



Risk assessment strategies for early detection and prediction of infectious disease outbreaks associated with climate change

EE Rees^{1*}, V Ng², P Gachon³, A Mawudeku⁴, D McKenney⁵, J Pedlar⁵, D Yemshanov⁵, J Parmely⁶, J Knox^{1,2}

Abstract

A new generation of surveillance strategies is being developed to help detect emerging infections and to identify the increased risks of infectious disease outbreaks that are expected to occur with climate change. These surveillance strategies include event-based surveillance (EBS) systems and risk modelling. The EBS systems use open-source internet data, such as media reports, official reports, and social media (such as Twitter) to detect evidence of an emerging threat, and can be used in conjunction with conventional surveillance systems to enhance early warning of public health threats. More recently, EBS systems include artificial intelligence applications such machine learning and natural language processing to increase the speed, capacity and accuracy of filtering, classifying and analysing health-related internet data. Risk modelling uses statistical and mathematical methods to assess the severity of disease emergence and spread given factors about the host (e.g. number of reported cases), pathogen (e.g. pathogenicity) and environment (e.g. climate suitability for reservoir populations). The types of data in these models are expanding to include health-related information from open-source internet data and information on mobility patterns of humans and goods. This information is helping to identify susceptible populations and predict the pathways from which infections might spread into new areas and new countries. As a powerful addition to traditional surveillance strategies that identify what has already happened, it is anticipated that EBS systems and risk modelling will increasingly be used to inform public health actions to prevent, detect and mitigate the climate change increases in infectious diseases.

Suggested citation: Rees EE, Ng V, Gachon P, Mawudeku A, McKenney D, Pedlar J, Yemshanov D, Parmely J, Knox J. Risk assessment strategies for early detection and prediction of infectious disease outbreaks associated with climate change. *Can Commun Dis Rep* 2019;45(5):119–26. <https://doi.org/10.14745/ccdr.v45i05a02>

Keywords: climate change, risk assessment, event-based surveillance systems, artificial intelligence, machine learning, natural language processing, risk modelling

Introduction

Climate warming trends have been accelerating over the last few decades. The world's nine warmest years in the time period from 1850 to 2017 have all occurred in the last twelve years, with a total increase of approximately 0.97°C in the average annual air temperature for the time period from 1880 to 2017 (1). This ostensibly small increase in average global temperature is nevertheless responsible for significant changes in the worldwide weather patterns and associated effects on society through sea level rise (and associated erosion) and increased frequency and intensity of flooding, droughts (with associated wildfires and crop failures) and freezing rain events (2). Of particular importance to Canada, climate warming is even more acute at higher latitudes and in the winter months (3). Over the past 70 years, the overall annual average temperature in Canada has increased by 1.8°C (4), with an average winter temperature increase of 3.4°C (4). In some areas in the northwest, this increase has been even higher. Because climate change affects not only temperatures but precipitation patterns, Canada is experiencing generally drier conditions in the west and above average precipitation in the east (4).

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Affiliations

¹ Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, St. Hyacinthe, QC

² Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, Guelph, ON

³ Centre pour l'Étude et la Simulation du Climat à l'Échelle Régionale (ESCLER), Université du Québec à Montréal (UQAM), Montréal, QC

⁴ Office of Situational Awareness and Operations, Centre for Emergency Preparedness and Response, Public Health Agency of Canada, Ottawa, ON

⁵ Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie, ON

⁶ Canadian Wildlife Health Cooperative, University of Guelph, Guelph, ON

***Correspondence:**
erin.rees@canada.ca



Climate-driven changes to temperature and precipitation are known to affect the risk of infectious disease transmission. Climate change is modifying range distributions of disease vectors (i.e. ticks and mosquitoes) and reservoir populations (i.e. birds, rodents and deer) that participate in the transmission of pathogens from ticks and mosquitoes to humans as climate suitability for vector and reservoir populations change (5,6). For example, the increase in cases of Lyme disease in Canada reflect the northward expansion of the range of the black-legged tick vector, *Ixodes scapularis*, in the United States (US) and into southern Canada, as climate change has made Canada more conducive to establishing tick populations (7,8). This expansion of the area where the vectors and their reservoirs can thrive means not only an increased risk of sporadic infectious disease but also an increased likelihood that these vectors, and the diseases that they carry, can become endemic (6,9–11).

