GLOBAL BIOSECURITY

Nair SP, Moa A & MacIntyre CR. Investigation of early epidemiological signals of COVID-19 in India using open source data. *Global Biosecurity, 2020; 1(4).*

RESEARCH ARTICLES

Investigation of early epidemiological signals of COVID-19 in India using open source data

Shruti Premshankar Nair[1], Aye Moa[1] & Raina Macintyre[1]

[1] Biosecurity Program, The Kirby Institute, University of New South Wales, Sydney, Australia

Abstract

Background: The pandemic of novel coronavirus disease (COVID-19) is worsening, with the widespread disease spread in most countries. Due to varied clinical characteristics of the disease and lack of access to testing, the true burden of disease may be unknown. Epidemiological data and early signals of COVID-19 infection are crucial for disease investigation.

Aim: To assess early signals of COVID-19 in India before official reporting of cases in the country and to compare epidemiological characteristics using different surveillance sources.

Methods: We used open-source data from November 2019 to April 2020 from the rapid intelligence surveillance tool Epiwatch to determine trends in " pneumonia of unknown causes" in India. COVID-19 line list was extracted from the crowdsourced database to determine the demographic characteristics of cases. Descriptive analysis was performed to assess the trend of pneumonia of unknown cause in India.

Results: Reporting of pneumonia of unknown cause increased in India from 24th January 2020. Before the first notification on 30th January 2020, four cases of pneumonia of unknown cause were identified in news reports.

Conclusion: The study findings suggest that COVID-19 may have been present in India before the first notified case. Rapid surveillance tools like Epiwatch can be a useful adjunct to traditional, validated surveillance in estimating the trends and burden of infectious diseases.

Keywords: COVID-19, pneumonia of unknown cause, India, Epiwatch, pandemic

Introduction

In December 2019, a cluster of pneumonia cases with an unknown origin was reported in several healthcare facilities in the Chinese city of Wuhan (Hubei province).[1] The clinical presentations of this unknown viral pneumonia resembled the symptoms of Severe Acute Respiratory Syndrome coronavirus which emerged in 2002 in Guangdong province of China.[2] On January 7th, the virus was identified and tentatively named as 2019-nCoV by the World Health Organization (WHO), who later renamed it as SARS-CoV-2.[3,4] The disease caused by the SARS-CoV-2 was termed as the coronavirus disease 2019 (COVID-19).[5] Coronaviruses are enveloped non-segmented positive-sense -RNA viruses that belong to the family Coronaviridae and the order Nidovirales.[6]

On March 11th, COVID-19 was declared a pandemic by the WHO.[7] As the pandemic is rapidly spreading across the globe, real-time analysis of epidemiological data is crucial to increase the situational awareness and to monitor trends. During the early stages of a new infectious disease outbreak, it is important to understand the transmission dynamics of the infections caused by the virus, disease severity, and the natural history of the emerging pathogen.[8,9] As of 26th July 2020, the outbreak of coronavirus disease 2019 (COVID-19) has resulted in 16,055,909 confirmed cases and more than 644,661 deaths across the globe.[10] India has reported 1,385,521 confirmed cases and 32,063 deaths till July 26th,2020.[11] The case fatality rate was 2.31%.

The first case in India was reported on 30th January 2020, in a person who had travelled from Wuhan, China.[12] India ranked 17th among the countries having the highest risk of importation of COVID-19 through air travel from China during the early stages of the pandemic.[13] A travel advisory was issued for those traveling to or from China. On 5th March, the number of cases increased as local transmission appeared to rise. Since most early cases in India had a travel history to China, the Indian government issued on February 5th had issued a travel advisory for all Indian citizens to refrain from traveling to China. On 19th March, international travel to/from India was suspended for all the countries till March 29th,2020. In July 2020 travel restrictions have been eased and a mandatory 14 days quarantine has been advised to all the

international travellers .[11]

