EHDI

Epidemiology & Health Data Insights
(eISSN 3080-8111)

Australasia Publishing Group

Review Article

Integrating Real-Time Genomic Surveillance (Next-Generation Sequencing) with Epidemiological Models for Infectious Disease Intervention Planning

Vivian Ukamaka Nwokedi[1], Patra Chisom Ezeamii[2], Adepeju Kafayat Olowookere[3], Oyebamiji Hafeezat Omolabake[4]

[1]*Department of Clinical Pharmacy, Faculty of Pharmacy, University of Benin, Benin City, Nigeria*

[2]*Department of Biostatistics, Epidemiology, and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA, USA*

[3]*Department of Nanoscience, University of North Carolina at Greensboro, Greensboro, NC, USA*

[4]*Department of Medicine and Surgery, College of Health Sciences, University of Ilorin, Llorin, Nigeria*

Received: Jan 11, 2026
Accepted: Feb 05, 2026

Corresponding author's email:
viviannwokedi250@gmail.com

Abstract:
Infectious disease surveillance has long been vital in public health, but traditional methods often fall short in detecting emerging threats and understanding pathogen evolution. Recent advances in Next-Generation Sequencing (NGS) have revolutionized genomic surveillance, enabling near real-time monitoring of pathogens at the genetic level. This study explores the integration of real-time genomic surveillance with epidemiological models to enhance disease intervention planning. We examine how combining genomic data with models like Susceptible-Infectious-Recovered (SIR) and Susceptible-Exposed-Infectious-Recovered (SEIR) improves outbreak forecasting, facilitates early detection of new variants, and provides actionable insights for targeted interventions. The integration of NGS data allows for more precise transmission network mapping, better-informed resource allocation, and dynamic policy adjustments. However, challenges persist, including technical limitations, data privacy concerns, and equity in global surveillance capacities. The findings suggest that genomic integration enhances epidemic prediction and response but requires robust policy frameworks, equitable data-sharing practices, and continuous capacity-building efforts in low- and middle-income regions. The future of infectious disease control hinges on advancing technologies like artificial intelligence (AI), cloud computing, and machine learning to improve predictive accuracy and support real-time decision-making. This review underscores the potential of genomic surveillance to transform public health strategies and outlines key steps for effective global collaboration.

Keywords: Genomic Surveillance; Next-Generation Sequencing; Epidemiological Models; Outbreak Forecasting; Public Health; Data Integration; Policy Frameworks

Introduction

Infectious disease surveillance has always been a cornerstone of public health practice, forming the basis for detecting, tracking, and mitigating outbreaks. Traditional surveillance approaches rely heavily on clinical reporting, laboratory confirmation, and epidemiological investigations. While valuable, these approaches often struggle with delays, underreporting, and limited resolution for characterizing pathogen diversity and transmission pathways [1]. Genomic surveillance, particularly through the application of high-throughput sequencing technologies, has emerged as a transformative tool that enhances the capacity of public health systems to observe pathogens at the molecular level and in near real time [2, 3]. By capturing the genetic blueprint of infectious agents, genomic surveillance enables health authorities to detect emerging variants, monitor transmission dynamics, and generate actionable insights that inform intervention strategies with unprecedented precision.

Next-Generation Sequencing (NGS) lies at the heart of contemporary genomic surveillance. NGS platforms allow comprehensive sequencing of pathogen genomes rapidly and at progressively lower costs compared with traditional methods, such as Sanger sequencing [4]. This high-resolution capacity is especially important for identifying genomic mutations that can alter virulence, transmissibility, and resistance to therapeutics or vaccines [2]. During the COVID-19 pandemic, for example, real-time NGS enabled the identification and tracking of SARS-CoV-2 variants, providing critical data that shaped global public health responses and vaccine development efforts [3]. These capabilities have accelerated a paradigm shift within infectious disease surveillance, transforming it from reactive and retrospective analysis to proactive, data-driven decision-making frameworks that can anticipate pathogen evolution and guide interventions.

The significance of integrating genomic data into disease monitoring becomes even clearer when we consider the limitations of classical surveillance alone. Traditional methods, while essential, often provide only coarse levels of resolution, frequently missing subtle but important evolutionary changes that affect outbreak dynamics [1]. Genomic sequencing, particularly whole-genome sequencing, enriches surveillance by pinpointing exact genetic differences between pathogen samples, enabling precise differentiation between outbreak strains and sporadic cases [5]. This high level of specificity is crucial for understanding patterns of pathogen spread, identifying sources of infection, and rapidly adjusting public health interventions. Moreover, genomic data facilitate the study of antimicrobial resistance evolution, helping to anticipate resistance trends and guide stewardship programs [2].

Epidemiological models — such as Susceptible-Infectious-Recovered (SIR) and its derivatives — are foundational tools for understanding and forecasting disease spread at the population level. These compartmental frameworks categorize individuals by infection status and describe transitions over time, providing estimates of key parameters like the basic reproduction number (R0), incidence rates, and outbreak duration [6, 7]. Such models inform decisions about vaccine allocation, social distancing policies, contact tracing, and other control measures by simulating the potential impact of interventions before they are implemented. However, when used in isolation, model predictions may lack the biological granularity needed to fully capture the heterogeneity of pathogen behavior in real populations. This depends on external data to inform parameters and validate results, limiting their utility when critical genomic information is absent.

The integration of real-time genomic surveillance with epidemiological models bridges this gap, creating synergistic insights that enhance intervention planning. Combining NGS data with dynamic models allows researchers to refine parameter estimates, improve accuracy in estimating transmission networks, and better understand how genetic variation influences epidemic trajectories [8, 9]. Studies integrating genomic and temporal data demonstrate how combined approaches can reconstruct transmission clusters and reveal the emergence of novel variants earlier than surveillance alone [8]. By incorporating both genomic and epidemiological inputs, integrated frameworks support more precise outbreak forecasting and provide evidence that can drive adaptive public health strategies, from targeted vaccination to localized non-pharmaceutical interventions.

Despite these advances, challenges remain that hinder widespread implementation of integrated surveillance and modeling systems, including technological barriers, data sharing constraints, and uneven capacity across regions. This narrative review aims to synthesize current knowledge on the integration of real-time genomic surveillance (NGS) with epidemiological models for infectious disease intervention planning. Specifically, it examines the methods and practical frameworks that combine genomic and epidemiological data, evaluates their contributions to public health decisionmaking, identifies limitations and gaps in current practice, and highlights opportunities for future research and policy development. Through an in-depth exploration of these intersections, this review seeks to offer researchers and practitioners a clearer understanding of how connected genomic and modeling approaches can improve outbreak prediction and intervention effectiveness in diverse epidemiological contexts.

Methodology

A systematic literature search was conducted in several electronic databases, including PubMed, Dimensions, and Google Scholar. The search strategy employed a combination of key terms such as "Next-Generation Sequencing", "genomic surveillance", "epidemiological models", "infectious disease intervention", and "outbreak forecasting" to capture relevant studies published between 2010 and 2025.

The inclusion criteria focused on peer-reviewed articles, reviews, and studies that discussed the application of NGS in infectious disease surveillance and its integration with epidemiological models. Exclusion criteria included studies that did not address NGS technologies or epidemiological modeling, and articles published in languages other than English.

Data extraction involved summarizing key findings from the included studies, including the methods of integration between genomic data and epidemiological models, their applications in outbreak forecasting, and their impact on intervention strategies. A thematic synthesis approach was used to categorize the findings into broader themes, such as enhanced outbreak prediction, variant tracking, and resource allocation.

The results were presented through a narrative synthesis, highlighting the strengths and challenges of integrating genomic surveillance with epidemiological models and offering insights into future directions for global health surveillance strategies.

Understanding Genomic Surveillance and NGS Technology

Next-Generation Sequencing (NGS) represents a fundamental shift in how scientists read, interpret, and employ genetic information for public health. Where DNA sequencing once required years of painstaking laboratory work and high cost, NGS allows millions of DNA fragments to be sequenced simultaneously, dramatically reducing both time and resources needed for whole genome analysis [10]. These parallel sequencing approaches have matured over three generations of sequencing technologies, beginning with the first-generation methods like Sanger sequencing and evolving to high-throughput second-generation platforms such as Illumina, Ion Torrent, and third-generation long-read technologies like Oxford Nanopore and Pacific Biosciences that can sequence entire molecules without amplification [10, 11]. This evolution underpins genomic surveillance as it is practiced today, enabling public health systems to observe pathogen genomes in real time and with deep resolution.

Genomic surveillance refers to the systematic collection, analysis, and interpretation of pathogen genome data to understand infectious disease dynamics at the molecular level. Unlike classical epidemiology, which draws primarily on case counts, symptom reports, and contact tracing, [3]. This molecular perspective enhances traditional surveillance by revealing transmission chains, detecting emerging variants before they become dominant, and uncovering subtle genetic changes linked to antimicrobial resistance or altered virulence [3]. With the advent of affordable and scalable sequencing technologies, genomic surveillance has emerged as an essential tool for monitoring diseases from foodborne outbreaks to global pandemics, as it provides both breadth and depth of information not achievable through conventional methods alone [12].

