Sarcouncil Journal of Internal Medicine and Public Health

ISSN(Online): 2945-3674

Volume- 05| Issue- 01| 2026

Review Article **Received:** 15-01-2026| **Accepted:** 10-02-2026 | **Published:** 24-02-2026

Integrated Epidemiologic Modeling for Next-Generation U.S. Infectious Disease Surveillance Systems: A Narrative Review

***Merrera Kebeba*[1] *and Emmanuel Amoako Agyei*[2]**

[1]*Santa Clara County Department of Public Health, USA*
[2]*Washington University in St. Louis, USA*

Abstract: Background: The surveillance system for infectious diseases in the United States relies on a complex system that utilizes case reporting systems that are limited by reporting delay, data infrastructure, and population coverage. The recent outbreak of the COVID-19 pandemic has underlined the importance of timeliness, sensitivity, and interoperability in the existing architectures for surveillance. **Objective:** This narrative review examines how next-generation infectious disease surveillance systems in the United States can be improved by incorporating heterogeneous data sources with advanced epidemiologic models. **Approach:** We discuss the integration of traditional case surveillance, syndromic surveillance, wastewater epidemiology, genomic epidemiology, and digital data sources into mechanistic, machine learning, and hybrid models, focusing on real-time assimilation, nowcasting in the presence of reporting delay, interoperability, governance, and privacy-preserving analytics. Our aim is to provide an updated literature synthesis from 2020 to present. **Findings:** Integrated modeling frameworks that infer epidemic dynamics from multiple data streams always improve timeliness and forecasting compared to single data streams. Despite this, there remain several challenges in data standardization, capacity, and governance that have limited operational use of such frameworks. **Conclusion:** To improve infectious disease surveillance in the United States, there is a need for coordinated investments in interoperable infrastructure, integrated epidemiologic models, and adaptive governance frameworks to facilitate effective public health decision-making.

Keywords: Infectious disease surveillance; integrated epidemiologic modeling; real-time data assimilation.

INTRODUCTION

Surveillance of infectious diseases as a systematically organized public health activity has a long history of development since its institutionalization in the early twentieth century(Li *et al.*, 2024), Ssemujju & Solomon, 2025). Although the underlying principle, namely unceasing gathering, examination, and dissemination of data on the prevalence of diseases, have stayed the same, technological and analytical instruments that facilitate surveillance have changed radically(Shen *et al.*, 2025). The modern surveillance in the United States is a multi-layered ecosystem of federal, state, and local systems with the wide variety of structural, interoperability and analytical sophistication (Ugwu *et al*, 2025). Key among them is the National Notifiable Diseases Surveillance System (NNDSS) whereby state health departments report on reportable diseases to the Centers for Disease Control and Prevention which received around 2.7 million reports of cases each year in 2013(CDC, 2025). Syndromic surveillance systems have been developed to complement the traditional case-based surveillance to enhance the timeliness in detecting an outbreak(Romano *et al.*, 2022). These systems would normally detect new threats a month or two before they are reported by the laboratory (Baum *et al.*, 2025). The U.S. surveillance is still largely disintegrated, regardless of these advances. Various jurisdictions use separate platforms that have little real-time interoperability, partial data integration and limited analytical potential (Zeba *et al.*, 2025). These weaknesses became more pronounced than ever by the COVID-19 pandemic. The situation awareness was weakened by delays in reporting, unequal standards among jurisdictions and the lack of ability to consolidate the various streams of data, thereby preventing prompt response in public health(McClellan *et al.*, 2025). One of the basic problems is the dissimilarity of the speed of the spread of the pathogen and the time accuracy of traditional surveillance. In the case of quickly spreading respiratory pathogens, test, diagnosis, and reporting delays imply that by the time the information reaches the decision-makers, the dynamics of the spread may have changed. Studies on modeling propose that reduction of the impact of an outbreak by 20-30 percent can be achieved through early outbreak detection, but actual real-time surveillance is not common practice (Kaur & Butt, 2025).

There are several evidence gaps that remain persistent and that restrict the effectiveness of the existing surveillance architectures. Though compartmental models, including Susceptible-Infectious-Recovered (SIR) and Susceptible-

***Corresponding Author:** Merrera Kebeba

DOI- https://doi.org/10.5281/zenodo.18761805

Exposed-Infectious-Recovered (SEIR) frameworks, continue to form the core of the comprehension of the transmission processes and the effects of interventions, they are seldom operationally combined with real-world surveillance data(Tang *et al*., 2020). Simultaneously, new sources of information such as wastewater-based epidemiology (WBE), genomic surveillance, syndromic indicators and digital signals based on social media) tend to exist as parallel systems with little formal integration (Ingelbeen *et al*., 2025). Empirical evidence shows that WBE can reveal SARS-CoV-2 circulation dozens of weeks before clinical cases are even confirmed,(Levy *et al*., 2023) but these indicators are rarely included in regular epidemiologic frameworks or governmental situational awareness data. Equally, machine learning and artificial intelligence techniques have verified potential in detecting and predicting outbreaks,(Ye *et al*., 2025) although these applications have been slowed by apprehensions about both the quality and usefulness of data, as well as by the absence of solid operational validation and a definitive understanding of which methods constitute an established practice and which are experimental research projects(Villanueva-Miranda *et al*., 2025). In addition to technical constraints, there are workforce and infrastructure constraints, which also hinder the modernization of surveillance. Most state and local health departments do not have adequate data scientists, quantitative epidemiologists and informaticians who can deploy and maintain integrated surveillance platforms(Akingbola *et al*., 2025). These diminishing points are augmented by systemic issues, such as the lack of a unique patient identifier nationally, erratic coding standards, privacy and security limitations through Health Insurance Portability and Accountability Act (HIPAA) and organizational silos that restrict real-time sharing of data(Dixon *et al*., 2020). These impediments together make it harder to transform more detailed data streams into public health actionable intelligence.

Notably, none of the surveillance modalities by itself can fully and timely capture the dynamics of infectious diseases. The case data based on laboratory confirmed cases are highly specific but an average of 7-10 days behind the true incidence is witnessed(Chen *et al*., 2010). Syndromic surveillance is useful in monitoring early health seeking behavior, but not diagnostically accurate(Adedire *et al*., 2024). Wastewater-based epidemiology allows detection in the population without regard to use of healthcare services(Giancotti *et al*., 2024), whereas genomic surveillance offers information on the development of pathogens and the emergence of variants but is still resource-intensive and has limited coverage(John *et al*., 2021). Online behavior produces digital indicators that can be used clinically several weeks in advance but are influenced by demographic bias and signal noise (Fallatah & Adekola, 2024). As a result, the combination of complementary data sources is becoming a more important contributor to better sensitivity, timeliness, and predictive accuracy.

Combining epidemiologic modeling systems with mechanistic transmission models coupled with artificial intelligence have specific potential. Hybrid models build on the interpretability and theoretical basis of mechanistic models and the adaptive pattern-recognition capabilities of machine learning(Gijón *et al*., 2025). Models that are concurrently fitted to laboratory-confirmed cases, syndromic indicators, wastewater pathogen loads, and genomic data give numerous constraints, which enhance the estimation of parameters and decrease the uncertainty in their predictions(Chowell & Skums, 2024). The COVID-19 pandemic has shown that the more integrated and real-time surveillance systems a jurisdiction had, the faster they detected outbreaks and subsequently allocated resources more effectively. The CDC has responded with efforts at modernization, such as the Public Health Data Strategy and the One CDC Data Platform (1CDP) that would bring together case, laboratory, and emergency department data into unified infrastructure. Nevertheless, these attempts are still not complete, and the systematic involvement of advanced modeling in decision-making is still not practiced.

Against this backdrop, the present narrative review synthesizes existing evidence on the next-generation U.S. infectious disease surveillance based on integrated epidemiologic modeling methods. It analyses the changing surveillance information environment, considers approaches to multi-source data integration, considers the impediments to interoperability, privacy, and workforce capacity as well as points to new practices with potential to move to more timely, sensitive, equitable and resilient surveillance systems.

