A MATHEMATICAL MODEL IN PUBLIC HEALTH EPIDEMIOLOGY: COVID-19 CASE RESOLUTION AMONG VETERANS USING PUBLIC-USE SURVEILLANCE DATA AND FIRST-ORDER DIFFERENTIAL EQUATIONS

GV Fant

US Department of Veterans Affairs, Veterans Health Administration, Washington, DC

ABSTRACT

Introduction: *This project applies linear first-order differential equations to model COVID-19 case resolutions within the Veterans Health Administration. This model uses cumulative public health data to estimate case resolution, offering a practical tool for managing COVID-19 and potentially other communicable disease outbreaks.*

Methods: *A two-part approach was adopted: Part A models cumulative COVID-19 recovery and death rates, while Part B incorporates vaccination data to refine these rates. The model is structured through first-order differential equations and a homogeneous mixing assumption, using a meticulously cleaned dataset to ensure accurate forecasting of disease outcomes across a healthcare system.*

Findings: *In Part A, baseline rates (recovery=0.9717, death=0.0267) provided a control for assessing vaccination impact. With cumulative vaccination data, in Part B, vaccination-adjusted rates showed enhanced recovery (1.5066) and reduced death (0.0193). This model uses public-use, cumulative public health surveillance data, enabling real-time adaptations with minimal computational complexity.*

Discussion: *We demonstrate that vaccination increases COVID-19 recovery rates and reduces mortality within a public healthcare system. Using a differential equation model with validated data protocols, this approach combines mathematical rigor and data integrity with actionable insights, offering a reliable foundation for strategic public health decisions.*

Conclusion: *COVID-19 case resolution in the Veterans Health Administration was modeled using a first-order differential approach. Vaccination-adjusted recovery and death rates showed improved outcomes, supporting resource planning and related public health efforts. A consistency check validated the stability of the model across datasets, supporting its reliability in public health decision-making.*

KEYWORDS

Public health epidemiology; decision-makers; COVID-19; Public health surveillance data; Mathematical model

1. INTRODUCTION

1.1. Mathematical Models in Public Health Decision-Making

Throughout the COVID-19 pandemic and in other communicable disease outbreaks, mathematical models have played a crucial role in guiding public health interventions, predicting disease outcomes, and supporting effective resources allocation. These models are essential for estimating disease spread, assessing the impact of a public health intervention, and helping to prepare a healthcare system for possible surge in disease case counts. The range of models used for such purposes [1-8] spans from simple first-order differential equations to highly complex simulations, including cohort models, decision-analysis frameworks, and network-based models. While more sophisticated models, such as the transmission dynamic models reviewed by Mac et al. [1], account for dynamic variables like vaccination rates and population interactions, the simplicity of a first-order equation offers distinct advantages. It provides public health decision-makers with timely, data-driven insights that require minimal computational resources and rely on regularly updated public-use surveillance data. Given these needs, a practical, first-order differential equation approach serves as a timely and reliable tool, balancing simplicity with effectiveness. While stochastic and network-based models offer detailed insights into disease transmission [9], they often require extensive data inputs and computational resources, limiting their utility in real-time public health applications. No prior instances were identified in PubMed (2020 to 2024) where simple, first-order differential equations were used to model case resolutions at the health system level for actionable public health decision-making.

1.2. Utility of First-Order Differential Equations

The review of literature includes several more advanced models that incorporate complex factors like co-infections [2] and stochastic processes [3] These models allow for a more detailed exploration of disease dynamics, particularly in cases where multiple health conditions or external interventions are involved. However, even the most advanced models depend heavily on the quality and timeliness of data [3]. The first-order differential equation approach used here emphasizes practicality and speed, providing public health and healthcare system administrators with timely, usable insights during a communicable disease outbreak. For example, while more complex models such as those leveraging network theory [5] provide valuable insights into how diseases spread through social connections, they often require extensive data inputs and longer processing times, which can delay the delivery of actionable insights.

Another important consideration is the role of simplicity in enhancing model transparency and usability. While simpler than more advanced options like the ARIMA forecasting models [6] or the Padé approximation SIR model [7] can be easily understood and applied by public health and healthcare decision-makers without requiring specialized training in mathematical epidemiology. The co-infection model [2] highlights the importance of capturing interactions between COVID-19 and other chronic conditions, the focus of this study remains on COVID-19-specific outcomes. By concentrating on a narrower set of parameters—recovery and death rates—the model ensures that it can deliver rapid, actionable insights without the need for excessive data or computational complexity.

