

## Commentary

# Preventing the next pandemic: Use of artificial intelligence for epidemic monitoring and alerts

C. Raina MacIntyre,<sup>1</sup> Samsung Lim,<sup>1,2</sup> and Ashley Quigley<sup>1,\*</sup><sup>1</sup>Biosecurity Program, The Kirby Institute, UNSW, Sydney, Australia<sup>2</sup>School of Civil & Environmental Engineering, UNSW, Sydney, Australia\*Correspondence: [ashley.quigley@protonmail.com](mailto:ashley.quigley@protonmail.com)<https://doi.org/10.1016/j.xcrm.2022.100867>

Emerging infections are a continual threat to public health security, which can be improved by use of rapid epidemic intelligence and open-source data. Artificial intelligence systems to enable earlier detection and rapid response by governments and health can feasibly mitigate health and economic impacts of serious epidemics and pandemics. EPIWATCH is an artificial intelligence-driven outbreak early-detection and monitoring system, proven to provide early signals of epidemics before official detection by health authorities.

True epidemic diseases are characterized by exponential growth. This means cases rise very rapidly over short periods of time, usually days or weeks.<sup>1</sup> It is this exponential growth that causes health system surges and compromises critical infrastructure. An endemic disease does not have these features, and changes, if any, occur over longer periods of time, usually years. Pandemic planning has historically been based on influenza, with the assumption that the most likely pandemic will be one emerging from a zoonotic influenza virus to affect humans.<sup>1</sup> The majority of investment in pandemic planning is around diagnostics, drugs, and vaccines. While essential, these typically are available late in the genesis of pandemics and after infections have spread widely. The economic cost of COVID-19 illustrates the devastation that pandemics can cause and the benefit of both early detection and prevention. The cost of the COVID-19 pandemic was estimated in 2020 alone to be more than \$16 trillion, and it is estimated to have more than doubled since then.<sup>2</sup>

Vast, open-source data can provide intelligence and early warning for pandemics and epidemics at a time when they have not yet spread beyond national borders. Signals that are detected early enough can feasibly prevent a pandemic by allowing early identification of a small outbreak, which can then be contained through isolation, contact tracing, and quarantine. Traditional public health surveillance relies on data being reported

from the health system or laboratories that is then validated to enable the monitoring of trends in infectious diseases. These data provide insights into trends over time, allow comparison between time periods, and can signal an unusual rise in a disease incidence. Trends can also help evaluate the success of public health interventions such as vaccination programs. Traditional surveillance is an essential public health tool, but it is substantially delayed and not timely enough to allow early detection of serious epidemics. Open-source rapid intelligence is not a replacement for traditional public health surveillance but an adjunct to enable early response and investigation of emerging outbreaks.

One of the earliest systems, ProMED Mail,<sup>3</sup> was developed in 1994 and relies on doctors and other health professionals to report unusual outbreaks. It remains an important early warning system. In the last decade, several other quantitative and automated Internet-based early warning systems have been developed. Yet, they remain niche tools that are either paywalled and open only to paying customers, restricted to selected users, or only used by a small minority of people. In routine public health practice at an operational level, very few people use open-source intelligence. The adoption of artificial intelligence (AI)-based technologies in public health is significantly less than in clinical medicine, and public health practitioners are wary of such tools.<sup>4</sup> For AI systems to have a prospect of prevent-

ing the next pandemic, they need to be widely used and easily accessible. This means the use of digital technology needs to be embedded in public health departments at the grassroots level and also in public health training so that trainees learn these methods as they develop their public health skills. It also needs to be equitable and affordable or provided through open-source licensing.

