19 in Indonesia on 2 March 2020. In the same period in the previous year, there were far fewer reports of pneumonia of unknown cause, with only

five *EpiWATCH*-reported cases of pneumonia of unknown cause from January to March 2019.

Almost half of all *EpiWATCH*-reported cases of pneumonia of unknown cause were from provinces in Java. East Java (39 cases; 12.8%) and Jakarta (37 cases; 12.1%) were the provinces with the highest number of *EpiWATCH*-reported cases, followed by Central Java (26 cases; 8.5%), West Java (17 cases; 5.6%), Yogyakarta (15 cases; 4.9%) and Banten (seven cases; 2.3%). Outside Java, Bali had the highest number of *EpiWATCH*-reported cases (28 cases; 9.2%) (Fig. 2 and Fig. 3).

Gender information was given for 157 of the 304 *EpiWATCH*-reported cases. Among those 157 cases, there was a higher proportion of males (96 cases; 61.1%) than females (61 cases; 38.9%), with a 1.5:1 ratio of males to females. Information on age was given for 138 of the 304 *EpiWATCH*-reported cases, with the highest proportion of cases being those aged 50–60 years.

## DISCUSSION

This study provides a descriptive analysis of cases of pneumonia of unknown cause that occurred in Indonesia from November 2019 to March 2020 as detected by the *EpiWATCH* observatory. These *EpiWATCH*-reported cases of pneumonia of unknown cause in Indonesia increased from late January 2020, which may reflect the presence of COVID-19 cases in the country before official identification of two cases at the start of March 2020.

Data from the Ministry of Health<sup>10</sup> show that, by 31 March 2020, Indonesia had 1528 laboratory-confirmed cases of COVID-19, with 136 documented deaths. Among these cases, almost half (48.9%) were identified in Jakarta, followed by other provinces in Java: West Java (11.9%), Banten (9.2%), East Java (5.9%), Central Java (5.2%) and Yogyakarta (1.2%).<sup>10</sup> At that time, Bali had reported only 19 confirmed cases of COVID-19 (1.2% of total cases).<sup>10</sup>

The spatial distribution of notified COVID-19 cases in Indonesia is similar to that of pneumonia of unknown cause from *EpiWATCH*, suggesting the potential use of pneumonia of unknown cause as a proxy for COVID-19 cases in Indonesia. However, there were proportionately

more notified cases from Jakarta compared with the spatial distribution of pneumonia of unknown cause from *EpiWATCH* in our study. This might suggest geographical differences in the ability to identify and report COVID-19 cases. Health infrastructure gaps have been cited as one of the most critical issues in the Indonesian health system.<sup>11</sup> Health facilities, including hospital and laboratory services, are more readily available in the urban Java region, where Jakarta, the capital city, is located.<sup>12</sup>

Provinces in Java, especially Jakarta, might have better capability for identifying cases of COVID-19. Meanwhile, other provinces may have underdetection of COVID-19 cases; for example, Bali notified a low number of confirmed COVID-19 cases compared with the *EpiWATCH*-reported cases of pneumonia of unknown cause. Of the 12 COVID-19 national reference laboratories in Indonesia, only three are located outside Java, which may hinder the ability of provinces outside Java to rapidly identify and respond to the presence of COVID-19.<sup>13</sup>

This study is the first of its kind to describe the epidemiological pattern of pneumonia of unknown cause reported from open-source intelligence before and after the official notification of COVID-19 cases in Indonesia. It shows that monitoring trends of pneumonia of unknown cause in open-source data could provide rapid, unvalidated information for early detection of COVID-19 outbreaks. Although the information is unvalidated, once a signal is detected, it could prompt further investigation and validation. Consequently, such information may supplement traditional surveillance, which is frequently subject to delays in reporting and validation by local health authorities.<sup>4</sup>

This study has several limitations. First, there is a possibility of reporting bias because we may have captured increasing media awareness rather than an actual increase in disease occurrence. The early signal was detected in late January 2020, when COVID-19 had been detected in many countries worldwide. Second, this study relies on an online news-based surveillance system, which is unvalidated and may include other etiologies of pneumonia that were not COVID-19.<sup>14</sup> Third, we only included pneumonia in our study; however, COVID-19 has a wide range of symptoms, with most of those infected having a mild illness.<sup>15</sup> Nevertheless, this study has highlighted the ability of open-source data to identify

Fig. 1. Number of *EpiWATCH* reported cases of pneumonia of unknown cause in Indonesia between 1 November 2019 and 31 March 2020

