RESEARCH ARTICLES

The Epidemiology of Unknown Disease Outbreak Reports Globally

Robin Jacob Mao[1], Aye Moa[1] & Abrar Chughtai[2]

[1] Biosecurity Program, Kirby Insitute, University of New South Wales, Sydney, Australia
[2] School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia

Abstract

Disease outbreaks can adversely impact the health and economic status of affected populations. The notification of disease outbreaks to public health authorities can take months, delaying efficient response to those outbreaks. However, disease outbreak data gathered from informal sources such as media reports may be an efficient and resource-light source of real-time data. This provides a faster option to conduct epidemiologic analyses on disease outbreaks globally. At the time this report was prepared, there were no epidemiological analyses of unknown disease outbreaks globally – outbreaks whose causes are not immediately known. EpiWATCH data was used to analyse the epidemiology of unknown disease outbreaks globally from 2016 to 2019. Descriptive analysis of EpiWATCH data was conducted. One hundred and nine reports of unknown outbreaks were found. Ninety-three reports were on human outbreaks and 17 were on animal outbreaks (one report also included human cases). Unknown disease outbreaks resulted in 6714 human cases. India and USA were responsible for 37% of outbreak reports. The year 2017 saw the greatest number of reports on unknown diseases published. There was a peak in reports in the month of August across all four years. Fever and vomiting were the most common symptoms reported. Where there was a cause identified, measles, Nipah virus, norovirus and influenza were the most common causes of unknown disease outbreaks. However, 76% of unknown disease outbreaks remain undiagnosed.

Keywords: unknown diseases, disease outbreaks, global outbreaks, descriptive analysis, EpiWATCH data

Background

The WHO [1] states that a disease outbreak is an excess occurrence of disease in a defined area or season. An outbreak may begin with a single case of disease that is uncommon or caused by an agent that is not common in the defined region. It may also include a single case of disease that was previously unknown. Outbreaks can adversely impact the health and economic status of the affected populations whether they are directly or indirectly affected (for example, through the decimation of livestock) [2, 3].

Even though there are official routes by which disease outbreaks are notified to the respective public health authorities, the collection of such data may not be available for weeks or months [4]. This results in delayed epidemiologic analysis and, therefore, delayed public health responses to the outbreak. However, from as early as 2001, more than 50% of disease outbreak data were collected through informal sources such as media reports [5]. This proved to be a resource-light source of real-time data that was deemed detailed enough for scientific analyses [4, 6].

Sometimes, the cause of disease outbreaks is not immediately known. For example, in the recent outbreaks of acute flaccid myelitis [7], the cause had not been confirmed. New, emerging infectious diseases also present as unknown outbreaks initially, for example, the first cases of AIDS, SARS, MERS Coronavirus and COVID-19. Similarly, recurrent epidemics of encephalitis in India have an unknown aetiology [8].

Lack of surveillance or delayed reporting can lead to inadequate public health responses, as was the case with COVID-19 in Wuhan [9]. At this moment, there are no epidemiological surveillances of unknown disease outbreaks globally. However, there are many informal sources of such data available online. The aim of this study is to analyse the epidemiology of unknown disease outbreaks and case reports globally.

Methods

EpiWATCH is an open source epidemic observatory, which contains outbreak data reported in news and social media from 2016 onward [10]. The EpiWATCH Outbreak Alerts database was used to collect data on reports dated between 2016-2019 (15-08-2019 to 30-09-2019). Disease keywords related to unknown disease outbreak were used to retrieve specific data for the analysis (See Table 1). Reports matching the keywords were analysed and were excluded if they were deemed inappropriate for analysis. A final list of eligible reports was compiled, and a descriptive analysis of those reports was conducted. All analyses were conducted using Microsoft Excel. The reports were scanned through manually and their relevant information was noted down. The country from which the report originated was first identified followed by the year and then the month in which it was published. Based on this information, an Excel graphical representation of a world map was created to showcase the countries that were responsible for the greatest number of outbreaks in the reporting