There are three systems we are aware of that provide open access to epidemic intelligence: ProMED Mail, HealthMap, and EPIWATCH. ProMED Mail is largely qualitative and has been at the forefront of early detection for decades.<sup>3</sup> HealthMap is not specific to epidemics but is responsive to emergencies—for example, providing a monkeypox dashboard in 2022.<sup>5</sup> EPIWATCH was developed at UNSW, Sydney from 2016 onward following extensive consultation with Australian and regional stakeholders involved in epidemic response.<sup>6</sup> EPIWATCH provides epidemic intelligence in a quantitative format with free public access.<sup>6</sup> This system offers an open-access, public dashboard accessed through the EPIWATCH public website with a sortable, searchable, filterable global map and table of epidemics with data visualization options limited to 30 days of data. The internal application contains full data access, dashboards with relevant visualizations (for mapping, statistical analytics, searching, or decision support), a user administration panel, and searchable access to the database. EPIWATCH is supplemented by the development of tools



such as a seasonal influenza forecasting tool (FLUCAST),<sup>7</sup> an epidemic risk analysis tool (EPIRISK),<sup>8</sup> and a tool for determining the origins of epidemics,<sup>9</sup> all of which have been tested through rigorous research.

We have tested the timeliness of EPIWATCH outbreak alerts and found that over 60% of public health stakeholders, including staff in health departments and primary care, reported that they were unaware of the outbreaks presented to them from EPIWATCH and that they would value a system like this.<sup>6</sup> We demonstrated that EPIWATCH detected global outbreaks not detected by other surveillance systems.<sup>10</sup> Our research on rapid epidemic intelligence using data algorithms for mining social media showed that the Ebola virus disease epidemic could have been detected in late 2013, months before the WHO was aware of the epidemic.<sup>11</sup> The WHO was notified about the West African Ebola epidemic in March 2014 but responded after a long delay, during which time the epidemic grew exponentially from a few hundred cases to over 28,000 cases.<sup>11</sup> We also showed retrospectively that EPIWATCH could have detected a signal for the early COVID-19 outbreak a month before it was officially reported<sup>12</sup> and identified unknown severe pneumonia in the Hubei province in November 2019.<sup>12</sup> At the time the system was unfunded, so no analysts were able to review the signals in real time. EPIWATCH therefore presents high value in epidemic intelligence collection that has been repeatedly tested, trained, evaluated, and researched. It is not directly comparable with ProMED Mail, which is qualitative in nature, but it has also provided important early warnings from observations of health professionals in the field.<sup>13</sup> There are few formal evaluations of other systems in the peer-reviewed literature. However, EIOS and Blue Dot were found to provide additional intelligence to the usual event-based surveillance used in Japan during the Tokyo Olympics.<sup>13</sup> ProMED and HealthMap have been able to quantify the risk of Ebola retrospectively in the 2013–2016 West African Ebola epidemic spread 1–4 weeks in advance.<sup>5</sup> More research is needed for continual improvement of such systems.

Open-source intelligence can result in unmanageable volumes of data. This makes it difficult for users to know which data they should review to quickly identify outbreaks of interest. EPIWATCH curates data using two AI sub-systems and human analysts. EPIWATCH uses a natural language processing (NLP) entity recognition AI sub-system and the ArcGIS location data API to identify the precise location and adds this and other metadata, such as disease and date, automatically to an outbreak report. It uses a second AI sub-system that automatically classifies articles into four priority quadrants (very high, high, medium, and low) with 88% accuracy. This priority can be double-checked by a human reviewer and can be corrected if deemed incorrect, creating a continuous improvement cycle. Due to the high-consequence nature of pandemics, EPIWATCH is tuned toward high sensitivity, which requires some degree of “over catch” of false positives, to reduce likelihood of missing valid signals. EPIWATCH also has clustering algorithms that can group articles from various sources when they contain information about the same outbreak and show how to best represent this in terms of both data storage and presentation.

Relevant articles that are automatically collected, processed, and prioritized by the EPIWATCH AI systems are added to a database for daily human review. A team of expert public health analysts review the data daily, remove irrelevant or duplicate articles, and enter relevant data into the final database. Analysts have a minimum of graduate public health qualifications and are trained to use a standard operating procedure to ensure consistency and quality. The data obtained after human review is kept in an internal database for the article classification and prioritization AI sub-system and helps improve the machine learning. All intelligence is then reviewed weekly at a meeting with analysts and more senior staff, and a weekly digest, EPISCOPE, is published, which summarizes new outbreaks and outbreaks of interest as well as “mystery” or unknown outbreaks.