To reduce the local transmission in India, a 14-hour lockdown was imposed in the country on 22nd March, 2020.[11] According to the WHO on March 5th, India had local transmission, following which complete lockdown was imposed for 21 days (March 24-April 14) to reduce chances of community transmission.[14] The lockdown extended until 31st May, 2020.[10] As of 26th July, 2020, partial lockdown has been imposed by few states/union territories.[11] A complete lockdown was imposed in containment zones until 31st July, 2020, and lockdown has been eased for areas outside containment zones .[15] As of 26th July, 2020, India has a cluster of cases and there was no community transmission according to WHO Situational Report-188.[16]

India has engaged in rigorous contact tracing, identification of COVID-19 hotspots, increased testing, health promotion regarding hand hygiene, home quarantine, social distancing, use of Aarogyasetu (syndromic surveillance) app, thermal screening at airports and seaports, and closed state and district borders to reduce local and community transmission.[11]

Studies involving real-time surveillance estimates using local or social media have suggested that there could be delayed reporting of cases which could lead to under-reporting or misdiagnosis of cases.[17,18].The present study aims to detect the early epidemiological signals of unknown pneumonia in India before the identification of the first case using open source surveillance data .[19]

Methods

This study used Epiwatch, a semi-automated outbreak data collection tool and analysis observatory, which monitors and provides a critical analysis of global outbreaks and epidemics of emerging infections using open-source data.[19] In the study, data from January 2019 to April 20th, 2020, was extracted from Epiwatch using the search terms: pneumonia, cough plus fever, flu/fever, severe acute respiratory infections (SARI), severe acute respiratory syndrome, severe chest infection, severe lung infection, Wuhan and China to identify early signals of COVID-19. As these keywords are related to respiratory illnesses, the use of such terms could potentially reflect undiagnosed COVID-19 infection. Then, we determined the trend of reporting of pneumonia of unknown cause during the study period i.e. from November 2019 to April 20th, 2020, to detect the early signals of COVID-19 infection. Reports of pneumonia of known cause were excluded.

The search terms were translated into Hindi, Gujarati, and Malayalam using Google translate to extract regional reports from open sources. The selection criteria were: inclusion of reports of pneumonia of unknown cause and confirmed cases of COVID-19, and the exclusion of cases of pneumonia with confirmed diagnosis (excluding COVID-19), non-pneumonia respiratory illnesses, chronic respiratory illness as well as news items not related to the topic of interest. The search terms were used to extract all news reports of confirmed COVID-19 cases, unknown pneumonia, influenza, and confirmed influenza H1N1pdm09 cases from the Epiwatch database to compare the number of cases in January, February, and March (2019 and 2020).

This study also uses the crowdsourced line list of COVID-19 cases to determine demographic characteristics.[20] The age-sex distribution, transmission stage, and nationality of the individuals of confirmed COVID-19 cases were collected and analysed from the line list.

WHO Situational Report and Press briefing reports by official government sources were used in this study to compare and evaluate the number of confirmed, recovered cases and deaths notified by these sources in comparison with the rapid surveillance sources like Epiwatch and line list.[21,15] Epiwatch data, line list, WHO Situational Reports and official Press Briefing by official government sources in India were tracked for "pneumonia of unknown cause" till April 20th,2020 to detect the early signals of COVID-19. Data were analysed using descriptive statistics.