METHODS

A narrative review methodology was adopted with an explicate but non-exhaustive search strategy to mitigate selection bias and ensure transparent documentation of evidence sources. Literature was identified through targeted searches PubMed, Scopus, Web of Science and Google Scholar. Search terms were selected iteratively to capture major themes in infectious disease surveillance, data integration, and epidemiologic modeling with emphasis on U.S focused systems and post 2020 literature. Boolean operators (AND/OR) were used to combine terms across thematic categories, and reference lists of key articles were manually screened to identify additional relevant studies. Core search terms included combinations of “infectious disease surveillance,” “syndromic surveillance,” “wastewater-based epidemiology,” “genomic surveillance,” and “digital epidemiology,” alongside modeling-related terms such as “epidemic modeling,” “SEIR models,” “forecasting,” and “data integration.” Additional searches incorporated terms related to governance and implementation, including “data interoperability,” “privacy-preserving analytics,” and “public health surveillance systems.”

Inclusion and Exclusion Criteria

The studies were incorporated in case of the following criteria: (1) peer-reviewed research article or extensive survey that is dealing with the integration of at least one surveillance data stream into analytic models, decision-support tools, or surveillance architectures; (2) explicit reporting of methods, evaluation metrics, or performance results; (3) concern infection disease surveillance. Articles that only studied non-communicable disease surveillance were eliminated without exception unless they had methodological information that was directly applicable to integrated infectious disease surveillance systems.

Peer-reviewed literature was used, but additionally, the grey literature was selected where necessary. This included technical reports and system documentation by the CDC and WHO, and initiatives by the government to modernize data, so long as the sources were describing functioning surveillance systems of operational surveillance of a validated system with documented performance characteristics, details of their implementation, or actual public health uses in the United States or abroad. The studies were eliminated because they met any of the following criteria: (1) opinion articles, editorials, or commentary articles with no empirical, methodological, or operational contribution; (2) purely clinical studies, which do not include any surveillance or population-level implications; (3) non-English articles, as it was necessary to arrange resource limitations; (4) review articles, which are concerned with different areas and do not involve epidemiologic systems or management of surveillance; or (5) review articles, which are not related to the research.

Screening and Study Selection

The titles and abstracts were first screened on a case-by-case basis by two reviewers to filter out studies that were obviously irrelevant, such as animal-only models, laboratory culture studies and weather prediction research, and non-epidemiologic applications. Articles stored at this phase were under a full-text review as they were reviewed to determine their eligibility according to the established inclusion criteria. Included studies lists were screened by hand to include other relevant publications that had not been retrieved by the initial database searches. With several articles on the same surveillance system, analytic tool or modeling approach, priority was assigned to original validation studies, systematic review or comprehensive benchmarking publications which had the most detailed description of their methodology and performance review. Due to the nature of the narrative review design and the heterogeneity of the types of studies, there was no formal risk-of-bias tool. However, more importance was given to the studies that revealed clear-cut data sources, clear and reproducible methodologies, quantitative measures of evaluation, and testing on independent or held-out data.

CURRENT U.S. SURVEILLANCE INFRASTRUCTURE AND DATA SOURCES

Architecture of the National Notifiable Diseases Surveillance System

The National Notifiable Diseases Surveillance System (NNDSS) is a key element of infectious disease monitoring in the United States that provides an interagency system between federal, state, local, territorial, and tribal health organizations (Majumder *et al*., 2023). The system works based on the use of standardized case definitions of nationally notifiable conditions identified by the Council of State and Territorial Epidemiologists (CSTE) as conditions that should be reported in a system approach (Rubis *et al*., 2024). At present, there exist over 140 names, the names of bacterial, viral, parasitic, and fungal pathogens, as well as some non-communicable non-epidemiological conditions that need monitoring (Rainey *et al*., 2024).

The NNDSS data is received in several channels. The primary sources are used, namely, physician and infection preventionist reports, laboratory reports, especially those of the public health laboratories and reportable disease laboratories, and today, electronic laboratory reporting (ELR) directly through the clinical laboratories and hospital information systems (Knicely *et al*., 2024). Information flows via jurisdiction-based surveillance systems to the NEDSS standardized system into which the data is transmitted, which then can be easily compatible and transferred to CDC (Hayman *et al*., 2023). The NEDSS architecture mandates compliance with the national consensus standards, such as the LOINC (Logical Observation Identifiers Names and Codes) of laboratory results, SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) of diagnoses and findings, and HL7 (Health Level Seven) of data exchange messaging standards (Chatterjee *et al*., 2022; Mishra *et al*., 2021; Osamika *et al*., 2021). The major aspect of this infrastructure is that NEDSS Base System (NBS) has been used in 22 health departments in various states and U.S. territories. NBS incorporates data about numerous sources and numerous public health conditions that enable the state and local public health authorities to address the reportable disease data, investigation and transfer of the epidemiologic, laboratory, and clinical data safely to CDC. The system aids in analyzing disease trends over time, finding clusters of cases, and following the results of investigation. Nevertheless, even with standardization frameworks, there are still major challenges in the real-world implementation- most state and local systems still run legacy databases that cannot accept NEDSS standards, and, because of this, still rely on manual data entry, transcription error, and delays in reporting (Zeba *et al*., 2025).

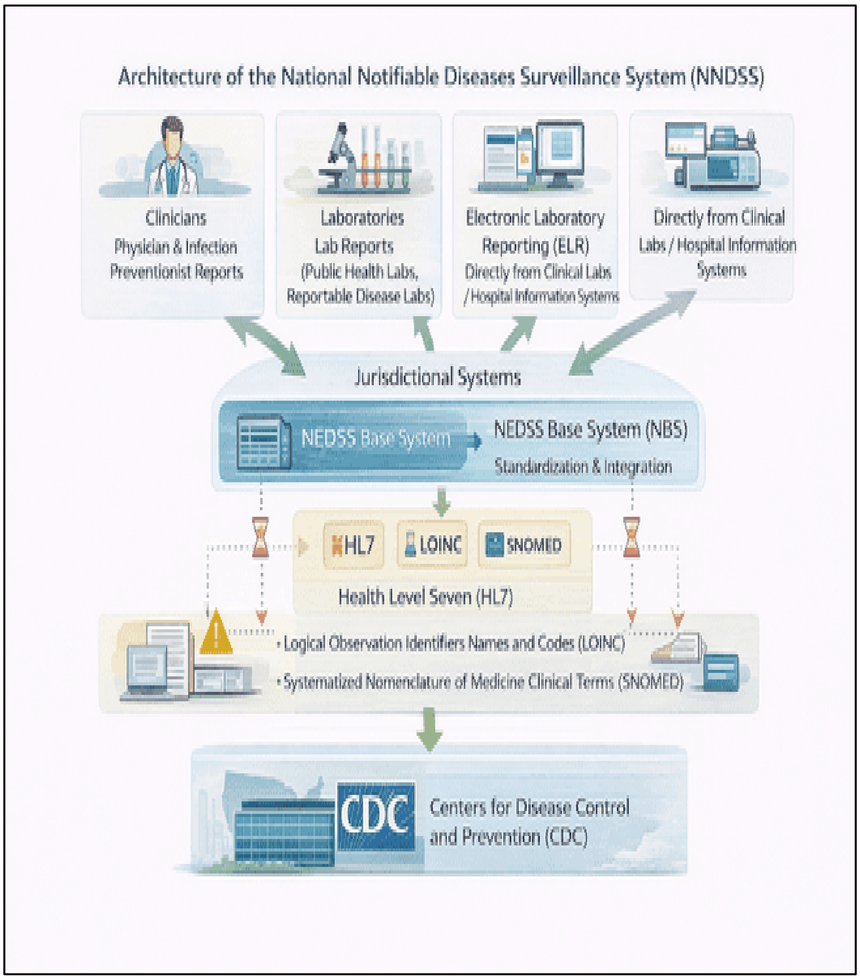

Figure 1: shows the NNDSS data flow from clinicians and laboratories through jurisdictional systems and the NEDSS Base System to the CDC, highlighting standardized data exchange (H7,LOINC, SNOMED) and remaining manual reporting bottlenecks.

