

Developing an Online Infectious Disease Outbreak Database: Codifying the World Health Organization's Disease Outbreak News reports

Rebecca Katz (✉ rebecca.katz@georgetown.edu)

Colin J Carlson

Georgetown University Medical Center

Munir C Pavez

Georgetown University Medical Center

Jessica Lin

Georgetown University Medical Center

Max A Palys

Georgetown University Medical Center

Matthew R Boyce

Georgetown University Medical Center

Research article

Keywords: Database, Global Health, Infectious Disease, Reporting, Surveillance System, World Health Organization

Posted Date: June 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-33473/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The World Health Organization's Disease Outbreak News (DONs) reports are the world's primary source of official information on global disease outbreaks. Access to this information is crucial for informing research analyses, global health priorities, and decision making. However, in its current form, the utility of the DONs reports for research and analysis is limited as a result of their reporting format. To this end, we designed a standardized methodology for codifying the data contained in DONs reports and created an online, searchable database.

Methods: We coded DONs reports published between the years 1996 and 2019, systematically collecting data from each individual report using a standardized methodology and tabulating data into a single spreadsheet. We created a Year-Pathogen-Country taxonomy to group related disease events and circumvent issues related to reporting inconsistencies in DONs reports.

Results: In total, we reviewed 2,806 DONs reports corresponding to 1,105 unique infectious disease outbreaks from 1996-2019. Overall, H5N1 represented the most frequently reported disease, while China was the country with the most reports. We observed the DONs reports to contain numerous issues relating to the standardization, accuracy, and transparency of reporting procedures.

Conclusions: Our database represents a new, accessible resource for research that improves the accessibility of the data contained in DONs reports. The World Health Organization should consider standardizing reporting practices, protocols, and procedures as a means of improving the reporting and transparency of infectious disease outbreaks.

Trial Registration: N/A

Introduction

Infectious disease outbreaks have significantly increased over the past few decades, possibly as a result of alterations in various climatic, environmental, biological, socioeconomic, and political factors [1–4]. Globalization has complicated this trend by accelerating the pace at which infectious diseases can spread around the world, as exemplified by the COVID-19 pandemic. These realities underscore the need for effective surveillance systems and rapid outbreak detection and reporting as a means of ensuring global health security.

Disease surveillance and reporting are mandated by the revised International Health Regulations (IHR) [5], and there is an abundance of publicly available data on disease events from Internet-based sources. These sources provide an open route for reporting that promotes transparency and include the International Society for Infectious Diseases' Program for Monitoring Emerging Diseases (ProMED-mail) [6], the Public Health Agency of Canada's Global Public Health Intelligence Network (GPHIN) [7], and HealthMap [8]. These sources allow for researchers to collect and translate these data sources into useable surveillance platforms [9]. The GIDEON database is another resource that provides access to

global infectious disease surveillance data [10], but is not freely available to the public. All of these sources are used for research and action but are unofficial records.

The World Health Organization's Disease Outbreak News (DONs) report catalog represents the only official, public, Internet-based source of surveillance data and is the world's primary source for World Health Organization (WHO)-confirmed infectious disease outbreaks [11]. This surveillance resource contains data dating back to 1996 on event-based outbreak information provided by countries, and other partners, which is then organized by the WHO into prose-based reports organized by date, country, and disease. The DONs reports serve as a tool for international actors to stay informed on health-related events and emergencies around the globe and include crucial information such as relevant dates, the status of laboratory confirmation and response activities, and other relevant contextual details.

Because the data contained in DONs reports represent the only official recordings of outbreaks from the WHO for the last 25 years, many researchers utilize them for their analyses [12–17]. However, in its current form, the analytic utility of the DONs database is severely hampered by its unstructured, prose-based system, which makes it infeasible to quickly access information or conduct analyses. This precludes it from being a useful research tool and makes it difficult to identify trends that could inform future outbreak policy, prevention, and response.

We endeavored to create an accessible database of DONs reports that systematically captures information in a non-prose format for every report from 1996–2019. In this paper, we discuss this process, address factors that limit the DONs from being a usable reference for international disease-related events, present recommendations for improving disease reporting, and offer the database of DONs reports to be used for future data analyses.

Methods

Previous work has proposed measures including the start, detection, notification, verification, and laboratory confirmation of an outbreak that could serve as metrics for monitoring infectious disease outbreaks [18]. Based on these measures, we created a standardized methodology for identifying, collecting, and reviewing the information contained in the DONs reports (Table 1). We then reviewed all of the reports contained in the WHO's catalog, strictly adhering to information reported in the text.

Table 1
Description of data extracted from DONs reports.