At its core, NGS technology revolutionized genomic surveillance through its ability to sequence vast amounts of nucleic acid simultaneously. Classical methods such as Sanger sequencing required targeted amplification of specific genomic regions and were limited by both scale and cost. In contrast, NGS platforms fragment genomes into millions of overlapping short reads that are sequenced in parallel and then reassembled into complete or near-complete genomes by bioinformatics tools[10, 13]. These workflows support multiple genomic approaches, including whole-genome sequencing (WGS), targeted sequencing, and metagenomic sequencing. WGS analyzes the entire genome of a pathogen, offering the highest resolution for characterizing genetic variations and phylogenetic relationships [14]. Targeted sequencing enriches for genomic regions of interest, enhancing sensitivity for known genes or markers, while metagenomic sequencing — especially useful for samples where the pathogen is unknown or present at low abundance — captures all genetic material in a specimen, including host and microbial DNA, without prior culture [13].

To appreciate how NGS fits within the broader landscape of genomic surveillance methods, Table 1 provides a comparative overview of major genomic surveillance techniques, including PCR-based tools and culture-based approaches. This table highlights differences in sensitivity, turnaround time, cost, and application scope between contemporary high-resolution genomic methods and traditional laboratory diagnostics. Understanding these distinctions is important for both researchers and public health practitioners as they design surveillance strategies tailored to specific pathogens, populations, and resource settings.

Table 1: Overview of Different Genomic Surveillance Methods

Method	Description	Sensitivity/Resolution	Time to Result	Cost	Usage Context
Whole-Genome Sequencing (WGS)	Comprehensive sequencing of pathogen genomes to detect mutations and strain differences	Very high genetic resolution, strain level	~24–72 h	Moderate to High	Outbreak investigation, transmission tracking, AMR surveillance
Targeted NGS	Sequencing of specific genomic regions using probes or multiplex PCR	High for chosen targets	~24–48 h	Moderate	Focused surveillance (e.g., resistance genes)
Metagenomic NGS (mNGS)	Unbiased sequencing of all DNA/RNA in a sample, culture-free	Broad detection of known/novel pathogens	~24–72 h	High	Unknown or polymicrobial infections, pathogen discovery
PCR-based Typing (e.g., MLST, rep-PCR)	Amplifies specific gene targets to infer strain differences	Moderate	Hours to days	Low to Moderate	Basic strain typing, small-scale surveillance
Culture-based Diagnoses	Traditional growth of pathogens in laboratory media	Variable; high for culturable organisms	Days to weeks	Variable	Clinical diagnosis, antimicrobial susceptibility phenotyping

[Sources: Compiled and synthesized from Illumina (n.d.) [12]; Papamentzelopoulou et al. (2025) [14]; Tiwari (2025) [3]; Cason et al. (2022) [15]; Lai et al. (2025) [16]]

Table 1 illustrates why NGS methods have become indispensable. Whole-genome sequencing, for example, delivers strain-level resolution that can trace transmission pathways across geographical and temporal scales. This depth is critical during outbreaks of rapidly mutating viruses or bacterial pathogens that develop resistance. Metagenomic NGS extends surveillance into cases where culture-dependent methods fail, such as detecting hard-to-grow pathogens directly from clinical or environmental samples [13]. In contrast, PCR and culture remain valuable for baseline surveillance and clinical diagnostics, especially where resources are limited, but they lack the comprehensive detection capacity and resolution of NGS-based approaches.

A defining strength of NGS in genomic surveillance is its ability to generate data that simultaneously inform multiple facets of disease dynamics. Because NGS does not require prior knowledge of the pathogen (unlike PCR, which needs target-specific primers), it can identify unexpected or emerging agents in a single run [16]. Moreover, combining sequencing with bioinformatics analysis supports identification of antimicrobial resistance genes, virulence factors, and phylogenetic relationships, thus informing both clinical and public health responses [14]. This versatility enables a shift from reactive outbreak confirmation to proactive surveillance and early warning systems, where changes in pathogen genomes can signal shifts in transmission potential or therapeutic vulnerability ahead of clinical trends.

Understanding genomic surveillance and NGS technology provides essential context for exploring how these systems integrate with epidemiological models in infectious disease intervention planning. NGS platforms and analytical strategies have matured to the point where they no longer serve solely as research tools but form the backbone of modern public health surveillance. When combined with other data sources and analytical frameworks, such as dynamic transmission models, genomic surveillance contributes not only to deeper biological understanding but also to evidence-based policy decisions that can save lives and guide resource allocation during health emergencies.

Epidemiological Models: An Overview

Epidemiological models stand at the heart of understanding and responding to infectious disease outbreaks. These models are mathematical constructs that simplify complex biological processes so that public health officials, researchers, and policymakers can anticipate disease trajectories, evaluate control strategies, and allocate limited resources more effectively. Among the earliest and most foundational of these models are compartmental frameworks, which segment a population into distinct categories or "compartments" based on disease status. The Susceptible-Infectious-Recovered (SIR) model, originally formalized by Kermack and McKendrick in the early 20th century, represents one of the most enduring and influential compartmental frameworks in epidemiology [17, 18]. In its simplest form, the SIR model depicts how susceptible individuals become infected and then recover, providing insights into dynamics such as the rate of spread and the eventual size of an outbreak.

Classical models such as SIR are grounded in differential equations that describe how individuals transition between states over time [17]. Susceptible individuals (S), who have not yet encountered the pathogen, may become infectious (I) after exposure; those who recover (R) are assumed to gain immunity or be removed from the pool of those capable of further transmission. The rates of these transitions depend on key parameters such as the transmission rate and recovery rate, which dictate how quickly a disease spreads and how long individuals remain infectious. The basic reproduction number, R0, emerges from this framework as a central concept defining whether a disease will expand or die out in a population: when R0 exceeds unity, each infected individual, on average, infects more than one other person, making sustained transmission likely [7, 17].

More sophisticated models have evolved from this foundational SIR structure to capture additional biological and social complexity. For instance, the Susceptible-Exposed-Infectious-Recovered (SEIR) model introduces an "exposed" compartment to account for incubation periods in which individuals have been infected but are not yet infectious [19]. This extension provides a more accurate representation of diseases with substantial latent periods — for example, influenza and COVID-19 — where there is a meaningful delay between exposure and onward transmission. Other variants include models that incorporate vaccinated individuals, age structure, spatial heterogeneity, and additional states reflecting hospitalization or quarantine. Such extensions reflect the recognition that real-world epidemics are influenced by many factors beyond simple infection and recovery transitions, including behavior, immunity, and public health interventions [19, 20].

While these compartmental frameworks have been invaluable for simulating disease progression, they traditionally depend on aggregate epidemiological data such as incidence, prevalence, and mortality counts. Outcomes from these models are primarily driven by changes in population-level metrics over time. However, when genomic data enters the equation, models gain a new dimension of biological specificity and predictive nuance. Integration of genomic sequencing information with classical epidemiological models — often termed phylodynamic modeling in the literature — allows models to simultaneously account for the evolutionary history of pathogens and their transmission dynamics in populations [8]. For example, methods such as birth–death SIR models combine phylogenetic inference with compartmental dynamics to jointly estimate epidemic parameters and the genealogical history of viral sequences, providing richer insights into how infection spread and diversification occurred [21].

Incorporating genomic data into epidemiological models enhances their precision in several ways. First, sequence variation acts as a molecular record of transmission events, enabling inference about who infected whom, even when epidemiological links are not directly observed in case reports [8]. Second, genomic data provide resolution for distinguishing between multiple introductions of a pathogen into a community versus sustained local transmission, which is critical for tailoring intervention strategies. Third, genetic signals can reveal changes in pathogen traits — such as increased transmissibility or immune escape — that influence model parameters and, consequently, projections of epidemic trajectories. These integrative approaches have been essential during recent outbreaks, including Ebola and COVID-19, where rapid sequencing informed dynamic estimates of transmission rates and guided targeted control efforts [8, 22].

To conceptualize how traditional epidemiological models intersect with genomic surveillance data, Figure 1 presents a flowchart depicting how NGS outputs are integrated with compartmental model frameworks. The chart begins with real-time sequencing of pathogen genomes collected from surveillance systems. Raw sequence data enter a bioinformatics pipeline for quality control, alignment, and variant calling. Cleaned genomic profiles then feed into phylogenetic analyses that reconstruct patterns of relatedness among isolates. These phylogenetic insights can be translated into quantities such as effective population size or transmission clusters. Meanwhile, temporal case data continue to populate the epidemiological model. At the convergence point, genomic and case data inform a hybrid model that adjusts compartment transition rates and estimates underlying parameters such as the time-dependent reproduction number (Rt). Outputs from this integrated model inform public health decisions about interventions such as social distancing guidelines, vaccination strategies, and travel restrictions.

The relevance of model parameters in infectious disease control cannot be overstated. Transmission rates reflect how frequently an infectious individual passes the pathogen to susceptible hosts; recovery rates determine how quickly individuals exit the infectious pool. Other parameters, such as latency periods in SEIR models, shape the timing of outbreaks and peak burden. When genome-informed, these parameters also reflect the genetic diversification of pathogens, which can affect transmissibility and immune recognition. The effective reproduction number, for instance, can vary over time as new variants arise or as population immunity shifts — whether through infection, vaccination, or behavioral change. Integrating genomic data enables models to dynamically adjust these parameters based on observed genetic changes, offering a more nuanced projection of epidemic evolution than models based solely on traditional epidemiological data.