Another method for more efficient and user-friendly open-source intelligence is automated red flagging (ARF) of epidemics using advanced geographic information system (GIS) methods. Such a tool

could automatically generate red flags on a map for epidemics that are deviating from baseline expected rates to enable rapid identification and response, improve the user experience, and reduce time spent trawling large volumes of data. ARF is based on spatial pattern recognition, cluster analysis, and space-time pattern analysis, and the automated generation of red flags can be designed as an AI sub-system. Spatial pattern recognition evaluates the data points and determines whether the data points are clustered or randomly distributed. Cluster analysis identifies hotspots (i.e., the locations of statistical significance). Pinpointing the locations of geospatial clusters is important as the whereabouts of disease outbreaks can often provide clues about their cause.<sup>14</sup> Space-time pattern analysis identifies the trends in the data points in a space-time cube and characterizes the trends as new hotspots, intensifying hotspots, or diminishing hotspots. These three tools are interdependent as the output of one tool becomes the data input to another and can add value to rapid epidemic intelligence.

The addition of risk analysis tools to open-source intelligence can add even more value for users. Toward this aim, we developed a suite of risk-prediction tools using input from stakeholders to provide optimal decision support for governments, private enterprises, and other stakeholders. Some of these tools are available on the EPIWATCH website, and some are in the internal dashboard. In the future, we will make them available as interactive tools in a live Decision Theater to provide real-time decision support for stakeholders.

For widespread uptake in public health, digital surveillance should be made easily available to public health stakeholders along with public health tools that are of value to these stakeholders. Finally, to ensure maximum impact of digital AI-based technology in public health, these tools need to be integrated into training of the public health workforce and into routine public health practice. The goal of preventing the next pandemic can never be achieved while open-source intelligence is a niche tool that is rarely used in daily public health practice or is only available to paying clients.

Translation of digital tools into routine practice is essential. Field epidemiology training programs are an ideal starting point for developing digital literacy among public health trainees. Epidemic investigation as an organized capability was pioneered in the United States when the Epidemic Intelligence Service (EIS) was developed by Alexander Langmuir in 1951.<sup>15</sup> The Center for Disease Control (CDC) EIS is an elite training program that sees trainees have short periods of technical learning in the classroom but spend most of their time in the field investigating outbreaks and applying that learning to real problems. The CDC EIS also spawned a global network of field epidemiology training programs, TEPHINET. Trainees learn the science of epidemic investigation and control. EPIWATCH is already working with selected field epidemiology training programs in digital surveillance methods to enhance capacity in rapid epidemic intelligence. Beyond this, public health training curricula should include digital surveillance and open-source intelligence to ensure public health workforces are familiar with the range of available tools for epidemic detection and prevention. This is as important as drugs and vaccines—enabling rapid response by governments and health systems will help to prevent or mitigate the health and economic impacts of serious epidemics and pandemics.

Emerging infections are a continual threat to our health and security, with an acceleration of serious epidemics in the last decade. The time is therefore ripe for utilizing rapid epidemic intelligence methods and vast open-source data to enable earlier detection of epidemics.

#### ACKNOWLEDGMENTS

The authors would like to thank the Department of Health of the Australian Government for funding

this research through the Medical Research Future Fund (MRFF) 2021 Frontier Health and Medical Research Grant (ID RFRHPI000280).

#### AUTHOR CONTRIBUTIONS

C.R.M. - conception and design of the study, manuscript drafting, critical revision of the article; S.L. - manuscript drafting and revision; and A.Q. - grammar checking, critical manuscript revision, manuscript submission.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

