

# Research Article Respiratory illness of unknown cause as an early signal of the COVID-19 pandemic

Rosalie Chen<sup>\*1</sup>, Mohana Kunasekaran<sup>1</sup>, Aye Moa<sup>1</sup>, Samsung Lim<sup>1,2</sup>, Abrar Ahmad Chughtai<sup>1,3</sup>, Elizabeth B Kpozehouen<sup>1</sup>, Xin Chen<sup>1</sup>, Ashley Quigley<sup>1</sup>, C Raina MacIntyre<sup>1,4</sup>

- 1. Biosecurity Program, The Kirby Institute, The University of New South Wales, Sydney, NSW, Australia
- 2. School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW, Australia
- 3. School of Population Health, The University of New South Wales, Sydney, NSW, Australia
- 4. Watts College of Public Service and Community Solutions, Arizona State University, Phoenix, AZ, United States

\*Corresponding author: Rosalie Chen, Biosecurity Program, Kirby Institute, The University of New South Wales Postal Address: Level 6, Wallace Wurth Building, Kensington, New South Wales 2052, Australia Email: <u>rosalie.chen@unsw.edu.au</u>; Phone: +61 2 9348 1252

# Abstract

**Background**: As the COVID-19 pandemic began, cases of respiratory illness of unknown cause could have presented early disease signals. The detection of such signals can provide useful clues regarding disease emergence and opportunities for early outbreak control. This study assessed trends in the global incidence of reports of respiratory illness of unknown cause as early indicators of COVID-19 in late 2019, using the early warning system, EPIWATCH, to investigate the disease's origins and early epidemiology.

**Methods**: Global news articles published between 1-Oct and 31-Dec in 2019 and collected by EPIWATCH, an open-source Artificial Intelligence (AI)-driven epidemic surveillance system, were searched for keywords relating to unknown respiratory illness. Control periods representing the pre-COVID period (1-Oct to 31-Dec for the years 2016-2018) were used to compare report trends over time. Data on report location, case demographics, outbreak setting, and type of illness, were extracted and descriptively analysed. Fisher's exact tests were conducted to assess proportional differences between years.

**Results**: Overall, 113 EPIWATCH reports of unknown respiratory illness were identified from 12 countries in 2019. Most reports (n=111 [98.2%]) were published in Northern hemisphere countries, in particular Russia (n=72 [63.7%]) and the United States (n=17 [15.0%]). China experienced a month-on-month increase in reports. Schools were the most reported outbreak setting (n=49 [43.4%]). The proportion of included reports in 2019 was significantly higher compared to 2016-2018 (p<.001).

**Conclusion**: Potential early COVID-19 signals were detected from various countries before the end of 2019. AIdriven open-source disease surveillance systems, such as EPIWATCH, can provide early warnings to enhance disease surveillance and improve outbreak prevention and response.

**Keywords**: artificial intelligence, COVID-19, disease surveillance, epidemiology, respiratory illness of unknown cause

# Introduction

Open-source epidemic intelligence systems (EISs) that use syndromic surveillance can identify early outbreak signals, which may trigger earlier response and prevent or mitigate epidemics or pandemics, such as COVID-19. On 31-Dec-2019, the World Health Organization (WHO) China Country Office was notified of a cluster of pneumonia of unknown cause cases linked to a seafood wholesale market in Wuhan, China [1]. On 7-Jan-2020, the causal pathogen was identified as a novel coronavirus, SARS-CoV-2 [2]. COVID-19 subsequently spread rapidly across China

and internationally, with cases reported across 18 other countries by end of January 2020 [1]. The substantial disease burden of the current COVID-19 pandemic highlights the threat that emerging infectious diseases (EIDs) can pose to global health security [3-5]. Understanding the circumstances surrounding COVID-19 emergence is crucial for understanding the disease and guiding future responses to SARS-CoV-2 and other emerging infectious diseases.



Chen R, Kunasekaran M, Moa A, Lim S, Chughtai AA, Kpozehouen EB, Chen X, Quigley A, MacIntyre CR. Respiratory illness of unknown cause as an early signal of the COVID-19 pandemic. Global Biosecurity. 2024; 6(1)

Rapid detection of early EID signals provides earlier opportunities for diagnostics, outbreak control and drug development. Digital surveillance can provide valuable information regarding early outbreak trends, to identify areas of concern and trigger earlier investigation than formal laboratory or health system reporting [6].

Open-source data obtained from EPIWATCH, an Artificial Intelligence (AI)-driven disease surveillance system, has proven capability in early outbreak detection [7-9]. A previous study indicated that the 2014 West African Ebola epidemic could have been identified months prior to WHO notification [7]. Unstructured open-source data contains valuable information that can be harnessed by EISs and used to detect early warnings of serious epidemics [10].