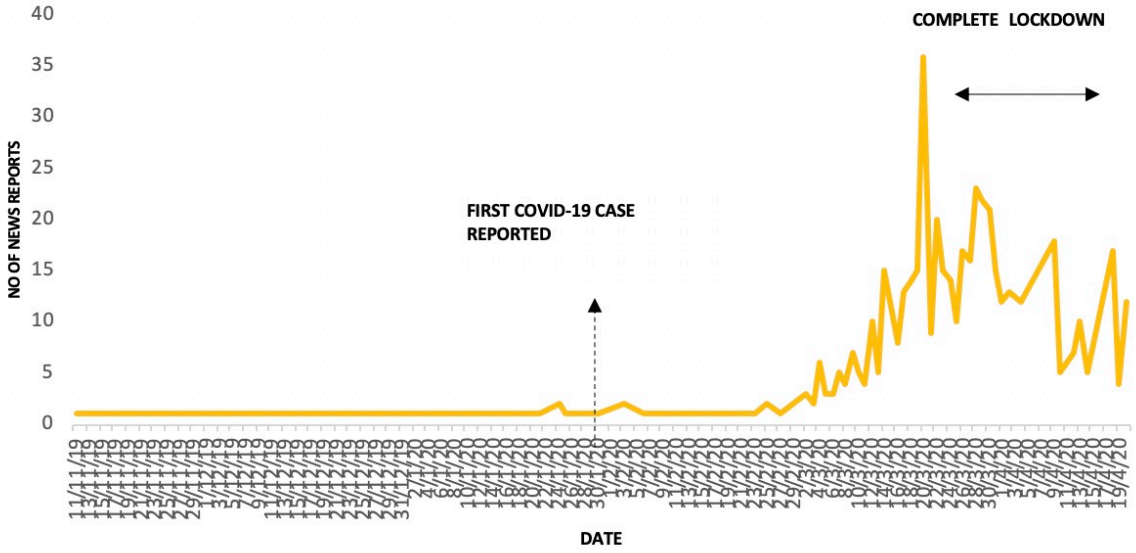
Results

A total of 556 reports were extracted from the Epiwatch database - 536 reports matched the selection criteria of this study; out of which 48 reports were excluded as they were duplicates of news articles. Thus, the study comprised of 488 reports (n=488).

Figure 1 shows the trend of news reports for unknown pneumonia in India from November 2019 to April 20th, 2020. The trend suggested that there were less than 5 reports of unknown pneumonia per day until the second week of March 2020. Pneumonia reports were lowest in November compared to the other months (December 2019 to March 2020). There were 9 reports of pneumonia in January which could have been the potential COVID-19 cases in India. These reports include cases of children who died due to unknown pneumonia. These were news coverages featuring the suspected COVID-19 cases from 26th January to 29th January, many of whom had returned from Wuhan, China.

In January and February 2019, there were less than 50 cases of unknown pneumonia according to Epiwatch data, but the number of unknown pneumonia cases significantly increased in March 2020. There were 42 cases of pneumonia of unknown cause identified in January 2020 and there was 644 pneumonia of unknown cause in March 2020. In 2019, Epiwatch data showed reports of confirmed influenza H1N1pdm09 cases across different parts of India. The news reports suggest that March 2019 recorded the highest number of confirmed Influenza H1N1pdm cases compared to January and February 2019. (Table 1)