period. Two other graphs were also created to showcase the number of outbreaks in each of the four years covered in this analysis as well as which month, throughout the reporting period, was responsible for how many outbreaks. The purpose of the latter graph is to identify potential seasonal variations in the number of unknown disease outbreaks every year. The reports were then further scanned to identify the number of cases arising in each outbreak over the reporting period. The number of cases were similarly categorised by year and month. This is to provide further graphical analysis of the seasonal variations in unknown disease outbreaks so that they may be used to identify trends in such outbreaks globally. Additionally, the distinction was made whether the cases were amongst children or adults. This information is sometimes ambiguously provided in media reports and was gathered to the best of the author's ability to identify child cases to adult cases as per the case details in each report analysed. Next, the reported symptoms in children and adults were noted. The former information was used to plot a graph to analyse the most common symptoms associated with unknown disease outbreaks in general, whereas the latter information was used to graphically plot the percentage of symptoms reported by children and adults. The latter graph helped identify the most common symptoms presented by children and adults separately in unknown disease outbreaks. Lastly, the reports were scanned to identify the frequency of

pertinent syndromes in unknown disease outbreaks. The syndromes were (1) fever of unknown origin (2) fever and rash (3) gastroenteritis (4) neurological syndromes (encephalitis, meningitis, neuropathy and myelitis) (5) respiratory syndrome and (6) others. All statistical information was tabulated in a Microsoft Excel sheet and the chart function of Excel was used to graphically represent the data collected.

Results

A total of 109 reports of unknown outbreaks were found between 2016-2019. Out of them, 17 reports were of disease outbreaks in non-human population (including one report that also included human cases). In those 93 reports that contained reports of human outbreaks, there were a total of 6714 cases (including outpatient cases, hospitalized cases and mortalities) described. Geographically, the highest number of reports emerged from India and USA (21 and 19 reports, respectively) (See Figure 1). The Democratic Republic of São Tomé and Príncipe only contributed one report between 2016-2019, however, it contributed to 2000 of the 6714 cases reported. South Africa and Liberia accounted for 5 reports each in the years analysed. Egypt and Nigeria accounted for the next highest number of reports, accounting for 4 reports each. A total of 41 countries provided reports on unknown disease outbreaks.

Table 1. Table of key words used for data retrieval from EpiWATCH

Illness outbreak	Undiagnosed mortality cluster
Lung disease related to vaping	Undiagnosed rash
Lung illness	Undiagnosed respiratory illness
Mass food poisoning	Undiagnosed urticaria reaction
Mass die-off	Undiagnosed viral disease
Mass Mortalities	Unexplained cluster of death
Mysterious disease	Unexplained deaths
Mysterious illness	Unexplained disease
Mysterious fever	Unexplained fatal illness
Mysterious lung disease	Unidentified disease
Mystery disease	Unidentified fatal illness
Mystery illness	Unidentified illness
Mystery illness causing fatalities	Unidentified respiratory disease
Mystery illness with fever causing fatalities	Unidentified severe illness
n.a.	Unknown disease
Not specified	Unknown bacteria
Not yet classified	Unknown bacteria infection
Unknown	Unknown fatalities
Unknown source	Unknown fever
Unknown virus	Unknown food-borne illness
Unclassified virus	Unknown illness
Undiagnosed illness	Unknown infection
Undiagnosed acute respiratory failure	Unknown mass mortalities
Undiagnosed animal death	Unknown respiratory illness
Undiagnosed bleeding disorder	Unknown viral infection
Undiagnosed death	Unknown virus
Undiagnosed febrile illness	Vaping illness
Undiagnosed haemorrhagic fever	Vaping-related lung disease
Undiagnosed mass mortalities	