Ultimately, epidemiological models remain the scaffolding upon which much of infectious disease forecasting and intervention planning is built. Advances in integrating real-time genomic surveillance transform these models from abstractions based purely on compartment counts into biologically grounded frameworks that capture the interplay between pathogen evolution and population dynamics. This evolution — from simple models like SIR to genomically enriched phylodynamic frameworks — underscores a broader shift toward precision public health, where data at multiple scales inform timely and effective responses to emerging infectious threats.

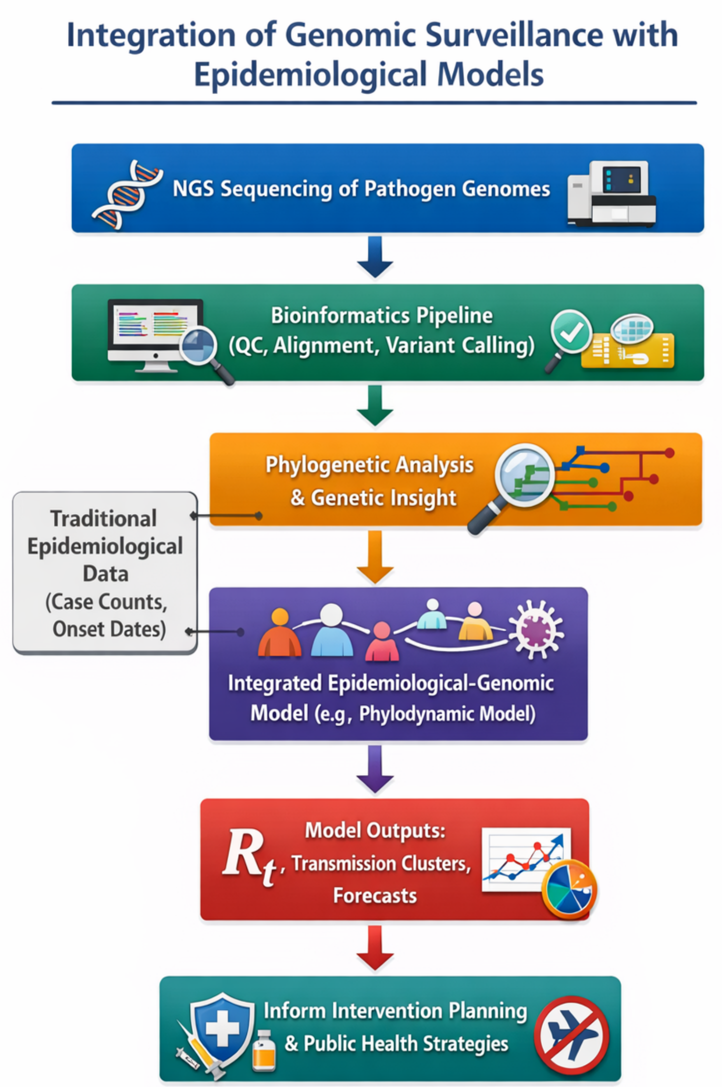

Figure 1: Flowchart of Integration of Genomic Surveillance in Epidemiological Models in Disease Surveillance

NGS Data Integration in Epidemiological Models

Integrating Next-Generation Sequencing (NGS) data into epidemiological modeling marks one of the most transformative advancements in infectious disease science over the last decade. At its essence, this integration merges genetic information from pathogens with traditional disease surveillance inputs — such as onset dates, case counts, geographic metadata, and contact histories — enabling models to capture not just "how many" cases are occurring but "how and why" infections spread in particular patterns [3, 23]. Rather than relying solely on temporal counts, integrated models leverage genomic sequences as additional evidence of transmission continuity, directionality, and evolutionary change, enriching model structures traditionally built on compartmental dynamics (SIR/SEIR) or statistical forecasting alone.

To understand how genomic data are embedded in epidemiological models, it helps to conceptualize the typical data pipeline. After pathogen genomes are sequenced via NGS technologies, raw reads undergo stringent quality control, alignment, and variant calling, producing high-confidence genomic profiles for each sample. These profiles — often represented as consensus sequences or SNP (single-nucleotide polymorphism) matrices — can be combined with sample metadata (e.g., collection date, location, host clinical information) in phylogenetic or phylodynamic models that reconstruct both evolutionary relationships and transmission pathways [8, 23]. In this framework, phylogenetic trees provide a scaffold for hypothesized genetic linkage among isolates, while epidemiological data anchor the timing, location, and context in which these infection events occurred [8, 24]. The result is a hybrid model that "triangulates" evidence from genetic relatedness, temporal progression, and case characteristics to infer quantities such as who infected whom, the effective reproduction number (R_t), and the relative contribution of different transmission pathways.

One compelling illustration of this integration is the use of transmission tree reconstruction algorithms that marry genomic and epidemiological signals. Duault and colleagues reviewed a family of methods specifically designed to infer transmission trees — detailed maps of infection event connections — by simultaneously incorporating genetic distances and case metadata (e.g., onset dates, removal or recovery times). Some approaches iterate between phylogenetic tree inference and transmission inference, while others embed both processes into a single probabilistic model that accounts for within-host diversity, mutation rates, and the timing of infections [8]. These transmission tree methods differ in complexity and assumptions but share the core principle that sequence data and epidemiological observations together produce far tighter estimates of transmission pathways than either data type alone. This integrated inference is especially pertinent in outbreaks where traditional epidemiological links (e.g., contact tracing records) are incomplete or ambiguous — a common scenario during large, fast-spreading epidemics.

The COVID-19 pandemic provided one of the most extensive real-world laboratories for integrating NGS data with epidemiological modeling. National and international sequencing consortia — including regional initiatives like the UK's COVID-19 Genomics UK Consortium (COG-UK) — generated hundreds of thousands of whole viral genomes, which were then coupled with detailed case metadata (collection dates, locations, patient demographics) to monitor the emergence and spread of key SARS-CoV-2 variants in near real time [23, 25]. By feeding time-stamped genomic data into phylodynamic models, researchers were able to estimate the effective reproductive number of emerging variants relative to ancestral strains, map their geographic expansion, and assess the impact of public health interventions such as lockdowns and travel restrictions on slowing transmission [23, 25]. Without such genomic integration, distinguishing between slowed transmission due to an intervention versus slowed reporting due to sampling lags would have remained far more uncertain.

Other infectious diseases have similarly benefited from this integrated modeling approach. A study reconstructing tuberculosis transmission in British Columbia, Canada, demonstrated that combining pathogen genomes with epidemiological metadata — including HIV status and social network information — yielded more accurate reconstructions of outbreak clusters than analyses relying only on genomic or case data alone [24]. The inclusion of detailed epidemiological metadata alongside sequence data allowed the model not only to infer probable transmission links with greater confidence but also to correlate transmission dynamics with host factors such as comorbidities, which can influence susceptibility and infectiousness. In avian influenza outbreaks, integrated analyses that used both spatial case data and genomic sequences helped delineate how geographic spread intersected with genetic diversification, offering insights into both transmission routes and source populations [24]. These practical instances demonstrate that genomic integration is not a theoretical enhancement but a crucial component of high-resolution outbreak understanding.

Despite these powerful outcomes, integrating NGS data into epidemiological models is beset with notable challenges. A recurring issue highlighted across multiple studies is data completeness and sampling bias. While genomic data are insightful, they typically represent only a subset of all infections, constrained by sequencing capacity, sampling design, and logistical limitations [24, 25]. When genome sequences are missing for key transmission links — a common occurrence during early or uncontrolled outbreaks — inferred transmission trees can be biased or uncertain, as the model must reconcile incomplete genetic evidence with observed case patterns. Even when genomic data are available, inconsistent metadata (e.g., missing onset dates, inconsistent geographic identifiers) degrade the ability of models to leverage combined signals effectively [24, 26].

Another set of challenges arises from computational complexity and model assumptions. Hybrid models that integrate genome sequences with epidemiological dynamics often require advanced statistical frameworks such as Bayesian inference, coalescent theory, or stochastic modeling, which are resource-intensive and sensitive to prior assumptions about pathogen mutation rates, transmission processes, and within-host evolution [8, 24]. For fast-evolving pathogens like RNA viruses, high mutation rates can yield complex phylogenies that, if modeled inadequately, may introduce inaccuracies in the inferred transmission relationships. Moreover, within-host diversity — the presence of multiple genetic variants within a single infected individual — poses additional complications for models that assume a single representative genome per case [26].

Integration across data systems and governance also remains a structural barrier in many settings. Genomic data are typically stored in specialized databases (e.g., GISAID, GenBank), while epidemiological case data reside in public health information systems. Harmonizing these data streams — particularly across institutional, national, or international boundaries — demands interoperable standards, shared metadata frameworks, and clear governance around data privacy and sharing agreements [3, 26]. Without such integration, efforts to link genomic sequences with case data can falter due to incompatible formats, missing identifiers, or regulatory constraints.

Finally, interpretation and actionable translation of integrated model outputs present another layer of challenge. While integrated models produce rich estimates — including reproduction numbers, transmission clusters, and lineage dynamics — translating these outputs into clear public health guidance (e.g., prioritizing vaccination, imposing restrictions, targeting contact tracing) requires sustained collaboration across disciplines. Public health decisionmakers must understand not only what the model predicts but also its limitations and uncertainty, especially when model outputs inform high-stakes decisions during rapidly evolving outbreaks.