Open-source data captured by EISs could also inform understanding of the origins of emerging diseases. COVID-19 is believed to have originated around the Huanan Seafood Market in Wuhan, China [11,12]. Recent genomic analyses of environmental samples obtained from the market have stirred controversy about the origin of COVID-19 [13,14]. However, current research has not identified a natural or intermediate host for SARS-CoV-2 or confirmed whether COVID-19 first emerged at the seafood market or elsewhere. Retrospective serological and wastewater sampling studies suggest that SARS-CoV-2 virus was already circulating in late 2019 in Europe and the Americas, much earlier than official reports of cases [15-22]. Other studies have also first retrospectively analysed potential early COVID-19 signals prior to December 2019 using EPIWATCH data [8,23,24].

The aim of this study was to identify early signals of potential COVID-19 globally, to further inform the question of the origin and early epidemiology of the virus.

#### Methods

#### Search strategy and screening process

EPIWATCH is an AI-driven disease surveillance system that harnesses open-source data to monitor EIDs [25]. EPIWATCH scans daily media articles and other Internet sources using two AI systems and prespecified search terms in 42 different languages (at the time of this study) for clinical syndromes and specific diseases worldwide [25], enabling rapid detection of early outbreak signals.

The EPIWATCH database was searched for reports of respiratory illness of unknown cause published during 1-Oct-2019 to 31-Dec-2019 as the period of interest for early signals of COVID-19, and during control periods representing the pre-COVID period (1-Oct to 31-Dec across 2016-2018). Keywords used by Kpozehouen et al. (2020) were adapted and expanded upon [23], and included words such as 'pneumonia', 'respiratory infection', 'ARVI' (acute respiratory viral infection), and 'SARI' (severe acute respiratory infection), among others (Supplementary Materials, Table S1). Due to evidence that SARS-CoV-2 virus was circulating in late 2019 [15-22], a start date of 1-Oct-2019 was chosen as a plausible time for the first potential early COVID-19 signals. To assess whether EPIWATCH could detect early COVID-19 signals prior to widespread news reporting, an end date of 31-Dec-2019 was chosen, as many countries began reporting their first suspected or official COVID-19 cases in January 2020. Unknown respiratory illness reports from 1-Oct to 31-Dec across 2016-2018 were analysed. Two reviewers (RC and MK) divided and analysed the EPIWATCH reports for cases or outbreaks with search terms relating to respiratory illness of unknown cause, unknown disease, and hospital outbreaks (Supplementary Materials Table S1), with gueries resolved through discussion.

# Inclusion/exclusion criteria

Reports were included if they provided recent information on cases or outbreaks of respiratory illness of unknown origin. Reports of unknown disease and hospital outbreaks were included if other keywords were present, or cases had evidence of respiratory symptoms. Non-English language reports were also included by translating them into English using Google Translate [26] and double-checked using Bing Microsoft Translator [27]. Reports were excluded if cause of illness was identified (e.g., pneumonia due to an identified bacterial or viral cause), if recent information on cases or outbreaks was not provided (e.g., historical trend data, advice on control measures, opinion pieces, etc.), and if they could no longer be accessed even with searching of web archives. Reports were also excluded if they did not provide any geospatial information that allowed identification of country location.

# Data extraction and analysis

Data were extracted from EPIWATCH reports. Reports were classified by country and publication date to assess geographic trends in unknown respiratory illness reports over time. Reports were further categorised by geolocation, case or outbreak setting, and age of cases. Children were defined as <18 years old, and adults were defined as ≥18 years [28]. Information on laboratory testing and respiratory illness type was also extracted.

To demonstrate the potential application of opensource surveillance data for describing early outbreak signals, data from the three high- or upper-middle income countries with the most reports were summarised and displayed as heatmaps to depict the geographic distribution and intensity of early outbreak signals across economically similar countries [29]. Choropleth maps were created using ArcGIS Pro software to depict the variations across the aggregated data points [30]. Report locations were further classified by school, city/town/village, district, and federal jurisdiction if possible. Reports could provide information on cases in both children and adults, and



list multiple case/outbreak settings or geographic locations.

Descriptive analyses were conducted, including tables, epidemic curves, and heatmaps to present the distribution of reports according to the categories mentioned above. Fisher's exact tests were also conducted using R4.2.1 [31] to assess for proportional differences in reports between the study and control periods. For reports discussing nation- or state- wide cases, no specific outbreak setting could be obtained. Such articles were still eligible for inclusion in descriptive and statistical analyses assessing proportions of included reports. A sub-analysis was conducted excluding reports that included only paediatric cases or that did not provide demographic information.

# Ethics

Ethics approval was not required for this study as no human or animal subjects were involved and all data were open-source and anonymous.