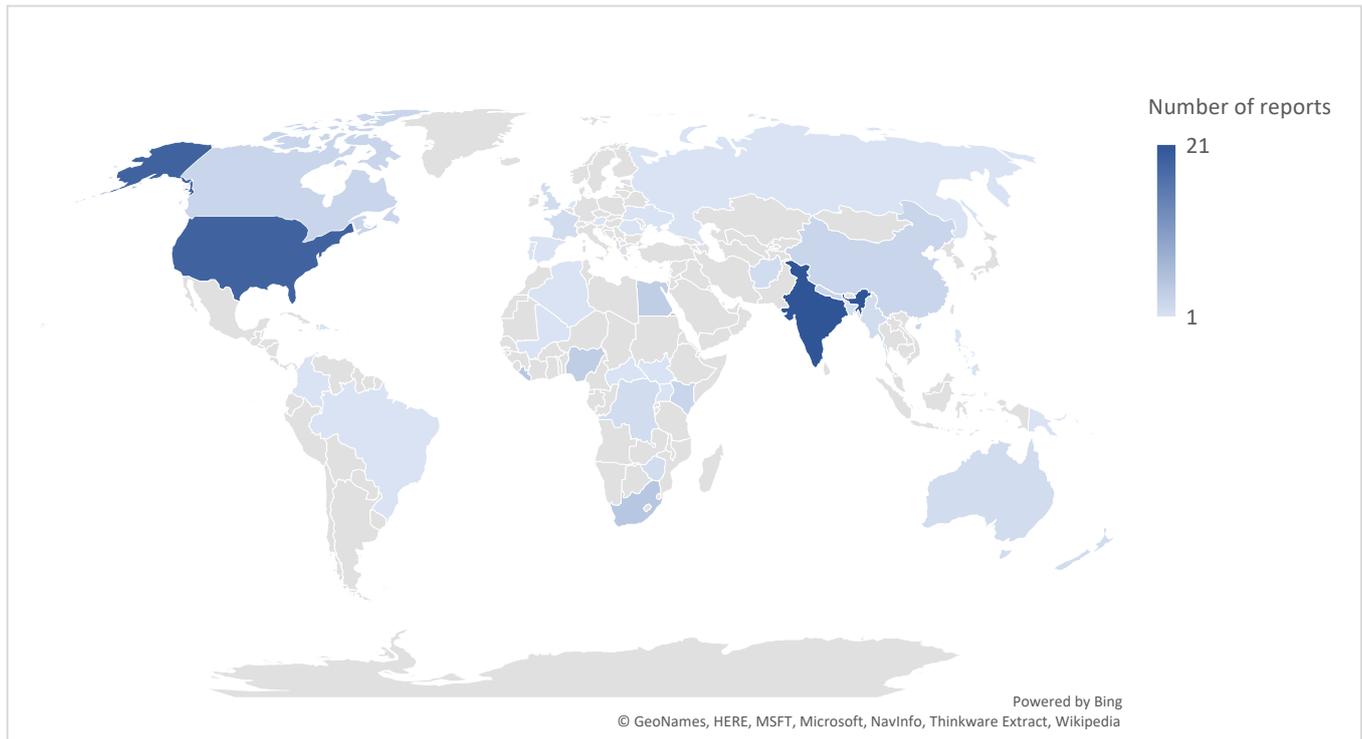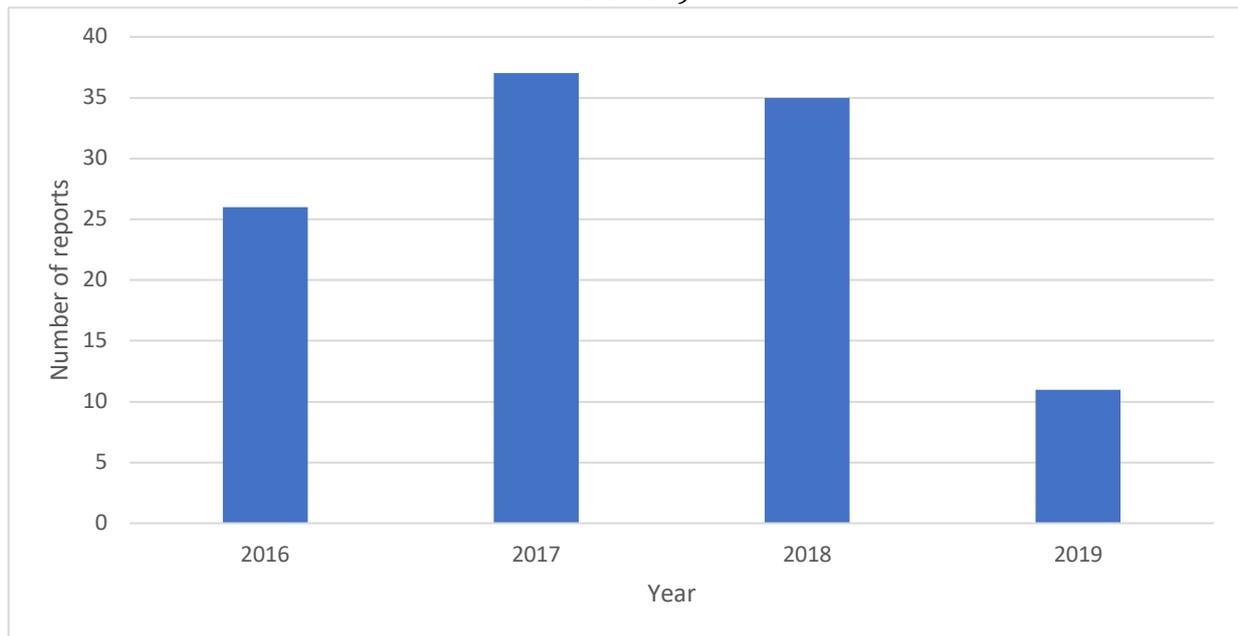**Figure 1.** Geographical distribution of unknown disease outbreak reports, 2016-2019