Data	Description
Year(s) of Outbreak	Year of publication
Date Recorded	Date of report publication
Disease/Pathogen Name	Disease/Pathogen specified in report
Country	Country reported in report
Event	Event reported, if subject of report is not properly captured by a pathogen/outbreak
Outbreak Epicenter	City, neighborhood, district, or region affected by outbreak
Outbreak Start	Reported date of index case presentation
Outbreak End	Reported date of the official declaration of the end of an outbreak by the World Health Organization
Outbreak Detection	Reported date outbreak was recorded
Outbreak Notification	Reported date local or national authorities were notified
Outbreak Notification to the World Health Organization	Reported date the World Health Organization was notified
Outbreak Verification	Reported date outbreak was verified through laboratory testing
Public Health Intervention	Public health interventions implemented to control the outbreak*
Public Communication	Release of information to the public by relevant national health authorities †
Antimicrobial Resistance	Reported antimicrobial resistance
Conflict	Reported conflict ‡
Community Resistance	Reported community actions hampering response
Climatic Factors	Reported climatic factors §
Confirmed Cases	Reported laboratory-confirmed cases
Probable Cases	Reported probable cases
Suspected Cases	Reported suspected cases
Total Cases	Reported total cases
Death Count	Reported fatalities
Case-Fatality Ratio	Reported case fatality ratio

Data	Description
Concurrent Outbreaks	Reported concurrent outbreaks
Displaced Persons or Migrants	Reported presence of displaced populations [¶]
Mass Gatherings	Reported mass gathering event
* Including: contact tracing, health monitoring, isolation, quarantine, health education and promotion, multimedia community sensitization, deploying experts to the field, national, multinational or cross-border meetings, vaccination, food, water and/or drug supplementation, or any other preventative or therapeutic interventions aimed at controlling the disease outbreak;	
† Including: press conferences, news briefings or other forms of written or oral media;	
‡ Including: armed conflict, skirmishes, clashes, war, civil war, civil unrest, militia, raid, unrest, civil unrest, hostilities, combat, confrontation, bloodshed, use of force, violence, instability, contested areas, insecurity, banditry, security compromised, security situation;	
§ Including: droughts, earthquakes, famines, fires, floods, heat, hurricanes, landslides, rain, snow, tornados, tsunamis, and effects from El Niño or La Niña;	
¶ Including: refugees, internally displaced persons, population displacements, cross-border migration.	

In addition to the data collected from the DONs itself, country reports were also matched with their corresponding 3-letter ISO country codes from the ISO 3166 Online Browsing Platform [19].

Following this compilation, we reviewed the rows of data to create a standardized ontology to group DONs reports of the same outbreak, address inconsistencies in reporting, and classify outbreaks according to pathogen etiology. Our labelling conventions followed the “Year-Pathogen-Country” format. In the event that one report included more than one outbreak or pathogen, we listed each outbreak separately but indicated that the information was obtained from the same report. In the event that reports for certain types of outbreaks spanned two or more calendar years, outbreaks were labelled using the start year of the multi-year event. In certain cases, if events spanned multiple countries, all countries were listed in the labelling; however, if the DONs report labeled an outbreak as a global outbreak, or more than 10 countries were listed in the DONs reports, we labelled the event as “Global.” Additional regional locations included in the DONs and captured in our ontology were West Africa, African Meningitis Belt, Asia, Northern Hemisphere, and Central America. We labelled travel cases based on the reported location, unless the genome sequencing of the pathogen, as described in the report, showed a clear origin of the outbreak. Table 2 details how we categorized diseases according to their pathogen etiology. Additional details on our ontology are available in an appendix (Additional File 1).

Table 2

Ontology classification of DONs report diseases and conditions according to pathogen etiology.

Pathogen Etiology	Included Diseases
Directly-Transmitted Human	Acute respiratory syndrome, Circulating vaccine-derived poliovirus, Diphtheria, Enterovirus, Gonococcal infection, Hand foot and mouth disease, Human immunodeficiency virus (HIV), Human coronavirus OC43, Measles, Meningitis, Meningococcal disease, Pertussis, Pneumonia, Smallpox, <i>Staphylococcus aureus</i> , Tuberculosis (XDR-TB)
Environmental/Foodborne	Acute watery diarrheal syndrome, Botulism, Cholera, Coccidioidomycosis, Creutzfeldt-Jakob disease, Diarrheal disease, Dysentery, Enterohemorrhagic E. coli, Food poisoning, Hemolytic uremic syndrome, Hepatitis A, Hepatitis E, Legionellosis, <i>Listeria</i> , <i>Pseudomonas aeruginosa</i> , <i>Salmonella</i> , <i>Shigella</i> , Transmissible spongiform encephalopathies, Typhoid fever, Wild type poliovirus
Influenza	Seasonal influenzas (H3, H7, H9), Pandemic influenzas (H1, H5)
Vectorborne	Chikungunya, Crimean-Congo hemorrhagic fever, Dengue fever, Elizabethkingia anopheles, Japanese encephalitis, Leishmaniasis, Malaria, O’Nyong-Nyong fever, Oropouche fever, Relapsing fever, Rift Valley fever, Saint Louis encephalitis, Tularemia, Typhus, West Nile virus, Yellow fever, Zika virus disease
Zoonoses, Non-Transmissible	Anthrax, Buffalopox, Ebola Reston virus, Hantavirus pulmonary syndrome, Leptospirosis, Rabies, <i>Streptococcus suis</i>
Zoonoses, Spillover	Arenaviridae virus, Ebola virus disease, Hendra virus, Lassa fever, Marburg hemorrhagic fever, Middle East respiratory syndrome coronavirus (MERS-CoV), Monkeypox, Nipah virus, Plague, Severe acute respiratory syndrome coronavirus (SARS-CoV-1), Viral hemorrhagic fever
Other	Bromide poisoning, Encephalitis, Guillain-Barré syndrome, Melamine contamination, Myocarditis, Silicon implants

