

The Role of Statistical Modeling in Predicting Disease Outbreaks: Applications and Challenges

Henry Ahumaraezemma Ogu ^{1,*}, Francis Ssenabulya Ssemujju ² and Eunice Abena Lettu ³

¹ Department of Mathematics and Statistics, North Carolina Agricultural and Technical State University, USA.

² Washington University in St. Louis, USA.

³ Department of Mathematics, Kwame Nkrumah University of Science and Technology, Ghana.

Magna Scientia Advanced Research and Reviews, 2026, 16(01), 152-158

Publication history: Received on 10 January 2026; revised on 15 February 2026; accepted on 18 February 2026

Article DOI: <https://doi.org/10.30574/msarr.2026.16.1.0025>

Abstract

This paper reviews the essential role of statistical modeling in predicting infectious disease outbreaks, highlighting the spectrum of methodologies to advanced machine learning techniques. It navigates how integrating multiple data sources improves outbreak forecasting accuracy and timeliness. The paramount applications in major epidemics such as COVID-19, Ebola, Influenza, and malaria illustrate model strengths and practical challenges. The paper also discusses crucial processes like model calibration and validation to improve reliability, as well as addressing challenges such as data quality, model interpretability, ethical concerns, and equity in access. This comprehensive review reinforces the importance of interdisciplinary collaboration, continuous technological innovation, and ethical frameworks in advancing data-driven public health strategies to better anticipate and mitigate future outbreaks.

Keywords: Infectious diseases; Public health; Epidemiology; Statistical models

1. Introduction

In recent years, the world has faced unpredicted and unprecedented outbreaks with the emergence and re-emergence of infectious diseases. The increasing recurrence and impact of these disease outbreaks demonstrate the crucial need for proactive measures in predicting and preventing diseases, and the integration of data-driven methods has significantly enhanced this pursuit. The complexity of infectious disease dynamics, influenced by factors ranging from human behavior and environmental conditions to microbial evolution, makes it necessary for a multidimensional approach to prediction. This accentuates the significance of leveraging data analytics, which harnesses the power of vast datasets and computational methodologies to discern patterns, trends, and early indicators of potential outbreaks [1], unlike traditional methods, which often fall short in capturing the intricate interplay of variables that contribute to outbreaks.

Statistical modeling is the method of using mathematical models that embody a set of statistical assumptions concerning the generation of sample data (and similar data from a larger population). It can also be defined as a mathematical relationship between one or more random variables and other non-random variables, which serves as a great tool for predicting diseases per the number of variables that may or may not lead to an outbreak due to its complexity. [2] With the emergence of machine learning and data analytics, a much better and cheaper alternative, there is a higher chance of predicting and preventing the outbreak of diseases with greater accuracy in this century than in previous decades.

* Corresponding author: Henry Ahumaraezemma Ogu

The first objective of this review is to present a view of the diverse models and tools used in the prediction of disease outbreaks. From statistical models to machine learning and sophisticated epidemiological frameworks, this overview encapsulates the spectrum of methodologies that data analytics encompasses. This research aims to give scientists, researchers, and practitioners insight into the limitations and effectiveness of current approaches in disease outbreak prediction. Real-life cases are also used in this study to show the practicality of utilizing the existing models, their strengths, and challenges for improvement. This study also lays bare the undeniable strength in technology, which contributes to shaping the agenda for future research and development in the realm of data analytics for public health. [3]

2. Theoretical and Methodological Foundations

Statistical and mathematical modeling are rudimentary to modern epidemiology, enabling scientists to analyze health data to understand disease patterns, predict outcomes, and test hypotheses across populations. These modeling techniques provide quantitative frameworks that couple biological mechanisms of infection with population-level outcomes and public health interventions.