Timeliness, Sensitivity, and Limitations of Traditional Surveillance

Delays in reporting and the time-constrained nature of traditional case-based infectious disease surveillance is a major inherent constraint to the existing method, especially when dealing with pathogens that spread rapidly (Kizza *et al*, 2025). Empirical studies during and following the COVID-19 pandemic reveal that confirmed case data are cumulatively vulnerable to cumulative delays caused by symptom onset, health-seeking behavior, laboratory turnaround times, laboratory reporting, and jurisdictional case verification of local and state results in total delays that typically extend to 7-14 days before confirmed cases may be utilized to support surveillance and decision-making (Günther *et al*., 2021; McGough *et al*., 2020). The work comparing U.S. and European surveillance systems indicates that this kind of latency is a massive impairment of real-time situational awareness, which results in a systemic underestimation of present incidence and late identification of epidemic breakout points, particularly at times of exponential growth (Gostic *et al*., 2021; Wu *et al*., 2020). Notably, delays in reporting are not always constant, but time-dependent, changing along with the availability of testing, governmental policy, and system burden along with population behavior, further complicating the interpretation of raw case counts in real-time (Charniga *et al*., 2024; Kretzschmar *et al*., 2020). Methodological studies with nowcasting and backfill-adjustment techniques applied to COVID-19 and influenza surveillance data show that the failure to explicitly model reporting delays the short-term forecast, overstates the transmission and lags the public health response by several days to weeks (Bracher *et al*., 2021; Funk *et al*., 2020). Taken together, this evidence confirms the fact that, although laboratory-confirmed case surveillance is very specific, it is constrained by time considerations, making it a less effective tool in its own right as a real-time signal of surveillance, supporting the importance of combined modeling systems that explicitly consider delays in reporting and include other data streams that arrive earlier.

A second inherent weakness of conventional case-based surveillance is sensitivity and case-finding bias: proven cases of disease only provide a portion of the actual number of infections within a population. Empirical research on the COVID-19 pandemic continuously reveals large under-ascertainment, which is caused by asymptomatic and mildly symptomatic infections, differences in healthcare-seeking behavior, unequal access to diagnostic testing, and changing testing eligibility criteria (Lau *et al*., 2020; R. Li *et al*., 2020; Reese *et al*., 2021). Research on large-scale seroprevalence studies conducted in the United States and Europe show that the reported case counts underestimated cumulative infections by a factor of five to over ten at several stages of the pandemic, and under-reporting was the most significant among younger people, socioeconomically marginalized groups, and underserved communities (Bajema *et al*., 2021; Bobrovitz *et al*., 2021). Not only, but also, the ascertainment rates do not remain constant with time; they change with the fluctuating testing capacity, societal risk perception, policy interventions, and health system burden, providing time-varying bias to incidence trends based on surveillance (Iuliano *et al*., 2021). The methodological studies indicate that the estimates of the strength of transmission may be biased by the failure to explicitly model imperfect and time-varying case detection, and that comparisons across jurisdictions may be distorted, and misleading the effectiveness of public health interventions (Gostic *et al*., 2021). This evidence, along with what follows, indicates that, though laboratory-confirmed case surveillance is characterized by high diagnostic specificity, its limited and patchy sensitivity restricts its capability to establish what the dynamics of population-wide transmission are, which underlines the necessity of frameworks of surveillance that are both more integrated combining case data with other sources that are less dependent on individual testing and healthcare interaction.

There is a significant difference between jurisdiction-based laboratory reporting standards. Electronic laboratory reporting (ELR) is required in some states and accepted in others; various jurisdictions have access to standardized reporting infrastructure, and others use paper-based processes (Mishra *et al*., 2021). Heterogeneity in that way poses serious obstacles to national surveillance and restricts the possibility to carry out similar analyses in the regions. State surveys of the highly developed ELR adopters indicate that despite certification assertions, even standardized HL7-compliant ELR messages often include non-standard state local codes instead of nationally standardized LOINC or SNOMED codes, which are inconsistent with operational reality (Sung *et al*., 2023).

An overview of academic medical centers that adopt LOINC revealed a significant divergence of approaches, with some variations in staffing, training, and mapping updates frequency of mapping changes, which makes it hard to predict costs and guarantee comparable data in all locations (Richesson *et al*., 2025). Research in different countries has found that the level of variability in the assignment of LOINC codes between different sites is high, and discrepancies have been reported in 17 percent of laboratory tests across different institutions, which creates a concern regarding the insufficient guidance on uniform implementation (Richesson *et al*., 2025; Sung *et al*., 2023). Equally, the deployment of SNOMED CT demonstrates disparity in the implementation of clinical terminology with continuing issues in realizing real semantic interoperability even with international standards implementation (Sung *et al*., 2023). These tenacious standardization lapses highlight the importance of well-defined and standardized practices of terminology implementation and maintenance to provide effective multi-site surveillance and research(Richesson *et al*., 2025).

Syndromic Surveillance Systems

Definition, Data Sources, and Operational Systems

Syndromic surveillance is a valuable addition to laboratory-confirmed case reporting and concentrates on early pre-diagnosis health indicators, which could indicate an outbreak of a disease prior to conducting the confirmatory testing (Burkom *et al*., 2021; Rahmon *et al*., 2024). Syndromic surveillance does not rely on the identification of diagnoses but instead monitors the trends of syndromes such as acute respiratory disease, acute gastrointestinal disease, or fever with rash as they emerge in healthcare facilities (Ageron *et al*., 2022; Shen *et al*., 2025). This method takes advantage of the fact that patients normally consult healthcare when they develop symptoms as leading indicators of the possibility of an outbreak. Syndromic surveillance allows the early identification of any abnormal health activities or outbreaks that may not be detected by the conventional indicator-based surveillance systems especially of emergent diseases where definite case definitions may not yet be available (Jean Louis *et al*., 2025).

National Syndromic Surveillance Program (NSSP) uses a cloud-based infrastructure called BioSense which is created and managed by the CDC. BioSense ingests electronic records of healthcare facilities in near-real-time, such as chief complaints that triage nurses identify and preliminary diagnoses that healthcare providers make before a definitive laboratory confirmation(Burkom *et al*., 2021). The facilities involved include emergency departments (primary source), urgent care facilities, outpatient facilities, emergency medical services, and some inpatient hospital units. Although NSSP receives approximately 80% of emergency departments in the U.S. submitting data to NSSP, most of these data is sent every day. NSSP usually has access to patient data within 24 hours of emergency room admission of a patient and can use this information to carry out near real-time surveillance and make more informed judgments about the problems of public health(Burkom *et al*., 2021).

The data items that are collected in syndromic surveillance comprise the time of attendance, the place of the facility, patient demographics such as age group, gender, chief complaint, which is coded into standard categories, and a preliminary diagnosis in ICD-10 codes (Burkom *et al*., 2021). The chief complaints are the main reason why the patient consults medical care in real-time, and they are typically captured as free text fields and can have misspellings and/or abbreviations, however, chief complaints can vary across healthcare facilities, thereby complicating the measurement of the exact burden of illness or injury in a community. The main tool of analysis is the Electronic Surveillance System of the Early Notification of Community-Based Epidemics (ESSENCE), which involves automatic daily analysis of syndromic data, based on the statistical process control methods (Burkom *et al*., 2021). ESSENCE identifies anomalies at rates higher than anticipated levels of cases on account of day of week effects and seasonality based on spatial and temporal statistical alerting algorithms involving moving average, cumulative sum (CUSUM), exponentially weighted moving average (EWMA) and Shewhart control charts (Burkom *et al*., 2021). It is analyzed at various levels of geography, such as individual facilities, state and national summary levels, and ESSENSE offers the functionality of custom querying, user-defined alerting, geographical mapping, remote data capture, and event reporting to enable users to compile and arrange information into a sensible perception of

Publisher: SARC Publisher

population health conditions (Burkom *et al*., 2021).