Therefore, integrating NGS data into epidemiological models has significantly advanced our capacity to interpret infectious disease dynamics, enabling fine-grained reconstructions of transmission pathways, real-time monitoring of variant emergence, and enriched forecasting of epidemic trajectories. Case studies in COVID-19, tuberculosis, and influenza illustrate both the potential power and practical complexity of these methods. Yet, realizing the full promise of genomic integration demands continued attention to data completeness, computational innovation, interoperable systems, and collaborative translation into public health practice. As genomic technologies advance further and integrated modeling frameworks continue to mature, their application promises to strengthen both the scientific understanding of pathogens and the strategic planning that protects populations from future infectious threats.

Applications in Infectious Disease Intervention Planning

When public health systems face the relentless challenge of infectious disease threats, the value of integrating Next-Generation Sequencing (NGS) with epidemiological models becomes acutely practical. Far beyond creating academic maps of pathogen evolution, this integration yields real-world insights that help planners decide when, where, and how to intervene. During the COVID-19 pandemic and beyond, NGS-informed models have shaped vaccination strategies, guided non-pharmaceutical interventions, and enhanced the responsiveness of surveillance systems to emerging risks. These applications illustrate how genomic information — interpreted not in isolation but within dynamic disease models — can transform raw data into actionable strategies that protect communities.

At the core of intervention planning lies the goal of identifying patterns of transmission with fine-grain clarity. Traditional models, which rely on case counts and reported exposures, are limited by reporting lags, incomplete contact histories, and the inability to distinguish between genetically distinct strains that behave differently in terms of transmissibility or immune escape. Genomic surveillance closes this gap by providing molecular signatures that signal how a pathogen is spreading and evolving in almost real time [2, 3]. For example, in the response to SARS-CoV-2, the continuous sequencing of viral genomes enabled public health authorities to detect "variants of concern" and "variants under investigation" promptly, informing decisions about border controls, resource allocation, and targeted testing campaigns [27]. By merging these genetic insights with temporal and geographic case data, models began to reflect which variant was contributing most to transmission at any given time, and how this influenced the course of local outbreaks.

Embedded within this practical framework is the power of genomic forecasting. Once genetic data are integrated into models, shifts in variant prevalence become leading indicators rather than trailing observations. In applications where models were calibrated to include mutation frequencies and strain-specific transmission parameters, it became possible to anticipate not just the next week's caseload but the potential impact of emergent mutations on future case trajectories [2]. These forecasts have direct consequences for vaccination planning. When models suggest that a new variant exhibits partial immune escape — that is, reduced vaccine effectiveness — planners can respond by adjusting vaccine composition, prioritizing booster campaigns, or intensifying non-pharmaceutical public health measures in vulnerable areas [2, 28]. This capacity to forecast variant impacts was unavailable to traditional models devoid of genomic input.

The growing field of precision public health exemplifies this translation of genomic data into actionable policy. Precision public health adapts principles from precision medicine — targeting the right intervention to the right population at the right time — to infectious disease control by leveraging granular data across diverse sources, including pathogen genomics, human behavior, and environmental context [17]. In the Netherlands, combined whole-genome sequencing and case data were used to characterize community transmission patterns of SARS-CoV-2, leading to more informed decisions about mass gatherings and school closures than would have been possible with classical surveillance alone [17]. In such examples, genomic data help distinguish true community spread from isolated clusters, shaping interventions that are proportionate and well-timed.

Another practical application of genomic integration appears in antimicrobial resistance (AMR) surveillance. By comparing pathogen genomes over time, NGS identifies resistance genes and tracks their prevalence within and across communities [3]. Models built around such genomic insights can alert health systems to spreading resistance faster than phenotypic testing alone. In response, planners might adjust antibiotic stewardship programs or deploy targeted infection-control measures in high-risk settings, such as hospitals or long-term care facilities, where resistant strains have taken hold [3, 29]. These responses, shaped by integrated models, strive not only to reduce current disease burden but to forestall future challenges that resistant pathogens may present.

Successful genomic-assisted intervention planning also depends on standardized tools and shared data ecosystems that support real-time analysis. Platforms like GISAID and Nextstrain play central roles in aggregating sequencing data from around the world and rendering phylogenetic analyses accessible to public health officials and researchers alike [30, 31]. With more than 17 million SARS-CoV-2 genomes shared globally, these platforms have enabled coordinated responses that transcend national boundaries, allowing countries to observe rising variant trends abroad and prepare their intervention strategies accordingly. Shared lineage nomenclatures, like the Pango system, help ensure that models interpret genomic signals consistently, strengthening comparability and coordination across regions [32].

Despite these advances, practical challenges remain in the implementation of integrated genomic intervention planning. Data gaps — arising from inequitable sequencing capacity, delayed sharing, or incomplete metadata — can skew model outputs and therefore the recommendations that flow from them. Resource constraints in low- and middle-income countries, where sequencing infrastructure and bioinformatics expertise are less developed, can leave critical blind spots in global surveillance networks and slow the detection of emergent variants [29, 33]. Addressing these disparities through capacity building and investment in genomic infrastructure remains essential for enhancing the effectiveness of intervention planning worldwide.

In summary, NGS-integrated models function as powerful tools for guiding infectious disease intervention strategies, enabling a nuanced understanding of transmission dynamics, informing vaccination decisions, and supporting outbreak predictions and resource deployments. By embedding genomic data into analytical frameworks that inform public health decisions, planners can navigate the complex terrain of evolving pathogens with evidence that is both timely and biologically precise. As sequencing becomes more widespread and integrated with ever-improving analytical models, this approach promises to make infectious disease intervention more anticipatory, tailored, and impactful.

Real-Time Genomic Surveillance for Emerging Infectious Diseases

Emerging infectious diseases present a persistent and dynamic threat to global health security as populations grow, travel increases, and ecological interfaces between humans and animals expand. In recent decades, outbreaks ranging from Ebola virus disease in West Africa to the COVID-19 pandemic have underscored the limitations of conventional epidemiological surveillance systems, which depend heavily on symptom reporting, laboratory confirmations, and contact tracing that often emerge too late to influence timely interventions. The advent of real-time genomic surveillance — enabled by rapid sequencing technologies and high-throughput data pipelines — has reshaped how outbreaks are detected, characterized, and contained, placing genetic evidence at the core of real-world public health decisionmaking.

Real-time genomic surveillance refers to the continuous generation and analysis of pathogen genetic sequences as outbreaks unfold. Unlike retrospective studies that analyze genomes after an epidemic wave has passed, real-time systems seek to deliver actionable insights within hours or days of sample collection. This capacity has profound implications for how emerging diseases are managed. During the Ebola outbreak in West Africa, for example, mobile sequencing platforms such as Oxford Nanopore's MinION were deployed directly in outbreak zones, enabling researchers to generate and analyze genomes in less than 24 hours after a positive sample was obtained. Such rapid turnaround permitted high-resolution views of the Ebola virus's evolution and provided invaluable information to outbreak response teams about transmission chains and viral diversification during the crisis, even in resource-limited settings where conventional laboratory infrastructure was sparse [34].

The COVID-19 pandemic amplified the visibility and relevance of real-time genomic surveillance on a global scale. Sequencing platforms and genomic consortia — including regional collaborations and global data repositories like GISAID — generated millions of SARS-CoV-2 genomes within months of the virus's emergence. This unprecedented volume of data allowed surveillance systems to track the appearance and spread of viral variants in near real time, rather than weeks or months later, providing early warnings when mutations of potential clinical or epidemiological importance emerged. Tools such as dynamic lineage classification systems (e.g., PANGO) and rapidly updated phylogenetic trees enabled public health officials to visualize variant emergence and geographic expansion as it occurred, boosting the speed and accuracy of their responses [30].

The impact of real-time genomic data on intervention strategies is multifaceted. First, it improves detection and characterization of novel pathogens and variants early in their transmission cycles, allowing health authorities to adjust diagnostic assays, refine case definitions, and update treatment protocols to align with the most current genetic information. Early identification of unique mutations can guide laboratory diagnostic design and help prevent false negatives in molecular tests — an essential feature in rapidly evolving outbreaks. In addition, because genomic data capture evolutionary changes that may influence traits like transmissibility or immune escape, they allow models to anticipate shifts in epidemic behavior before they manifest as clinical surges [3].

Second, real-time genomic surveillance enhances epidemiological inference by differentiating between multiple introductions of a pathogen and sustained local transmission. For example, genomic analyses of imported COVID-19 cases in Beijing revealed patterns of transmission that could not have been distinguished based on case counts alone — insights that helped evaluate the effectiveness of border screening policies and targeted quarantines. By classifying sequences into discrete genetic clusters linked to known exposure histories, models could attribute cases to international introductions versus community spread, refining public health strategies accordingly [35].

Third, real-time sequencing supports variant monitoring and risk assessment. As viral genomes accumulate mutations, some changes can confer biological advantages — such as increased transmissibility or reduced sensitivity to antibodies. When genomic surveillance detects these variants early and in sufficient concentration, intervention strategies can adapt preemptively: vaccines can be updated, booster campaigns can be timed and targeted, and non-pharmaceutical measures like mask mandates or social distancing can be calibrated in anticipation of waves driven by more transmissible strains. Without timely sequence data, such pivotal decisions risk lagging behind the pathogen's spread, leading to preventable morbidity and mortality [2].