The full database of DONs reports is freely available online as a part of the Georgetown Infectious Disease Atlas [20]. We plan to update this database regularly as new disease reports are produced by the WHO.

Results

We reviewed a total of 2,806 DONs reports from 1996–2019. The number of reports published annually ranged from 59 reports in 2011 to 205 reports in 2014 (Fig. 1). The average number of annual DONs reports between 1996 and 2019 was 117 and the median number of reports was 114.5. The three most commonly reported diseases were H5N1 influenza, Ebola virus disease, and MERS-CoV, which had totals of 453, 296, and 291 reports, respectively.

Using our ontology to organize the DONs reports resulted in a total of 1,105 unique infectious disease outbreaks. The number of unique outbreaks reported annually between 1996 and 2019 ranged from 30

unique outbreaks in 2007 to 77 unique outbreaks in 1998 (Fig. 2). The average number of unique outbreaks between 1996 and 2019 was 46 per year and the median was 45 per year.

The 2014 west African Ebola epidemic was the outbreak with the greatest number of DONs reports, with a total of 172 reports; the 2009 H1N1 influenza pandemic was the outbreak with the second most DONs reports, with a total of 110 reports; the 2003 severe acute respiratory syndrome (SARS) outbreak had the third most reports, with a total of 95. Each unique outbreak received 3 DONs reports, on average.

Classifying the outbreaks by pathogen etiology, 223 (20%) were caused by directly-transmitted human pathogens, 118 (11%) were caused by seasonal and pandemic influenza strains, 228 (21%) were caused by vector-borne pathogens, 312 (28%) were caused by environmental or foodborne pathogens, 29 (3%) were caused by non-transmissible zoonotic pathogens, 169 (15%) were spillover events caused by transmissible zoonotic pathogens, and 26 (2%) were caused by other pathogens or events (Fig. 3).

Geographically, China, Saudi Arabia, the Democratic Republic of the Congo, Indonesia, and Egypt were the five countries most frequently affected by outbreaks reported in the DONs (Fig. 4). From 1996–2019, a total of 262 reports were published on outbreaks occurring in China, 194 on outbreaks in Saudi Arabia, 180 on outbreaks in the Democratic Republic of the Congo, 147 on outbreaks in Indonesia, and 115 on outbreaks in Egypt.

Discussion

This work improves the accessibility of data of the WHO's DONs reports and promotes future research efforts focused on analyzing WHO-confirmed infectious disease outbreaks. The availability of the data will allow for detailed analyses including descriptive work examining correlations between infectious disease outbreaks and other contextual factors, such as climatic events or conflict. Such analyses could help to predict future infectious disease outbreaks.

There are several notable shortcomings related to consistency, standardization practices, accuracy, and transparency that should be addressed. Regarding consistency, there is no clear format or structure for DONs reports. The reports and information included therein seem to be contingent on the specific disease, the information provided by the source or country, and even the report author. Reports can vary greatly in length and amount of detail provided. The inclusion of data tables, charts or links in reports can provide useful information, but their inclusion is infrequent and does not adhere to specific patterns or methodologies. While the lack of structure may improve flexibility and reflect the information reported to the WHO, these inconsistencies make it difficult to quickly identify important information.

Attempting to characterize the reports also presented a challenge while creating our database. While we strove to adhere strictly to the information contained in DONs reports, the reports do not employ a standardized list of pathogens or naming convention. For example, various DONs reports relating to Ebola referred to the disease as 'Ebola,' 'Ebola Haemorrhagic Fever,' 'EHV,' 'Ebola Virus Disease,' and 'EVD.' Additionally, the DONs reports often switch between spellings of cities or provinces; for example, certain

cities in Egypt appeared in consecutive reports with several different spellings. This oversight could make outbreaks appear to be more severe to those not intimately familiar with the geography of the outbreak by suggesting that the disease is spreading to other localities.

To this end, we believe that the WHO could improve DONs reports by implementing a consistent and systematized reporting format. An appropriate place to start is with the criteria put forward by Smolinski and colleagues [18]; case total tables organized by defined probable, suspected and confirmed cases; and subsections for appropriate or priority contextual factors, such as meteorology or climate hazards, community resistance, conflict, migration, or mass gatherings. Doing so would make information easier to find, ensure consistent reporting, and could allow for the reports to become machine-readable, and thus a more accessible source of information.