In addition, climate change is influencing the mobility patterns of people and goods. An increase in “climate refugees”, people displaced when their lives and/or livelihoods are at risk from extreme weather events, is expected (11). Refugees, often from geographical areas where infectious diseases are more common and with different vaccination schedules and practices, may inadvertently bring these diseases into Canada (12). Tourism is also affected by climate change, as changes in both home and travel destinations influence the push and pull of factors motivating people to travel and the potential for disease spread (13–15). Vectors and pathogens can inadvertently be transported through shipments by air, land and sea (16–18). Land and sea containers are known to support the invasion of mosquitoes because larva can develop in trapped standing water, and if no water exists, eggs can withstand desiccation for weeks to months (19,20). Air travel has also been responsible for travellers carrying infections into new areas. In Canada, returning travellers have brought with them the Zika virus and have also sparked an outbreak of severe acute respiratory syndrome (SARS) coronavirus (15,21,22).

Thus, the increased risks of infectious diseases with climate change pose important public health risks and work is underway to monitor, assess and predict the impact of these risks. In the past, public health management has depended on notifiable disease reporting surveillance systems to detect outbreaks, monitor disease progression and inform prevention and mitigation policies. However, traditional surveillance systems are typically characterized by delays in the reporting and analysis of the data and the communication of the results.

To address the need for closer to real-time surveillance of emerging issues and earlier insight on potential health impacts, two risk assessment strategies have been, and are being, developed: event-based surveillance (EBS) systems, which increasingly incorporate artificial intelligence; and risk modelling. The objective of this overview is to describe these two risk assessment strategies and how they can inform public health actions to prevent, detect and mitigate the climate change increases in infectious diseases.

Event-based surveillance systems

Event-based surveillance systems use a variety of open-source internet data and assessment techniques to identify disease threats (23,24). Typical open-source internet data include online newswires, social media and other internet data streams, in multiple languages, to detect early-warning signals of threats to public health. These systems have proven to be more timely in comparison with conventional surveillance data sources from laboratory results or hospitals (25), and can be used in conjunction with conventional surveillance systems to enhance early warning of public health threats (26). The more quickly signals from an evolving outbreak are identified, the more quickly the outbreak can be tracked and a public health response can be planned and implemented (27).

There are three types of EBS systems: moderated; partially moderated; and fully automated (28). The level of automation influences how the information flow in EBS systems is managed from the open-source internet data from news aggregators (e.g. Factiva, Google News, Moreover Baidu), Rich site summary (RSS) and social media feeds from official and unofficial sources (e.g. Twitter for US Centers for Disease Control and general public), and validated official reports (e.g. World Health Organization, US Centers for Disease Control). The Program for Monitoring Emerging Diseases (ProMED) is an example of a moderated system and was on the forefront of EBS development over 25 years ago (29,30). ProMED is run by volunteer analysts (who are expert curators) who monitor and choose news articles, validate the content and notify subscribers of noteworthy infectious disease events. Strengths of this system include having a low signal-to-noise ratio, being open access and having a broad reach. However, volunteers do not cover all populations at risk, volunteer biases can influence the moderation of events and volunteers do not have the resources (nor are they expected) to provide detailed information giving situational awareness for assessing the threat level (29).

The Global Public Health Intelligence Network (GPHIN) is a partially moderated system that was developed by the Government of Canada, in collaboration with the World Health Organization, four years after ProMED (31–33). GPHIN access is restricted to agencies with health-related mandates. Artificial intelligence (AI) algorithms in GPHIN automate a stream of two to three thousand news articles per day that are moderated by 12 expert analysts who identify and issue alerts for threats using tacit contextual information (e.g. historic context, market trends, travel bans and climate anomalies). An example of the usefulness of GPHIN dates back to early 2003 when analysts identified reports from China referring to increased sales of antiviral therapies just before the global onset of the SARS epidemic (34). Unlike ProMED, GPHIN benefits from multi-staged filtering using AI and trained analysts. Artificial intelligence enables processing of a larger data stream, and analysts have the resources to provide information for situational awareness. Both ProMED and GPHIN can function in multiple languages; however, it is



expensive for GPHIN to add in other languages because of the cost to hire analysts with language fluency (33).

Fully automated systems include the European Commission's Medical Information System (MedISys), Pattern-based Understanding and Learning System (PULS) and HealthMap. These systems are open to the public, but also have restricted access to serve the needs of health agencies such as private discussion forums, increased functionality and data processing of commercial sources (35,36). Fully automated systems are faster at processing data and less expensive to operate than moderated systems. The main drawback is the higher signal-to-noise ratio meaning that there is an increased risk of identifying false threats (37,38). The EBS systems can be connected in synergistic ways to address this risk (39). For example, MedISys uses low signal-to-noise ratio data from ProMED and GPHIN, and uses more advanced language processing algorithms from PULS. The PULS extracts information about events identified in the MedISys stream and then returns these data back to MedISys (36,40). The different types of EBS systems are summarized in **Table 1**.