2.1. Statistical Modeling in Epidemiology

Statistical models depend on data to discover patterns, correlations, and probabilistic relationships between variables that may be influencing disease spread. They are mainly data-driven, aiming to estimate how risk factors, demographic conditions, and interventions are impacting infection outcomes. Some of these techniques, such as regression and time-series models, have become important analytical tools and are contributing enormously to the field of epidemiology. [4]

2.2. Mathematical Modeling in Epidemiology

Mathematical epidemiology provides causal descriptions of how diseases propagate through populations, often formulating systems of equations that capture infection dynamics. These models, such as compartmental models (SIR, SEIR, SIS), are deterministic, and, on a more modern side, stochastic models describe causal mechanisms rather than purely statistical associations. The key quantity in these models is the basic reproduction number (The expected number of secondary infections produced by a single case in a susceptible population. When > 1 , the disease tends to spread; when < 1 , it typically dies out.[5]

3. Statistical and Mathematical Time Series Methods

One prevalent method for predicting and identifying epidemics or the spread of diseases is time series analysis. A time series is a sequence of numerical data points in successive order. Time series have been used to track data trends, such as stock prices and the number of infected people over a specific period. Many countries conducted infectious disease surveillance in order to detect epidemics at an early stage.

As the field of epidemiology continues to evolve, researchers and practitioners are exploring innovative approaches and technologies to improve the accuracy, timeliness, and usability of predictive models. The integration of deep learning techniques, a subset of machine learning, has improved and keeps improving pattern recognition in complex datasets. Deep neural networks can automatically learn intricate relationships with epidemiological, environmental, and social data, potentially uncovering hidden patterns that traditional statistical and mathematical models might miss. [6]

Machine learning is one of the most advanced concepts of artificial intelligence and provides a strategic approach to developing automated, complex, and objective algorithm techniques for multimodal and dimensional biomedical or data analysis. [7]. The ML algorithms can read and modify their structure based on a set of observed data, with adaptation done by optimizing over a cost function or an objective. [8-9] ML has already proven its ability for diagnosing, detecting, containing, and therapeutic monitoring of many diseases and its potential outbreaks. [10-12]

Also, automatic methods for disease outbreak detection and prediction have proven effective in the early detection of naturally occurring outbreaks or those related to bioterrorism. These methods aim to minimize the time between the appearance of strains and the identification of the outbreak. This reduced time gap allows for more efficient investigation and intervention in controlling the disease. [13]

Model calibration is crucial for optimizing the performance of complex models across disciplines. In this era of change, symbolizing rapid technological advancement through the integration of advanced digital technologies into industrial processes, model calibration plays a key role in advancing digital twin technology, ensuring alignment between digital

representations and real-world systems. Model calibration, a fundamental process in computational science and engineering, involves the adjustment of model parameters to optimize the model's performance and enhance its predictive capabilities. These model parameters are often referred to as calibration parameters. [14]

The continuous improvement in computational power and data collection is leading to the development of increasingly complex models, which integrate multiple coupled physical and engineering processes to achieve a better representation of the modeled system. Most of the parameters included in models are difficult to measure directly, so they must be estimated from collected data through calibration procedures like Bayesian inference, maximum likelihood estimation, and variational data assimilation. Furthermore, when models are used to make forecasts about future or hypothetical scenarios, it is important to bound the uncertainty of their results. Therefore, the application of systematic approaches for parameter estimation to integrate data and models and quantify potential errors is more necessary now since well-estimated parameters reduce structural uncertainty and ensure that the model captures both mechanistic fidelity and dynamic variability present in complex systems. [15]

4. Validation of predictive accuracy

Validation is the assessment that the model reliably links the measurements and the predictive variables. [16] Model validation tests a calibrated model's capability to make accurate and precise predictions for new, unseen data or different operating conditions. Predictive algorithms and multivariable models must be validated in a suitable and comprehensive way in order to establish the robustness of the model in terms of accuracy, predictive ability, reliability, and generalizability. These models must be validated in order not to predict incorrectly. [17]

Data Sources and Applications of Statistical Modeling

Machine learning and related technologies are already supporting human decision-making in various fields like economics, medicine, and social science, and thus have the potential to transform the scope and power of infectious disease epidemiology. ML methods can be applied to routinely collected disease surveillance data. [18]

4.1. Data Sources

Data can be collected from various sources. The types of data used for modeling are traditional, which include case counts and surveillance, and emerging data sources like social media, climate, genomics, and mobility.