Figure 2. Graphical representation of number of unknown disease outbreak reports published every year between 2016-2019

The number of reports published globally in each of the four years analysed was estimated (See Figure 2) The greatest number of reports (37) were published in 2017. In 2018, 35 reports were published. The years 2016 and 2019 were responsible for 26 and 11 reports, respectively. Next, the monthly variation of reports published in the four years was calculated (See Figure 3). Over the four years, the greatest number of reports were published in the month of August (15 reports) and April (13 reports). May, September and October gave rise to 11 reports each, followed by March with 9 reports, June and November with 8 reports each, January, February and July with 6 reports each, and December with 5 reports.

Figure 3. Graphical representation of number of unknown disease outbreak reports published every month
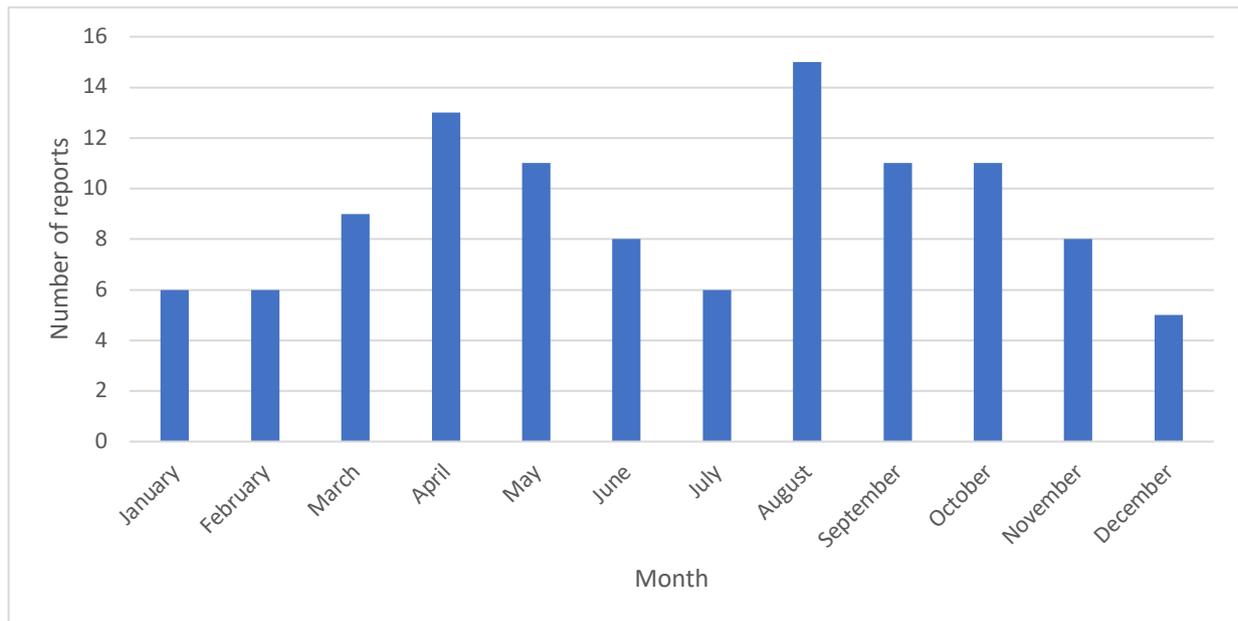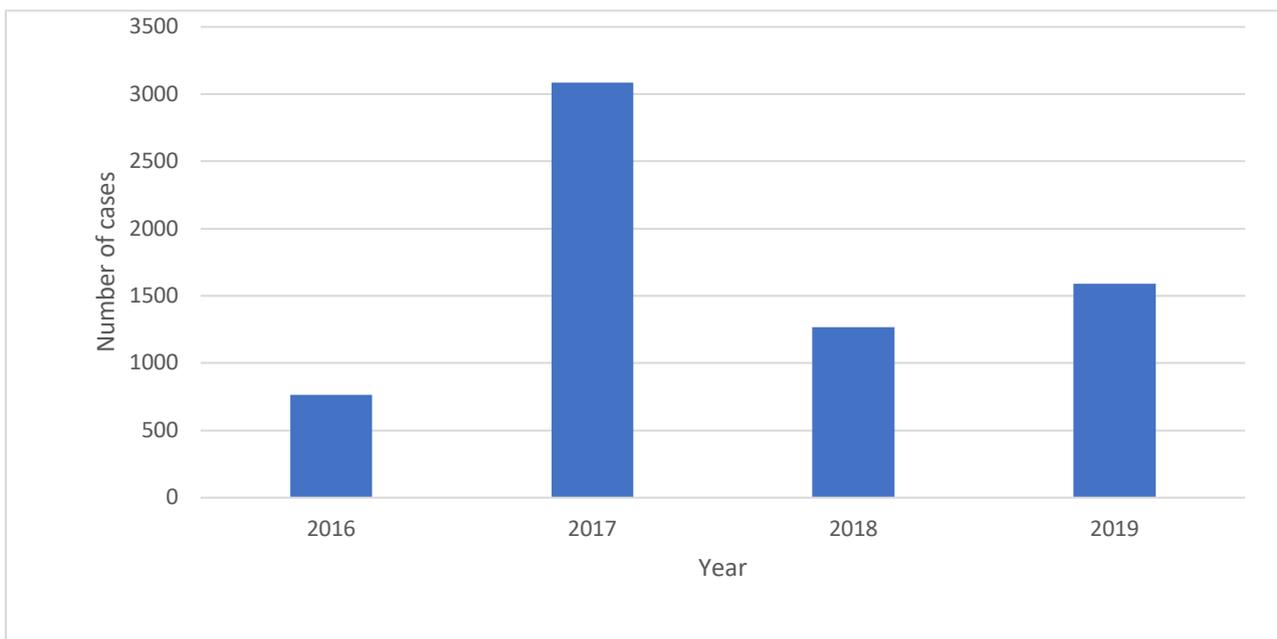between 2016-2019

Figure 4. Graphical representation of number of cases arising every year between 2016-2019

The number of unknown disease outbreak cases arising globally was then estimated annually (See Figure 4). The year 2017 was responsible for the greatest number of cases (3088 cases) amongst the other three years but 2000 cases can be attributed to one outbreak that occurred in the Democratic Republic of São Tomé and Príncipe in 2017. The year 2019 was responsible for 1591 cases, out of which, 805 cases can be attributed to vaping-related lung diseases that occurred in USA. The years 2018 and 2016 gave rise to 1269 and 766 cases, respectively. The monthly variation of case representation over the last four years was estimated (See Figure 5). February (2031 cases) and August (1251 cases) saw the greatest number of cases arise over the last four years. However, 2000 cases in February can be attributed to the Democratic Republic of São Tomé and Príncipe outbreak. One thousand and sixty-eight cases arose in the month of September over the last four years. Once again, however, 805 cases can be attributed to vaping-related lung diseases. December was responsible for 510 cases over four years. October was responsible for the least number of cases over four years (194 cases) followed closely by March (203 cases).

Next, the symptom profile of these outbreaks was estimated. The symptoms that the cases presented with were tabulated against the frequency with which they appeared in reports (See Table 2). Fever and vomiting were the most frequent symptoms that afflicted cases between 2016-2019.