Table 1: Summary of some event-based surveillance systems

| Type | Example | Establishment | Public availability |
|---|--|--|--|
| Moderated system ^a | Program for Monitoring Emerging Disease (ProMED) (29,30) | In 1994 as a nonprofit organization | Yes |
| Partially moderated system ^b | Global Public Health Intelligence Network (GPHIN) (31–33) | In 1998 through partnership between the Government of Canada and World Health Organization | No; available to partnered health agencies |
| Fully automated system ^c | Medical Information System (MedISys) (36,41,42) | In 2004 by the European Commission | Yes |
| | HealthMap (35,38,40,43) | In 2006 by Boston Children's Hospital | Yes |
| | Pattern-based Understanding and Learning System (PULS) (36,44,45) | In 2007 by the Department of Computer Science, University of Helsinki, Finland | Yes |

^a A moderated system: volunteer expert-curators identify, review and validate sources and create the reports

^b A partially-moderated system: automatically acquires, categorizes, and filters sources. Expert-curators moderate the subset of sources and create the reports

^c A fully-automated system: automatically acquires, categorizes, filters and reports the health-related sources

Artificial intelligence applications

The ability of EBS systems to quickly and accurately detect threats (such as outbreaks of infectious diseases) has been revolutionized by artificial intelligence applications for data processing. Open-source internet data are considered “unstructured” in the sense that news articles, blogs, tweets, etc., provide a narrative describing an event. The text, numbers and dates are not organized in a data model, such as a database, that can be used for automated event detection and risk modelling; therefore, open-source data must be processed to extract and structure information about what happened, where it happened, when it happened and to whom it happened. The EBS systems use natural language processing (NLP) methods to process and understand event narratives (46–48). Natural language processing is a field of research dedicated to understanding human discourse (49). Early methods include the sub-language approach, where rules and patterns are used to interpret and classify vocabulary, syntax and semantics of the unstructured narrative. The EBS systems have taxonomies of terms to match predefined terms and their synonyms to those found in the data sources. Much like with a conventional literature search, taxonomic classification of narratives can identify health-related articles by searching for related terms (e.g. human influenza A synonyms include H1N1, swine flu, California flu, human influenza and influenza A) (50). The sublanguage approach for identifying health-related data in EBS systems is effective but also has drawbacks. Taxonomies are not easily generalizable and must be developed for each disease being monitored and kept up-to-date as language evolves and new discoveries about diseases are made. In this light, NLP has established a strong foundation in using machine learning (ML) methods.

Machine learning is a subset of AI that uses algorithms, such as statistical models, to perform a specific task without using explicit instructions; instead, relying on patterns and inference. The EBS systems gather open-source internet data (feeds and web queries) and then filter these data through a combination of the sublanguage approach and ML methods, where the latter is used to perform more complex tasks for analysing syntax, semantics, morphology, pragmatics and discourse (51). For example, ML methods can be used to determine the difference between non-health related articles (e.g. “Bieber fever” refers to avid supporters of Justin Bieber) and those discussing an infectious disease outbreak (43,51,52). Machine learning methods can also be used to distinguish between ambiguities in dates and locations, such as past and present outbreaks in articles that discuss historical context (53,54). Novel applications for ML methods are also being developed, such as structuring disease case information into epidemiological line lists (a listing of individuals affected by the disease and related information; i.e. health status, sex, location, date of onset, hospitalized) that can be used in outbreak investigations and risk modelling (55). Once the information from open-source internet data has been processed into a data model, the event can then be reviewed and reported, as appropriate; furthermore, additional data

**Table 2: Information flow from open-source internet data in event-based surveillance systems**

| EBS | Data collection | Data processing | Data analytics | Reporting |
|---|---|--|---|--|
| Moderated systems | Human analysts search and identify open-source internet data for health-related concern | Human analysts review, filter and designate the threat level of the event | None | Reports on health-related threats are communicated through email and posted on EBS system website |
| Partially moderated and fully automated systems | Automated feed of open-source internet data | Taxonomic classification and ML algorithms filter and classify events based on their metadata (e.g. type of threat, location and date). ML algorithms score the level of relevancy. In partially moderated systems, highly scored data sources are curated by human analysts | Analytic techniques evolve with time and differ among EBS systems. Current techniques include the following: mapping of geo-tagged events; bar plots showing changes over time to keyword counts, number of identified articles and expected and observed number of disease cases; word clouds showing importance of keyword terms; alert notices given sudden increases to case counts, reliability of sources and/or number of unique sources | Reports on health-related threats are communicated through email and posted on EBS system website and notified to appropriate web application user communities |

Abbreviations: EBS, event-based surveillance; ML, machine learning

analytics can be performed to communicate the current and predicted impact of the health threat. A summary of information flow from data collection, processing, analytics and reporting for EBS systems is presented in **Table 2**.