Timeliness Advantages and Operational Applications

The ability to derive timeliness benefits and operational uses is a strength of syndromic surveillance systems since it gathers pre-diagnostic healthcare utilization indicators, which are conveyed much quicker than lab-diagnosed case information(Rock *et al*., 2021). Syndromic indicators of respiratory and influenza-like illness based on the emergency department have shown that before epidemics grew, early alerts were enacted by emergency department-based syndromic indicators, often confirmed by confirmed case surveillance one to two weeks prior to the peak of rapid transmission(Ageron *et al*., 2022). Empirical tests have been conducted in the United States and Europe indicating that these early warning signals enhanced real-time situational awareness, hospital surge planning, and short-term prediction of healthcare demand, especially in situations where the laboratory capacity was limited or reporting delays were long. Syndromic surveillance has also been found useful in tracking strain in the healthcare system, and trends in emergency department respiratory hospital visits were strongly predictive of changes in hospital admissions and intensive care utilization(Horng *et al*., 2021). Taken together, this suggests that syndromic surveillance systems are timelier than the classical case-based surveillance systems and have a critical operational role in the early response during an outbreak and preparedness when incorporated in an integrated surveillance architecture (Adam *et al*., 2022).

Wastewater-Based Epidemiology: Principles and Early Detection Value

Wastewater-based epidemiology (WBE) is an epidemiological method that makes use of the identification and measurement of genetic material of pathogens in municipal wastewater to deliver population-wide monitoring with no dependence on individual healthcare- seeking behavior or access to diagnostic testing(Larsen & Wigginton, 2020). Since patients can release viral RNA in feces before they have shown symptoms or even independent of the severity of the illness, wastewater indicators combine symptomatic, asymptomatic, and unreported infections, and provide a more complete and less biased picture of community transmission compared to case-based surveillance alone(Larsen & Wigginton, 2020). Progress since 2020 has made sampling strategies, normalization strategies based on fecal indicators, and quantitative workflows of reverse transcription PCR (qRT) and sequencing more consistent to allow comparisons across sites and over time (Wolfe *et al*., 2021). Notably, WBE data are often made available a few days following collection of a sample and are also not subject to reporting lag that characterises case surveillance through laboratory confirmation, thus making them especially apt in near-real-time tracking of epidemic dynamics at the community level (W. Ahmed *et al*., 2020).

The empirical data of several nations in the COVID-19 pandemic indicates that the wastewater viral loads were often one to three weeks ahead of reported cases, hospitalization, and deaths, proving WBE an effective early warning indicator of a new transmission (Peccia *et al*., 2020). Longitudinal studies in large metropolitan regions in the United States demonstrated high correlations between wastewater SARS-CoV-2 RNA patterns and the following clinical outcomes despite lapses in testing and adjustments in case ascertainment (Sims & Kasprzyk-Hordern, 2020). In addition to early detection, WBE has been useful in monitoring the dynamics of variants emergence and replacement by sequencing wastewater, which can also be used in population-scale genomic surveillance to supplement clinical sampling (Crits-Christoph *et al*., 2021). Taken together, these observations suggest that wastewater-based surveillance can offer real-time, population-representative indications that improve outbreak detection and situational awareness especially when used in conjunction with clinical, syndromic, and genomic data streams in cohesive epidemiologic modelling frameworks.

Genomic Surveillance: Whole-Genome Sequencing and Targeted Traveler-Based Applications

Genomic surveillance relies on next-generation sequencing to produce full or nearly full pathogen genomes, allowing for the detection of genetic variants associated with shifts in transmissibility, virulence, and immune evasion. In the SARS-CoV-2 pandemic, whole genome sequencing (WGS) has played a key role in the tracking of SARS-CoV-2 evolution, allowing for the rapid evaluation of the potential impact of new variants (Oude Munnink *et al*., 2020; Tegally *et al*., 2021). By providing insight into the total mutational profile, WGS enables the rapid detection of new variants before they become dominant in clinical surveillance. Phylogenetic and phylodynamic analysis can also trace the pathways of transmission and allow for the estimation of key epidemiologic parameters, providing information that can complement traditional clinical surveillance (Volz *et al*., 2021). Genomic surveillance has thus become a critical tool in the modern infectious disease landscape, particularly when coupled with epidemiologic data streams. In addition to the standard clinical sequencing, targeted genomic surveillance approaches have proven highly useful for the early detection of variants. The establishment of traveler-based genomic surveillance initiatives at entry points around the world has demonstrated that the sequencing of samples from arriving international travelers is highly effective in detecting new variants earlier than would be possible through community sampling alone, particularly in areas that do not have the capacity to conduct extensive in-country sequencing efforts (Gonzalez-Reiche *et al*., 2020; Lemey *et al*., 2021). Modeling and empirical data suggest that sequencing at highly connected travel nodes could provide early detection of new variants in a manner that is more efficient and less resource-intensive than untargeted approaches (Chinazzi *et al*., 2020). Experience from several countries during the current pandemic has shown that targeted genomic surveillance initiatives offer the opportunity for earlier lead-time compared to localized clinical detection and therefore offer improved preparedness through the earlier assessment and planning that is possible through earlier risk estimation and planning for response (Grubaugh *et al*., 2018).

Machine Learning-Enhanced Outbreak Detection and Short-Term Forecasting: Evidence, Limitations, and Hybrid Integration

Machine learning (ML) can extend mechanistic epidemiologic models by discovering patterns from high-dimensional surveillance data without requiring the specification of transmission dynamics. Nevertheless, the applicability of ML differs dramatically between operational systems and research prototypes(Lin *et al*., 2021). A prominent example of an ML-related outbreak detection method with operational applicability is the use of supervised hidden Markov models, which use labeled outbreak data to deduce underlying “outbreak versus baseline” states from surveillance time series data. Analyses based on real Salmonella and Campylobacter surveillance data from Germany showed that supervised HMMs perform better than unsupervised methods and on par with or better than the best available outbreak detection systems operating in European surveillance systems (Zacher & Czogiel, 2022). These observations also reflect that performance depends on the data and the reporting process. As has been seen in other outbreak detection studies, performance estimates based on retrospective analysis may overestimate performance, and little information exists on performance with prospective use, fatigue, and decision thresholds.

Although more applications of deep learning models have occurred in the area of short-term forecasting, rather than the operational identification of outbreaks, the literature shows that models based on LSTMs have the potential to provide accurate predictions in the near term, based on case data, but that performance is highly variable depending on the scenario and that accuracy diminishes with increasing forecast horizon or changes in the underlying data-generating process such as interventions, behavior change, emergence of new strains (Chandra *et al*., 2022; Kumar *et al*., 2021). This has driven the development of hybrid models that seek to leverage the strengths of both mechanistic and learning models, adding adaptive capabilities to mechanistic frameworks in a way that retains epidemiologic interpretability, for example, Epi-DNN models, which merge the structure of compartment models with the capabilities of deep neural networks to estimate time-dependent elements and better align model predictions with empirical epidemic dynamics (Ning *et al*., 2023). There is sufficient evidence to support the addition of ML capabilities to anomaly detection and short-term predictive tools, but the boundary for the combination of surveillance capabilities would be around mechanistic-AI modeling.