Fever, alone, was a symptom in 32 out of 93 reports. Vomiting featured in 26 reports. Diarrhoea was the third most common symptom, featuring 20 times. Twenty-five other symptoms were identified that occurred in lower frequencies.

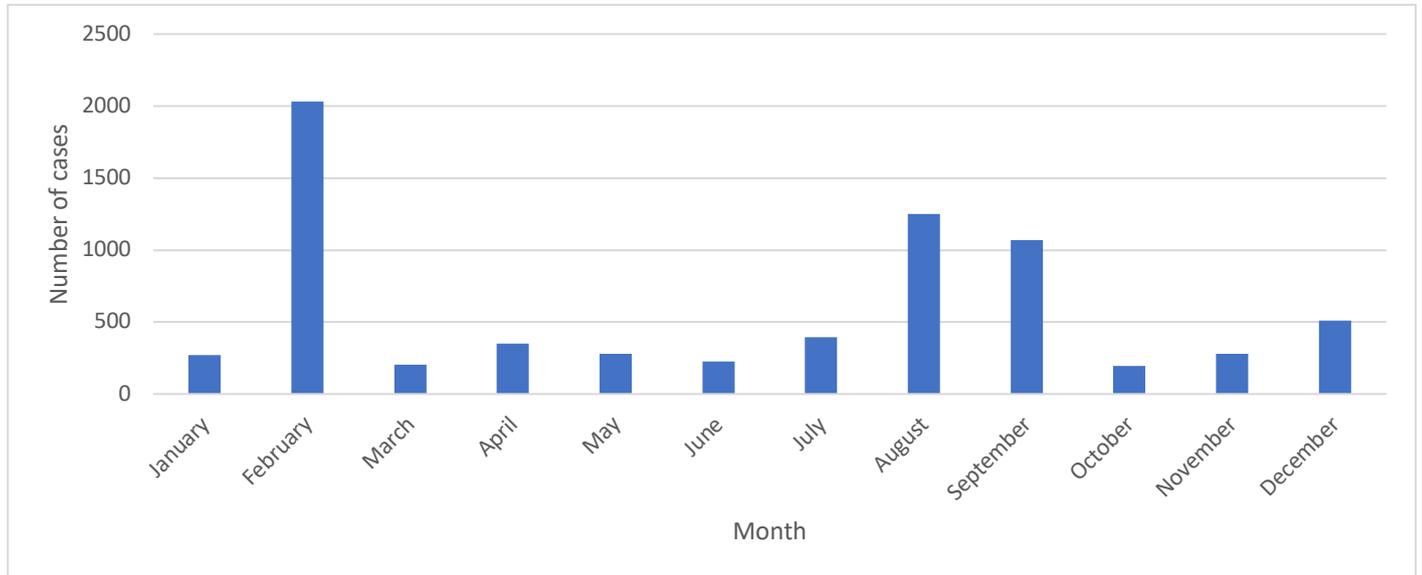**Figure 5.** Graphical representation of number of cases arising every month between 2016-2019

Table 2. Tabular representation of symptom frequency in outbreak reports between 2016-2019

Symptom	Frequency (in numbers)
Fever	32
Vomiting	26
Diarrhoea	20
Headache	7
Coughing	6
Abdominal pain	6
Fainting	5
Rash	5
Breathing difficulties	4
Leg swelling	4
Fatigue	3
Seizures	2
Chest pain	2
Nosebleed	2
Paralysis	2
Body ache	2
Throat infection	1
Muscle pain	1
Black urine	1
Blood in urine	1
Bone pain	1
Swellings	1
Swollen head	1

Table 3. Tabular representation of syndrome frequency in outbreak reports between 2016-2019

Syndrome	Frequency (in numbers)
Fever of unknown origin	20
Fever and rash	0
Gastroenteritis	30
Neurological syndromes	1
Respiratory syndrome	6
Other	28