Figure 1. Trend of reports of pneumonia in India, November 2019 – March 2020



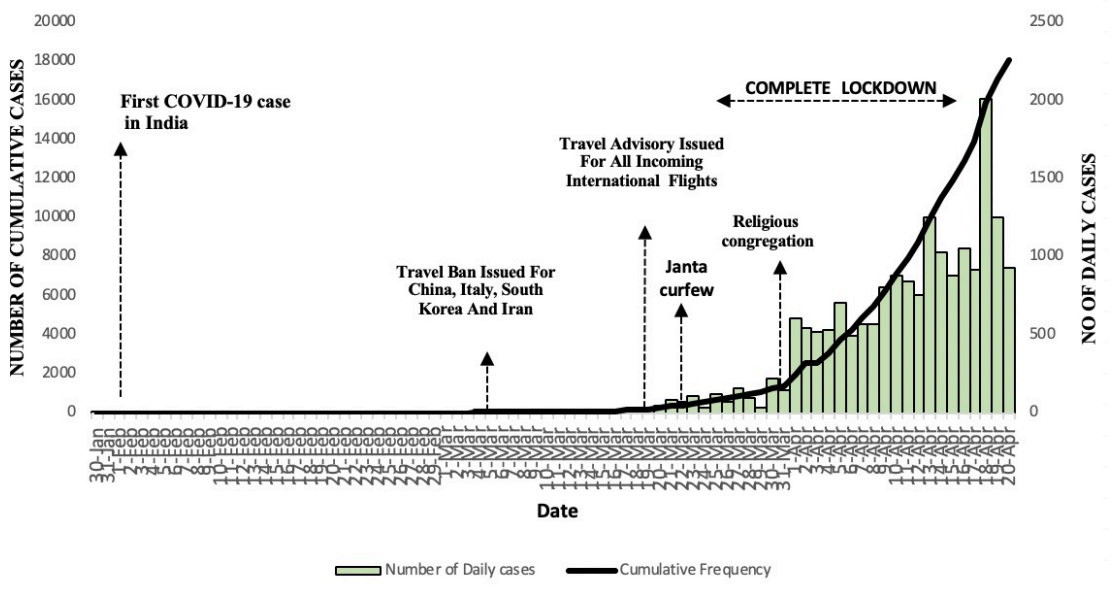
**Source: Epiwatch*

Table 1. Number of cases of pneumonia of unknown cause, influenza and confirmed influenza h1n1pdm based on news reports as of 31st march,2020

	Unknown Pneumonia	Influenza	Influenza H1N1pandemic
January			
2019	2	976	3226
2020	42	7	20
February			
2019	Nil	129	866
2020	10	8	973
March			
2019	2	6	4977
2020	644	5	17

**Source: Epiwatch Data*

Figure 2. Epidemic curve of covid-19 cases in India based Epiwatch data till 20th April,2020



The number of daily cases increased in India after March 20th, with 227 new cases were recorded on March 30th is the highest number of cases in a day during March. (Figure 2).

The epidemic curve of COVID -19 cases in India according to Epiwatch data suggested that first confirmed case was notified by the official government sources in India on January 30th, The number of cases did not rise till March 5th.The number of daily cases increased in India after March 20th. 2018 new cases were recorded on 18th April which was the highest number of single day case reported as of April 20th. On March 5th, a travel ban was issued for countries like China, Italy, South Korea and Iran in order to reduce the imported cases in India as these countries were most affected countries. A 14-hour lockdown was imposed on March 22nd followed by which a complete lockdown was enforced from March 25th to 14th April in order to reduce local transmission of COVID-19 in India. The number of cases in India had significantly increased after 31st March amongst people who had attended religious congregation, which reportedly increased the locally transmitted cases in India.

Line List

According to the line list, there were 18,544 confirmed cases, 593 deaths and 3,373 recovered cases in India as of April 20th (Table 2).

The line list found that males were more susceptible to COVID-19 compared to females among known cases. Majority of the known cases (n=307) were reported among men who were aged 31-40 years compared to other age groups. Women aged 21-30 years were more affected compared to the other age groups. The line list suggested that there were 606 imported cases, around 11,030 cases were notified as locally transmitted cases, and rest of the cases were under investigation. Majority of the COVID-19 cases in India (n=18,410) were among Indian citizens and there were few cases which were reported amongst citizens of Indonesia, Italy, Malaysia, Myanmar, Philippines, Thailand, Tibet, United Kingdom and United States of America.

Maharashtra reported the highest number of cases (n=4,666) followed by Delhi (n=2081) and Gujarat (n=1,939) compared to the other Indian states/Union Territories. Around 4,01,586 people were tested in India for COVID-19 till 20th April.

World Health Organization Situational Reports

According to WHO Situational Reports there were 17,625 confirmed cases ,1553 recovered cases and 543 deaths.

Official Government Sources in India (Press Briefing Report)

17,656 confirmed cases, 2567 recoveries and 543 deaths were reported by the official government sources in the Press briefing (Figure 4)

Comparison of the Epiwatch Data with WHO Situational Report and Official government sources in India (Press Briefing Report)

The number of confirmed cases, deaths, and recovered cases from Epiwatch, line list, WHO situational report and official government sources in India are presented in (Table 2).

The Epiwatch & open source line list reported more confirmed cases compared to the other sources. Epiwatch has also reported more deaths compared to the WHO situational report and Press briefing report by official government sources.