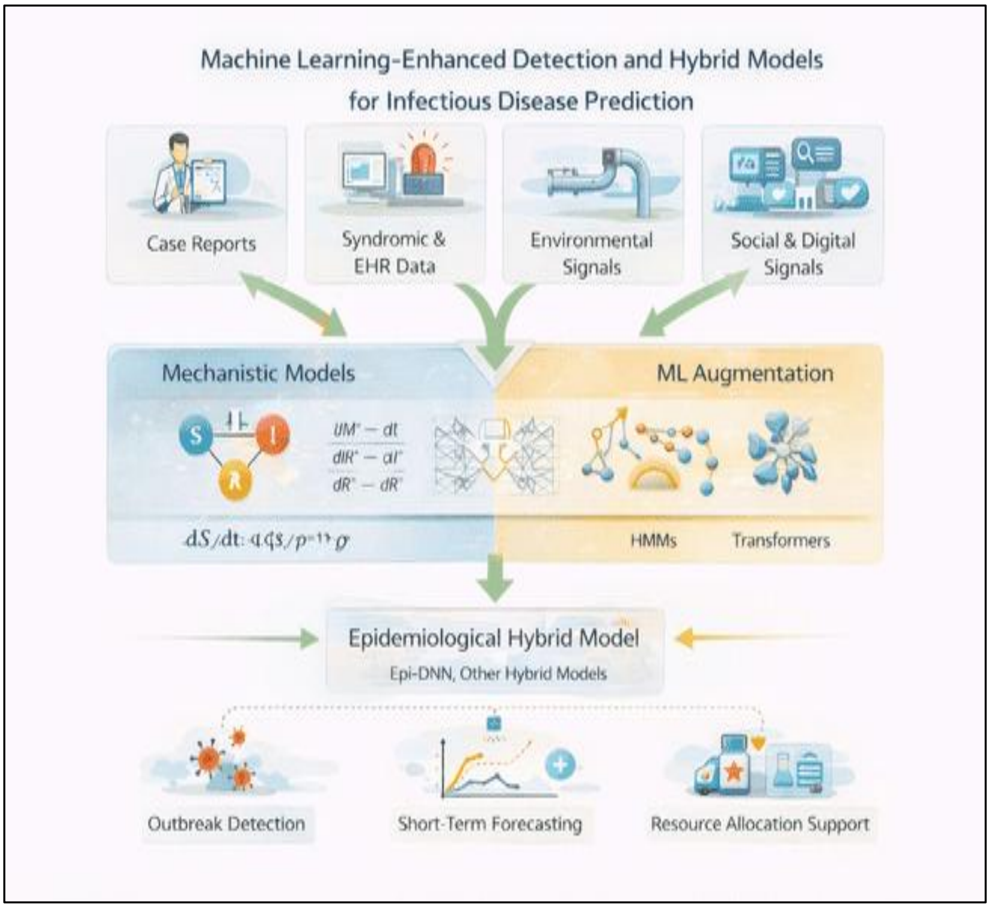

Figure 2: shows the integration of mechanistic epidemiological models with machine learning to enhance outbreak detection and short-term forecasting.

REAL-TIME DATA ASSIMILATION AND NOWCASTING UNDER REPORTING DELAYS

Data assimilation into real-time systems refers to the continuous updating of model states and parameters based on the availability of new surveillance data, thereby improving situational awareness during rapidly changing epidemic situations and data flows (Biegel, 2020). This essentially translates to the assimilation model reformulating the surveillance process as an inference problem in the context of state-space theory, where model forecasts are updated based on new data availability, while propagating uncertainty; recent work in this area has utilized ensemble methods (Kalman filter-type methods) and Monte Carlo methods to infer the calibration of model parameters based on the availability of common COVID-19 surveillance indicators (Papageorgiou & Tsaklidis, 2023). This work highlights the importance of the inference model being capable of handling the count process as a non-Gaussian process, handling time-varying parameters, and handling the availability of delayed data, which is common in the case of laboratory, syndromic, wastewater, and genomic surveillance data, respectively (Yang *et al*., 2022). This, in turn, has led to the recent focus of assimilation work being on the importance of incorporating latency into the inference model, rather than the development of new algorithms, as the key requirement for the successful implementation of real-time multi-stream surveillance models in the context of public health surveillance (Arroyo-Marioli *et al*., 2021).

Nowcasting, or inferring current incidence after accounting for reporting delays, is a closely related task and a central operational application of integrated surveillance models, as reported today is seldom contemporary with occurrence today during outbreak conditions (McGough *et al*., 2020). Peer-reviewed research on nowcasting during the COVID-19 pandemic suggests that modeling the reporting delay distribution, including dynamics, is essential to improve current incidence estimation of epidemic curves, as well as related quantities such as growth rates or reproductive ratios, compared to raw reporting data alone (Günther *et al*., 2021). Further, related work incorporating leading indicators improves nowcasting, especially when traditional reporting is irregular or volatile, thereby justifying integration of multiple streams of data for example combining reporting of new cases with other proximal indicators (Bergström *et al*., 2022). In line with this thinking, incidence modeling based on wastewater signals, along with incorporation of relevant context such as testing rates or circulating lineages, illustrates community incidence detection despite changes to clinical reporting, and particle-filtered mechanistic models based on wastewater have provided actionable lead time to predict outbreak growth rates (McManus *et al*., 2023). In aggregate, this body of research provides clear support for a single, comprehensive framework: namely, real-time integration and nowcasting are most effective as implemented as part of integrated frameworks accounting for delay reporting, latency, and uncertainties among diverse streams of surveillance data (Yang *et al*., 2022).

INTEROPERABILITY STANDARDS AND MULTI-SOURCE DATA FUSION

The ability to have an integrated infectious disease surveillance in the United States is hampered by the persistent technical heterogeneity in the healthcare delivery, laboratory, and public health information systems, which vary in their data models, terminologies, and reporting processes (Idahor *et al*., 2025). Despite the existence of national interoperability standards, such as the HL7 messaging standards, as well as the FHIR data model, that ensure a standardized structure for the exchange of clinical and public health data, the empirical assessment of their implementation in real-world settings shows that the process of data exchange is still incomplete due to the lack of full semantic standardization, as well as the continued dependence on the existing legacy systems (Ayaz *et al*., 2021). Laboratory data have been shown to have an irregular application of standardized terms, despite the existence of regulatory drivers, thereby contributing to the delays in the process of data exchange, as well as the potential loss of analytic fidelity for surveillance (Roy *et al*., 2025). This, therefore, means that the challenges of interoperability are not just technical in nature but have a direct impact on the ability to have real-time modeling, due to the potential latencies, incompleteness, as well as the existence of biases that must be corrected in the analytic frameworks.

From a methods perspective, a data fusion paradigm that can accommodate and reconcile differences in timeliness, specificity, and measurement error between different sources of surveillance data, including laboratory confirmation, syndromic surveillance, and environmental surveillance, has been developed (Ning *et al*., 2023). Bayesian hierarchical modeling has emerged as a popular and rigorously justifiable paradigm that models observed data as conditionally independent manifestations of latent epidemiologic processes, with each data stream being able to contribute information based on its known quality and bias characteristics (Bracher *et al*., 2021). More recently, machine-learning paradigms of data fusion that are suited to integrating high-dimensional spatially structured surveillance data, including graph neural networks that model geographic areas as interlinked nodes that are connected by mobility or transmission networks, have been investigated (S. F. Ahmed *et al*., 2024). While they have been demonstrated to improve short-term predictive accuracy in retrospective analyses, recent analyses highlight that their use in surveillance must be accompanied by mechanistic constraint and attention to data quality and interpretability, and that standards-based interoperability and statistically informed data fusion remain essential in scalable multi-source surveillance systems (McGough *et al*., 2020).

INTEROPERABILITY, GOVERNANCE, AND PRIVACY CONSTRAINTS IN INTEGRATED SURVEILLANCE

Despite the availability of national interoperability standards, technical fragmentation persists as a barrier to effective infectious disease surveillance integration across the United States. Legacy surveillance systems installed before the adoption of current interoperability models continue to be used by state or local public health agencies, posing compatibility issues that are expensive and difficult to overcome (Zeba *et al*., 2025). Empirical studies of electronic laboratory reporting systems during the COVID-19 pandemic show a lack of semantic interoperability between HL7 messaging systems, where a large percentage of laboratory reports fail to adhere to standardized test and results codes, making it necessary for manual integration by public health personnel (Dixon *et al*., 2020). Apart from technological limitations, legal and organizational issues further impede data integration, and although U.S. laws allow data sharing for public health reasons, the healthcare sector often follows a conservative approach to data sharing, and the lack of standardized data use agreements and the difficulty of connecting data from different systems without a nationwide patient ID further hinders data integration (McGraw & Mandl, 2021). To address such challenges, there has been a growing focus on the design of privacy-preserving analysis infrastructure and governance frameworks that make possible the integration of data across jurisdictions without the need for central sharing of microdata. Federated learning infrastructure facilitates the distributed training of models across jurisdictions without the central storage of data, and their applicability within the realm of healthcare to minimize privacy threats with the preservation of predictive accuracy for surveillance and forecasting analysis (Rieke *et al*., 2020). Supporting infrastructure designs within the realm of secure research environments and access models make possible the use of governed environments where approved analyses are allowed with the implementation of high technical security measures with the possibility of surveillance and monitoring by governance authorities, thus making possible collaboration across institutions without the need for unfettered pooling of data. Differential privacy tools and fairness monitoring are increasingly being considered within the realm of limiting the possibility of re-identification of microdata with the monitoring of possible performance disparities within models across jurisdictions or populations with the use of surveillance within the realm of public health monitoring (Ficek *et al*., 2021).

