International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | Apr-2023 SJIF Rating: 8.176 ISSN: 2582-3930

Mathematical Techniques for Epidemiological Formulation-A Review

Kamaljeet Kaur

Assistant Professor, Guru Nanak College, Sri Muktsar Sahib, Punjab.

Abstract

Mathematical epidemiology provides a robust framework for modelling disease dynamics, analysing the progression of outbreaks, and informing public health policy. This review examines the evolution of these techniques, from foundational, knowledge-driven models to modern, data-intensive approaches. The field's roots lie in the pioneering work of Daniel Bernoulli in 1760 and the seminal compartmental models of Kermack and McKendrick, which use ordinary differential equations to describe the flow of a population between health-based groups. While these models are computationally efficient, they assume homogeneous mixing, a limitation that has driven the development of more advanced paradigms.

To address these limitations, modern approaches include network models, which account for heterogeneous contact structures, and agent-based models, which simulate disease spread at the individual level. These methods offer greater realism but come with significant computational demands. The paper highlights the importance of nuanced interpretation of key metrics like the Basic Reproduction Number (R_0) and the Effective Reproduction Number (R_e), which are often misunderstood. Finally, it addresses the burgeoning role of modern computational techniques, including machine learning and artificial intelligence, which are increasingly used to process high-dimensional data and enhance predictive accuracy, often in combination with traditional models to create more robust and adaptable hybrid systems.

Keywords

Epidemiological models, Compartmental models, Agent-based models, Network models, Reproduction number, Machine learning.

1. Introduction

1.1. Historical Overview

The application of mathematics to the study of epidemics has a rich history that predates modern computing. The earliest known example dates back to 1760 with the work of Daniel Bernoulli, a physician and mathematician who developed a mathematical model to defend the practice of smallpox inoculation. Bernoulli's work established a precedent for using mathematical models as tools for comparing different health strategies and predicting disease outcomes.

The early 20th century saw the emergence of the foundational theories that define mathematical epidemiology today. In the 1920s, the field of compartmental models was born with the work of William Hamer, Ronald Ross, and, most notably, William Kermack and A.G. McKendrick. Their 1927 Kermack–McKendrick epidemic model was a landmark achievement, successfully predicting the behaviour of outbreaks and establishing the principles that govern the relationship between susceptible, infected, and immune individuals in a population.

Over time, the utility of these models has been repeatedly demonstrated in real-world scenarios. The 2001 foot-and-mouth disease outbreak in Great Britain and the 2002–2003 Severe Acute Respiratory Syndrome (SARS) epidemic served as key catalysts, reviving interest in mathematical models as indispensable tools for forecasting disease progression and evaluating the effectiveness of control strategies. More recently, the COVID-19 pandemic underscored the vital role of these models, as they were widely used by researchers and decision-makers to estimate outbreak parameters, inform public health policies, and predict future trends.

1.2. Foundational Concepts and Terminology

At its core, mathematical epidemiology is the science of modelling diseases, a process that is often underpinned by a set of well-defined concepts and terminology. A central paradigm is the compartmental model, which simplifies a complex population by dividing it into distinct subgroups or "compartments" based on their infection status. The dynamics of an epidemic are then represented by the flow of individuals between these compartments over time. For example, the classic SIR model uses three such compartments: Susceptible (S), Infectious (I), and Recovered (R). The change in the size of each compartment is expressed as a derivative, forming a system of differential equations.

Another critical distinction in modelling is between **deterministic** and **stochastic** approaches. Deterministic models, often used for large populations, assume that the epidemic process is predictable and that changes in a compartment's population can be calculated using only the history that was used to develop the model. In contrast, stochastic models incorporate random variables and chance variations in inputs, allowing for a probability distribution of potential outcomes. They are particularly valuable for modelling small populations or the early stages of an outbreak.

A vital concept derived from these models is the **Basic Reproduction Number (R_0)**, the average number of new infections caused by a single infectious person in an entirely susceptible population. This number serves as a critical threshold: if $R_0>1$, the disease will spread and die out otherwise. The related **Effective Reproduction Number (R_e or R_t)** provides a more dynamic, real-time measure, representing the number of new infections caused by one person at any given time, accounting for evolving factors like population immunity from prior infection or vaccination.