Figure 3. Epidemic curve, India, from WHO situation report data to April 20 2020

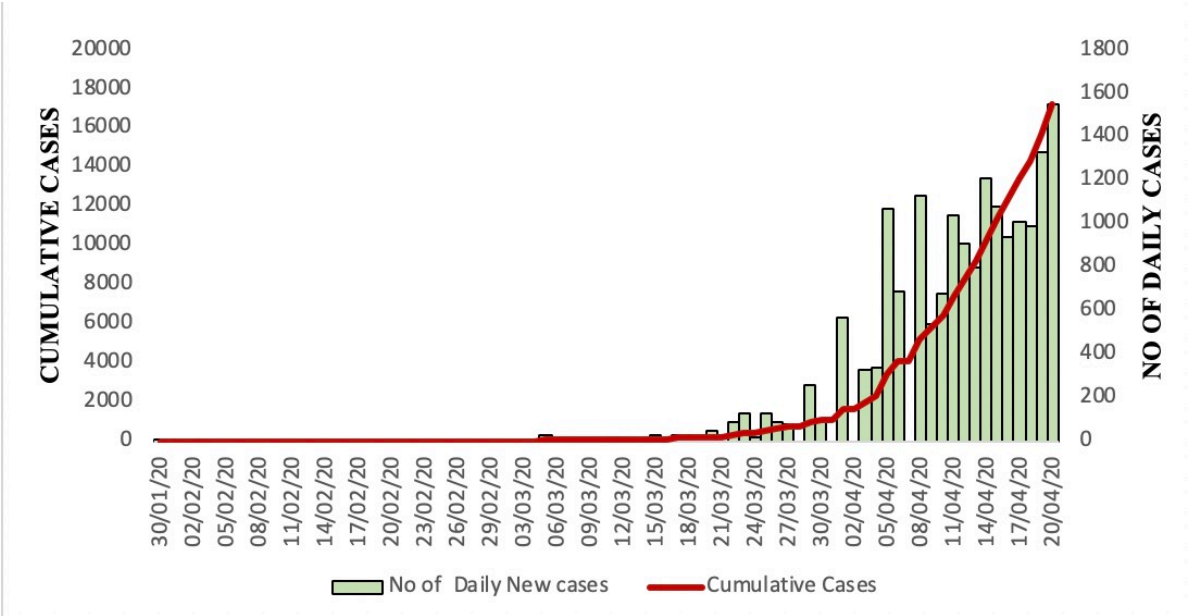


Figure 4. Epidemic curve of COVID-19 cases in India based on the press briefing by official government sources

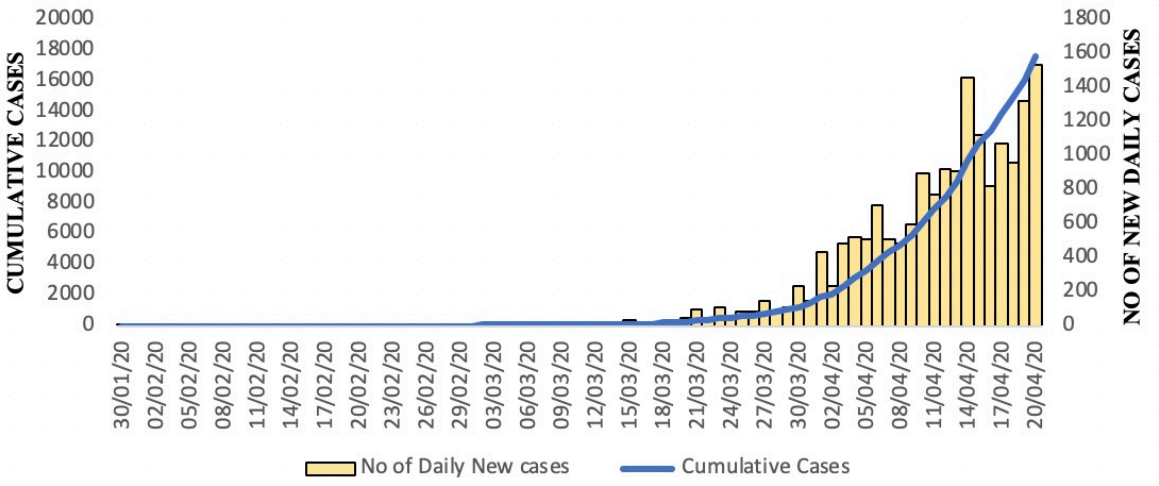


Table 2. Comparison of the cases based on the frequency of confirmed, recovered cases and deaths in India based on the Epiwatch data, line list, WHO Situational Report and Press Briefing report by official government sources in India as of 20th April,2020