#### Results

### Study period, October-December 2019

Between 1-Oct-2019 and 31-Dec-2019, 1,518 reports of various news articles were published. Of the 360 reports matching our search terms, 190 were excluded because they did not meet inclusion criteria (133 reported a diagnosis, 33 described diseases with no respiratory symptoms, 15 described animal diseases, five did not provide current case information, and four were no longer accessible). Overall, 113 undiagnosed respiratory illness reports from 12 countries were included (Supplementary Materials, Tables S3 and S4). Of these, 111 reports were from ten northern hemisphere (98.2%) countries: Russia, the United States (US), China, India, the United Kingdom (UK), Malaysia, Nigeria, Ireland, Sudan, and Trinidad and Tobago; while two (Kenya and Indonesia) were from equatorial countries (Supplementary Materials, Figure S1). Russia had the

most reports (n=72 [63.7%]), followed by the US (n=17[15.0%]), China (n=6 [5.3%]), India (n=6 [5.3%]), and the UK (n=5 [4.4%]), while the other countries each had one report (0.9%). Forty-four reports mentioned laboratory testing but did not provide results. Several reports used the terms SARI or ARVI interchangeably. Most reports described SARI/ARVI cases (n=50 [44.2%]), followed by pneumonia (n=39 [34.5%]), influenza-like illness (n=27 [23.9%]), fever (n=8 [7.1%]), upper respiratory tract infection (URTI) (n=2 [1.8%]), bronchiolitis (n=1 [0.9%]), and lung injury (n=1 [0.9%]). Of reports with demographic information, 67 (59.3%) described respiratory illnesses in children, while 35 reports (31.9%) described cases in adults. Schools were the most reported outbreak setting (n=49 [43.4%]), while two reports (1.8%), both published on 31-Dec-2019, described cases linked to the Wuhan Seafood Market. There were single reports (0.9%) of cases associated with an aged care facility and a boat, and remaining reports did not specify outbreak setting.

Russia, the US, and China were the three high- or upper-middle-income countries with the most reports (a summary of case/outbreak characteristics are provided in Table 1). In Russia, 136 locations were mentioned across 38 federal subjects, with 16 reports (22.2%) listing multiple locations (Figure 1; for an animated map, please see Supplementary Materials). Penza Oblast was the most common location (n=25 [34.7%]), followed by the Komi Republic (n=18 [25.0%]) and Rostov Oblast (n=7 [9.7%]). In the US, report locations were distributed across 30 states with most reports from Louisiana (n=5 [29.4%]), while Alabama, California, South Carolina, and Texas had four reports (23.5%) each (Figure 2). Five reports (29.4%) described nationwide cases. In China, the most reported location was Hubei province (n=4 [50.0%]), followed by Macau SAR (Special Administrative Region) (n=3 [37.5%]) and Taiwan (n=1 [12.5%]) (Figure 3).

Table 1. Ch	naracteristics of EPIWATCH	reports of resp	oiratory illness	of unknown	cause for 1 October to 31
	December 2019	for Russia, the	e United States	s, and China	

Case/Outbreak Information	Number of EPIWATCH reports of respiratory illness of unknown cause				
	Oct-2019	Nov-2019	Dec-2019	Total	
Russia					
<b>Overall Reports</b>	22	27	23	72	
Demographic					
Children	14	20	12	46	
Adults	4	9	5	18	
Not stated	8	7	11	26	
Setting					
Community-wide	16	19	18	53	
School	10	13	14	37	
Not stated	-	2	-	2	
Type of Illness					
ARVI or SARI*	15	15	19	49	
Pneumonia	11	16	6	33	



Influenza-like illness	3	1	1	5
Fever	-	1	-	1
Laboratory Testing				
Reported	5	8	10	23
United States				
Overall reports	8	4	5	17
Demographic				
Adults	3	2	4	9
Children	3	1	3	7
Not stated	3	1	1	5
Setting				
Community-wide	4	3	4	11
School	3	1	-	4
Not stated	3	-	1	4
Type of Illness				
<b>ARVI</b> or SARI*	3	2	5	10
Pneumonia	1	1	-	2
Lung injury	1	-	-	1
Laboratory Testing				
Reported	4	3	3	10
China				
<b>Overall Reports</b>	-	2	4	6
Demographic				
Adults	-	2	-	2
Children	-	1	1	2
Not stated	-	-	3	3
Setting				
Community-wide	-	-	2	2
Seafood market	-	-	2	2
Aged care	-	1	-	1
School	-	1	-	1
Type of Illness				
Pneumonia	-	2	2	4
Influenza-like illness	-	1	2	3
Laboratory Testing				
Reported	-	1	3	4



**Figure 1.** Locations of reports of respiratory illness of unknown cause from Russia between October – December 2019





Figure 2. Locations of reports of respiratory illness of unknown cause from the United States between October – December 2019