The DONs reports also occasionally contained errors that could call the validity of reported results into question. For example, a report on avian influenza in Vietnam published on January 11, 2009 reports that an individual first developed symptoms on January 28, 2009 and was hospitalized on January 31, 2009, several weeks after the report was published [21]. Other reports contain similar errors and chronological inconsistencies. These include the same report published twice on consecutive days [22], and conflicting information regarding the subject of the report [23].

Closely related to the issues surrounding inconsistencies and errors in the DONs reports, transparency regarding how the reports are compiled presents an additional concern. The WHO does not publish any secondary information about the DONs reports or catalog. Many questions arise in the absence of this information – namely with regard to the author of reports, the information included in the reports, and the prioritization of pathogens or geographies. We posit that this information isn't included not out of enigmatic intentions, but rather as a result of the aforementioned inconsistencies and reporting errors. We recommend the WHO publish secondary information detailing how reports are prepared and compiled. Establishing these parameters and guidelines would improve the transparency and standardization of the reporting process, could act to improve confidence in the reports themselves as the single authoritative collection of disease outbreaks and might allow for it to be more easily adapted and used for analytic purposes.

Additionally, while reviewing DONs reports, it became clear that there exist significant discrepancies between the coverage allocated to various diseases. Some diseases, such as H5N1 influenza, Ebola virus disease, and MERS-CoV enjoy consistent, precise, and timely coverage. Still, reports on other infectious diseases may not materialize into a DONs report until months after the WHO is notified. Cholera, for example, is the fourth most commonly reported disease in the DONs, but some reports are not published until months after initial reporting of the outbreak to relevant authorities [24]. Furthermore, we identified multiple events that received support from the Contingency Fund for Emergencies (CEF) that are not included in the DONs. For example, a 2017 outbreak of dengue fever in Pakistan received funding from the CEF, as did a 2018 outbreak of malaria in Nigeria, but both were omitted from the DONs reports. On the premise that these events constitute important health threats that warrant the allocation of millions of

dollars for response efforts, their notable absence in DONs reports is surprising, and ultimately supports the conclusion that, though the WHO's catalog of DONs reports represents the only official source of surveillance data, it is far from a comprehensive summary of priority outbreaks.

To remedy these challenges, we recommend the WHO decide upon and publish a set of criteria outlining what events merit a DONs report. Acknowledging that a majority of emerging infectious diseases are zoonotic in nature [2, 4], discussions surrounding pathogen detection in animal and non-human reservoirs must also be a part of this conversation. At present, some DONs reports contain information regarding the detection of pathogens in animals in the absence of any human cases, but these are not consistent nor comprehensive.

We believe the standardized methodology we provide addresses many of these limitations while maintaining the integrity of the data in the DONs reports and hope that our efforts to create a searchable, standardized database of DONs reports will inform important global health analyses and policy decisions.

Conclusions

This work improves the accessibility of data of the WHO's DONs reports – the only official and public source for WHO-confirmed infectious disease outbreaks. The resulting dataset ultimately addresses several limitations in the current reporting practices used by the WHO and will allow for important and informative analyses in the future, including descriptive work examining correlations between infectious disease outbreaks and other contextual factors.

Abbreviations

CEF – Contingency Fund for Emergencies

DONs – Disease Outbreak News

HIV – Human immunodeficiency virus

MERS-CoV – Middle East respiratory syndrome coronavirus

SARS – Severe acute respiratory syndrome

WHO – World Health Organization

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed are free available online as a part of the Georgetown Infectious Disease Atlas (<https://gida.ghscosting.org/>).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the Open Philanthropy Project. The funders had no role in the study design; data collection, analysis, and interpretation; preparation of the manuscript; decision to publish.

Authors' contributions

RK conceptualized the project; CJC, JL, and MRB contributed to project design and project oversight; JL, MAP, and MCP conducted data collection; MRB conducted data visualization; JL, MCP, and MRB drafted the manuscript; all authors contributed to reviewing and editing the final manuscript submitted for publication.

Acknowledgements

We are immensely grateful to Dr. Alexandra Phelan for her help supporting this work, and sincerely thank her for the guidance and advice provided throughout the project.

References

1. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature*. 2004;430:242-9. DOI:10.1038/nature02759.
2. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451:990-3. DOI:10.1038/nature06536.
3. Morse SS. Factors and determinants of disease emergence. *Rev Sci Tech*. 2004;23:443-51. DOI:10.20506/rst.23.2.1494.