The foundation of compartmental models is built on a simplifying assumption that populations mix homogeneously. This implies that any individual is equally likely to contact any other individual. This approach provides a mathematically elegant and powerful framework for capturing the essential dynamics of disease spread. However, its simplicity is also its most significant limitation. The real world is not a well-mixed system; individuals form clusters, and their contacts are not uniform.

Parameter	Symbol	Definition
Basic Reproduction Number	R_0	The average number of new infections caused by a single infectious person in a fully susceptible population.
Effective Reproduction Number	R_e or R_t	The average number of new infections caused by a single infectious person at any specific time.
Transmission Rate	β	The rate at which an infected individual spreads the disease.

Recovery Rate	γ	The rate at which infected individuals recover.

2. Classical Compartmental Models

2.1. The SIR Model: Formulation and Key Equations

The SIR model is the cornerstone of compartmental epidemiology, providing a foundational framework for understanding the spread of infectious diseases. Developed by Kermack and McKendrick, it divides a fixed population (N) into three mutually exclusive compartments: Susceptible (S), Infectious (I), and Recovered (R). This model is particularly suited for diseases that confer lifelong immunity upon recovery, such as measles, mumps, and rubella.

The dynamics of the SIR model are described by a system of three coupled ordinary differential equations (ODEs), which represent the rate of change of the population within each compartment over time (t).

The System of Ordinary Differential Equations (ODEs)

$$\frac{dS}{dt} = \frac{-\beta SI}{N}$$
$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

A fundamental quantity derived from these equations is the Basic Reproduction Number, R_0. It is defined by the ratio of the transmission rate to the recovery rate: $R_0 = {}^{\gamma}/_{\beta}$. This value is paramount because it encapsulates the potential for an epidemic to spread. If $R_0>1$, the disease will spread, whereas if $R0_0<1$, it will die out.

2.2. Extensions of the SIR Model

The simplicity of the SIR model makes it a powerful starting point, but its assumptions do not apply to all diseases. Over time, the core compartmental framework has been adapted to accommodate more complex biological realities, leading to a family of models that build upon the SIR foundation.

2.2.1. The SIRS Model

The SIR model assumes that once an individual recovers, they have lifelong immunity. However, for a class of diseases like seasonal influenza, immunity wanes over time, allowing for reinfection. The **SIRS (Susceptible-Infectious-Recovered-Susceptible)** model addresses this by adding a new transition from the recovered compartment back to the susceptible compartment. This transition is governed by a new parameter, ξ, which represents the rate at which recovered individuals lose immunity and return to a susceptible state.

2.2.2. The SEIR Model

Many infectious diseases, such as mumps and COVID-19, have a non-trivial incubation period—a time lag between infection and the onset of infectiousness. The SIR model cannot account for this latent period. The **SEIR (Susceptible-Exposed-Infectious-Recovered)** model was developed as an extension to address this limitation. It introduces a new compartment, "Exposed" (E), for individuals who have been infected but are not yet infectious. In this model, susceptible individuals transition to the exposed compartment upon infection, and

after a period determined by the incubation rate σ, they move on to the infectious compartment.

2.2.3. Other Complex Variants

The compartmental modelling approach is highly flexible and can be expanded to capture an increasing number of real-world details. The **SIDARTHE** model, proposed for the COVID-19 pandemic, is a prime example of this. It expands the SIR model by introducing multiple compartments to differentiate infected individuals based on their diagnosis status and the severity of their symptoms. The model's name, SIDARTHE, stands for Susceptible, Infected, Diagnosed, Ailing, Recognized, Threatened, Healed, and Extinct (deaths), which allowed it to model the importance of combining public health interventions like social distancing with contact tracing and testing.

2.2.4. The SEIQR Model and Quarantine

The compartmental modelling framework can be further expanded to explicitly account for public health interventions such as quarantine and isolation. An extension of the SEIR model, the**SEIQR (Susceptible-Exposed-Infectious-Quarantined-Recovered)** model, introduces a new compartment, **Q**, for individuals who have been quarantined. Quarantine is a key public health strategy used to detach seemingly healthy but potentially infected individuals from the general population to prevent further transmission. Similarly, isolation is used for confirmed infected individuals.