Risk modelling

An important advancement for risk assessment is increasing the variety of data being used in modelling approaches. Risk modelling in the context of infectious diseases is the process of identifying and characterizing factors in individuals or populations that increase their vulnerability to contracting disease (e.g. age, proximity to outbreak). Statistical inference is a well-grounded and informative risk modelling approach that includes regression analysis. This method is used to determine how risk factors (explanatory variables) are associated with the outcome of interest (e.g. number of reported cases). Regression models, and statistical inference in general, are developing to include information from open-source internet data. An early example was the inclusion of search query engine data from Google Flu Trends as a predictor for the outcome of the number of reported physician visits for flu-like illnesses (56). The resulting model was then used to predict the number of seasonal influenza cases one to two weeks into the future; however, this approach was not as effective in predicting outbreaks outside of the traditional flu season because of associations being identified with search query trends not related to seasonal influenza (e.g. winter basketball season) (57). Subsequent work improved the accuracy of predicting seasonal influenza flu trends by using additional sources of open-source data (e.g. Twitter) and expanding the regression method to benefit from ML algorithms that can find complex associations among the outcome and explanatory variables (58). Furthermore, regression modelling for the risk of infection has improved by including, in addition to open-source internet data, additional explanatory variables

(e.g. climate and meteorological data from satellite imagery) that account for the presence, movement and distribution of pathogens, vectors, reservoir populations and infected people (59,60). For example, in China, the expected number of cases of hand, foot and mouth disease in children was best predicted by including data on weekly temperature and precipitation as well as data on hand, foot and mouth disease-related queries from the Chinese Baidu search engine (61).

Another dominant risk modelling approach is the use of compartmental models to mathematically simulate transmission dynamics of a population; that is, the flow of individuals among health states, such as susceptible (S), infectious (I) and recovered (R). For example, SIR models require defining parameters for the infectious rate (or inversely, the infectious period) and the rate of infectious contacts. It is then possible to estimate if an infected population will become epidemic, and to characterize the prevalence of a disease over time. The compartmental modelling approach has more recently developed to simulate transmission dynamics among multiple populations (meta-populations). This requires the inclusion of mobility data to define the rate of individuals moving among populations (62). Human mobility at a meta-population level can be considered as the movement of people in a connected network of cities and countries. These data can be obtained from mobile phone call records and air traffic passenger volumes (63,64). Through meta-population modelling, it is possible to identify the travel routes through which pathogens may spread or be carried to Canada, as well as to determine the likelihood of these events (65,66). For example, the Zika virus is estimated to have first appeared in Brazil between August 2013 and April 2014 by infected travellers entering the country at Rio de Janeiro, Brasilia, Fortaleza and/or Salvador; and this introduction was followed by epidemics in Haiti, Honduras, Venezuela and then Colombia (21).



Discussion

There is uncertainty as to how climate change will affect the many factors related to the occurrence and spread of infectious diseases. These factors will undoubtedly include changes to the distributions of vector and reservoir populations, and changes to the mobility of people and goods and potential transport of pathogens, with subsequent impacts on exposure and transmission risks. To monitor infectious disease outbreaks in an effective and timely manner, public health professionals need better access to up-to-date surveillance data. To achieve this, conventionally-obtained data, such as that from existing notifiable disease reporting surveillance systems, are increasingly being augmented by EBS systems. The EBS systems are benefiting from ML and NLP methods to more fully exploit the available data; however, challenges remain (59). There are issues of data sharing and privacy that need to be resolved. For example, at what level can personal data be used and disclosed in the detection of health-related events? Both Google and Twitter provide their data freely to the public as finely aggregated per week and city; however, more precise information on the timing and location of the source would enable more comprehensive event detection (26). Also, there are differences where and how people use the internet and social media around the world: there are gaps in internet and mobile phone use in Africa (67); Baidu, rather than Google, is the predominant search engine in China (61); and the propensity of people using Twitter to report illnesses is dependent on age and socioeconomic status (68).