#### REFERENCES

1. MacIntyre, C.R., and Bui, C.M. (2017). Pandemics, public health emergencies and antimicrobial resistance - putting the threat in an epidemiologic and risk analysis context. *Arch Public Health* 75, 54. <https://doi.org/10.1186/s13690-017-0223-7>.
2. Bruns R, Teran N. Weighing the Cost of the Pandemic. Knowing what We Know Now, How Much Damage Did COVID-19 Cause in the United States? Institute for Progress. Published online 2022.
3. Rolland, C., Lazarus, C., Giese, C., Monate, B., Travert, A.S., and Salomon, J. (2020). Early detection of public health emergencies of International concern through Undiagnosed disease reports in ProMED-Mail. *Emerg. Infect. Dis.* 26, 336–339. <https://doi.org/10.3201/eid2602.191043>.
4. Muscatello, D.J., Chughtai, A.A., Heywood, A., Gardner, L.M., Heslop, D.J., and MacIntyre, C.R. (2017). Translation of real-time infectious disease Modeling into routine public health practice. *Emerg. Infect. Dis.* 23, e161720. <https://doi.org/10.3201/eid2305.161720>.
5. Bhatia, S., Lassmann, B., Cohn, E., Desai, A.N., Carrion, M., Kraemer, M.U.G., Herringer, M., Brownstein, J., Madoff, L., Cori, A., and Nouvellet, P. (2021). Using digital surveillance tools for near real-time mapping of the risk of infectious disease spread. *NPJ Digit Med* 4, 73. <https://doi.org/10.1038/s41746-021-00442-3>.
6. Hii, A., Chughtai, A.A., Housen, T., Saketa, S., Kunasekaran, M.P., Sulaiman, F., Yanti, N.S., and MacIntyre, C.R. (2018). Epidemic intelligence needs of stakeholders in the Asia-Pacific region. *Western Pac Surveill Response J* 9, 28–36. <https://doi.org/10.5365/wpsar.2018.9.2.009>.
7. Moa, A., Muscatello, D., Chughtai, A., Chen, X., and MacIntyre, C.R. (2019). Flucast: a real-time tool to Predict Severity of an influenza Season. *JMIR Public Health Surveill* 5, e11780. <https://doi.org/10.2196/11780>.
8. Lesmanawati, D.A.S., Veenstra, P., Moa, A., Adam, D.C., and MacIntyre, C.R. (2020). A rapid risk analysis tool to prioritise response to infectious disease outbreaks. *BMJ Glob Health* 5, e002327. <https://doi.org/10.1136/bmjgh-2020-002327>.
9. Chen, X., Chughtai, A.A., and MacIntyre, C.R. (2020). Application of a risk analysis tool to Middle East Respiratory Syndrome Coronavirus (MERS-CoV) outbreak in Saudi Arabia. *Risk Anal.* 40, 915–925. <https://doi.org/10.1111/risa.13472>.
10. Puca, C., and Trent, M. (2020). Using the surveillance tool EpiWATCH to rapidly Detect global Mumps outbreaks. *Global Biosecurity* 1. <https://doi.org/10.31646/gbio.54>.
11. Ajisegiri, W.S., Chughtai, A.A., and MacIntyre, C.R. (2018). A risk analysis Approach to Prioritizing epidemics: Ebola virus disease in west Africa as a case study. *Risk Anal.* 38, 429–441. <https://doi.org/10.1111/risa.12876>.
12. Kpozehouen, E.B., Chen, X., Zhu, M., and MacIntyre, C.R. (2020). Using open-source intelligence to Detect early signals of COVID-19 in China: Descriptive study. *JMIR Public Health Surveill* 6, e18939. <https://doi.org/10.2196/18939>.
13. Kasamatsu, A., Ota, M., Shimada, T., Fukusumi, M., Yamagishi, T., Samuel, A., Nakashita, M., Ukai, T., Kurosawa, K., Urakawa, M., et al. (2021). Enhanced event-based surveillance for imported diseases during the Tokyo 2020 Olympic and Paralympic Games. *Western Pac Surveill Response J* 12, 13–19. <https://doi.org/10.5365/wpsar.2021.12.4.903>.
14. Quigley, A.L., Nguyen, P.Y., Stone, H., Lim, S., and MacIntyre, C.R. (2020). Cruise Ship Travel and the spread of COVID-19 – Australia as a case study. *Int J Travel Med Glob Health* 9, 10–18. <https://doi.org/10.34172/ijtmgh.2021.03>.
15. Schultz, M.G., and Schaffner, W. (2015). Alexander Duncan Langmuir. *Emerg. Infect. Dis.* 21, 1635–1637. <https://doi.org/10.3201/eid2109.141445>.