FUTURE DIRECTIONS AND ADVANCING THE FRONTIER

One of the major aims of next-generation infectious disease surveillance is therefore the creation of real-time and dynamic epidemiological modeling systems that can adjust according to changes in outbreak scenarios. The major problem with current forecasting models is that they are developed based on retrospective training and are used in a static manner when an outbreak occurs, thereby limiting their ability to remain accurate in changing outbreak scenarios due to human behavior and pathogen mutations. The next generation of surveillance systems should therefore be able to update and modify their parameters and structures with new streams of data available online, and this is indicated by recent advancements in online inference and validation pipelines.

Another area that is connected to the former is the detection and handling of regime shifts in epidemic cases that are sudden in nature and are caused by factors like the development of new variants of the virus or widespread behavioral changes. Models that have state switching or mixture modeling components have the potential to address these regime shifts by adjusting the assumptions in accordance with these transitions in order to counteract the effects of degradation in forecasting. It is important to ensure that the development of adaptive surveillance systems is supported by strong governance and validation practices to ensure that these systems are transparent and valid in accordance with public health policy decisions.

CONCLUSION

The status of infectious disease surveillance in the USA is at a crossroads. While the traditional case-based surveillance system serves as the cornerstone, its limitations are becoming more apparent in terms of its ability to offer timely and sensitive surveillance data that would enable the pre-detection of emerging threats. Simultaneously, new advances in mathematical modeling, machine learning, whole genome sequencing, wastewater surveillance, and real-time analytics offer great opportunities for improving pre-detection surveillance. The challenge has moved from the availability of tools to their effective combination. The review makes it clear that a more effective system of integration must involve more than technological innovation. The establishment of interoperable data structures, workforce capabilities, and frameworks of governance that promote and support privacy, accountability, and trust are fundamental preconditions of this goal. It is important that a clear line be drawn between tested and effective methods of operation and new and innovative approaches in order to avoid premature investment in methods that are as yet untested and unproven. The COVID-19 experience has made it clear that surveillance is of value and that current systems are inadequate; it is a rapidly closing opportunity to capitalize on lessons that have been learned and incorporate them into improved systems of surveillance that are real-time and trustworthy and that serve as a direct determinant of early detection and preparedness in any future public health crisis.

REFERENCES

1. Adam, C. M., Borroto, R., Thomas, E., Tuttle, J., Pavlick, J., & Drenzek, C. L. "Syndromic surveillance as a tool for case-based varicella

reporting in Georgia, 2016-2019." *Public Health Reports* 137.6 (2022): 1070-1078.

2. Adedire, O., Love, N. K., Hughes, H. E., Buchan, I., Vivancos, R., & Elliot, A. J. "Early detection and monitoring of gastrointestinal infections using syndromic surveillance: a systematic review." *International journal of environmental research and public health* 21.4 (2024): 489.
3. Ageron, F. X., Hugli, O., Dami, F., Caillet-Bois, D., Pittet, V., Eckert, P., & Carron, P. N. "Lessons from COVID-19 syndromic surveillance through emergency department activity: a prospective time series study from western Switzerland." *BMJ open* 12.5 (2022): e054504.
4. Ahmed, S. F., Kuldeep, S. A., Rafa, S. J., Fazal, J., Hoque, M., Liu, G., & Gandomi, A. H. "Enhancement of traffic forecasting through graph neural network-based information fusion techniques." *Information Fusion* 110 (2024): 102466.
5. Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., & Mueller, J. F. "First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community." *Science of the total environment* 728 (2020): 138764.
6. Akingbola, A., Adegbesan, A., Tundealao, S., Benson, A. E., Makinde, A. O., Shekoni, M., & Fakiyesi, T. "Post-pandemic era: global trends, benefits, and barriers in integrating artificial intelligence into public health education." *Frontiers in Public Health* 13 (2025): 1648970.
7. Arroyo-Marioli, F., Bullano, F., Kucinskas, S., & Rondón-Moreno, C. "Tracking R of COVID-19: A new real-time estimation using the Kalman filter." *PloS one* 16.1 (2021): e0244474.
8. Ayaz, M., Pasha, M. F., Alzahrani, M. Y., Budiarto, R., & Stiawan, D. "The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities." *JMIR medical informatics* 9.7 (2021): e21929.
9. Bajema, K. L., Wiegand, R. E., Cuffe, K., Patel, S. V., Iachan, R., Lim, T., & Edens, C. "Estimated SARS-CoV-2 Seroprevalence in the US as of September 2020." *JAMA internal medicine* 181.4 (2021): 450-460.
10. Baum, J. H. J., Dörre, A., Boender, T. S., Heldt, K., Wilking, H., Drynda, S., & Schranz, M. "Establishing and validating syndromic surveillance of gastrointestinal infections using routine emergency department data, Germany, 2019–2023." *Scientific Reports* 15.1 (2025): 38281.
11. Bergström, F., Günther, F., Höhle, M., & Britton, T. "Bayesian nowcasting with leading indicators applied to COVID-19 fatalities in Sweden." *PLOS Computational Biology* 18.12 (2022): e1010767.
12. Biegel, H. “Near Real-Time Forecasting of Epidemics Using Data Assimilation with Simple Models.” *Diss. The University of Arizona*, (2020).
13. Bobrovitz, N., Arora, R. K., Cao, C., Boucher, E., Liu, M., Donnici, C., & Cheng, M. P. "Global seroprevalence of SARS-CoV-2 antibodies: a systematic review and meta-analysis." *PloS one* 16.6 (2021): e0252617.
14. Bracher, J., Ray, E. L., Gneiting, T., & Reich, N. G. "Evaluating epidemic forecasts in an interval format." *PLoS computational biology* 17.2 (2021): e1008618.
15. Burkom, H., Loschen, W., Wojcik, R., Holtry, R., Punjabi, M., Siwek, M., & Lewis, S. "Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE): overview, components, and public health applications." *JMIR public health and surveillance* 7.6 (2021): e26303.
16. CDC. “National Notifiable Diseases Surveillance System (NNDSS).” *National Notifiable Diseases Surveillance System (NNDSS).* (2025)
17. Chandra, R., Jain, A., & Singh Chauhan, D. "Deep learning via LSTM models for COVID-19 infection forecasting in India." *PloS one* 17.1 (2022): e0262708.
18. Charniga, K., Park, S. W., Akhmetzhanov, A. R., Cori, A., Dushoff, J., Funk, S., & Abbott, S. "Best practices for estimating and reporting epidemiological delay distributions of infectious diseases." *PLoS computational biology* 20.10 (2024): e1012520.
19. Chatterjee, A., Pahari, N., & Prinz, A. "HL7 FHIR with SNOMED-CT to achieve semantic and structural interoperability in personal health data: a proof-of-concept study." *Sensors* 22.10 (2022): 3756.
20. Chen, H., Zeng, D., & Yan, P. "Syndromic surveillance data sources and collection strategies." *Infectious Disease Informatics: Syndromic Surveillance for Public Health and BioDefense*. New York, NY: Springer US, 2010. 33-48.

21. Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., & Vespignani, A. "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak." *Science* 368.6489 (2020): 395-400.
22. Chowell, G., & Skums, P. "Investigating and forecasting infectious disease dynamics using epidemiological and molecular surveillance data." *Physics of Life Reviews* 51 (2024): 294-327.
23. Crits-Christoph, A., Kantor, R. S., Olm, M. R., Whitney, O. N., Al-Shayeb, B., Lou, Y. C., & Nelson, K. L. "Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants." *MBio* 12.1 (2021): 10-1128.
24. Dixon, B. E., Zhang, Z., Arno, J. N., Revere, D., Joseph Gibson, P., & Grannis, S. J. "Improving notifiable disease case reporting through electronic information exchange–facilitated decision support: A controlled before-and-after trial." *Public Health Reports* 135.3 (2020): 401-410.
25. Du Plessis, L., McCrone, J. T., Zarebski, A. E., Hill, V., Ruis, C., Gutierrez, B., & Pybus, O. G. "Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK." *Science* 371.6530 (2021): 708-712.
26. Ficek, J., Wang, W., Chen, H., Dagne, G., & Daley, E. "Differential privacy in health research: A scoping review." *Journal of the American Medical Informatics Association* 28.10 (2021): 2269-2276.
27. Funk, S., Abbott, S., Atkins, B. D., Baguelin, M., Baillie, J. K., Birrell, P., & ISARIC4C Investigators, "Short-term forecasts to inform the response to the Covid-19 epidemic in the UK." *MedRxiv* (2020): 2020-11.
28. Giancotti, M., Lopreite, M., Mauro, M., & Puliga, M. "Innovating health prevention models in detecting infectious disease outbreaks through social media data: an umbrella review of the evidence." *Frontiers in Public Health* 12 (2024): 1435724.
29. Gijón, A., Eiraudo, S., Manjavacas, A., Schiera, D. S., Molina-Solana, M., & Gómez-Romero, J. "Integrating physics and data-driven approaches: An explainable and uncertainty-aware hybrid model for wind turbine power prediction." *Computer Physics Communications* (2025): 109761.
30. Gonzalez-Reiche, A. S., Hernandez, M. M., Sullivan, M. J., Ciferri, B., Alshammary, H., Obla, A., & van Bakel, H. "Introductions and early spread of SARS-CoV-2 in the New York City area." *Science* 369.6501 (2020): 297-301.
31. Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C., & Cobey, S. "Practical considerations for measuring the effective reproductive number, R t." *PLoS computational biology* 16.12 (2020): e1008409.
32. Grubaugh, N. D., Ladner, J. T., Lemey, P., Pybus, O. G., Rambaut, A., Holmes, E. C., & Andersen, K. G. "Tracking virus outbreaks in the twenty-first century." *Nature microbiology* 4.1 (2019): 10-19.
33. Günther, F., Bender, A., Katz, K., Küchenhoff, H., & Höhle, M. "Nowcasting the COVID-19 pandemic in Bavaria." *Biometrical Journal* 63.3 (2021): 490-502.
34. Panel, O. H. H. L. E., Hayman, D. T., Adisasmito, W. B., Almuhairi, S., Behravesh, C. B., Bilivogui, P., & Koopmans, M. "Developing one health surveillance systems." *One Health* 17 (2023): 100617.
35. Horng, S., O'Donoghue, A., Dechen, T., Rabesa, M., Shammout, A., Markson, L., & Stevens, J. P. "Secondary use of COVID-19 symptom incidence among hospital employees as an example of syndromic surveillance of hospital admissions within 7 days." *JAMA Network Open* 4.6 (2021): e2113782.
36. Idahor, C. O., Esomu, E. J. O., Ogbonna, N., Momoh, Z., Ogbeide, O. A., Ikhu-Omoregbe, O., & Oronsaye, N. J. "Infectious Disease Surveillance in the Era of Big Data and AI: Opportunities and Pitfalls." *Cureus* 17.10 (2025).
37. Ingelbeen, B., van Kleef, E., Mbala, P., Danis, K., Macicame, I., Hens, N., & van der Sande, M. A. "Embedding risk monitoring in infectious disease surveillance for timely and effective outbreak prevention and control." *BMJ global health* 10.2 (2025).
38. Iuliano, A. D., Chang, H. H., Patel, N. N., Threlkel, R., Kniss, K., Reich, J., & Reed, C. "Estimating under-recognized COVID-19 deaths, United States, march 2020-may 2021 using an excess mortality modelling approach." *The Lancet Regional Health–Americas* 1 (2021).
39. Louis, F. J., Nichols, L., de la Torre, C., & Dahourou, A. G. "Strengthening outbreak detection in Africa to achieve the 7-1-7 global framework: challenges and opportunities." *Public health reviews* 46 (2025): 1608039.

40. John, G., Sahajpal, N. S., Mondal, A. K., Ananth, S., Williams, C., Chaubey, A., & Kolhe, R. "Next-generation sequencing (NGS) in COVID-19: a tool for SARS-CoV-2 diagnosis, monitoring new strains and phylodynamic modeling in molecular epidemiology." *Current issues in molecular biology* 43.2 (2021): 845-867.
41. Kaur, J., & Butt, Z. A. "AI-driven epidemic intelligence: the future of outbreak detection and response." *Frontiers in Artificial Intelligence* 8 (2025): 1645467.
42. Kizza, T., Aduampong, M. J. K., & Kaiser, F. “Survival analysis in U.S. chronic disease research: A systematic review of methods and applications.” *International Journal of Frontline Research in Life Science*, 3.2 (2025): 26–34.
43. Knicely, K., Loonsk, J. W., Hamilton, J. J., Fine, A., & Conn, L. A. "Electronic case reporting development, implementation, and expansion in the United States." *Public Health Reports®* 139.4 (2024): 432-442.
44. Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C., Van Boven, M., van de Wijgert, J. H., & Bonten, M. J. "Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study." *The Lancet Public Health* 5.8 (2020): e452-e459.
45. Kumar, S., Sharma, R., Tsunoda, T., Kumarevel, T., & Sharma, A. "Forecasting the spread of COVID-19 using LSTM network." *BMC bioinformatics* 22.Suppl 6 (2021): 316
46. Larsen, D. A., & Wigginton, K. R. “Tracking COVID-19 with wastewater.” *Nature Biotechnology*, 38.10 (2020): 1151–1153.
47. Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. "The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China." *Journal of travel medicine* 27.3 (2020): taaa037.
48. Lemey, P., Ruktanonchai, N., Hong, S. L., Colizza, V., Poletto, C., Van den Broeck, F., & Dellicour, S. "Untangling introductions and persistence in COVID-19 resurgence in Europe." *Nature* 595.7869 (2021): 713-717.
49. Levy, J. I., Andersen, K. G., Knight, R., & Karthikeyan, S. "Wastewater surveillance for public health." *Science* 379.6627 (2023): 26-27.
50. Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)." *Science* 368.6490 (2020): 489-493.
51. Li, Z., Meng, F., Wu, B., Kong, D., Geng, M., Qiu, X., & Liu, S. "Reviewing the progress of infectious disease early warning systems and planning for the future." *BMC Public Health* 24.1 (2024): 3080.
52. Lin, H. J., Wang, X. L., Tian, M. Y., Li, X. L., & Tan, H. Z. "Machine learning and its epidemiological applications." *Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi* 42.9 (2021): 1689-1694.
53. Majumder, M. S., Cusick, M., & Rose, S. "Measuring concordance of data sources used for infectious disease research in the USA: a retrospective data analysis." *BMJ open* 13.2 (2023): e065751.
54. McClellan, M., DeSalvo, K. B., Benjamin, G. C., Cerise, F. P., Choucair, B., Del Rio, C., & Zink, A. "Updating US public health for healthier communities: article examines future directions for US public health." *Health Affairs* 44.2 (2025): 148-155.
55. McGough, S. F., Johansson, M. A., Lipsitch, M., & Menzies, N. A. "Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking." *PLoS computational biology* 16.4 (2020): e1007735.
56. McGraw, D., & Mandl, K. D. "Privacy protections to encourage use of health-relevant digital data in a learning health system." *NPJ digital medicine* 4.1 (2021): 2.
57. McManus, O., Christiansen, L. E., Nauta, M., Krogsgaard, L. W., Bahrenscheer, N. S., von Kappelgaard, L., & Ethelberg, S. "Predicting COVID-19 incidence using wastewater surveillance data, Denmark, October 2021–June 2022." *Emerging Infectious Diseases* 29.8 (2023): 1589.
58. Mishra, N., Duke, J., Karki, S., Choi, M., Riley, M., Ilatovskiy, A. V., ... & Lenert, L. "A modified public health automated case event reporting platform for enhancing electronic laboratory reports with clinical data: design and implementation study." *Journal of Medical Internet Research* 23.8 (2021): e26388.
59. Ning, X., Jia, L., Wei, Y., Li, X. A., & Chen, F. "Epi-DNNs: Epidemiological priors informed deep neural networks for modeling COVID-19 dynamics." *Computers in biology and medicine* 158 (2023): 106693.
60. Osamika, D., Adelusi, B. S., Kelvin-Agwu, M. T. C., Mustapha, A. Y., Forkuo, A. Y., & Ikhalea, N. "A critical review of health data