Risk modelling provides a means of estimating the health impacts of emerging infectious diseases. Advances in risk modelling approaches include integrating open-source internet and climate data to inform these estimates, and accounting for the mobility of humans to spread infectious diseases globally. As with EBS systems, risk modelling approaches are limited by the availability of the data that can be obtained. For example, mobile phone call records and air traffic data provide information to the nearest cell phone tower and airport respectively, but more precise location data are available, granted privacy concerns, through the global position system in mobile phones. Information at the individual level could greatly increase our understanding of the factors affecting disease occurrence and pathogen spread, for example, the role of certain people to drive the 2003 SARS outbreak (69).

Conclusion

Advances in assessing changes to vector and reservoir populations and human activity, and their impacts on infectious diseases, are now being monitored by a number of different surveillance and analytical strategies. Event-based surveillance systems use open-source data to gather information relating to infectious diseases. These systems can be moderated, partially moderated or fully automated, and each type of system has advantages and disadvantages. There is a growing trend

towards automation because of the ability to process high volumes of data, and the accuracy of ML and NLP methods to identify events are improving and may one day surpass the ability of human moderators. Risk modelling to understand and predict the health impacts of infectious diseases is commonly performed using statistical inference and compartmental modelling approaches. These methods are advancing the ability to identify populations at risk to emerging diseases, and forecast health impacts and determine pathways of disease spread by integrating open-source internet data and human mobility data, along with more traditional data variables from climate data and infectious disease outbreak data. The methods we have presented here are promising new developments that will increase our capacity to deal with evolving disease threats as the climate changes. Having more information (and more accurate information) sooner will make it possible for public health professionals to confirm and evaluate potential infectious disease outbreaks faster and thus to develop and commence treatment and other mitigation strategies in a more timely fashion.

Authors' statement

EER — Conceptualization, investigation, writing—original draft, supervision and project administration
 VN — Investigation, writing—review and editing
 PG — Writing—review and editing
 AM — Writing—review and editing
 DM — Writing—review and editing
 JP — Writing—review and editing
 DY — Writing—review and editing
 JK — Investigation, writing—review and editing

Conflict of interest

None.

Funding

This work was supported by the Public Health Agency of Canada.