4. Smith KF, Goldberg M, Rosenthal S, Carlson L, Chen J, Chen C, et al. Global rise in human infectious disease outbreaks. *J R Soc Interface*. 2014;11:20140950. DOI:10.1098/rsif.2014.0950.
5. World Health Organization. International Health Regulations (2005). Geneva (Switzerland): World Health Organization; 2005.
6. Madoff LC. ProMED-mail: An early warning system for emerging diseases. *Clin Infect Dis*. 2004;39:227-32. DOI:10.1086/422003.
7. Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and early warning outbreak detection: A Canadian contribution to global public health. *Can J Public Health*. 2006;97:42-4. DOI:10.1007/BF03405213.
8. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med*. 2008;5:e151. DOI:10.1371/journal.pmed.0050151.
9. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med*. 2009; 360:2153-5. DOI:10.1056/NEJMp0900702.
10. Toovey S. Gideon database. *Travel Med Infect Dis*. 2010;8:40. DOI:10.1016/j.tmaid.2009.11.005.
11. World Health Organization. Disease Outbreak News. [cited 2020 May 15]. <http://www.who.int/csr/don/en>.
12. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, Keller M, et al. Global capacity for emerging infectious disease detection. *Proc Natl Acad Sci USA*. 2010;107:21701-6. DOI:10.1073/pnas.1006219107.
13. Chan EH, Scales DA, Brewer TF, Madoff LC, Pollack MP, Hoen AG, et al. Forecasting high-priority infectious disease surveillance regions: a socioeconomic model. *Clin Infect Dis*. 2013;56:517-24. DOI:10.1093/cid/cis932.
14. Kluberg SA, Mekaru SR, McIver DJ, Madoff LC, Crawley AW, Smolinski MS, et al. Global Capacity for Emerging Infectious Disease Detection, 1996-2014. *Emerg Infect Dis*. 2016;22:E1-6. DOI:10.3201/eid2210.151956.
15. Poletto C, Boëlle P-Y, Colizza V. Risk of MERS importation and onward transmission: a systematic review and analysis of cases reported to WHO. *BMC Infect Dis*. 2016;16:448. DOI:10.1186/s12879-016-1787-5.
16. Aghazadeh-Attari J, Mohebbi I, Mansorian B, Ahmadzadeh J, Mirza-Aghazadeh-Attari M, Mobaraki K, et al. Epidemiological factors and worldwide pattern of Middle East respiratory syndrome coronavirus from 2013 to 2016. *Int J Gen Med*. 2018;11:121-5. DOI:10.2147/IJGM.S160741.
17. Talisuna AO, Okiro EA, Yahaya AA, et al. Spatial and temporal distribution of infectious disease epidemics, disasters and other potential public health emergencies in the World Health Organisation Africa region, 2016-2018. *Glob Health*. 2020;16:9. DOI:10.1186/s12992-019-0540-4.
18. Smolinski M, Crawley AW, Olsen JM. Finding Outbreaks Faster. *Health Secur*. 2017;15:215-20. DOI:10.1089/hs.2016.0069.

19. International Organization for Standardization. ISO 3166 Country Codes. [cited 2020 May 15]. <https://www.iso.org/iso-3166-country-codes.html>.
20. Center for Global Health Science & Security. Georgetown Infectious Disease Atlas. [cited 2020 May 15]. <https://gida.ghscosting.org/>.
21. World Health Organization. Avian influenza – situation in Viet Nam - update. 2009 Feb 11 [cited 2020 May 15]. https://www.who.int/csr/don/2009_02_11/en/.
22. World Health Organization. Influenza A(H1N1) - update 39. 2009 May 27 [cited 2020 May 15]. https://www.who.int/csr/don/2009_05_27/en/
23. World Health Organization. Test Rift valley. 2007 Sep 25 [cited 2020 May 15]. <https://www.who.int/csr/don/testpicturenov07/en/>.
24. World Health Organization. Cholera outbreaks in the Democratic Republic of Congo (DRC) and the Republic of Congo. 2011 Jul 22 [cited 2020 May 15]. https://www.who.int/csr/don/2011_07_22/en/.