The speed and accuracy afforded by real-time genomic approaches also enhance predictive modeling and public health preparedness. Integrated epidemiological models that incorporate real-time genomic data generate more precise estimates of key transmission parameters — including reproductive numbers (R_t), infection growth rates, and the effective impact of control measures. These parameters are essential for modeling scenarios under different intervention strategies, such as vaccination prioritization or targeted closure policies. Models informed by real-time genomic signals typically outperform those relying solely on case counts, as genetic variation carries granular clues about recent transmissions and emerging clusters that traditional metrics may obscure. Such precision supports evidence-based decisions at the strategic and operational levels of outbreak response [2].

Despite these strengths, real-time genomic surveillance still faces implementation challenges that temper its promise. Many low- and middle-income countries, particularly in sub-Saharan Africa and parts of Asia and Latin America, lack sufficient sequencing infrastructure, trained personnel, and data analytics capacity to produce and interpret genomic data on a continuous basis. While more than half of African countries have established in-country sequencing capabilities since the COVID-19 pandemic, gaps remain in data integration, sustainable funding, and equitable access to bioinformatics tools. These disparities can slow the generation of real-time insights and limit the global surveillance network's responsiveness to emerging threats [33]

Interoperability and data-sharing policies also influence the utility of real-time genomic surveillance. Global repositories like GISAID have been central to pandemic response by aggregating and disseminating sequence data from around the world, but challenges related to governance, data ownership, and equitable participation persist. Ensuring that genomic data are shared rapidly and responsibly — with appropriate protections for patient privacy and ethical use — remains a priority for sustainable global surveillance [30].

Finally, real-time genomic surveillance is most effective when coupled with robust epidemiological context and public health infrastructure. Sequenced genomes need associated metadata on case onset dates, geographic location, clinical severity, and exposure history to be most informative. Without these contextual layers, the interpretive power of genomic data is weakened, and models may misattribute patterns or overlook critical outbreak dynamics. Building systems that link genomic, clinical, and epidemiological data in real time across sectors and institutions continues to be a technical and collaborative challenge for surveillance networks worldwide.

In sum, the integration of real-time genomic surveillance into outbreak detection and response systems has revolutionized the management of emerging infectious diseases. From rapid identification of pathogens and variants to enhanced precision in forecasting and intervention planning, genomic data have accelerated the speed and accuracy of decision-making during crises like COVID-19 and Ebola. As genomic technologies continue to evolve and global capacity expands, real-time surveillance will remain a cornerstone of resilient public health systems that can anticipate, track, and counter future biological threats.

Challenges and Limitations of Real-Time Genomic Surveillance in Epidemiology

Real-time genomic surveillance has emerged as a powerful pillar of modern infectious disease epidemiology, offering high-resolution insight into pathogen evolution, transmission dynamics, and outbreak trends. Yet beneath its promise lies a complex web of challenges that shape how effectively genomic data can be used in real-world settings. These challenges span technical constraints, ethical questions around data use, structural limitations in global surveillance capacities, barriers to cross-border data sharing, and analytical dilemmas that affect the accuracy of models and the interpretation of genomic signals. Understanding these limitations is essential for both practitioners and policymakers as they seek to build resilient surveillance systems capable of supporting rapid and equitable responses to emerging infectious threats.

Technical hurdles form a foundational challenge in real-time genomic surveillance. High-throughput sequencing platforms and bioinformatics pipelines demand significant infrastructure, sustained funding, and specialized workforce skills that are still scarce in many regions. Even in well-resourced settings, gaps in expertise for quality control, sequence assembly, variant calling, and downstream interpretation can create bottlenecks that delay data availability or compromise data quality. A review of global genomic surveillance strategies during the COVID-19 pandemic observed that limitations in genotyping standardization and bioinformatic expertise complicate data processing and interpretation, leading to delays in reporting and reducing the utility of datasets for timely public health action [36]. These delays matter because the value of genomic surveillance depends on the speed with which data can feed back into decision-making; delays transform "real-time" into "near retrospective," weakening the intended agility of surveillance systems.

Closely intertwined with technical challenges are issues of data privacy and ethical use. Genomic data, by its nature, is deeply personal and by extension ethically sensitive. Although the primary focus in pathogen surveillance is on microbial genomes rather than human genomes, linkage of pathogen sequences with clinical, demographic, or location information can inadvertently reveal personal or community health details. The risk of re-identification, particularly when genomic data are combined with other health records, raises legitimate privacy concerns. Discussions in the genomics ethics literature emphasize the need to balance the open sharing of data for public health benefit with robust protections for individual privacy and consent frameworks [37]. Without thoughtful governance structures, well-intended surveillance systems risk eroding public trust, particularly in contexts where communities have historically experienced misuse of health or genomic data.

A related limitation lies in the accessibility and equity of genomic surveillance across different regions and populations. While high-income countries have rapidly scaled sequencing capacity and analytical infrastructure, low- and middle-income countries often remain underrepresented in global genomic datasets due to resource constraints. This inequity creates "blind spots" in global surveillance that can hinder early detection of variants or emerging pathogens in regions with weaker capacity. Researchers have highlighted that lack of sequencing equity not only limits the global understanding of pathogen evolution but also hampers local response efforts where they may be most needed [38]. These disparities raise ethical questions about the distribution of scientific tools and benefits, and call for sustained investments in capacity building to make genomic surveillance an inclusive global public good.

Cross-border data sharing stands as another formidable challenge. Effective real-time surveillance depends on the rapid exchange of sequencing data, associated metadata, and analytical results across national, institutional, and disciplinary boundaries. However, practical and political barriers often slow or distort this exchange. For example, platform governance, legal frameworks around data ownership, intellectual property concerns, and varying levels of digital infrastructure can fragment how sequence data are shared and re-used. Studies of viral data sharing during COVID-19 pointed out that data were sometimes released too late, incompletely, or without sufficient metadata, undermining their potential to inform outbreak responses internationally. Meanwhile, systems such as the GISAID database — popular for sharing pathogen genomic data — require identity verification and specific access agreements precisely to protect data contributors, yet these protections can also complicate rapid open access in emergency settings [39, 40].

Even with high-quality and timely data, model accuracy and genomic interpretation pose additional limitations. Integrated genomic-epidemiological models rely on assumptions about mutation rates, sampling representativeness, and the relationship between genetic change and phenotypic traits like transmissibility or virulence. When sampling is biased — for example, when convenience sampling overrepresents certain regions or clinical settings — computational models can produce distorted estimates of variant prevalence or epidemiological parameters. Reviews of genomic surveillance practice have shown that inconsistencies in clinical and demographic metadata can slow the compilation of comprehensive datasets, reducing confidence in model outputs and complicating interpretation. These analytic uncertainties underscore that genomic data alone are not self-explanatory; they must be embedded within robust epidemiological contexts and interpreted with an understanding of underlying biases and limitations [36].

Moreover, misinterpretation of genomic signals can lead to either under- or over-estimation of public health risks. Complex phenomena such as within-host diversity, recombination events, or sequencing artifacts can be misread as indicators of new variants or transmission clusters when they are not, generating false alarms or obscuring true transmission paths. Addressing these analytical challenges requires continuous methodological advances — such as improved error models, rigorous phylogenetic inference techniques, and better standards for metadata quality — as well as ongoing training for analysts who must interpret genomic outputs for public health use.

Finally, ethical considerations extend beyond individual privacy to questions about data ownership, transparency, and governance. Pathogen genomic data can be politically sensitive, particularly when associated with outbreaks that have economic or diplomatic ramifications. Countries may hesitate to share data rapidly if they fear stigma, travel restrictions, or loss of control over how data are used. Overcoming this requires trust-building, enforceable governance frameworks that harmonize national interests with global public health benefits, and international agreements that balance openness with appropriate protections and recognition of contributions.

In summary, while real-time genomic surveillance has revolutionized the capacity to detect and respond to infectious threats, it is constrained by a constellation of technical, ethical, structural, and interpretive barriers. Tackling these challenges will require not only technological innovation but also equitable investment in global infrastructure, thoughtful governance frameworks for privacy and data sharing, and sustained efforts to strengthen the accuracy and interpretability of integrated genomic-epidemiological models. Only by addressing these limitations can the full potential of genomic surveillance be realized in guiding effective, timely, and ethical public health action worldwide.

Future Directions and Opportunities

As genomic surveillance solidified its role in public health during recent pandemics, researchers and decision-makers are increasingly looking ahead to the next frontier of infectious disease monitoring. While real-time sequencing has shifted the paradigm from retrospective analysis to near-real-time insight, the future promises even deeper integration of genomic data, predictive analytics, and advanced technologies that enhance both the speed and accuracy of epidemic forecasting. These developments will not only improve outbreak detection but will help transform genomic surveillance from a reactive tool to a proactive backbone of global health strategy.