Source of Data	Confirmed Cases	Recovered Cases	Deaths
Epiwatch	18,047	924	592
Line list	18,544	3373	593
WHO situational report	17,625	1553	543
Press Briefing	17,656	2567	543

Discussion

The use of open source data found early signals of pneumonia of unknown cause prior to the first confirmed case in India. The first cases of COVID-19 in China may have occurred in November or even earlier.[22] Most countries are now finding that COVID-19 was circulating prior to the first official detections. In Spain, SARS-COV-2 was isolated in wastewater in January 2020, before it was officially identified in Spain, and even in samples from 2019. [23]

Given Wuhan is a major hub of international commerce and trade, it is likely the virus had already spread outside of China by January, including to India, which is consistent with our findings.

The other major cause of pneumonia reports from Epiwatch was H1N1pdm09, which has caused substantial morbidity and mortality in India.[24] The search terms were able to differentiate pneumonia of unknown cause, influenza, influenza H1N1pdm09, and COVID-19.

The epidemic curve of COVID-19 based on the Epiwatch data has suggested that the cumulative confirmed cases in India (18,047) till April 20th. Epiwatch data notified 9961 confirmed COVID-19 cases till 13th April which is much higher than the cases predicted in the modelling study conducted by Pandey et al.[25]

This study has found that most of the reported cases of COVID-19 were younger, in the 21-30 and 31-40 years-old age groups, which is similar to the findings of several studies .[26,27,28] Infection in this age group is more likely to be mild or asymptomatic, which may be a challenge for epidemic control.

The use of social media and local media reports have proven useful in the rapid investigation of COVID-19 cases. Our data has also used three regional languages to track the news reported in the regional areas. A similar study was conducted using a different methodology by Sun et al in China.[26] Their study also used crowdsourced data (line list) from social media/ local media news reports to investigate the cases in different provinces of China. However, contrary to our results, their findings suggested that due to manual editing and reporting of line list the cases reported in the crowdsourced data were significantly lower than the cases reported by the government sources.[26]

The cumulative cases in India are low compared to many developed countries. The public health interventions used by India are - social distancing (1-meter distance), home isolation/self-quarantine, closed public places, imposing a travel ban on all incoming international commercial flights, hand stamping and thermal screening at airports to track the incoming travelers, closed the state and district borders across India, stopped passenger rail services,

identified hotspot areas, and a14-hour lockdown (Janta curfew) followed by complete lockdown (till 3rd May).[11,15] A modelling study assessed the impact of 21 days lockdown (March 25th- April 14th) on COVID-19 transmission in India; the study had predicted 378,036 COVID-19 cases without the 21-days lockdown and predicted 70,424 with lockdown. India had reported 10,363 cases till April 14th which suggests that the 21-day lockdown was effective in reducing the COVID-19 cases in India. The model had predicted that implementation of a strict lockdown for a period of 21 days would reduce the transmission of COVID-19 and suggested that further extension of up to 42 days would be required to significantly reduce the transmission of COVID-19 in India.[29]

Although the public health interventions may have been effective in controlling the community transmission in India, it is still widely believed that there is limited testing in India, which could result in under-reporting or misdiagnosis of the exact number of cases.[30] In order to combat this issue, India has scaled up testing by increasing the number of laboratories and by providing free testing for COVID-19.[12,16] However, the majority of Indians live in rural areas, where healthcare and testing is less accessible, which may make it difficult to ascertain the true burden of disease in India. In addition, outbreaks in urban slums have also been reported and are challenging to control.