Figures

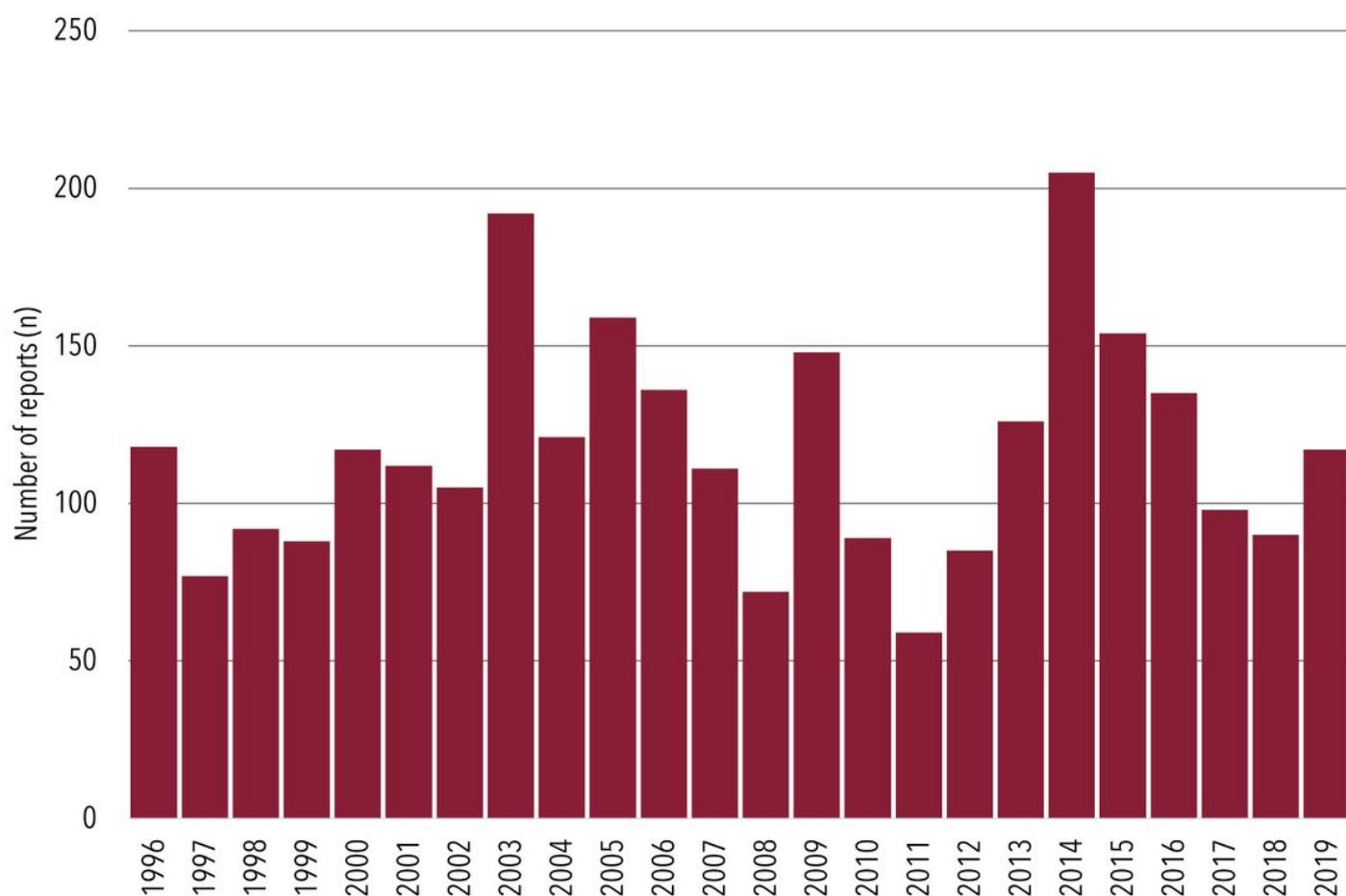


Figure 1

Temporal distribution of reports published in the WHO's Disease Outbreak News, 1996–2019.