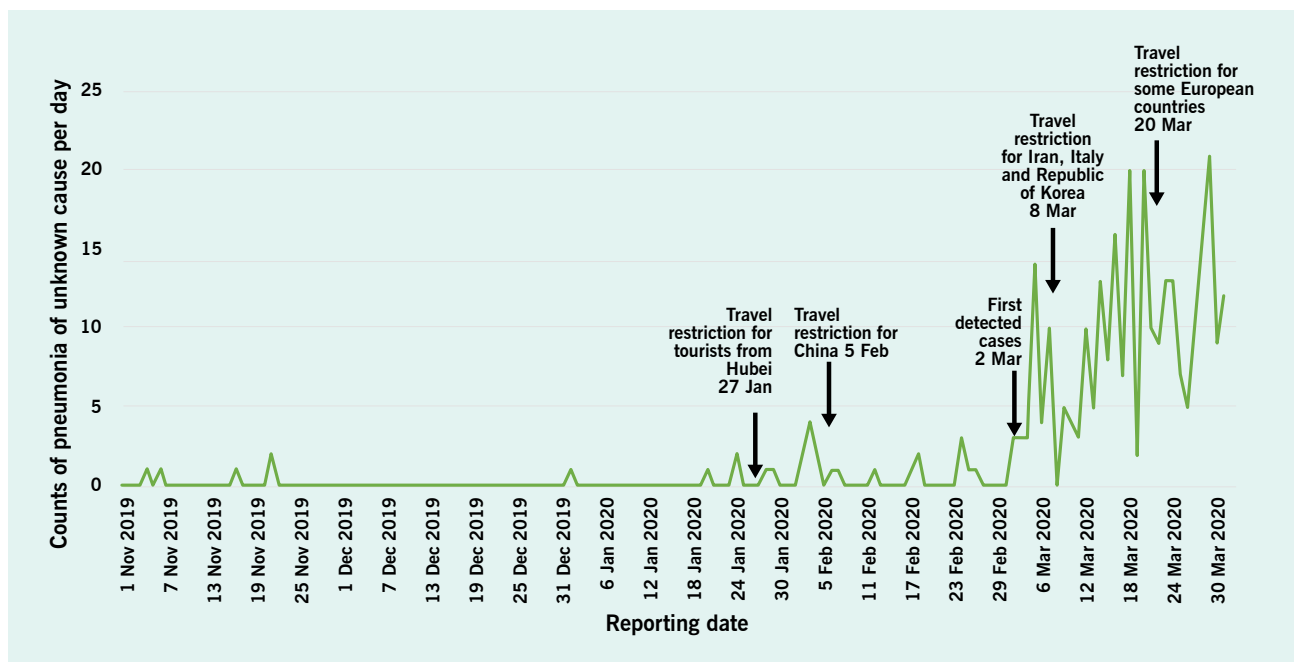
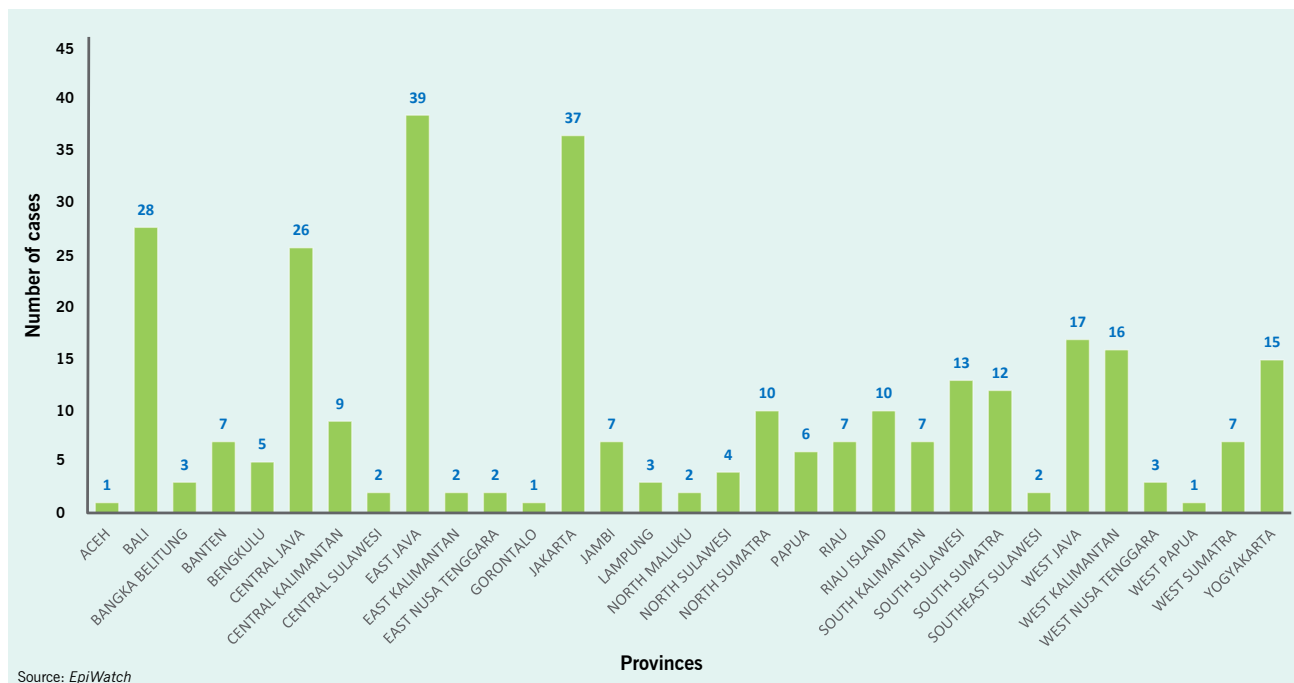