Figure 6. Graphical representation of the percentage of individual symptoms reported by children and adult
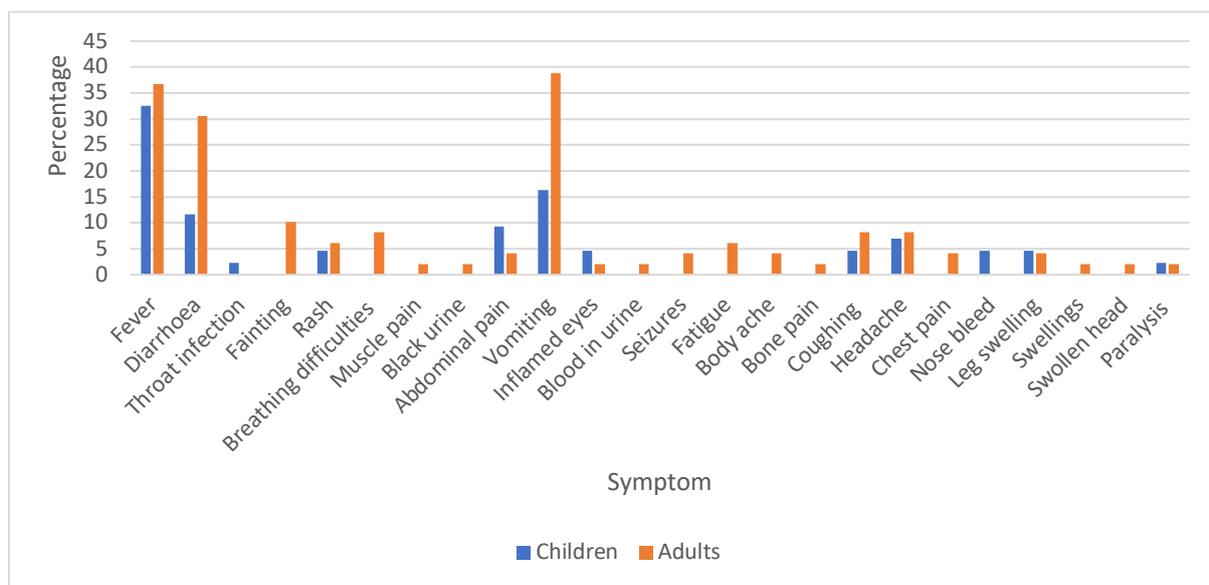
In addition to individual symptoms, the syndrome profile of the outbreaks was also estimated. The syndromes chosen for the analysis are fever of unknown origin; fever and rash; gastroenteritis; neurological syndromes (encephalitis, meningitis, neuropathy, and myelitis); respiratory syndrome; and Other (See Table 3). Gastroenteritis was the most common syndrome, appearing in 30 of the reports collected. The 'Other' category, which includes any collection of symptoms that do not make up any of the syndromes considered, included the next highest number of outbreak reports (28). 'Fever of unknown origin' was a syndrome associated in 20 of the reports analysed. Respiratory syndromes were reported in 6 outbreaks in the reporting period. One report concerning encephalitis (neurological syndrome) was identified. No outbreak reported fever and rash occurring simultaneously in the period analysed.

Cases in children and adults were analysed. Outbreaks amongst children comprised 35 out of 93 reports. These 35 reports accounted for 1431 out of 6714 (27%) cases. A graph was constructed to showcase the percentage of individual symptoms exhibited by children and adults, individually, in the reports analysed (See Figure 6). The symptoms that children presented with were estimated

to check if there were any obvious differences between them and the adults. Children had fever as their most common symptom, followed by nausea/vomiting (all reports concerning children reported vomiting alongside nausea) and diarrhoea. This is not dissimilar to the group statistics. However, the graphical representation of adult symptoms shows that vomiting occurs more frequently than fever. Fever and diarrhoea are the most common symptoms in the adults after vomiting.

Seventeen out of 109 reports dealt with non-human species. Unknown disease outbreaks occurred most commonly in birds and cattle (4 outbreaks each). Additionally, there were two outbreaks in goats and dogs. A cause was eventually found in unknown outbreaks in 20/109 (18%) cases. A cause could only be ascertained in 1/17 animal outbreaks – an outbreak in birds in the Democratic Republic of the Congo in 2017, which was found to be caused by avian influenza. Among reports of human cases, the causes of 19/93 (20%) unknown outbreaks was ascertained. Measles, Nipah virus, norovirus and influenza were each twice responsible for an unknown outbreak. However, 83/109 (76%) of outbreaks remain undiagnosed in the reporting period 2016-2019.