The model includes a new flow from the Exposed (E) or Infectious (I) compartments into the Quarantined (Q) compartment. Specific parameters, such as the quarantine rate ε or the fraction of quarantined individuals q, are introduced to represent this transition. The quarantined individuals may then recover (R) or, in some cases, die (D). SEIQR and similar quarantine models are crucial tools for evaluating the impact of interventions on disease dynamics. For example, they can be used to quantify how measures like quarantine and contact tracing reduce R_0.

The progression from simple models like SIR to more complex variants like SIRS, SEIR, and SIDARTHE illustrates a core developmental trend in the field. The evolution is a direct response to the biological realities of diseases that do not conform to the simple assumptions of the SIR model. For example, the recognition of waning immunity for influenza required the addition of the R→S transition in the SIRS model, while the long incubation period of mumps necessitated the creation of the E compartment in the SEIR model.

3. Advanced Modelling Paradigms

The classical compartmental models, while foundational, are built on the significant simplifying assumption of homogeneous mixing, where every individual in a population is considered equally likely to contact every other. To overcome this limitation and better reflect the complexity of real-world interactions, more advanced modelling paradigms have emerged.

3.1. Network and Graph-Based Models

Network models represent a shift from a top-down, population-level approach to a more granular, interaction-based framework. In this representation, individuals are conceptualized as "nodes" or "vertices," and the contacts that can transmit a disease are the "links" or "edges" between them. This approach allows modelers to move beyond the unrealistic assumption of a uniformly mixed population. By explicitly modelling the contact network, these models can capture heterogeneous social features that profoundly influence disease propagation, such as an individual's sociality, location, or wealth. These models can describe complex network topologies, including "small-world" and "scale-free" networks, which are common in real-world social structures. Crucially, scale-free networks feature "hubs"—a small number of highly connected nodes—that can act as "super-

spreaders," rapidly transmitting a disease to a large number of contacts and significantly altering the course of an epidemic.

3.2. Agent-Based Models (ABMs)

Agent-based models (ABMs) adopt a bottom-up approach by simulating disease spread at the individual level, where each agent has unique traits, behaviors, and disease status. Their main strength lies in flexibility, enabling detailed representation of factors like age, mobility, and social interactions. This allows ABMs to model complex scenarios such as contact tracing, targeted interventions, and household or school networks—beyond the scope of compartmental models. However, this granularity makes ABMs highly complex, computationally demanding, slower to implement, and heavily reliant on detailed data for parameterization and validation.

3.3. A Comparison of Paradigms

Paradigm	Fundamental Approach	Assumptions	Computational Needs	Data Requirements
Compartmental	Top-down, population-based	Homogeneous mixing; fixed population-level rates. [6]	Low	Minimal, aggregated data. [6]
Agent-Based	Bottom-up, individual-based	Unique agent characteristics; explicit interactions; non-uniform mixing. [6]	High	Detailed, granular data for each agent. [6]
Network	Interaction-based	Models connections between individuals; allows for heterogeneous contact structures.	Moderate to High	Network topology data; contact patterns. [3]

4. Interpretation, Limitations, and Challenges

4.1. The Nuances of the Reproduction Number (R_0)

The Basic Reproduction Number (R_0) is a widely referenced term, but its interpretation is often misunderstood. Formally, R_0 is a theoretical, population-averaged value that describes the expected number of infections generated by a single case in a population that is *entirely susceptible*. It is a critical metric for understanding a disease's potential for spread but should not be mistaken for a fixed, biological constant. Its estimated value can vary depending on the model used and the data that informs it.

It is essential to distinguish R_0 from the **Effective Reproduction Number (R_e or R_t)**. While R_0 is a static, theoretical value, R_e is a dynamic, real-time metric. It represents the number of new infections caused by a single person at any specific time, taking into account the proportion of the population that is already immune due to prior infection or vaccination. The relationship between the two is expressed as

$$R_e = R_0(1 - P_i),\ P_i$$

is the proportion of the immune population. A high level of immunization can bring a high R_0 value down to a sub-threshold R_e <1, which is necessary to prevent the sustained spread of an infection.