One promising direction lies in improving epidemic forecasting through more sophisticated integration of genomic signals with epidemiological and environmental data. Traditional models often treat genomic information as an added layer of evidence in understanding transmission chains, but emerging methods envision genomic forecasting, where sequence variation itself becomes a predictive input. Genetic changes in a pathogen reflect evolutionary pressures, host immunity, and viral fitness, all of which shape how outbreaks unfold in real populations. Early evidence from machine learning studies suggests that models incorporating evolutionary signals outperform conventional time series models, detecting subtle shifts in variant prevalence that presage broader epidemiological changes [41, 42]. As computational methods become more refined, the capacity to predict not just outbreak magnitude but also variant-specific risk profiles will be crucial in deploying targeted control measures before case surges occur. This represents a fundamental shift from "respond when it happens" to "anticipate before it intensifies."

Cutting-edge technologies on the horizon are already accelerating this evolution. Artificial intelligence (AI) and machine learning (ML) stand out as engines of transformation, capable of parsing enormous, heterogeneous datasets that include genomic sequences, epidemiological records, mobility patterns, and socio-environmental indicators. AI-driven surveillance systems — particularly those leveraging deep learning and predictive analytics — can detect patterns invisible to traditional algorithms and human analysts. These tools have shown promise in early outbreak detection, anomaly recognition, and trend forecasting, effectively functioning as automated early-warning systems that flag aberrations in complex data ecosystems [42, 43]. For example, AI can process raw genomic sequences alongside non-traditional data streams such as social media trends or environmental sensors, flagging signals that might correlate with unusual transmission activity long before clinical case counts rise.

Machine learning specifically offers the capacity for self-improving models that refine predictions as more data accrue over time. Unlike static statistical models, ML algorithms "learn" from successive sequencing cycles and outbreak patterns, enabling dynamic re-estimation of parameters such as transmissibility, variant emergence likelihood, or regional risk profiles. This iterative learning enhances the resilience of surveillance systems, allowing them to adapt to new pathogens or shifting epidemiological landscapes [44]. As large-scale training datasets become available — especially from shared international repositories like GISAID — these models will become more robust and generalizable, providing crisis managers with tools that guide resource allocation and intervention planning before outbreaks reach crisis levels.

Cloud computing and big data analytics also play a vital role in this future ecosystem. Genomic data are inherently vast and complex, often requiring terabytes of storage and significant processing power to analyze effectively. Cloud platforms offer scalable infrastructure that can handle these demands while facilitating rapid data sharing across countries and institutions. By hosting analytical pipelines in the cloud, public health laboratories everywhere — including those in resource-limited regions — can access state-of-the-art analytics without investing in prohibitively expensive local infrastructure. Moreover, cloud-based systems enable collaborative analysis, where genomic and epidemiological data from diverse settings are integrated and visualized in near real time, fostering a shared situational awareness among global partners.

Other emerging technologies — such as blockchain for secure data sharing and real-time data streaming from Internet-of-Things (IoT) sensors — promise to enhance trust, transparency, and promptness in surveillance infrastructure. Blockchain can provide tamper-proof records of sequence metadata and analysis results, addressing concerns about provenance and data integrity that sometimes slow cross-border cooperation. Meanwhile, IoT-based streaming of sequencing outputs or environmental genomic signals (for example, from wastewater monitoring) can feed directly into predictive models, generating a continuous flow of actionable insight rather than batch-processed snapshots.

While the technological horizon is rich with possibility, advancing policy frameworks that support integration of genomic data into global disease monitoring remains essential. International governance structures must adapt to ensure equitable data access, standardized sequencing protocols, and interoperable metadata standards. The success of repositories like GISAID during the COVID-19 pandemic — where millions of SARS-CoV-2 genomes were rapidly shared — illustrates the power of coordinated data ecosystems [30]. However, this model must evolve to include clear policies on data privacy, ethical use of sensitive metadata, and equitable sharing of benefits from genomic insights. Policies that strengthen capacity building in low- and middle-income countries are also critical, ensuring that all regions can contribute to and benefit from global surveillance networks.

Governments and multilateral institutions should also incentivize cross-sector collaborations that bridge public health, academia, industry, and community stakeholders. Public-private partnerships can support the development of open-source tools and affordable sequencing technologies, while academic collaborations can ensure that model development remains transparent and scientifically robust. At the same time, public health agencies must invest in formal training programs that build the bioinformatics and data science workforce needed to operate these future systems effectively.

Taken together, these advancements and policy shifts point toward a future where genomic surveillance is not just a research pursuit but an embedded, proactive component of global health architecture — one that anticipates threats, informs policy in real time, and ensures that interventions are as precise as the data that drive them. Below is Figure 2, illustrating how emerging technologies could transform surveillance and modeling.

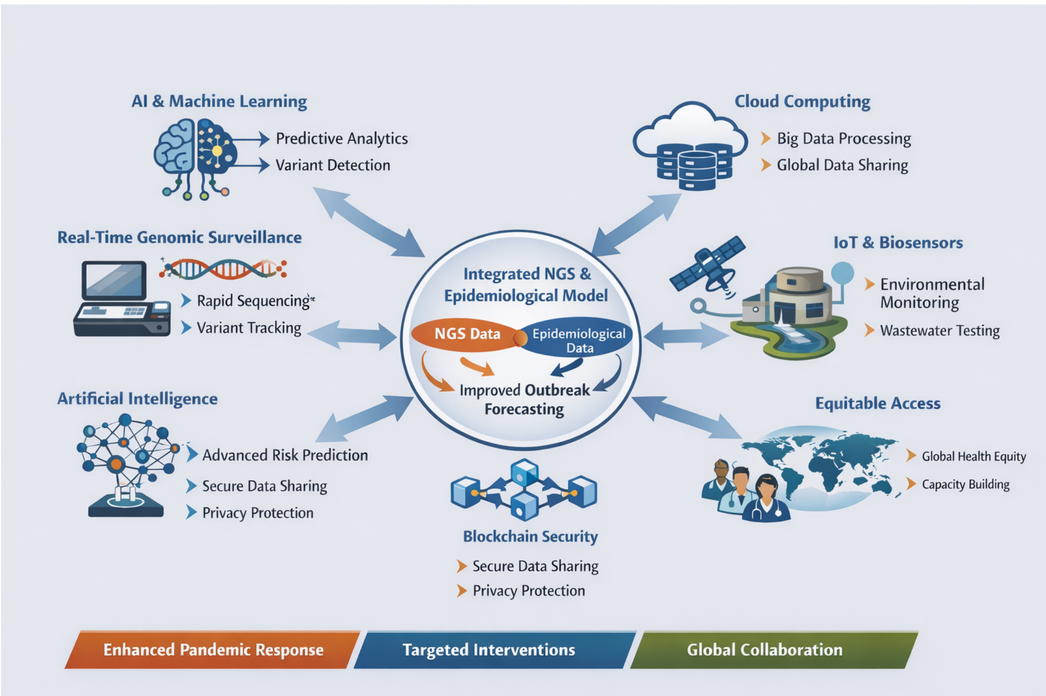

Figure 2: Potential Future Technologies Enhancing NGS-Epidemiological Model Integration

Ethical and Policy Implications of Integrated Surveillance

As real-time genomic surveillance becomes anchored in public health practice, the ethical and policy dimensions of its use come sharply into focus. The power to sequence pathogens continuously, share data rapidly, and embed genomic insights into decision-making creates extraordinary opportunities to protect populations from emerging threats. Yet it also raises deeply human questions about privacy, consent, equity, trust, and governance. These ethical considerations must be balanced with the public health benefits of surveillance, and policy strategies must be crafted to support responsible use of genomic data while safeguarding individual and community rights.

A central ethical issue in genomic surveillance arises from the nature of the data themselves. Pathogen genomes are not inherently human DNA, yet they carry links to individuals — through metadata such as sample origin, clinical history, or geographical location — that may inadvertently reveal sensitive personal information if mishandled [45]. International principles drafted by the World Health Organization emphasize informed consent, privacy protection, and transparency in data collection, access, and use, underscoring the importance of public trust in genomic initiatives [45]. Without clear safeguards, individuals may feel reluctant to participate in surveillance efforts, especially when data could be re-identified or used for purposes beyond immediate public health needs. This tension between utility and privacy echoes broader debates in health informatics, where ethical frameworks seek to maximize public benefit without compromising individual rights [46].

Another layer of ethical complexity involves equity of access and benefit sharing. The rapid sharing of genomic data during the COVID-19 pandemic highlighted glaring disparities between high-income and low-income regions. When the Omicron variant was first identified and shared by scientists in southern Africa, the ensuing travel bans imposed by many countries illustrated how data sharing can produce unintended economic and social harms that fall disproportionately on researchers and populations who contribute data [46]. Such reactions can undermine trust and discourage participation in future genomic surveillance, particularly in regions still building scientific capacity. Ethical frameworks therefore call not only for equitable access to genomic technologies but also for mechanisms that ensure fair distribution of benefits arising from shared data, including support for local research and public health infrastructure [45].

This imperative for fairness extends into global governance and the design of data-sharing policies. The longstanding Bermuda Principles and subsequent agreements like the Fort Lauderdale Agreement established norms for rapid data release within scientific communities, particularly in genomics research, to advance collective knowledge [47, 48]. Yet translating these norms into pathogen surveillance — where data may have immediate implications for outbreak trends and policy decisions — requires careful governance that balances rapid dissemination with accountability and respect for national sovereignty. Frameworks such as those proposed by the Africa Centres for Disease Control and Prevention aim to guide member states in developing national pathogen genomic policies that are locally relevant while supporting cross-border collaboration [49]. These policy approaches emphasize shared vision, contextualized strategies, and sustained investment in capacity building — all essential to ethical and resilient genomic surveillance systems.