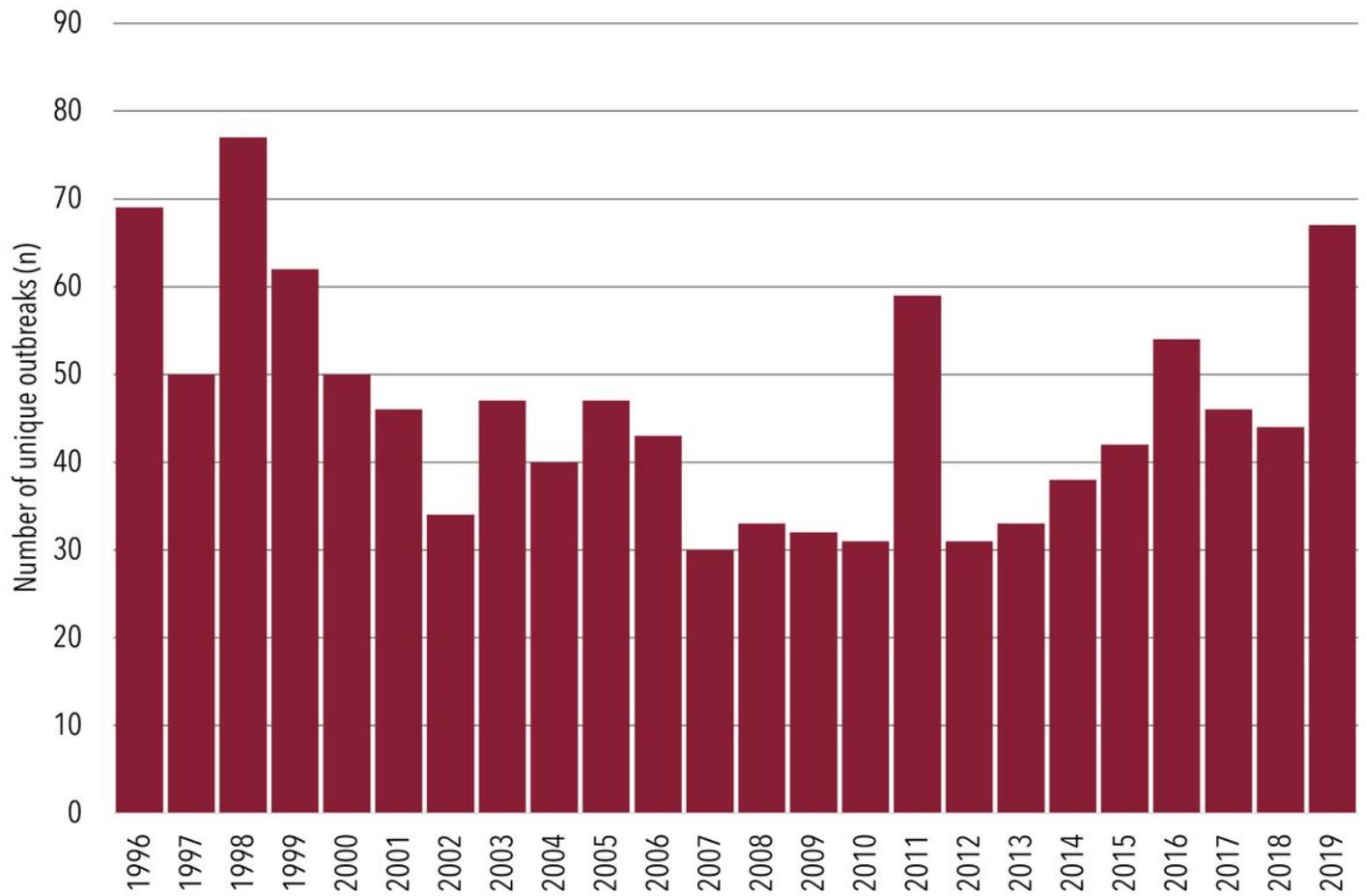


Figure 2

Temporal distribution of unique outbreaks reported in the WHO's Disease Outbreak News, 1996–2019.