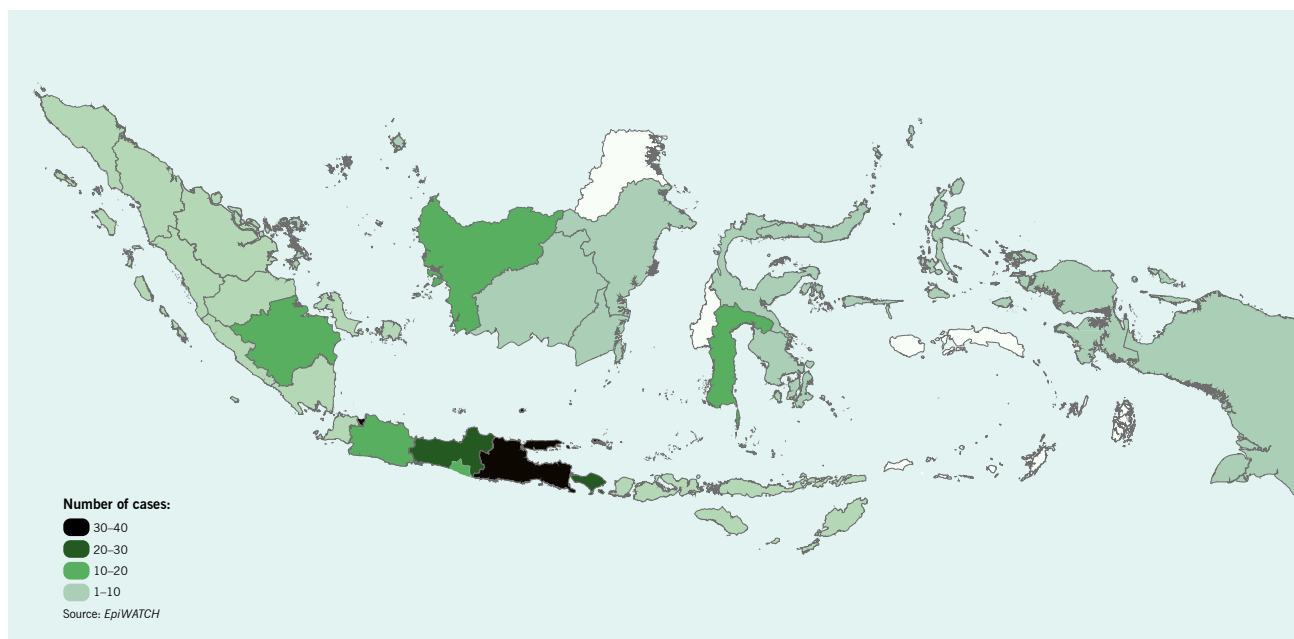