Policy development must also address public-private partnerships and the social contract between citizens and institutions. As genomics initiatives increasingly involve commercial entities — for analytical platforms, cloud infrastructure, or AI tools — questions arise about who benefits from access to genomic information and how private interests intersect with public health goals. Research on ethical and social implications of such partnerships suggests that transparency, robust governance, and explicit benefit-sharing agreements are critical to maintaining public trust and ensuring that commercial collaborations serve the public interest rather than narrow commercial ends [50]. Public policy strategies thus need to articulate clear rules for data access, use constraints, and safeguards that align commercial activities with ethical public health missions.

To bring these ethical and policy strands together into a coherent structure, Figure 3 synthesizes the key elements required for governance and oversight (see below). The figure illustrates how ethical principles provide the foundation, guiding core policy domains including privacy and consent, equitable access and benefit sharing, data governance and interoperability standards, and international cooperation mechanisms. These policy domains are supported by implementation pillars such as workforce training, technological infrastructure, legal frameworks, and ongoing stakeholder engagement. The framework highlights feedback loops that ensure policy evolves with technological advances and emerging ethical insights, and underscores the need for monitoring and accountability measures that assess performance against both public health outcomes and ethical standards.

Importantly, genomic surveillance does not operate in isolation. The integration of pathogen genomics with epidemiological and clinical data requires interoperable systems that respect privacy while enabling efficient data flow. Standardization efforts such as those championed by the Genomic Standards Consortium promote common metadata standards that facilitate data exchange and quality assurance across platforms (Genomic Standards Consortium, 2025). Such standards must be embedded within policy frameworks that clearly define data access rights, custodianship, and security protocols to protect against misuse or unauthorized access.

Global collaboration is equally vital. Initiatives like the WHO's pathogen genomic data sharing principles and regional policy frameworks encourage international cooperation on ethical surveillance practices, capacity building, and equitable benefit realization. These collaborative structures must be grounded in mutual respect for sovereignty and cultural values, and tailored to diverse public health systems and regulatory environments. Only through alignment of ethical frameworks, legal instruments, and technical standards can genomic surveillance achieve the dual goals of protecting individual rights and enhancing collective health security.

Therefore, the ethical and policy implications of real-time genomic surveillance are multifaceted, touching on privacy, equity, trust, governance, and international cooperation. Navigating these issues requires policy approaches that are both principled and pragmatic — grounded in ethical values that protect individuals and communities, and designed to support the effective, equitable use of genomic data for public health. By embedding strong ethical safeguards and adaptive governance structures into surveillance systems, policymakers can foster public trust and international solidarity while harnessing the power of genomic science to anticipate, understand, and mitigate infectious disease threats.

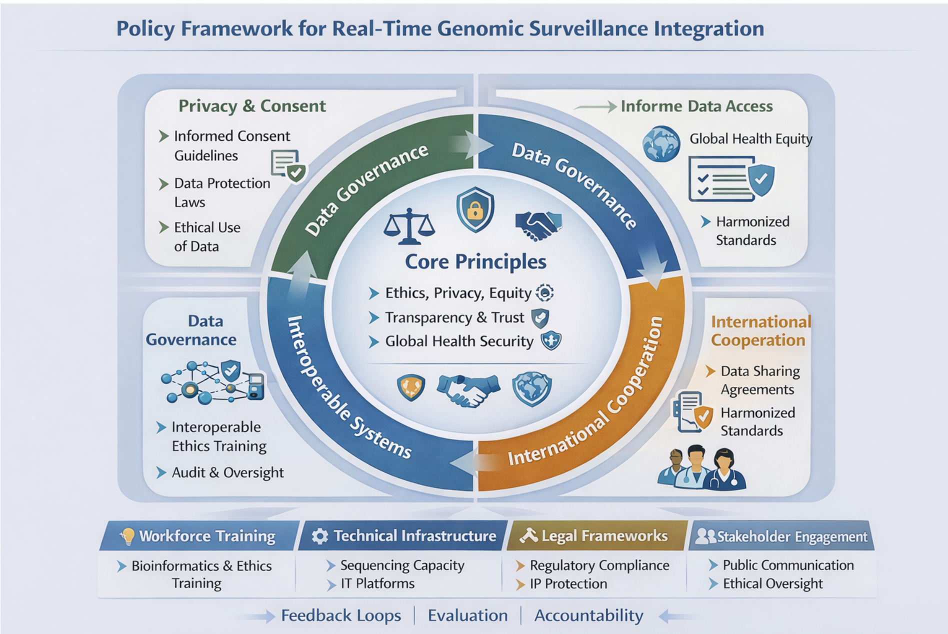

Figure 3: Policy Framework for Real-Time Genomic Surveillance Integration

Conclusion

This study highlights the transformative potential of integrating real-time genomic surveillance, through Next-Generation Sequencing (NGS), with epidemiological models in infectious disease intervention planning. By combining genomic and epidemiological data, health systems can enhance outbreak forecasting, better track pathogen evolution, and implement more precise interventions. However, challenges such as data gaps, equity issues, and the need for stronger global collaboration persist. Moving forward, strengthening infrastructure in low-resource settings, investing in data-sharing frameworks, and fostering international partnerships are critical to maximizing the effectiveness of this integrated approach. Future research should explore further advancements in AI-driven forecasting and real-time data applications for global health security.

Acknowledgments

Competing Interests: All the authors declare that they have no conflict of interest.

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: All the resources consulted in the review are provided in the reference section.

Funding: No funding of any kind was received.

Authors' Contributions: This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

References

1. Mami DM, Romba M, Patrick M. The Surveillance in Infectious Disease Control Role of Genomics. International Journal of Advanced Multidisciplinary Research and Studies. 2025 Aug 13;5(4):1608-15. Available from: https://www.multiresearchjournal.com/arclist/list-2025.5.4/id-4803
2. Struelens MJ, Ludden C, Werner G, Sintchenko V, Jokelainen P, Ip M. Real-time genomic surveillance for enhanced control of infectious diseases and antimicrobial resistance. Front Sci. 2024 Apr 25;2:1298248. https://doi.org/10.3389/fsci.2024.1298248
3. Tiwari S, Dhakal T, Kim BJ, Jang GS, Oh Y. Genomics in Epidemiology and Disease Surveillance: An Exploratory Analysis. Life. 2025 Dec 1;15(12):1848. https://doi.org/10.3390/life15121848
4. Osei Sekyere J. Next-Generation Sequencing in Infectious-Disease Diagnostics: Economic, Regulatory, and Clinical Pathways to Adoption. Microbiol Open. 2025 Dec;14(6):e70104. https://doi.org/10.1002/mbo3.70104
5. Revez J, Espinosa L, Albiger B, Leitmeyer KC, Struelens MJ, ECDC National Microbiology Focal Points and Experts Group. Survey on the use of whole-genome sequencing for infectious diseases surveillance: rapid expansion of European national capacities, 2015–2016. Front Public Health. 2017 Dec 18;5:347. https://doi.org/10.3389/fpubh.2017.00347
6. Centers for Disease Control and Prevention. Technical explainer: Infectious disease transmission models [Internet]. 2025 Jan 6 [cited 2025 Oct 7]. Available from: https://www.cdc.gov/cfa-modeling-and-forecasting/about/explainer-transmission-models.html
7. World Bank. An introduction to deterministic infectious disease models [Internet]. Working Paper No. 161374, Vol. 1. 2021 Jul 2 [cited 2025 Oct 7]. Available from: http://documents.worldbank.org/curated/en/888341625223820901
8. Duault H, Durand B, Canini L. Methods combining genomic and epidemiological data in the reconstruction of transmission trees: a systematic review. Pathogens. 2022 Feb 15;11(2):252. https://doi.org/10.3390/pathogens11020252
9. Contreras S, Oróstica KY, Daza-Sanchez A, Wagner J, Dönges P, Medina-Ortiz D, et al. Model-based assessment of sampling protocols for infectious disease genomic surveillance. Chaos Solitons Fractals. 2023 Feb 1;167:113093. https://doi.org/10.48550/arXiv.2301.07951
10. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-generation sequencing technology: current trends and advancements. Biology. 2023 Jul 13;12(7):997. https://doi.org/10.3390/biology12070997
11. Wikipedia. Nanopore sequencing [Internet]. [cited 2025 Oct 7]. Available from: https://en.wikipedia.org/wiki/Nanopore_sequencing