References

1. National Oceanic and Atmospheric Administration. National Centers for Environmental Information. Climate at a Glance: Global Mapping. NOAA; 2018 (Accessed 2018-11-14). www.ncdc.noaa.gov/cag/
2. O'Neill BC, Oppenheimer M, Warren R, Hallegatte S, Kopp RE, Pörtner HO, Scholes R, K, van Ypersele JP, Yohe G. IPCC reasons for concern regarding climate change risks. *Nature Climate Change* 2017;7:28–37. DOI
3. Environment Canada and Climate Change Canada. Climate data and scenarios for Canada: synthesis of recent observation and modelling results. Gatineau(QC); 2016. www.canada.ca/en/environment-climate-change/services/climate-change/publications/data-scenarios-synthesis-recent-observation.html
4. Blunden J, Anrdt DS, Hartfield G, editors. State of the Climate in 2017. *Bull Am Meteor Soc.* 2018;99(8):Si-S332. DOI
5. Myers P, Lundrigan BL, Hoffman SM, Haraminac AP, Seto SH. Climate-induced changes in the small mammal communities of the Northern Great Lakes Region. *Glob Change Biol* 2009;15(6):1434–54. DOI
6. Roy-Dufresne E, Logan T, Simon JA, Chmura GL, Millien V. Poleward expansion of the white-footed mouse (*Peromyscus leucopus*) under climate change: implications for the spread of lyme disease. *PLoS One* 2013 Nov;8(11):e80724. DOI PubMed
7. Gasmi S, Ogden NH, Lindsay LR, Burns S, Fleming S, Badcock J, Hanan S, Gaulin C, Leblanc MA, Russell C, Nelder M, Hobbs L, Graham-Derham S, Lachance L, Scott AN, Galanis E, Koffi JK. Surveillance for Lyme disease in Canada: 2009-2015. *Can Commun Dis Rep* 2017 Oct;43(10):194–9. DOI PubMed
8. McPherson M, García-García A, Cuesta-Valero FJ, Beltrami H, Hansen-Ketchum P, MacDougall D, Ogden NH. Expansion of the Lyme disease vector *Ixodes Scapularis* in Canada inferred from CMIP5 climate projections. *Environ Health Perspect* 2017 May;125(5):057008. DOI PubMed
9. Ogden NH, Mechai S, Margos G. Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. *Front Cell Infect Microbiol* 2013 Aug;3:46. DOI PubMed
10. Lafferty KD. The ecology of climate change and infectious diseases. *Ecology* 2009 Apr;90(4):888–900. DOI PubMed
11. Berchin II. Climate change and forced migrations: an effort towards recognizing climate refugees. *Geoforum* 2017;84:147–50. DOI
12. McMichael C. Climate change-related migration and infectious disease. *Virulence* 2015;6(6):548–53. DOI PubMed
13. Scott D, McBoyle G. Using a 'tourism climate index' to examine the implications of climate change for climate as a tourism resource. In: Matzarakis A, de Freitas CR, editors. *Proceedings of First International Workshop on Climate Tourism and Recreation.* 2001 Oct 5- Oct 10; Porto Carras, Greece. International Society of Biometeorology; 2001, p. 69–88. www.academia.edu/2876399/Using_a_tourism_climate_index_to_examine_the_implications_of_climate_change_for_climate_as_a_tourism_resource
14. Scott D, McBoyle G, Schwartzentruber M. Climate change and the distribution of climatic resources for tourism in North America. *Clim Res* 2004;27(2):105–17. DOI
15. Ogden NH, Fazil A, Safronetz D, Drobot MA, Wallace J, Rees EE, Decock K, Ng V. Risk of travel-related cases of Zika virus infection is predicted by transmission intensity in outbreak-affected countries. *Parasit Vectors* 2017 Jan;10(1):41–9. DOI PubMed
16. Lounibos LP. Invasions by insect vectors of human disease. *Annu Rev Entomol* 2002;47(1):233–66. DOI PubMed
17. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Adv Parasitol* 2006;62:293–343. DOI PubMed
18. Tuite AR, Thomas-Bachli A, Acosta H, Bhatia D, Huber C, Petrasko K, Watts A, Yong JHE, Bogoch II, Khan K. Infectious disease implications of large-scale migration of Venezuelan nationals. *J Travel Med.* 2018;25(1). DOI
19. Manguin S, Boete C. The Importance of Biological Interactions in the Study of Biodiversity. Rijeka, Croatia: InTech; 2011. Chapter 3, Global Impact of Mosquito Biodiversity, Human Vector-Borne Diseases and Environmental Change; p. 27–50. DOI
20. Egizi A, Kiser J, Abadam C, Fonseca DM. The hitchhiker's guide to becoming invasive: exotic mosquitoes spread across a US state by human transport not autonomous flight. *Mol Ecol* 2016 Jul;25(13):3033–47. DOI PubMed
21. Zhang Q, Sun K, Chinazzi M, Pastore Y Piontti A, Dean NE, Rojas DP, Merler S, Mistry D, Poletti P, Rossi L, Bray M, Halloran ME, Longini IM Jr, Vespignani A. Spread of Zika virus in the Americas. *Proc Natl Acad Sci USA* 2017 May;114(22):E4334–43. DOI PubMed
22. Ruan S, Wang W, Levin SA. The effect of global travel on the spread of sars. *Math Biosci Eng* 2006 Jan;3(1):205–18. DOI PubMed
23. Barboza P, Vaillant L, Le Strat Y, Hartley DM, Nelson NP, Mawudeku A, Madoff LC, Linge JP, Collier N, Brownstein JS, Astagneau P. Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. *PLoS One* 2014 Mar;9(3):e90536. DOI PubMed
24. Hartley D, Nelson N, Walters R, Arthur R, Yangarber R, Madoff L, Linge J, Mawudeku A, Collier N, Brownstein J, Thinus G, Lightfoot N. Landscape of international event-based biosurveillance. *Emerg Health Threats J* 2010;3(1):7096. DOI PubMed



25. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, Eysenbach G, Brownstein JS. Use of unstructured event-based reports for global infectious disease surveillance. *Emerg Infect Dis* 2009 May;15(5):689–95. [DOI PubMed](#)
26. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014 Mar;343(6176):1203–5. [DOI PubMed](#)
27. Heymann DL. SARS and emerging infectious diseases: a challenge to place global solidarity above national sovereignty. *Ann Acad Med Singapore* 2006 May;35(5):350–3. [PubMed](#)
28. Linge JP, Steinberger R, Weber TP, Yangarber R, van der Goot E, Al Khudhairi DH, Stilianakis NI. Internet surveillance systems for early alerting of health threats. *Euro Surveill* 2009 Apr;14(13):19162. [PubMed](#)
29. Carrion M, Madoff LC. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. *Int Health* 2017 May;9(3):177–83. [DOI PubMed](#)
30. Yu VL, Madoff LC. ProMED-mail: an early warning system for emerging diseases. *Clin Infect Dis* 2004 Jul;39(2):227–32. [DOI PubMed](#)
31. 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 Jan-Feb;97(1):42–4. [PubMed](#)
32. Mawudeku A, Blench M, Boily L, St. John R, Andraghetti R, Ruben M. *Infectious Disease Surveillance*. 2nd ed. New York, NY: John Wiley & Sons; 2013. Chapter 31, The Global Public Health Intelligence Network; p. 457–69. www.wiley.com/en-us/Infectious+Disease+Surveillance%2C+2nd+Edition-p-9780470654675
33. Dion M, AbdelMalik P, Mawudeku A. Big Data and the Global Public Health Intelligence Network (GPHIN). *Can Commun Dis Rep* 2015 Sep;41(9):209–14. [DOI PubMed](#)
34. Knobler S, Mahmoud A, Lemon S, Sivitz L, Oberholtzer K, editors. *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary* (Institute of Medicine). Washington, DC: The National Academies Press; 2004. p. 376. www.nap.edu/catalog/10915/learning-from-sars-preparing-for-the-next-disease-outbreak-workshop
35. Harris JK, Hinyard L, Beatty K, Hawkins JB, Nsoesie EO, Mansour R, Brownstein JS; Louis Department of Health. Evaluating the implementation of a twitter-based foodborne illness reporting tool in the city of St. *Int J Environ Res Public Health* 2018 Apr;15(5):833. [DOI PubMed](#)
36. Linge JP, Steinberger R, Fuat F, Bucci S, Belyaeva J, Gemo M, Al-Khudhairi D, Yangarber R, van der Goot E. *Advanced ICTs for Disaster Management and Threat Detection: Collaborative and Distributed Frameworks*. Hershey, PA: IGI Publishing; 2011. Chapter: 9, MedISys: Medical Information System; p.131-142. [DOI](#)
37. Mantero J, Belyaeva J, Linge JP. How to maximise event-based surveillance web- systems : the example of ECDC/JRC collaboration to improve the performance of MedISys. Luxembourg; European Commission Joint Research Centre; 2011. Report No.: JRC 63805. <http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/16206/1/lb-na-24763-en-c.pdf>
38. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 2008 Jul;5(7):e151. [DOI PubMed](#)
39. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneau P; Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. *PLoS One* 2013;8(3):e57252. [DOI PubMed](#)
40. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 2008 Mar-Apr;15(2):150–7. [DOI PubMed](#)
41. Rortais A, Belyaeva J, Gemo M, van der Goot E, Linge JP. MedISys: an early-warning system for the detection of (re-) emerging food- and feed-borne hazards. *Food Res Int* 2010;43(5):1553–6. [DOI](#)
42. Alomar O, Batlle A, Brunetti JM, García R, Gil R, Granollers A, Jimenez S, Lavina A, Linge JP, Pautasso M, Reverte C, Riudavets J, Rortais A, Sancanelli G, Volani S, Vos S. Development and testing of the media monitoring tool MedISys for early identification and reporting of existing and emerging plant health threats. *Bull OEPP* 2015;45(2):288–93. [DOI](#)
43. Hawkins JB, Tuli G, Kluberg S, Harris J, Brownstein JS, Nsoesie E. A Digital Platform for Local Foodborne Illness and Outbreak Surveillance. *Online J Public Health Inform* 2016;8(1):e60. [DOI](#)
44. Hartley DM, Nelson NP, Arthur RR, Barboza P, Collier N, Lightfoot N, Linge JP, van der Goot E, Mawudeku A, Madoff LC, Vaillant L, Walters R, Yangarber R, Mantero J, Corley CD, Brownstein JS. An overview of internet biosurveillance. *Clin Microbiol Infect* 2013 Nov;19(11):1006–13. [DOI PubMed](#)
45. Yangarber R, Jokipii L, Rauramo A, Huttunen S. Extracting information about outbreaks of infectious epidemics. In: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*; 2005 Oct 6 - Oct 8; Vancouver, Canada. Stroudsburg, PA: Association for Computational Linguistics; 2005. p. 22–3. <https://dl.acm.org/citation.cfm?id=1220575&picked=prox>
46. Zeng Z, Shi H, Wu Y, Hong Z. Survey of Natural Language Processing Techniques in Bioinformatics. *Comput Math Methods Med* 2015;2015:674296. [DOI PubMed](#)
47. Iroju OG, Olaleke JO. Information Technology and Computer Science. *Inf Technol Comput Sci*. 2015;08:44–50. www.mecspress.org/DOI:10.5815/ijitcs.2015.08.07