4.2. Challenges in Integrating Human Behavior

A major limitation of traditional epidemiological models, such as the SIR model, is their inability to capture the complexities of human behavior. These models assume fixed parameters, yet real epidemics involve dynamic feedback loops where disease spread influences behavior and behavior alters transmission. For example, rising cases heighten perceived risk, prompting protective actions, while declining cases reduce caution and accelerate spread. Ignoring such adaptive responses creates omitted variables bias: changes in dynamics are wrongly attributed to biological parameters like transmission (β) or recovery (γ). This misrepresentation produces biased estimates and reduces the predictive accuracy of epidemic models.

4.3. The Role of Uncertainty

Epidemiological modelling is inherently approximate, with uncertainty arising from human and pathogen behaviour, data quality, and random transmission. Stochastic models can capture this uncertainty by projecting outcome ranges, but effectively communicating such probabilistic results to policymakers and the public is challenging. Policymakers often prioritize worst-case scenarios, and oversimplifying complex forecasts risks eroding public trust and weakening the implementation of public health measures.

5. Conclusion and Future Directions

The mathematical formulation of epidemics has evolved from simple, ODE-based compartmental models to complex, data-driven agent-based and machine learning systems.[3] This progression is a direct response to the limitations of simpler models, particularly their inability to capture the profound impact of behavioral, social, and spatial heterogeneity on disease spread.

The analysis in this paper confirms that no single modelling technique is universally superior. The choice of a method depends on the specific public health question at hand, the availability of data, and the computational resources accessible to the modeler.[6] Compartmental models remain valuable for quickly evaluating disease dynamics with limited data, while agent-based and network models are essential for capturing individual-level variability and complex social structures.[6] The integration of machine learning offers a powerful complement, capable of handling high-dimensional data and adapting to dynamic changes in pathogens.[18]

The future of mathematical epidemiology lies in the continued development of hybrid approaches that combine the theoretical underpinnings of mechanistic models with the predictive capabilities of machine learning, all while placing a renewed emphasis on critical real-world factors. The primary challenges to be addressed include:

- **Data and Collaboration:** There is a persistent need for more diverse and high-quality data, especially concerning human behavior and social dynamics. Overcoming this requires enhanced interdisciplinary collaboration, bridging the gaps between behavioral, biological, and data sciences to create a shared language and standardized methods.

- **Interpretability and Trust:** The "black box" nature of many modern models must be addressed.

Developing methods that make these models more transparent will be crucial for building trust with policymakers and the public and ensuring that these powerful tools are used effectively to inform evidence-based policies and interventions.

References

1. Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir.
2. Brauer, F., Castillo-Chavez, C., & Feng, Z. (2008). Mathematical epidemiology. Lecture notes in mathematics.
3. Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics.
4. Reed, L. J., & Frost, W. H. (1928). The Reed-Frost epidemic model.
5. Louz, G. (2010). Emergence of a social network.
6. Mikolajczyk, A. (2009). Influenza and public health.
7. Perra, F. P. et al. (2011). Adaptive Human Behavior in Epidemiological Models.
8. Yildirim, A. (2012). An analytical approach to transmission dynamics of infectious diseases with waning immunity.
9. El-Sayed, A. M., Scarborough, P., Seeman, L., & Galea, S. (2012). Social network analysis and agent-based modelling in social epidemiology.
10. Luke, D. A., & Stamatakis, K. A. (2012). Systems science methods in public health: dynamics, networks, and agents.
11. Barabási, A.-L. (2016). Network Science.
12. Wu, Y. et al. (2018). Deep learning in medical image analysis.
13. Rhodes, C. J., & Anderson, R. M. (2008). Contact rate calculation for a basic epidemic model.
14. Giordano, G. et al. (2020). SIDARTHE model for COVID-19 pandemic.
15. Deng, X. et al. (2020). Colagnn: A graph neural network for collaborative filtering.
16. Song, X. et al. (2020). Reinforced infection control modelling.
17. Science. (2020)..
18. Rahmandad, F. et al. (2021)..
19. Almeida, L. (2021). Final size and convergence rate for an epidemic in heterogeneous populations.
20. Gao, Q. (2021). Transmission dynamics and quarantine control of COVID-19 in cluster community.
21. Shorten, C. (2021). Deep learning in epidemiology.
22. Ferguson, N. et al. (2021). Communicating uncertainty in epidemic models.