12. Illumina. NGS methods to identify and track infectious disease threats [Internet]. [cited 2025 Oct 7]. Available from: https://www.illumina.com/areas-of-interest/microbiology/public-health-surveillance/genomic-surveillance.html
13. Jabłońska A, Sadkowski A, Richert-Przygońska M, Styczyński J. Next-generation sequencing for infectious disease diagnostics in pediatric patients with Malignancies or after hematopoietic cell transplantation: A systematic review. J Clin Med. 2025 Sep 12;14(18):6444. https://doi.org/10.3390/jcm14186444
14. Papamentzelopoulou M, Vrioni G, Pitiriga V. Comparative Evaluation of Sequencing Technologies for Detecting Antimicrobial Resistance in Bloodstream Infections. Antibiotics. 2025 Dec 12;14(12):1257. https://doi.org/10.3390/antibiotics14121257
15. Cason C, D'Accolti M, Soffritti I, Mazzacane S, Comar M, Caselli E. Next-generation sequencing and PCR technologies in monitoring the hospital microbiome and its drug resistance. Front Microbiol. 2022 Jul 28;13:969863. https://doi.org/10.3389/fmicb.2022.969863
16. Lai LM, Chen QG, Liu Y, Zhao R, Cao ML, Yuan L. The value of metagenomic next-generation sequencing in the diagnosis of fever of unknown origin. Sci Rep. 2025 Jan 14;15(1):1963. https://doi.org/10.1038/s41598-025-86295-2
17. Tolles J, Luong T. Modeling epidemics with compartmental models. JAMA. 2020 Jun 23;323(24):2515-6. https://doi.org/10.1001/jama.2020.8420
18. Wikipedia. Kermack–McKendrick theory [Internet]. [cited 2025 Oct 15]. Available from: https://en.wikipedia.org/wiki/Kermack%E2%80%93McKendrick_theory
19. Milgroom MG. Epidemiology and sir models. In: Milgroom MG, editor. Biology of Infectious Disease: From Molecules to Ecosystems. Cham: Springer International Publishing; 2023. p. 253-68. https://doi.org/10.1007/978-3-031-38941-2_16
20. Mirsaeedi F, Sheikhalishahi M, Mohammadi M, Pirayesh A, Ivanov D. Compartmental models in epidemiology: bridging the gap with operations research for enhanced epidemic control. Ann Oper Res. 2025 Oct 20:1-58. https://doi.org/10.1007/s10479-025-06893-1
21. Kühnert D, Stadler T, Vaughan TG, Drummond AJ. Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth–death SIR model. J R Soc Interface. 2014 May 6;11(94):20131106. https://doi.org/10.1098/rsif.2013.1106
22. Glennon EE, Bruijning M, Lessler J, Miller IF, Rice BL, Thompson RN, et al. Challenges in modeling the emergence of novel pathogens. Epidemics. 2021 Dec 1;37:100516. https://doi.org/10.1016/j.epidem.2021.100516
23. Hasan GM, Mohammad T, Shamsi A, Sohal SS, Hassan MI. Applications of Genome Sequencing in Infectious Diseases: From Pathogen Identification to Precision Medicine. Pharmaceuticals. 2025 Nov 7;18(11):1687. https://doi.org/10.3390/ph18111687
24. Carson J, Keeling M, Ribeca P, Didelot X. Incorporating epidemiological data into the genomic analysis of partially sampled infectious disease outbreaks. Mol Biol Evol. 2025 Apr;42(4):msaf083. https://doi.org/10.1093/molbev/msaf083
25. Hill V, Ruis C, Bajaj S, Pybus OG, Kraemer MU. Progress and challenges in virus genomic epidemiology. Trends Parasitol. 2021 Dec 1;37(12):1038-49. https://doi.org/10.1016/j.pt.2021.08.007
26. Oltean-Parke H. Enhancing public health surveillance: integrating genomic and epidemiologic data to inform public health action and one health progress [dissertation]. Seattle (WA): University of Washington; 2023.

Available from: http://hdl.handle.net/1773/51162
27. Wikipedia. COVID-19 Genomics UK Consortium [Internet]. [cited 2025 Oct 22]. Available from: https://de.wikipedia.org/wiki/COG-UK
28. Carter LL, Yu MA, Sacks JA, Barnadas C, Pereyaslov D, Cognat S, et al. Global genomic surveillance strategy for pathogens with pandemic and epidemic potential 2022–2032. Bull World Health Organ. 2022 Apr 1;100(4):239. https://doi.org/10.2471/BLT.22.288220
29. World Bank. Strengthening genomic surveillance in Africa: Capacity building beyond COVID-19 [Internet]. 2025 Apr 7 [cited 2025 Nov 7]. Available from: https://www.worldbank.org/en/news/feature/2025/04/07/strengthening-genomic-surveillance-in-afe-africa-capacity-building-beyond-covid-19
30. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data–from vision to reality. Euro Surveill. 2017 Mar 30;22(13):30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494
31. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018 Dec 1;34(23):4121-3. https://doi.org/10.1093/bioinformatics/bty407
32. O'Toole Á, Pybus OG, Abram ME, Kelly EJ, Rambaut A. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. BMC Genomics. 2022 Feb 11;23(1):121. https://doi.org/10.1186/s12864-022-08358-2
33. Ochola R. The Case for Genomic Surveillance in Africa. Trop Med Infect Dis. 2025 May 8;10(5):129. https://doi.org/10.3390/tropicalmed10050129
34. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016 Feb 11;530(7589):228-32. https://doi.org/10.1038/nature16996
35. Du P, Ding N, Li J, Zhang F, Wang Q, Chen Z, et al. Genomic surveillance of COVID-19 cases in Beijing. Nat Commun. 2020 Oct 30;11(1):5503. https://doi.org/10.1038/s41467-020-19345-0
36. Ling-Hu T, Rios-Guzman E, Lorenzo-Redondo R, Ozer EA, Hultquist JF. Challenges and opportunities for global genomic surveillance strategies in the COVID-19 era. Viruses. 2022 Nov 16;14(11):2532. https://doi.org/10.3390/v14112532
37. Wang S, Jiang X, Singh S, Marmor R, Bonomi L, Fox D, et al. Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States. Ann N Y Acad Sci. 2017 Jan;1387(1):73-83. https://doi.org/10.1111/nyas.13259
38. Zeghbib S, Kemenesi G, Jakab F. The importance of equally accessible genomic surveillance in the age of pandemics. Biol Futura. 2023 Jun;74(1):81-9. https://doi.org/10.1007/s42977-023-00164-5
39. UK Government Department of Business, Energy and Industrial Strategy. A case study of viral genomic data sharing during the COVID-19 pandemic [Internet]. 2022 Nov [cited 2025 Nov 10]. Available from: https://assets.publishing.service.gov.uk/media/63766ddf8fa8f57721e34e30/intelligent-open-science.pdf
40. Staunton C, Edgcumbe A, Abdulrauf L, Gooden A, Ogendi P, Thaldar D. Cross-border data sharing for research in Africa: An analysis of the data protection and research ethics requirements in 12 jurisdictions. J Law Biosci. 2025 Jan;12(1):lsaf002. https://doi.org/10.21203/rs.3.rs-4217849/v1
41. Idahor CO, Esomu EJ, Ogbonna N, Momoh Z, Ogbeide OA, Ikhu-Omoregbe O, et al. Infectious Disease Surveillance in the Era of Big Data and AI: Opportunities and Pitfalls. Cureus. 2025 Oct

6;17(10):e93929. https://doi.org/10.7759/cureus.93929
42. Villanueva-Miranda I, Xiao G, Xie Y. Artificial Intelligence in Early Warning Systems for Infectious Disease Surveillance: A Systematic Review. Front Public Health. 2025 Jun 23;13:1609615. https://doi.org/10.3389/fpubh.2025.1609615
43. da Silva Mendes VI, Milhomens de Ferreira Mendes B, Moura RP, Mota Lourenço I, Ferreira Alves Oliveira M, Lee Ng K, et al. Harnessing artificial intelligence for enhanced public health surveillance: a narrative review. Front Public Health. 2025 Jul 30;13:1601151. https://doi.org/10.3389/fpubh.2025.1601151
44. Pinto AD, Birdi S, Durant S, Rabet R, Parekh R, Ali S, et al. Machine Learning Applications in Population and Public Health: Guidelines for Development, Testing, and Implementation. JMIR Public Health Surveill. 2025 Oct 24;11:e68952. https://doi.org/10.2196/68952
45. World Health Organization. WHO releases new principles for ethical human genomic data collection and sharing [Internet]. 2024 Nov 20 [cited 2025 Nov 18]. Available from: https://www.who.int/news/item/20-11-2024-who-releases-new-principles-for-ethical-human-genomic-data-collection-and-sharing
46. Moodley K, Cengiz N, Domingo A, Nair G, Obasa AE, Lessells RJ, et al. Ethics and governance challenges related to genomic data sharing in southern Africa: the case of SARS-CoV-2. Lancet Glob Health. 2022 Dec 1;10(12):e1855-9. https://doi.org/10.1016/S2214-109X(22)00417-X
47. Wikipedia. Bermuda Principles [Internet]. [cited 2025 Nov 18]. Available from: https://en.wikipedia.org/wiki/Bermuda_Principles
48. Wikipedia. Fort Lauderdale Agreement [Internet]. [cited 2025 Nov 18]. Available from: https://en.wikipedia.org/wiki/Fort_Lauderdale_Agreement
49. Africa CDC. Pathogen genomic surveillance policy framework [Internet]. 2025 [cited 2025 Nov 18]. Available from: https://africacdc.org/download/pathogen-genomic-surveillance-policy-framework/
50. Horn R, Merchant J. Ethical and social implications of public–private partnerships in the context of genomic/big health data collection. Eur J Hum Genet. 2024 Jun;32(6):736-41. https://doi.org/10.1038/s41431-024-01608-9