**Figure 3.** Locations of reports of respiratory illness of unknown cause from China between October – December 2019



#### Comparison of 2016-2018 and 2019 EPIWATCH reports

Overall, 25 reports of respiratory illness of unknown cause were identified during the control periods across previous years (1-Oct to 31-Dec, 2016-2018), for an annual average of 8.33 reports (for a summary of report characteristics, see Supplementary Materials, Tables S5 and S6). Most unknown respiratory illness reports were from northern hemisphere countries (n=24 [96.0%]), while one (4.0%) was from an equatorial country (Colombia). Most reports were from the US, with eight reports in 2016 and three in 2017. China had one report of six deaths during the influenza season in October 2016 and two reports in 2017. India had two reports in 2016 and one report of a death due to fever in October 2018. The following countries had one report each during the control period: Myanmar (published 5-Oct-2016), Colombia (12-Oct-2016), Canada (15-Nov-2016), the UK (7-Dec-2016), France (27-Dec-2016), Sweden (29-Dec-2016), Italy (5-Dec-2017) and Mexico (17-Dec-2017) [32]. Seven reports (28.0%) described adult cases, while 11 (44.0%) described paediatric cases, and the rest did not specify. Nine reports mentioned

laboratory testing being conducted. In total, 18 (72.0%) reports described influenza-like illness cases, three (12.0%) each described pneumonia and fever cases, two (8.0%) each reported on cases of ARVI/SARI and URTI cases. Nine reports (32.1%) provided case/outbreak setting, with six (21.4%) reporting on schools, and one (3.6%) each reporting on aged care facilities, a household, and a water park. Remaining reports did not state a setting.

The US, China, India, and the UK were the only countries with EPIWATCH reports of respiratory illness of unknown cause during both 2019 and the control periods, and each experienced an increase in reports in 2019. The annual average for 2016-2018 was 13.6 times fewer than the same period in 2019. The proportions of included reports for 2016, 2017, and 2018, were 17/655 (2.60%), 7/425 (1.65%), and 1/494 (0.20%), respectively. The proportion of included reports in 2019 was significantly different from reports in 2016 (two-tailed p<.001), 2017 (two-tailed p<.001) and 2018 (two-tailed p<.001). The US and China both experienced an increase in reports, while no reports were identified for Russia during 2016-2018 (Figure 4).





Figure 4. Trends in reports of respiratory illness of unknown cause between October-December, 2016-2018 compared to 2019, in Russia, the United States, and China.

# Sub-analysis excluding EPIWATCH reports with paediatric cases

Further sub-analysis was performed excluding reports that described paediatric cases. During 2019, 36 reports described adult cases only (18 from Russia, 9 from the US, three from India, two from China, and one each from Indonesia, Malaysia, Sudan, and the UK). Most reports (n=30) described community-wide cases. Specific outbreak settings included: schools (n=11) in Russia (n=8), China (n=1), the UK (n=1), and the US (n=1); an aged care facility (n=1) in Macau SAR; and a boat (n=1) in Malaysia. Reports described cases of ARVI/SARI (n=13), pneumonia (n=13), influenza-like illness (n=9), fever (n=4), and a single report each described URTI cases and cases of lung injury being investigated for possible associations with vaping. In previous control periods, there were six reports in 2016, one report in 2017 and none in 2018 (Table 2). There was a significant difference between the proportions of reports in 2019 compared to the control periods (Table 2).



**Table 2.** Comparison of EPIWATCH reports of respiratory illness of unknown cause between October-December2019 and control periods in 2016-2018 for adult cases.

	Number of Reports	Fisher's exact test p-value*
2016	6	0.026
2017	1	0.002
2018	0	<.001

\*Compared to 36 reports in 2019.

### Discussion

demonstrates that open-source This studv intelligence could be used for syndromic surveillance to flag the emergence of new diseases. Potential early COVID-19 signals were detected across 12 countries between October to December 2019, prior to first official notification. Most reports were from the northern hemisphere, in particular Russia and the US, while China had a month-on-month increase in reports. We also identified a report of a severe pneumonia case in a 63vear-old man on 22-Nov-2019 from Xiangvang, Hubei, who was medically evacuated to Wuhan. The data identified schools as the most common outbreak setting, which suggests other respiratory viruses in these cases, as children generally experience milder COVID-19 illness compared to adults, though moderate or severe COVID-19 disease and school outbreaks can still occur in children [33-37]. These cases could also reflect lack of testing for known pathogens. Early in the pandemic, it was assumed that transmission of SARS-COV-2 was rare in children and schools, but subsequently, many school outbreaks have been described, especially during times of high community transmission [38-40]. In a closed indoor setting such as a classroom, long-range airborne SARS-CoV-2 transmission may occur, particularly if ventilation is poor [41]. However, the first recognition of an unknown pneumonia during the pandemic was in adults in Wuhan. Even when reports describing paediatric cases were excluded, the proportion of unknown respiratory illness reports in late 2019 significantly increased compared to previous years.