1.3. Infectious Disease Modeling Principles

The practical utility of a simpler, first-order differential equation to deliver timely insights is described [8]. Simpler models are often more accessible for decision-making and can be particularly useful when policymakers require rapid forecasts for resource allocation and public

health interventions. These models align with recommendations to base predictions on accurate, current data, ensuring that healthcare administrators receive relevant and actionable information; leveraging real-world public health surveillance data supports this usage. Furthermore, while more advanced models offer greater granularity, the trade-off in computational complexity and data requirements reinforces the choice to employ a first-order differential equation is not only a practical one but also reflects the real-world need for models that can adapt quickly to changing conditions using readily available data from public health surveillance systems.

Project aim: This project deliberately focuses on first-order differential equations due to their accessibility and practicality. These equations rely on principles from applied calculus, a topic commonly covered in courses for business, management, economics, social sciences, and life sciences. As a result, public health epidemiologists and health administrators are more likely to possess foundational knowledge of the utility of these equations, making this approach feasible for health system decision-making. In contrast, more complex models often require advanced training, limiting their usability in real-world public health applications.

The linear first-order differential equation employed in this study strikes a balance between simplicity and utility. It may provide public health and healthcare administrators with a practical tool for making informed decisions based on public-use surveillance data. This study uses data from the U.S. Department of Veterans Affairs (VA) which oversees the Veterans health Administration (VHA)—the largest integrated healthcare system in the United States, serving over six million Veterans annually. The aim of this project is to apply linear first-order differential equations with cumulative public data to estimate long-term resolution of COVID-19 cases. By leveraging cumulative data from Veterans Health Administration (VHA) healthcare facilities, including fully vaccinated counts, this model provides valuable insights that can support both public health and healthcare system administrators. The use of a public-use dataset from VHA highlights its potential utility as a resource for tracking and forecasting infectious disease outcomes across integrated healthcare systems.

2. METHODS

2.1. Study Design

This secondary data analysis study uses a mathematical modeling approach to estimate the resolution of cumulative active COVID-19 cases in Veterans Health Administration (VHA) facilities, focusing on recovery and death outcomes. The model is divided into **Part A**, which estimates cumulative recovery and death rates without vaccination data, and **Part B**, which adjusts these rates based on the cumulative counts of fully vaccinated individuals, including Pfizer/Moderna (Dose 2 of 2) or Janssen (Dose 1 of 1) data. Both parts apply first-order differential equations to describe the progression of COVID-19 cases, using a homogeneous mixing assumption, which simplifies interactions in epidemiological modeling [9]. A detailed explanation of first-order differential equations in public health epidemiology provides a foundational overview of mathematical modeling [10].

2.2. Scope

The **COVID-19 National Summary** data from the Department of Veterans Affairs (VA), Veterans Health Administration (VHA), dated October 22, 2024, provides an in-depth view of cumulative COVID-19 cases and vaccination efforts across the VA healthcare system (from 2020 to the Present). As the largest integrated health system in the U.S., serving over six million

veterans annually, VHA's summary data encompass cumulative cases, active cases, recoveries, and known deaths.

The report provides vaccination statistics, showing over 5.4 million fully vaccinated individuals, with breakdowns by vaccine type (Pfizer/Moderna and Janssen) and recipient groups (veterans, VHA employees, and federal partners). Data accessibility features allow for filtering by facility, patient category, and demographic factors, presenting a detailed view of COVID-19's impact across the VHA network. Accurate and comprehensive data are crucial for reliable mathematical modeling in public health epidemiology.

2.3. Data Accuracy and Updates

The report notes that on September 29, 2021, VA definitions were updated to align with CDC standards, adjusting historical cumulative case and death counts. Regular updates enhance data accuracy, though some VHA facilities in the process of transitioning to the Cerner Millennium electronic health record system may have temporary data gaps.

The "At a Glance" section in the report provides an accessible, current summary of COVID-19's impact across VHA medical facilities. This document establishes the scope and reliability of data used in this project, emphasizing VA's commitment to comprehensive coverage, accuracy, and structured COVID-19 reporting.