48. Jordan S, Hovet S, Fung I, Liang H, Fu KW, Tsz Ho Tse Z. Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. *Data (Basel)* 2018;4(1):6. DOI
49. Liu F, Weng C, Yu H. *Clinical Research Informatics, Health Informatics*. pringer International Publishing; 2019. Chapter 17, Advancing Clinical Research Through Natural Language Processing on Electronic Health Records: Traditional Machine Learning Meets Deep Learning; p. 357–78. <https://link.springer.com/book/10.1007/978-3-319-98779-8>
50. Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals in Twitter. In: *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From linguistic signal to clinical reality*; June 2014; Baltimore, MD: Association for Computational Linguistics; 2014. p. 51-60. DOI
51. Şerban O, Thapen N, Maginnis B, Hankin C, Foot V. Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification. *Inf Process Manage* 2019;56(3):1166–84. DOI
52. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J Biomed Inform* 2015 Feb;53:196–207. DOI PubMed
53. Sprugnoli R, Caselli T, Tonelli S, Moretti G. The Content Types Dataset: a New Resource to Explore Semantic and Functional Characteristics of Texts. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*; 2017 July 30 - Aug 4; Vancouver Canada. ACL; 2017. p. 260–6. <http://acl2017.org/>
54. Gritta M, Pilehvar MT, Collier N. *A Pragmatic Guide to Geoparsing Evaluation: Toponyms, Named Entity Recognition and Pragmatics*. 2018. Language Technology Lab, Department of Teoretical and Applied Linguistics, University of Cambridge, Cambridge UK. www.academia.edu/38110706/A_Pragmatic_Guide_to_Geoparsing_Evaluation
55. Ghosh S, Chakraborty P, Lewis BL, Majumder M, Cohn E, Brownstein JS, Marathe M, Ramakrishnan N. GELL: Automatic extraction of epidemiological line lists from open sources. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2017 Aug 13- Aug 17; Halifax, Canada. New York USA: Association for Computing Machinery; 2017. p. 1477–86. DOI
56. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb;457(7232):1012–4. DOI PubMed
57. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) Google Flu Trends? *Am J Prev Med* 2014 Sep;47(3):341–7. DOI PubMed
58. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLOS Comput Biol* 2015 Oct;11(10):e1004513. DOI PubMed
59. Pollett S, Althouse BM, Forshey B, Rutherford GW, Jarman RG. Internet-based biosurveillance methods for vector-borne diseases: are they novel public health tools or just novelties? *PLoS Negl Trop Dis* 2017 Nov;11(11):e0005871. DOI PubMed
60. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of Internet-based dengue query data: Google Dengue Trends. *PLoS Negl Trop Dis* 2014 Feb;8(2):e2713. DOI PubMed
61. Huang DC, Wang JF. Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors. *Sci Total Environ* 2018 Jan;612:1293–9. DOI PubMed
62. Ajelli M, Gonçalves B, Balcan D, Colizza V, Hu H, Ramasco JJ, Merler S, Vespignani A. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect Dis* 2010 Jun;10:190. DOI PubMed
63. Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJ. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *J Infect Dis* 2016 Dec;214 suppl_4:S414–20. DOI PubMed
64. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci USA* 2009 Dec;106(51):21484–9. DOI PubMed
65. Arino J, Khan K. *Analyzing and Modelling Spatial and Temporal Dynamics of Infectious Disease*. Hoboken NJ: John Wiley & Sons; 2014. Chapter 5, Using Mathematical Modeling to Integrate Disease Surveillance and Global Air Transportation Data; p. 1–14.
66. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science* 2013 Dec;342(6164):1337–42. DOI PubMed
67. Bornman E. Information society and digital divide in South Africa: results of longitudinal surveys. *Inf Commun Soc* 2016;19. DOI
68. Nsoesie EO, Flor L, Hawkins J, Maharana A, Skotnes T, Marinho F, Brownstein JS. Social Media as a Sentinel for Disease Surveillance: What Does Sociodemographic Status Have to Do with It? *PLoS Curr* 2016;8. DOI PubMed
69. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005 Nov;438(7066):355–9. DOI PubMed