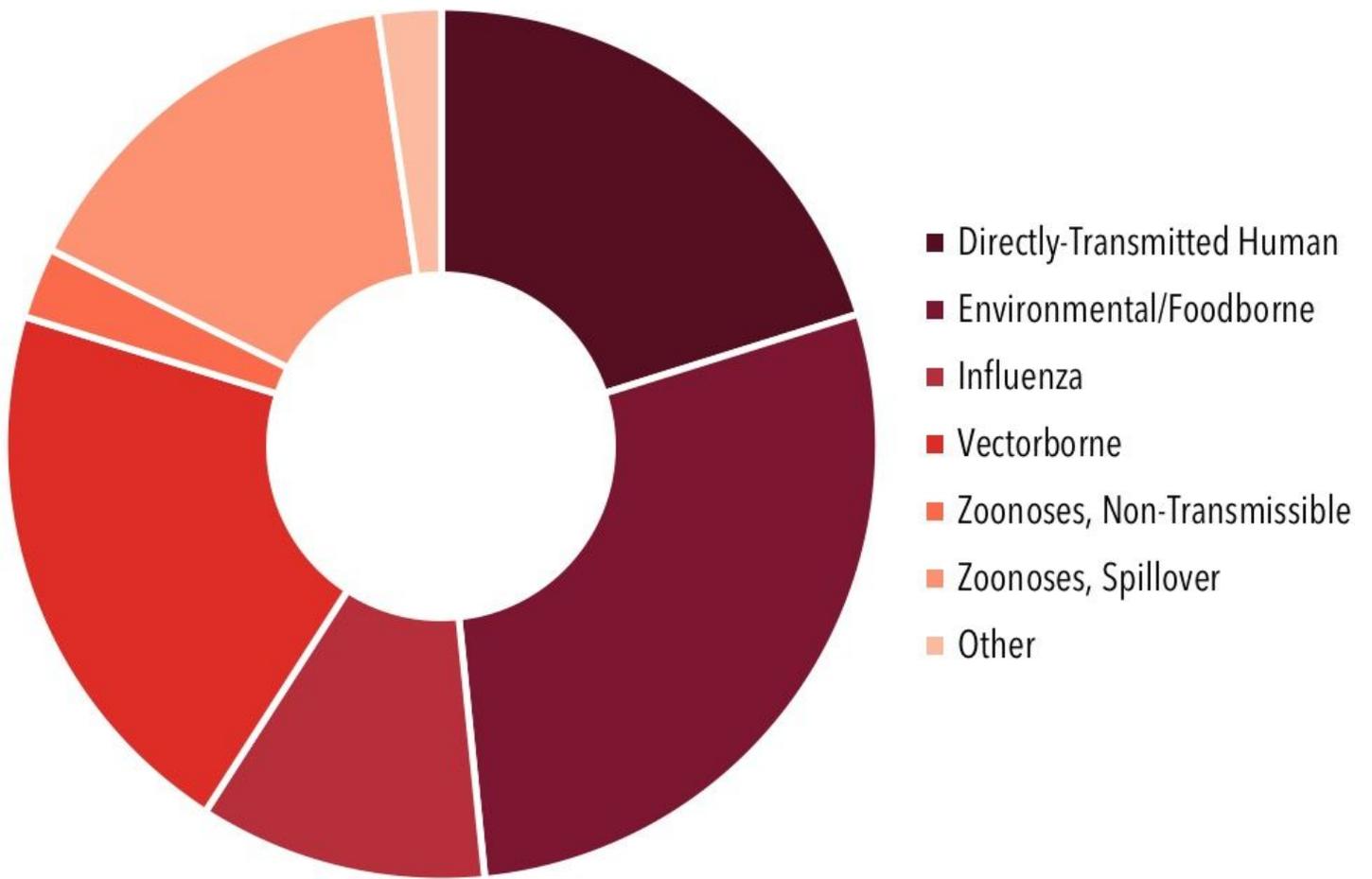


Figure 3

Unique outbreaks in the WHO's Disease Outbreak News classified by pathogen etiology, 1996–2019.

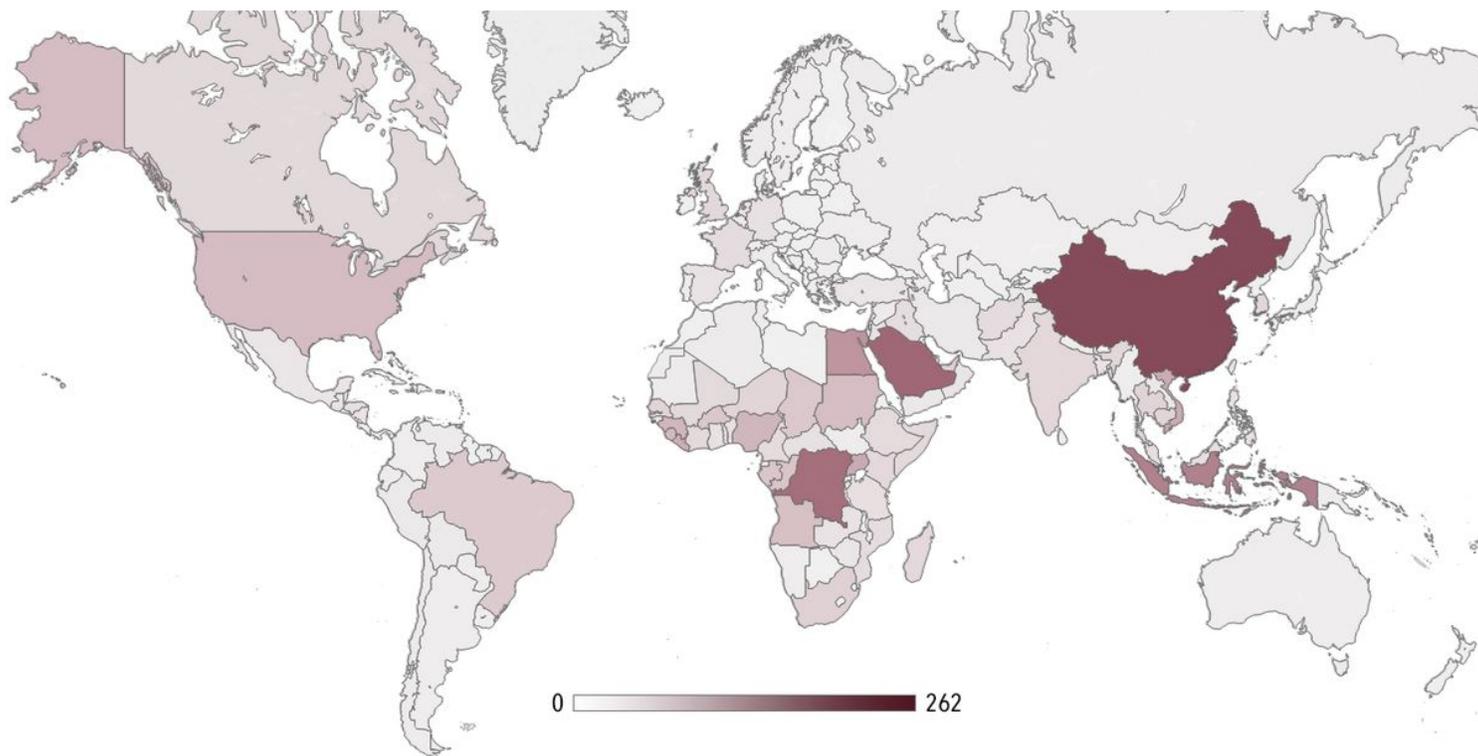


Figure 4

Number and geographic distribution of reports in the WHO's Disease Outbreak News, 1996-2019. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement8.docx](#)