interoperability standards: FHIR, HL7, and beyond." *World Scientific News* 203 (2025): 194-233.

61. Oude Munnink, B. B., Nieuwenhuijse, D. F., Stein, M., O'Toole, Á., Haverkate, M., Mollers, M., & Koopmans, M. "Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands." *Nature medicine* 26.9 (2020): 1405-1410.
62. Papageorgiou, V. E., & Tsaklidis, G. "An improved epidemiological-unscented Kalman filter (hybrid SEIHCRDV-UKF) model for the prediction of COVID-19. Application on real-time data." *Chaos, Solitons & Fractals* 166 (2023): 112914.
63. Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., & Omer, S. B. "Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics." *Nature biotechnology* 38.10 (2020): 1164-1167.
64. Rahmon, I., Bosmans, M., Baliatsas, C., Hooiveld, M., Marra, E., & Dückers, M. "COVID-19 health impact: a use case for syndromic surveillance system monitoring based on primary care patient registries in the Netherlands." *JMIR public health and surveillance* 10.1 (2024): e53368.
65. Rainey, J. J., Lin, X. M., Murphy, S., Velazquez-Kronen, R., Do, T., Hughes, C., & Gundlapalli, A. V. "Deployment of the National Notifiable Diseases Surveillance System during the 2022–23 mpox outbreak in the United States—Opportunities and challenges with case notifications during public health emergencies." *Plos one* 19.4 (2024): e0300175.
66. Reese, H., Iuliano, A. D., Patel, N. N., Garg, S., Kim, L., Silk, B. J., & Reed, C. "Estimated incidence of coronavirus disease 2019 (COVID-19) illness and hospitalization—United States, February–September 2020." *Clinical Infectious Diseases* 72.12 (2021): e1010-e1017.
67. Richesson, R. L., Campion, T. R., Knosp, B. M., & Hanauer, D. A. "LOINC implementation approaches in academic medical research centers–results from a survey of CTSA sites." *Journal of clinical and translational science* 9.1 (2025): e223.
68. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., & Cardoso, M. J. "The future of digital health with federated learning." *NPJ digital medicine* 3.1 (2020): 119.
69. Rock, P. J., Quesinberry, D., Singleton, M. D., & Slavova, S. "Emergency medical services and syndromic surveillance: a comparison with traditional surveillance and effects on timeliness." *Public Health Reports* 136.1_suppl (2021): 72S-79S.
70. Romano, S., Yusuf, H., Davis, C., Thomas, M. J., & Grigorescu, V. "An Evaluation of Syndromic Surveillance–Related Practices Among Selected State and Local Health Agencies." *Journal of Public Health Management and Practice* 28.2 (2022): 109-115.
71. Roy, S., Mandal, B., & Das, B. K. "Quality assurance in laboratory practices and equipment maintenance: Ensuring precision, reliability, and compliance." *Laboratory Techniques for Fish Disease Diagnosis*. Singapore: Springer Nature Singapore, 2025. 79-103.
72. Ssemujju, F. S., & Solomon, D. "Integrating Spatial Modeling and Machine Learning for Infectious Disease Surveillance in US Urban and Rural Settings." *Journal Of Internal Medicine And Public Health* 5.1 (2026): 1-12.
73. Shen, Y., Liu, Y., Krafft, T., & Wang, Q. "Progress and challenges in infectious disease surveillance and early warning." *Medicine Plus* 2.1 (2025): 100071.
74. Sims, N., & Kasprzyk-Hordern, B. "Future perspectives of wastewater-based epidemiology: monitoring infectious disease spread and resistance to the community level." *Environment international* 139 (2020): 105689.
75. Sung, S., Park, H. A., Jung, H., & Kang, H. "A SNOMED CT mapping guideline for the local terms used to document clinical findings and procedures in electronic medical records in South Korea: methodological study." *JMIR Medical Informatics* 11.1 (2023): e46127.
76. Tang, L., Zhou, Y., Wang, L., Purkayastha, S., Zhang, L., He, J., & Song, P. X. K. "A review of multi-compartment infectious disease models." *International Statistical Review* 88.2 (2020): 462-513.
77. Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., & De Oliveira, T. "Detection of a SARS-CoV-2 variant of concern in South Africa." *Nature* 592.7854 (2021): 438-443.

78. Ugwu, D., Amofah, A. D., & Kaiser, F. "Enhancing the detection and response to infectious disease outbreaks." *International Medical Science Research Journal* 5.2 (2025).
79. Villanueva-Miranda, I., Xiao, G., & Xie, Y. "Artificial intelligence in early warning systems for infectious disease surveillance: a systematic review." *Frontiers in public health* 13 (2025): 1609615.
80. Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, Á., & Allan, J. "Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity." *Cell* 184.1 (2021): 64-75.
81. Wolfe, M. K., Archana, A., Catoe, D., Coffman, M. M., Dorevich, S., Graham, K. E., & Boehm, A. B. "Scaling of SARS-CoV-2 RNA in settled solids from multiple wastewater treatment plants to compare incidence rates of laboratory-confirmed COVID-19 in their sewersheds." *Environmental science & technology letters* 8.5 (2021): 398-404.
82. Wu, S. L., Mertens, A. N., Crider, Y. S., Nguyen, A., Pokpongkiat, N. N., Djajadi, S., & Benjamin-Chung, J. "Substantial underestimation of SARS-CoV-2 infection in the United States." *Nature communications* 11.1 (2020): 4507.
83. Yang, X., Wang, S., Xing, Y., Li, L., Xu, R. Y. D., Friston, K. J., & Guo, Y. "Bayesian data assimilation for estimating instantaneous reproduction numbers during epidemics: Applications to COVID-19." *PLoS computational biology* 18.2 (2022): e1009807.
84. Ye, Y., Pandey, A., Bawden, C., Sumsuzzman, D. M., Rajput, R., Shoukat, A., & Galvani, A. P. "Integrating artificial intelligence with mechanistic epidemiological modeling: a scoping review of opportunities and challenges." *Nature Communications* 16.1 (2025): 581.
85. Zacher, B., & Czogiel, I. "Supervised learning using routine surveillance data improves outbreak detection of Salmonella and Campylobacter infections in Germany." *PloS One* 17.5 (2022): e0267510.
86. Zeba, Z., Lartey, S. T., Durneva, P., Roy, S., Jha, N., Ofori, M. A., & Joshi, A. "Best practices for data modernization across the United States public health system: scoping review." *Journal of Medical Internet Research* 27 (2025): e70946.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:

Kebeba, M. & Agyei, E. A. "Integrated Epidemiologic Modeling for Next-Generation U.S. Infectious Disease Surveillance Systems: A Narrative Review." *Sarcouncil Journal of Internal Medicine and Public Health* 5.1 (2026): pp 48-63.