Several countries (Russia, the US, China, India, the UK) experienced an increase in unknown respiratory illness reports in late 2019 compared to previous years. In Russia and the US, respiratory illnesses had reportedly increased earlier than expected for the season, with more severe disease, and unusually high numbers. Several factors could explain these cases, such as seasonal trends in common respiratory illnesses [42], lack of laboratory testing or reporting of diagnostic results, or the gas explosion on 16-Sep-2019 at the State Research Centre of Virology and Biotechnology (Vector Institute) in Siberia, which sparked concerns of accidental release of infectious pathogens [43]. There were reports of several respiratory illness cases that could not be diagnosed despite laboratory testing, including one report from Kemerovo Oblast in Siberia, dated six weeks after the Vector lab explosion [32]. In the US, vaping-related lung injuries could have contributed to the large number of reports, as the country experienced an outbreak which peaked in September 2019 and caused 2,807 cases or deaths as of mid-February 2020 [44]. However, there have been claims that several of these cases with symptom-onset prior to 2020 could have been COVID-19 and were misdiagnosed [45]. One US study retrospectively detected anti-SARS-CoV-2 antibodies in blood donations collected in mid-December 2019, prior to the country's first official reported case on 21-Jan-2020 [46]. Similar studies in France, Italy, and Brazil conducted retrospective testing of serum samples, frozen respiratory samples and wastewater, and also found evidence of SARS-CoV-2 RNA or antibodies from November onward in 2019 [15-18,20-22]. However, retrospective serological testing has limitations, as early serological tests may have lacked specificity [17,19,21]. In China, there were claims that COVID-19 emerged in connection with the Wuhan World Military Games between 18- to 27-Oct-2019, with reports athletes experiencing undiagnosed respiratory of illnesses [47]. Open-source EISs can provide early warnings of serious emerging or re-emerging infections, given their ability to quickly capture local news and community chatter prior to official signal detection by authorities.

This study was not without limitations. Firstly, non-COVID causes of unknown respiratory illness cases may not have been reported, or there may have been lack of testing, therefore presenting challenges in identifying whether early signals were due to an unknown or novel disease. The study is retrospective. Although EPIWATCH existed in 2019, at the time COVID-19 was emerging, it was unfunded, hence no analysts were available to review the collected data. Other limitations include changes in language searching in EPIWATCH. The Russian language was only added to the EPIWATCH system in September 2019, so the increase in reports in late 2019 for Russia may reflect increased detection due to the addition of the Russian language. Language bias could also influence the number of reports gathered. English-language bias may be more likely to occur, as many countries report news in both their native language(s) and English [48]. However, in 2019, most reports were from Russia, which could be due to media bias, as Russia has a greater number of media outlets compared even to more densely populated countries such as China and the US [49-51]. In countries with fewer media resources, early signals may not be reported in news outlets or social media. Nonetheless, 73% of intelligence captured by EPIWATCH is in non-English languages. Sociocultural-specific disease reporting (e.g., use of colloquialisms, local terminology) [52] and sociopolitical factors (e.g., deliberate under-reporting,



censorship) may also influence disease reporting. Our study did not search for colloquialisms or individual symptoms which may have limited the reports identified from countries. Despite these limitations, this study demonstrates the potential applications of open-source disease surveillance systems for detecting early disease signals and analysing outbreaks.

The data captured by open-source EISs can trigger earlier public health investigation, diagnostics, and response to future outbreaks and pandemics. Disease surveillance is therefore more robust when a combined approach of open-source digital surveillance and traditional surveillance is used. However, there are still improvements to be made, including the use of large language models in low-resource languages. Transparent, timely dissemination of information is essential, as delayed official reporting allows disease spread. With SARS-CoV-2, early identification of cases focused on severe disease presenting as pneumonia of unknown cause, but mild cases were not included in early case surveillance in Wuhan, and asymptomatic or presymptomatic SARS-CoV-2 transmission was not recognised [53]. Equity in surveillance efforts should be considered, given the likelihood of under-reporting and lack of diagnostics in low-income countries [54]. Lastly, appropriate physical, political, and human resources must be available for acting on early warnings to ensure timely outbreak containment. With the often unique and unforeseen challenges presented by (re)emerging infectious diseases, robust open-source disease surveillance systems could provide a crucial tool for early outbreak detection and response, particularly in lowresource settings [55,56].

#### Conclusion

This study demonstrates that open-source digital disease surveillance systems, such as EPIWATCH, can provide early warning signals which can prompt earlier diagnosis and mitigation of outbreaks with pandemic potential. Whilst our findings do not provide definitive evidence of the presence of SARS-CoV-2 prior to official notification on 31-Dec-2019, the available open-source data suggests signals of COVID-19 may have been detected earlier. Such methods can provide a complementary approach to traditional surveillance for providing early warning of outbreaks, and for analysis of disease origins and emergence. Open-source epidemic intelligence may be of particular value in low-income countries with weak surveillance systems.

#### Acknowledgments

We thank Nicholas Ojo for his assistance with the EPIWATCH system and data collection during their time as a student with the Biosecurity Program, the Kirby Institute.