Traditional Data Sources consist of case counts and incidence rates, which are collected from hospital records, lab-confirmed cases, and mortality data. Traditional data may include public health surveillance systems like syndromic surveillance, sentinel sites, World Health Organization, and the Centers for Disease Control and Prevention reports. Also, demographic and census data like population density, age distribution, and healthcare access all fall under traditional data sources. With the advancement of technology and Internet of Things (IoT), there are emerging data sources such as social media and web data, which include platforms like Twitter, Google Trends, Facebook health searches, i.e., early signals of outbreaks. Secondly, the climate and environmental data, that are temperature, rainfall, and humidity, determine the outbreak of diseases such as malaria, cholera, dengue, and flu during winter. Furthermore, genomic and molecular data were used during the COVID-19 pandemic using pathogen sequencing for tracking mutations of the SARS-CoV-2 variants much more quickly. Lastly, the use of human mobility data through airline travel, mobile phone geolocation, migration used to determine the spread prediction as well as IoTs and wearable health devices gives data like real-time vitals and remote monitoring.

5. Statistical Applications

5.1. Covid-19 (2020-2023)

COVID-19 has been considered the most significant threat since World War II and the greatest global health disaster of the century. Wuhan City, Hubei Province, China, reported 27 cases of novel pneumonia on December 31, 2019. The cause and origin of the novel pneumonia were unknown. The Coronavirus Disease 2019 (COVID-19) has been named by the World Health Organization (WHO). Across the globe, it was spreading rapidly, posing significant health, economic, and social challenges for everyone. [19] Sooner or later, this novel disease had infected more than 550 million people. Confirmed cases per an article published by the BBC [20] in July 2022 and the WHO in February 2024, released that close to 21 million people had died globally from the direct or indirect effects of COVID-19, with majority dying in the first two years of the pandemic.

Many country-level and global-level institutions used statistical modeling, such as time-series forecasting methods, for COVID-19 predictions. [21] Using differencing and scaling, the statistical time-series method modeled future values based on past values and errors. Also, the Facebook Prophet is an open-source forecasting tool developed by Facebook that models time-series data by combining variables such as trend, seasonality, and holiday effects. The adaptability of the Prophet improved predictability by being flexible, accommodating irregular observations, and fitting curves quickly. [22]

Throughout the pandemic, researchers also used mathematical models to analyze COVID-19 data for a better understanding of transmission patterns, monitoring disease severity, anticipating future epidemic outcomes [23], and justifying the adoption of intervention measures. [24] Among these mathematical models, compartmental models describing the disease as a sequence of different stages encountered upon infection to recovery, such as the *SEIR* (Susceptible-Exposed-Infectious-Recovered) model, have been generally adopted to forecast or simulate future transmission trajectories. [25-26]

Compartmental models like SIR, SEIR, or their variants have been widely used to reenact the transmission and forecast cases, hospitalizations, and deaths. The epidemiological compartmental models capture the dynamics of disease spread and prediction. Real-time challenges were underreporting, rapidly evolving variants, and policy changes affecting model accuracy due to changing variables.

5.2. Ebola (West Africa, 2014-2016; DRC, 2018-2020)

The Ebola virus disease, formerly known as Ebola hemorrhagic fever, is a severe and often fatal multifaceted infectious zoonosis in humans. The 2013-2016 West African Ebola outbreak had distinctive characteristics, and it was the largest and most complex epidemic since the virus discovery in 1976. Although the 2018-2020 Ebola outbreak in the Democratic Republic of the Congo had many similarities, there were additional challenges due to the presence of armed rebel groups at the epicenters of the epidemic. [27] The outbreak of Ebola that started in West Africa in late 2013 has caused at least 28,000 illnesses and 11,000 deaths. As the outbreak progressed, global and local public health authorities scrambled to contain the spread of the disease by isolating those who were ill, putting in place infection control processes in health care settings, and encouraging the public to take steps to prevent the spread of the illness in the community. It took a massive investment of resources and personnel from many countries to eventually bring the outbreak under control.