There are a few limitations to this study. The first limitation is that this study uses data reported by crowdsourced line lists, social media and local news reports. This is unvalidated data as it was unclear from most of the news reports if these cases have been confirmed after the laboratory testing as these reports broadly mention the symptoms and the confirmed diagnosis of these cases. However, the data are useful for monitoring trends and detecting early signals to estimate the start of COVID-19 circulation in India.

In conclusion, the study findings suggested that COVID-19 may have been circulating in India prior to the first official reported case.

References

1. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol. 2020;92(4):401–2.
2. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses. 2020;12(2).
3. Chen X, Yu B. First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model. Glob Health Res Policy. 2020;5:7.
4. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3.
5. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses. 2020;12(2).
6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020;395(10223):497-506.
7. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 2020.Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
8. Christaki E. New technologies in predicting, preventing and controlling emerging infectious diseases. Virulence. 2015;6(6):558-65.
9. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. The Lancet Infectious Diseases. 2020.
10. Johns Hopkins Coronavirus Resource Center [Internet]. Johns Hopkins Coronavirus Resource Center. 2020 . Available from: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6.
11. Ministry of Health and Family Welfare. COVID-19 INDIA.2020. Available from : https://www.mohfw.gov.in/.
12. Consumer News and Business Channel (CNBC). India confirms its first coronavirus case.2020.Available from: https://www.cnbc.com/2020/01/30/india-confirms-first-case-of-the-coronavirus.html.
13. Mandal S, Bhatnagar T, Arinaminpathy N, Agarwal A, Chowdhury A, Murhekar M, et al. Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach. Indian J Med Res. 2020.
14. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 45. 2020.Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4.
15. Press Information Bureau Government of India. Press briefing on the actions taken, preparedness and updates on COVID-19.2020.Available on: https://pib.gov.in/Webcast/WebcastDetail.aspx?webcast_tempID=382&MinID=31&d=0&m=3&y=2020.

16. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 188. 2020.Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200726-covid-19-sitrep-188.pdf?sfvrsn=f177c3fa_2.
17. Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: A systematic review. Milbank Q. 2014;92(1):7–33.
18. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. The Lancet Infectious Diseases. 2020.
19. Epi-watch [website]. Sydney: University of New South Wales; 2018 Available at: (https://sphcm.med.unsw.edu.au/centres-units/centre-research-excellence-epidemic-response/epi-watch).
20. COVID19INDIA.2020.Available at : https://www.covid19india.org/.
21. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Reports . 2020.Avialable at : https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
22. 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.
23. Chavarria-Miró G, Anfruns-Estrada E, Guix S, Paraira M, Galofré B, Sáanchez G, et al. Sentinel surveillance of SARS-CoV-2 in wastewater anticipates the occurrence of COVID-19 cases. medRxiv. 2020:2020.06.13.20129627.
24. Adam DC, Scotch M, MacIntyre CR. Phylodynamics of Influenza A/H1N1pdm09 in India Reveals Circulation Patterns and Increased Selection for Clade 6b Residues and Other High Mortality Mutants. Viruses. 2019;11(9).
25. Pandey G, Chaudhary P, Gupta R, Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. medRxiv. 2020;2020.04.01.20049825.
26. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. The Lancet Digital Health. 2020;2(4):e201-e8.
27. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. J Infect. 2020.
28. Du RH, Liang LR, Yang CQ, Wang W, Cao TZ, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. Eur Respir J. 2020;55(5).
29. Ambikapathy B, Krishnamurthy K. Mathematical Modelling to Assess the Impact of Lockdown on COVID-19 Transmission in India: Model Development and Validation. JMIR Public Health Surveill. 2020;6(2):e19368.
30. Al Jazeera. India's poor testing rate may have masked coronavirus cases [Internet]. Aljazeera.com. 2020 .Available from: https://www.aljazeera.com/news/2020/03/india-poor-testing-rate-masked-coronavirus-cases-200318040314568.html.

How to cite this article: Nair SP, Moa A & MacIntyre CR. Investigation of early epidemiological signals of COVID-19 in India using open source data. *Global Biosecurity, 2020; 1(4).*

Published: August 2020

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