#### **Author Contributions**

RC: study design, analysis and interpretation of data, manuscript drafting and revision, manuscript submission MK: study design, analysis and interpretation of data, manuscript drafting and revision

AM: study design, manuscript revision

SL: analysis of data, manuscript revision

AAC: manuscript revision

EBK: acquisition of data, manuscript revision

XC: analysis of data, manuscript revision

AQ: study design, manuscript revision

CRM: conceived study, initial study design, manuscript revision

#### **Declarations of Interest**

The authors declare that there is no conflict of interest.

#### **Funding Statement**

The following authors are supported by the Balvi Filantropic Fund: RC, MK, SL, EBK, XC, and AQ. CRM is supported by a NHMRC Investigator Grant [2016907] and MK is supported by a NHMRC Centre for Research Excellence BREATHE grant [APP2006595]. The funding sources had no role in: study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

#### References

- 1. Archived: WHO timeline COVID-19. World Health Organization. Updated 2020. https://www.who.int/news/item/27-04-2020who-timeline---covid-19
- 2. Zheng J. SARS-CoV-2: An emerging coronavirus that causes a global threat. Int J Biol Sci. 2020;16(10):1678-1685. doi:10.7150/ijbs.45053
- 3. WHO coronavirus (COVID-19) dashboard. World Health Organization. Updated 2023. https://covid19.who.int/
- 4. Pak A, Adegboye OA, Adekunle AI, Rahman KM, McBryde ES, Eisen DP. Economic consequences of the COVID-19 outbreak: The need for epidemic preparedness. Perspective. Front Public Health. 2020-May-29 2020;8. doi:10.3389/fpubh.2020.00241
- The unequal impact of COVID-19: A spotlight on frontline workers, migrants and racial/ethnic minorities. OECD. Updated 2022. https://read.oecdilibrary.org/view/?ref=1133\_1133188lq9ii66g9w&title=The-unequal-impact-of-COVID-19-A-spotlight-on-frontline-workersmigrants-and-racial-ethnic-minorities
- Aiello AE, Renson A, Zivich PN. Social mediaand Internet-based disease surveillance for public health. Annu Rev Public Health. Apr 2 2020;41:101-118. doi:10.1146/annurevpublhealth-040119-094402
- 7. MacIntyre CR, Lim S, Quigley A. Preventing the next pandemic: Use of artificial intelligence for epidemic monitoring and alerts. Cell Rep Med.



Chen R, Kunasekaran M, Moa A, Lim S, Chughtai AA, Kpozehouen EB, Chen X, Quigley A, MacIntyre CR. Respiratory illness of unknown cause as an early signal of the COVID-19 pandemic. Global Biosecurity. 2024; 6(1)

Dec 2022;3(12):100867. 20 doi:10.1016/j.xcrm.2022.100867

- 8. Thamtono Y, Moa A, MacIntyre CR. Using opensource intelligence to identify early signals of COVID-19 in Indonesia. Western Pac Surveill Response 02/172021;12(1):6. J. doi:10.5365/wpsar.2020.11.2.010
- Puca C, Trent M. Using the surveillance tool 9. EpiWATCH to rapidly detect global mumps outbreaks. Global Biosecurity. 2020;2. doi:10.31646/gbio.54
- 10. Meckawy R, Stuckler D, Mehta A, Al-Ahdal T, Doebbeling BN. Effectiveness of early warning systems in the detection of infectious diseases outbreaks: A systematic review. BMC Public Health. 2022/11/29 2022;22(1):2216. doi:10.1186/s12889-022-14625-4
- 11. Worobey M, Levy JI, Malpica Serrano L, et al. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. Science. Aug 26 2022;377(6609):951-959. doi:10.1126/science.abp8715
- 12. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020/04/01 2020;26(4):450-452. doi:10.1038/s41591-020-0820-9
- 13. Gao G, Liu W, Liu P, et al. Surveillance of SARS-CoV-2 in the environment and animal samples of the Huanan Seafood Market. Research Squarecris. 2022. doi:10.21203/rs.3.rs-1370392/v1
- 14. Crits-Christoph A, Gangavarapu K, Pekar JE, et al. Genetic evidence of susceptible wildlife in SARS-CoV-2 positive samples at the Huanan Wholesale Seafood Market, Wuhan: Analysis and interpretation of data released by the Chinese Center for Disease Control. Zenodo. 2023. doi:10.5281/zenodo.7754299
- 15. Trombetta CM, Marchi S, Viviani S, et al. A serological investigation in Southern Italy: Was SARS-CoV-2 circulating in late 2019? Hum Vaccines Immunother. 2022/11/30 2022;18(5):2047582.

doi:10.1080/21645515.2022.2047582

- 16. La Rosa G, Mancini P, Bonanno Ferraro G, et al. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. Sci Total Environ. Jan 2021;750:141711. 1 doi:10.1016/j.scitotenv.2020.141711
- 17. Carrat F, Figoni J, Henny J, et al. Evidence of early circulation of SARS-CoV-2 in France: from the population-based Findings "CONSTANCES" cohort. Eur J Epidemiol. 2021/02/01 2021;36(2):219-222. doi:10.1007/s10654-020-00716-2
- 18. Deslandes A, Berti V, Tandjaoui-Lambotte Y, et al. SARS-CoV-2 was already spreading in France in late December 2019. Int J Antimicrob Agents.