Various modeling approaches were and can be used in modeling Ebola. From this paper [28], of the 125 models reported across the studies, 74% included mechanistic assumptions about disease transmission (e.g., compartmental, agent-based, or phylogenetic models), while 26% were purely phenomenological. These models used estimates of the generation time, case fatality rate, or other epidemiological parameters as model inputs. Some challenges were limited surveillance and cultural burial practices, which further complicated data collection. [29] Therefore, an additional parameter to further improve accuracy was to incorporate socio-cultural factors and mobility patterns into predictive frameworks. [30]

5.3. Influenza (Seasonal & Pandemic Flu)

Influenza is an acute viral respiratory infection that affects all age groups and is associated with high mortality during pandemics, epidemics, and sporadic outbreaks. Nearly 10% of the world's population is affected by influenza annually, with about half a million deaths each year. [31] Influenza predictive models use the approach of combining mechanistic modeling that is SIR, with statistical learning from historical data, using Bayesian filter-like particle and Kalman filter variants for inference and forecasting of parameter estimation to improve forecast accuracy and peak predictions over standard methods. [32]

6. Malaria

Malaria is named after the Italian term "mal'aria", which means "bad air" to represent the association of the disease with marshy areas. Malaria is a global public health burden with an estimated 229 million cases reported worldwide in 2019. About 94% of reported cases were recorded in the African region. The Southeast Asia region accounts for about 3% of the burden of malaria cases globally. In the same year, an estimated 409 000 deaths from malaria occurred worldwide, of which 94% happened in Africa. About 200 different species of protozoa have been identified so far, and among them, at least 13 species are known to be pathogenic to humans. The life cycle of the malaria parasite is a complex process comprising an Anopheles mosquito and a vertebrate host. [33] Mathematical modeling for malaria transmission started from the basic Ross model and has now progressed to more complex compartmental models (Macdonald and Anderson-May) that incorporate more variables, such as the latency in mosquitoes and humans, age, immunity, and environments

like rainfall and humidity. The primary modeling technique is deterministic differential equation-based epidemiological compartment models applied to both human and mosquito populations. Variations may include parameters like mosquito-to-human ratios. Biting rates, recovery rates, and latency periods describe malaria dynamics, estimate transmission intensity, and assess intervention strategies. [34]

7. Challenges and limitations

Though disease outbreak prediction has significant strides, especially with the integration of diverse data sources, it faces ongoing challenges and limitations that require attention to maximize its impact on public health. In refining and fine-tuning models to improve the robustness of predictions, the models rely heavily, if not entirely, on input data. Therefore, incomplete or inaccurate data can lead to biased predictions and hinder the effectiveness of early warning systems as well as reduce reliability. Imputation or Bayesian priors are needed to handle missing or noisy data, especially in the instance of Ebola data gaps; smoothing techniques were heavily relied on. [35] In addressing data quality, the processing of data has to be cross-checked. Every data collected must go through validation and integration processes to ensure its quality. Also, timely access to data is critical for effective disease outbreak prediction. Relying wholly on case counts limits the predictive strength of the models. Combining traditional surveillance with IoT, climate, or genomic data enhances robustness. Influenza models improved significantly when mobility and weather data were integrated. [36]

The transparency and interpretability of models matter. Black-box predictions hinder trust and adoption by policymakers. Models that are interpretable gain more acceptance than opaque machine learning models, especially in relation to public health. In the instance of simple regression-based forecasts were more widely adopted than low-resource malaria prediction than complex ML approaches. [37]. When models are developed without decision-making processes, even the most accurate models will be limited. Co-designing with public health agencies ensures models are actionable. During the COVID-19 pandemic, forecasting agencies such as the U.S. COVID-19 Forecast Hub provided ensemble models tailored for policymakers. [38]

The integration of various datasets from different sources raises concerns about privacy and security. It is important to ensure the responsible handling of sensitive health and demographic data. Anonymizing data, obtaining consent, and adhering to privacy regulations are crucial steps to safeguarding individual privacy. [39] The deployment of predictive technologies should prioritize equitable access, ensuring that benefits are distributed across diverse populations. Avoiding biases in data collection and model training, and addressing disparities in healthcare infrastructure, is crucial for preventing the exacerbation of existing health inequities. As disease outbreak prediction continues to evolve, addressing these challenges and considerations is imperative for advancing the field responsibly. Ongoing research, collaborative initiatives, and a commitment to ethical practices are essential components in navigating the complexities of predicting and mitigating the impact of infectious diseases on global health.

8. Conclusion

The journey to predicting disease outbreaks is, at its core, about protecting people, families, communities, and entire nations from the uncertainty and disruption that infectious threats bring. This review has illuminated the many ways humanity is rising to meet these challenges, blending scientific ingenuity with real concern for human well-being. By bringing together skilled minds from different disciplines and harnessing the latest technologies in prediction, societies have found new hope in anticipating and responding to epidemics.