Fig. 2. Number of *EpiWATCH* reported cases of pneumonia of unknown cause in Indonesia, by province (November 2019 to March 2020)



Source: *EpiWatch*

Fig. 3. Geographical distribution of *EpiWATCH* reported cases of pneumonia of unknown cause in Indonesia (November 2019 to March 2020)



early alerts of pneumonia before the initial confirmation of COVID-19 cases in Indonesia, making such data a promising option to enhance epidemic surveillance.

## CONCLUSION

In this study, we detected possible early signals of the COVID-19 outbreak in Indonesia using an online news-based surveillance system that used Bahasa Indonesia through *EpiWATCH*. The earliest signals from the *EpiWATCH* observatory of pneumonia of unknown cause in Indonesia were in late January 2020, indicating possible unrecognized circulation of COVID-19 cases in Indonesia before the country's official notification of cases. The observed spatial pattern of cases between *EpiWATCH*-reported pneumonia of unknown cause and officially confirmed COVID-19 cases in Indonesia was similar.

Monitoring trends in open-source data can provide rapid, unvalidated information for the early detection of outbreaks. Although unvalidated, such information may be used to supplement or trigger investigation and testing. As *EpiWATCH* sources global information, this methodology can be repeated for other diseases and other countries within the Western Pacific Region.

## Conflicts of interests

The content of this manuscript is solely the responsibility of the authors, and we have no conflicts of interest to declare. We verify that all authors have seen and approved the final manuscript and declare the manuscript content herein has not been submitted for consideration or published elsewhere.

## References

1. Coronavirus disease (COVID-19) outbreak. World Health Organization; 2020. Available from: <https://www.who.int/westernpacific/emergencies/covid-19>, accessed 1 April 2020.
2. Confirmed imported case of novel coronavirus infection in Singapore; multi-ministry taskforce ramps up precautionary measures. Ministry of Health Singapore; 2020. Available from: <https://www.moh.gov.sg/news-highlights/details/confirmed-imported-case-of-novel-coronavirus-infection-in-singapore-multi-ministry-taskforce-ramps-up-precautionary-measures>, accessed 1 April 2020.
3. Kenyataan Akhbar KPK 26 Januari 2020 – Situasi Terkini Jangkitan Novel Coronavirus di Malaysia [Press statement 26 January 2020 – Current update on novel coronavirus situation in Malaysia]. Director-General of Health Malaysia; 2020. Available from: <https://kpkeshatan.com/2020/01/26/kenyataan-akhbar-kpk-26-januari-2020-situasi-terkini-jangkitan-novel-coronavirus-di-malaysia/>, accessed 1 April 2020.
4. May L, Chretien J-P, Pavlin JA. Beyond traditional surveillance: applying syndromic surveillance to developing settings—opportunities and challenges. *BMC Public Health*. 2009 Jul 16;9(1):242. doi:10.1186/1471-2458-9-242

5. Collier N. Uncovering text mining: a survey of current work on web-based epidemic intelligence. *Glob Public Health*. 2012;7(7):731–49. doi:10.1080/17441692.2012.699975
6. Epidemic intelligence – systematic event detection. Geneva: World Health Organization; 2016. Available from: <https://www.who.int/csr/alertresponse/epidemicintelligence/en/>, accessed 3 April 2020.
7. NHMRC Centre for Research Excellence, Integrated Systems for Epidemic Response. Available from: <https://iser.med.unsw.edu.au/epiwatch>, accessed 3 April 2020.
8. Kpozehouen EB, Chen X, Zhu M, Macintyre CR. Using open-source intelligence to detect early signals of COVID-19 in China: descriptive study. *JMIR Public Health Surveill*. 2020 Sep 18;6(3):e18939. doi:10.2196/18939
9. Harikasidi. Toddler from China treated in West Nusa Tenggara Hospital is not a coronavirus patient. *iNews Bali*; 2020. Available from: <https://bali.inews.id/berita/balita-asal-china-dirawat-di-rsud-ntb-negatif-virus-korona>.
10. COVID-19 update until April 1st 2020, 09.00 GMT+7.0. Indonesia Ministry of Health; 2020. Available from: <https://infeksiemerging.kemkes.go.id/category/situasi-infeksi-emerging/#.XoQ3x4gzZyw>, accessed 1 April 2020.
11. Mahendradhata Y, Trisnantoro L, Listyadewi S, Soewondo P, Marthias T, Harimurti P, et al. The Republic of Indonesia Health System Review. Vol 7, No 1. New Delhi: WHO Regional Office for South-East Asia; 2017.
12. Nababan HY, Hasan M, Marthias T, Dhital R, Rahman A, Anwar I. Trends and inequities in use of maternal health care services in Indonesia, 1986–2012. *Int J Womens Health*. 2017 Dec 29;10:11–24. doi:10.2147/IJWH.S144828
13. Pemerintah Perluas Lokasi Pemeriksaan Covid-19, Berikut Daftarnya [Government decided to expand the location for COVID-19 testing, here is the list]. Indonesia Ministry of Health; 2020. Available from: <http://sehatnegeriku.kemkes.go.id/baca/rilis-media/20200320/3333441/pemerintah-perluas-lokasi-pemeriksaan-covid-19-berikut-daftarnya/>, accessed 4 April 2020.
14. Aiello AE, Renson A, Zivich PN. Social media- and Internet-based disease surveillance for public health. *Annu Rev Public Health*. 2020 Apr 2;41(1):101–18. doi:10.1146/annurev-publ-health-040119-094402, accessed 29 October 2020.
15. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Geneva: World Health Organization; 2020. Available from: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>, accessed 1 April 2020.