Jun

2020;55(6):106006. doi:10.1016/j.ijantimicag.2020.106006

- 19. Basavaraju SV, Patton ME, Grimm K, et al. Serologic testing of US blood donations to identify Severe Acute Respiratory Syndrome (SARS-CoV-2)-reactive Coronavirus 2 antibodies: December 2019–January 2020. Clin Infect Dis. 2020;72(12):e1004-e1009. doi:10.1093/cid/ciaa1785
- 20. Fongaro G, Stoco PH, Souza DSM, et al. The presence of SARS-CoV-2 RNA in human sewage in Santa Catarina, Brazil, November 2019. Sci Total Environ. 2021/07/15/ 2021;778:146198. doi:https://doi.org/10.1016/j.scitotenv.2021.14 6198
- 21. Apolone G, Montomoli E, Manenti A, et al. Unexpected detection of SARS-CoV-2 antibodies in the prepandemic period in Italy. Tumori. 2021;107(5):446-451. doi:10.1177/0300891620974755
- 22. Gragnani L, Monti M, Santini SA, et al. SARS-CoV-2 was already circulating in Italy, in early December 2019. Eur Rev Med Pharmacol Sci. 2021;25(8):3342-3349. Apr doi:10.26355/eurrev\_202104\_25746
- 23. 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. 2020/9/18 2020;6(3):e18939. doi:10.2196/18939
- 24. Nair SP, Moa A, MacIntyre CR. Investigation of early epidemiological signals of COVID-19 in India using outbreak surveillance data. Global Biosecurity. 2020;2(1). doi:10.31646/gbio.72
- 25. MacIntyre CR, Chen X, Kunasekaran M, et al. Artificial intelligence in public health: the potential of epidemic early warning systems. J Int Med Res. Mar 2023;51(3):3000605231159335. doi:10.1177/03000605231159335
- 26. Google Translate. Google. https://translate.google.com/
- 27. Bing Microsoft Translator. Microsoft. https://www.bing.com/translator
- 28. Australian legal definitions: When is a child in protection? need of https://aifs.gov.au/resources/resourcesheets/australian-legal-definitions-when-childneed-protection
- 29. The World Bank. World Bank country and lending groups. Internet Archive. https://web.archive.org/web/20190321074546 /https://datahelpdesk.worldbank.org/knowled gebase/articles/906519-world-bank-countryand-lending-groups
- 30. About ArcGIS. Esri. https://www.esri.com/enus/arcgis/about-arcgis/overview
- 31. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2022. https://www.R-project.org/



- 32. AI-driven open-source outbreak observatory. The Kirby Institute, UNSW Sydney. https://www.epiwatch.org/about
- 33. Forrest CB, Burrows EK, Mejias A, et al. Severity of acute COVID-19 in children <18 years old March 2020 to December 2021. Pediatrics. 2022;149(4):e2021055765.

doi:10.1542/peds.2021-055765

- 34. Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. Acta Paediatr. 2020/06/01 2020;109(6):1088-1095. doi:https://doi.org/10.1111/apa.15270
- 35. Martin B, DeWitt PE, Russell S, et al. Characteristics, outcomes, and severity risk factors associated with SARS-CoV-2 infection among children in the US National COVID Cohort Collaborative. JAMA Netw Open. 2022;5(2):e2143151-e2143151. doi:10.1001/jamanetworkopen.2021.43151
- 36. Vosoughi F, Makuku R, Tantuoyir MM, et al. A systematic review and meta-analysis of the epidemiological characteristics of COVID-19 in BMC Pediatr. children. 2022/10/22 2022;22(1):613. doi:10.1186/s12887-022-03624-4
- 37. Niño-Serna LF, López-Barón E, Maya Ángel IC, Tamayo-Múnera C. Clinical characteristics of children with SARS-CoV-2 infection in a hospital in Latin America. Brief Research Report. Front Pediatr. 2022-June-09 2022;10. doi:10.3389/fped.2022.921880
- 38. Stein-Zamir C, Abramson N, Shoob H, et al. A large COVID-19 outbreak in a high school 10 davs after schools' reopening, Israel, May 2020. Euro Surveill. Jul 2020;25(29).doi:10.2807/1560-7917.Es.2020.25.29.2001352
- 39. Otte Im Kampe E, Lehfeld AS, Buda S, Buchholz U, Haas W. Surveillance of COVID-19 school outbreaks, Germany, March to August 2020. Euro Surveill. Sep 2020;25(38). doi:10.2807/1560-7917.Es.2020.25.38.2001645
- 40. Ismail SA, Saliba V, Lopez Bernal J, Ramsay ME, Ladhani SN. SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. Lancet Infect 2021;21(3):344-353. Dis Mar doi:10.1016/s1473-3099(20)30882-3
- 41. Palmer JC, Duval D, Tudge I, et al. Airborne transmission of SARS-CoV-2 over distances greater than two metres: A rapid systematic review. medRxiv. 2021:2021.10.19.21265208. doi:10.1101/2021.10.19.21265208
- 42. Murdoch KM, Mitra B, Lambert S, Erbas B. What is the seasonal distribution of community acquired pneumonia over time? A systematic review. Australas Emerg Nurs J. Feb 2014;17(1):30-42.doi:10.1016/j.aenj.2013.12.002