Discussion

This report provides a descriptive analysis of unknown disease outbreaks that occurred between 2016-2019. In 20% of these, a cause was later identified, and causes included vaping disease in the US, measles, Nipah virus and influenza. However, in the majority of the outbreaks, no cause was identified. The study did not cover the period of the likely emergence of COVID-19, but unknown outbreaks could signal newly emerged infections. Most of the outbreaks were reported in India and USA. The greatest number of reports were published in the years 2017 and 2018. This can be since collection of data only took place after August in 2016 and stopped in September in 2019. The year 2017 had the greatest number of cases, although this can be attributed to a major outbreak in the Democratic Republic of São Tomé and Príncipe in that year. Similarly, the greatest number of cases seem to arise in the month of February every year, but the previously mentioned outbreak happened to occur in that month, thereby tilting the overall balance in February's favour. The greatest number of reports published every year seem to be in the months of April and August, suggesting a seasonal cause that may be recurring. Twenty-seven percent of all cases are children and their most common symptom is fever. The predominance of children presenting with fever agrees with the fact that the most common ailment that makes a child visit the hospital is fever [11]. In adults, the most common symptom is vomiting. Following close behind, however, are the symptoms fever and diarrhoea. The ambiguity of the symptoms and the regularity with which they are found when a person is in a diseased state makes it difficult to use the above findings to pinpoint a particular infection or disease. Additionally, as this analysis is conducted over a 4-year span, large outbreaks make it difficult to interpret the temporal findings of this report. Finally, outbreaks reported by the media rarely have any follow-up making it difficult to ascertain a cause to them.

This report is the first of its kind to describe the epidemiology of unknown outbreaks globally. It has shown that there is potential to analyse and plot geographic, temporal and personal data of these outbreaks if a suitable timeframe is used. It has also highlighted the ability of media reports to rapidly identify unknown outbreaks globally and flag potentially serious events. Through continued analysis of such outbreaks, stronger commonalities can be identified within them to better inform and gauge public health responses.

References

1. Disease outbreaks: World Health Organization; 2019 [Available from: https://www.who.int/environmental_health_emergencies/disease_outbreaks/en/.
2. Constenla D, Carvalho A, Alvis Guzmán N. Economic Impact of Meningococcal Outbreaks in Brazil and Colombia. Open Forum Infectious Diseases. 2015;2(4).
3. Yoon H, Jeong W, Han J-H, Choi J, Kang Y-M, Kim Y-S, et al. Financial Impact of Foot-and-mouth disease outbreaks on pig farms in the Republic of Korea, 2014/2015. Preventive Veterinary Medicine. 2018;149:140-2.
4. Chunara R, Andrews JR, Brownstein JS. Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. The American Journal of Tropical Medicine and Hygiene. 2012;86(1):39-45.
5. Heymann DL, Rodier GR. Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. The Lancet Infectious Diseases. 2001;1(5):345-53.
6. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, et al. Use of unstructured event-based reports for global infectious disease surveillance. Emerg Infect Dis. 2009;15(5):689-95.
7. Hopkins SE, Elrick MJ, Messacar K. Acute Flaccid Myelitis—Keys to Diagnosis, Questions About Treatment, and Future Directions. JAMA Pediatrics. 2019;173(2):117-8.
8. Narain J, Dhariwal A, MacIntyre C. Acute encephalitis in India: An unfolding tragedy. Indian Journal of Medical Research. 2017;145(5):584-7.
9. Niehus R, De Salazar PM, Taylor A, Lipsitch M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. medRxiv. 2020:2020.02.13.20022707.
10. Hii A, Chughtai AA, Housen T, Saketa S, Kunasekaran MP, Sulaiman F, et al. Epidemic intelligence needs of stakeholders in the Asia-Pacific region. Western Pac Surveill Response J. 2018;9(4):28-36.
11. Barbi E, Marzuillo P, Neri E, Naviglio S, Krauss BS. Fever in Children: Pearls and Pitfalls. Children (Basel). 2017;4(9):81