As emerging analytical instruments like artificial intelligence and genomic monitoring broaden our insights, the world moves ever closer to forecasts that are not only rapid but empathetic, giving healthcare workers precious lead time and aiding decision-makers in crafting prudent policies. Nevertheless, forecasting the unknown remains a mission filled with intricacy: diseases propagate across tangled networks of people and animals, shaped by surroundings that can shift as swiftly as human habits.

Meeting these challenges calls for collective effort and lasting commitment from building systems that include everyone, to designing ethical policies that honor each person's privacy and dignity. It means facing the future with transparency, fairness, and open dialogue, never losing sight of the individuals whose lives are at the heart of each statistic.

Looking ahead, success will depend on learning together, adapting quickly, and fostering trust. The lessons of recent pandemics show that when researchers, healthcare providers, governments, and citizens join forces, the outcome is not just a more resilient public health system, but a more caring and united world. The greatest promise of predictive

analytics is not only technical progress, but the chance to protect, uplift, and empower communities everywhere, ensuring that no one faces an outbreak alone, and that the benefits of modern science truly reach all.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Salathe, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., ... & Vespignani, A. (2012). Digital epidemiology.
- [2] Wikipedia definition
- [3] Khan, N., Bano, A., & Zandi, P. (2018). Effects of exogenously applied plant growth regulators in combination with PGPR on the physiology and root growth of chickpea (*Cicer arietinum*) and their role in drought tolerance. *Journal of plant interactions*, 13(1), 239-247
- [4] Banerjee, S. K., & Andrews, B. (2023). Mathematical Analysis of Regression model Epidemiology. *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, 93(1), 39-49.
- [5] Kretzschmar, M., & Wallinga, J. (2009). Mathematical models in infectious disease epidemiology. In *Modern infectious disease epidemiology: Concepts, methods, mathematical models, and public health* (pp. 209-221). New York, NY: Springer New York.
- [6] Rahmani, A. M., Gia, T. N., Negash, B., Anzanpour, A., Azimi, I., Jiang, M., & Liljeberg, P. (2019). Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach. *Future Generation Computer Systems*, 98, 128-144.
- [7] Paul S. Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng*. 2006;8(1):537-65.
- [8] Jebara T. Machine learning: discriminative and generative. Norwell: Springer; 2003.
- [9] Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;7(369):m1328. <https://doi.org/10.1136/bmj.m1328>.
- [10] Haruna AA, Muhammad LJ, Yahaya BZ, et al. (2019) An improved C4.5 data mining driven algorithm for the diagnosis of coronary artery disease. In: International Conference on Digitization (ICD), Sharjah, United Arab Emirates, pp 48-52.
- [11] Hussain S et al. (2019) Performance evaluation of various data mining algorithms on road traffic accident dataset. In: Satapathy S, Joshi A. (eds) Information and communication technology for intelligent systems. Smart innovation, systems and technologies, 106
- [12] Ishaq FS, Muhammad LJ, Yahaya BZ, et al. Fuzzy based expert system for diagnosis of diabetes mellitus. *Int J Adv Sci Technol*. 2020;136:39-50.
- [13] Babanejaddehaki, G., An, A., & Papagelis, M. (2025). Disease outbreak detection and forecasting: A review of methods and data sources. *ACM Transactions on Computing for Healthcare*, 6(2), 1-40.
- [14] Sung, C. L., & Tuo, R. (2024). A review on computer model calibration. *Wiley Interdisciplinary Reviews: Computational Statistics*, 16(1), e1645.
- [15] Herrera, P. A., Marazuela, M. A., & Hofmann, T. (2022). Parameter estimation and uncertainty analysis in hydrological modeling. *Wiley Interdisciplinary Reviews: Water*, 9(1), e1569.
- [16] Lopez, E., Etxebarria-Elezgarai, J., Amigo, J. M., & Seifert, A. (2023). The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples. *Analytica Chimica Acta*, 1275, 341532.
- [17] Staffa, S. J., & Zurakowski, D. (2021). Statistical development and validation of clinical prediction models. *Anesthesiology*, 135(3), 396-405.
- [18] Kraemer, M. U., Tsui, J. L. H., Chang, S. Y., Lytras, S., Khurana, M. P., Vanderslott, S., ... & Bhatt, S. (2025). Artificial intelligence for modelling infectious disease epidemics. *Nature*, 638(8051), 623-635