- 0239-9 https://www.bbc.com/news/world-europe-
  - 50. Thomala LL. Number of newspapers in China from 2011 to 2021. Statista. Updated 2023. https://www.statista.com/statistics/279182/nu mber-of-newspapers-in-china/
  - 51. Number of daily newspapers in the United States from 1970 to 2018. Statista. Updated 2023. https://www.statista.com/statistics/183408/nu mber-of-us-daily-newspapers-since-1975/
  - 52. Sulaiman FB, Yanti NKS, Lesmanawati DAS, Trent M, Chughtai AA, MacIntyre CR. Language specific gaps in identifying early epidemic signals – A case study of the Malay language. Global Biosecurity. 2019. doi:10.31646/gbio.33
  - 53. Casey-Bryars M, Griffin J, McAloon C, et al. Presymptomatic transmission of SARS-CoV-2 infection: A secondary analysis using published data. BMJ Open. 2021;11(6):e041240. doi:10.1136/bmjopen-2020-041240
  - 54. Whittaker C, Walker PGT, Alhaffar M, et al. Under-reporting of deaths limits our understanding of true burden of covid-19. BMJ. 2021;375:n2239. doi:10.1136/bmj.n2239
  - 55. Xu L, Zhou C, Luo S, Chan DK, McLaws M-L, Liang W. Modernising infectious disease surveillance and an early-warning system: The need for China's action. Lancet Reg Health -

Chen R, Kunasekaran M, Moa A, Lim S, Chughtai AA, Kpozehouen EB, Chen X, Quigley A, MacIntyre CR. Respiratory illness of unknown cause as an early signal of the COVID-19 pandemic. Global Biosecurity. 2024; 6(1)

- 43. MacIntyre CR, Chen X, Kunasekaran M, et al. Tailored intelligence to detect unusual epidemic activity following the explosion at Vector, Global Biosecurity. Russia. 2020;2. doi:10.31646/gbio.85
- 44. Outbreak of lung injury associated with the use of e-cigarette, or vaping, products. Centers for Disease Prevention and Control. Updated 2021. https://www.cdc.gov/tobacco/basic\_informati on/e-cigarettes/severe-lung-disease.html
- 45. Exclusive: Some 2019 EVALI patients in the US may have been infected with COVID-19: sources. Global Times. https://www.globaltimes.cn/page/202107/123 0143.shtml
- 46. First travel-related case of 2019 novel coronavirus detected in United States. Centers for Disease Prevention and Control. https://www.cdc.gov/media/releases/2020/p0 121-novel-coronavirus-travel-case.html
- 47. Squitieri T. Did the Military World Games spread COVID-19? The American Prospect. Updated 2020. https://prospect.org/coronavirus/did-themilitary-world-games-spread-covid-19/
- 48. Hamborg F, Meuschke N, Gipp B. Bias-aware news analysis using matrix-based news aggregation. Int J Digit Libr. 2020/06/01 2020;21(2):129-147. doi:10.1007/s00799-018-
- 49. Russia media guide. BBC News. Updated 2023. 17840134



West. 2022;23. doi:10.1016/j.lanwpc.2022.100485

56. Kostkova P, Saigí-Rubió F, Eguia H, et al. Data and digital solutions to support surveillance strategies in the context of the COVID-19 pandemic. Review. Front Digit Health. 2021-August-06 2021;3. doi:10.3389/fdgth.2021.707902

**How to cite this article**: Chen R, Kunasekaran M, Moa A, Lim S, Chughtai AA, Kpozehouen EB, Chen X, Quigley A, MacIntyre CR. Respiratoryillness of unknown cause as an early signal of the COVID-19 pandemic. *Global Biosecurity.* 2024; 6(1)

#### Published: May 2024

**Copyright:** Copyright © 2024 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <u>http://creativecommons.org/licenses/by/4.0/</u>.

Global Biosecurity is a peer-reviewed open access journal published by University of New South Wales.