- [19] Naseer, S., Khalid, S., Parveen, S., Abbass, K., Song, H., & Achim, M. V. (2023). COVID-19 outbreak: Impact on global economy. *Frontiers in public health*, 10, 1009393.
- [20] <https://www.bbc.com/news/world-51235105>
- [21] Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Caution warranted: Using the Institute for Health Metrics and Evaluation model for predicting the course of the COVID-19 pandemic. *Annals of Internal Medicine*, 173(3), 226–227
- [22] Kumar, N., & Susan, S. (2020, July). COVID-19 pandemic prediction using time series forecasting models. In 2020 11th international conference on computing, communication and networking technologies (ICCCNT) (pp. 1-7). IEEE.
- [23] Nixon K, Jindal S, Parker F, Marshall M, Reich NG, et al. (2022) Real-time COVID-19 forecasting: challenges and opportunities of model performance and translation. *Lancet Digit Health* 4: e699–e701. pmid:36150779
- [24] Adam D (2020) Special report: The simulations driving the world’s response to COVID-19. *Nature* 580: 316–318. pmid:32242115
- [25] Holmdahl I, Buckee C (2020) Wrong but Useful—What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *N Engl J Med* 383: 303–305. pmid:32412711
- [26] Wang P, Zheng X, Liu H (2022) Simulation and forecasting models of COVID-19 taking into account spatio-temporal dynamic characteristics: A review. *Front Public Health* 10: 1033432. pmid:36330112
- [27] Guetiya Wadoun, R. E., Sevalie, S., Minutolo, A., Clarke, A., Russo, G., Colizzi, V., ... & Montesano, C. (2021). The 2018–2020 Ebola outbreak in the Democratic Republic of Congo: a better response had been achieved through inter-state coordination in Africa. *Risk management and healthcare policy*, 4923-4930.
- [28] Chretien, J. P., Riley, S., & George, D. B. (2015). Mathematical modeling of the West Africa Ebola epidemic. *Elife*, 4, e09186.
- [29] Chowell, G., & Viboud, C. (2016). Modeling the spatial and temporal spread of Ebola virus disease: A review. *BMC Medicine*, 14(1), 210.
- [30] Camacho, A., et al. (2015). Potential for large outbreaks of Ebola virus disease. *Epidemics*, 9, 70–78.
- [31] Javanian, M., Barary, M., Ghebrehewet, S., Koppolu, V., Vasigala, V., & Ebrahimpour, S. (2021). A brief review of influenza virus infection. *Journal of medical virology*, 93(8), 4638-4646.
- [32] Andronico, A., Paireau, J., & Cauchemez, S. (2024). Integrating information from historical data into mechanistic models for influenza forecasting. *PLOS Computational Biology*, 20(10), e1012523.
- [33] Fikadu, M., & Ashenafi, E. (2023). Malaria: an overview. *Infection and Drug Resistance*, 3339-3347.
- [34] Mandal, S., Sarkar, R. R., & Sinha, S. (2011). Mathematical models of malaria-a review. *Malaria journal*, 10(1), 202.
- [35] Chowell, G., & Viboud, C. (2016). Modeling the spatial and temporal spread of Ebola virus disease: A review. *BMC Medicine*, 14(1), 210. <https://doi.org/10.1186/s12916-016-0724-8>
- [36] Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50), 20425–20430. <https://doi.org/10.1073/pnas.1208772109>
- [37] Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50), 20425–20430. <https://doi.org/10.1073/pnas.1208772109>
- [38] Ray, E. L., Wattanachit, N., Niemi, J., et al. (2020). Ensemble forecasts of COVID-19 in the U.S. *medRxiv*. <https://doi.org/10.1101/2020.08.19.20177493>
- [39] Gronvall, G. K., Trent, D., Borio, L., Bouri, N., Boyce, R., Larson, H., ... & Usenza, C. (2013). Developing a framework for evaluating ethical outcomes of a public health surveillance system. *Health Security*, 11(1), 4-9.