2.4. Data Source

Data for this study were sourced from publicly available cumulative surveillance data available on the public VA website. The initial data file (File A) was comprised of the following:

- **Active cases (A)**: Cumulative active COVID-19 cases.
- **Convalescent cases (C)**: Cumulative total of recovered cases.
- **Known deaths (D)**: Cumulative deaths attributed to COVID-19.
- **Vaccination data**: Cumulative counts of individuals vaccinated with Pfizer/Moderna (Dose 2 of 2) or the Janssen vaccine (Dose 1 of 2), which together represent fully vaccinated individuals.

The data set contained 322 records representing VHA facilities across the U.S., covering a broad span of active cases, recoveries, and vaccination statuses. A final, cleaned dataset of 165 records (File B) was used for modeling, ensuring that all facilities had complete key metrics:

- **Total Confirmed Cases (A+C+D)**: 998,748
- **Active Cases (A)**: 1,655
- **Recovered Cases (C)**: 970,474
- **Known Deaths (D)** : 26,619
- **Baseline Recovery Rate (γ_1)**: 0.9717
- **Baseline Death Rate (γ_2)**: 0.0267

The dynamic public health surveillance data can be found at this location: https://www.accesstocare.va.gov/Healthcare/COVID19NationalSummary

2.5. Data Processing

The transition from the initial data file (File A) to the cleaned dataset (File B) followed a structured data processing protocol to ensure data quality and data readiness for mathematical modeling. Processing was driven by the need for comprehensive cumulative vaccination counts, as some facilities lacked data in this area, and complete vaccination counts were essential for subsequent modeling.

1. **Data Ingestion & Initial Review**: Public-use data, sourced from an Agency website, included 322 records with cumulative COVID-19 case and vaccination details across VA facilities. Preliminary data inspection ensured all fields were properly structured for analysis.
2. **Data Cleaning & Filtering**: Vaccination fields initially stored as text were converted to integers, and duplicate facility records were removed. Facilities with incomplete key metrics --active cases, recoveries, deaths, or vaccination counts--were excluded, resulting in a focused set of 165 records. This step was necessary to ensure that each facility in the dataset provided full information, especially on cumulative Dose 2 of 2 and Dose 1 of 1 counts of vaccinations, crucial for modeling recovery and death rates in vaccinated individuals.
3. **Handling Missing Values**: Missing data in the Dose 1 of 1 (Janssen) column were treated as zeroes, reflecting unreported vaccinations without altering otherwise complete records.
4. **Validation**: While data were assumed accurate as public-use figures, logical checks ensured consistency—non-negative counts, realistic total counts for COVID-19 cases, and cross-verification of cumulative vaccination data to flag any anomalies. These validations helped confirm the integrity of records, particularly for facilities with complete Dose 2 of 2 and Dose 1 of 1 vaccination data, essential for the modeling objectives.

This process yielded a clean, validated dataset (File B) optimized for modeling purposes in the study, with a particular focus on facilities that provided comprehensive cumulative vaccination data. Using the prior activities, we proceed to Part A of mathematical modeling, which addresses baseline recovery and death rates.

Aggregate data were used to model cumulative COVID-19 outcomes across VHA healthcare system, aligning with traditional compartmental models, specifically first-order differential equations, that focus on system-level trends. Such data capture population dynamics efficiently without requiring individual-level detail [9].

2.6. Data Management and Analysis Protocol

This study followed the Structured Data Management and Analysis Protocol (see Appendix A) to ensure accuracy and transparency in data handling. Data were ingested, cleaned, and validated to ensure consistency. Exploratory data analysis (EDA) was conducted to understand trends, and first-order differential equations were applied to model case progression. Both Dose 2 of 2 and Dose 1 of 1 data were combined to represent fully vaccinated individuals in Part B, ensuring comprehensive adjustments for cumulative vaccination impact. GenAI/ChatGPT Data Analyst was used to support data management and analysis, ensuring the process was efficient and reproducible, while avoiding black-box methodologies.

2.7. Mathematical Modeling: Part A

In Part A, first-order differential equations were used to model cumulative recovery and death outcomes based on active cases, recoveries, and deaths. This approach assumes constant rates for recovery and death [3], representing long-term averages rather than moment-to-moment changes,

and uses a homogeneous mixing assumption across the population--a common practice in simplified epidemiological models [9].

These equation sestablish a baseline for modeling cumulative outcomes, specifically recovery and death rates, across the healthcare facilities:

- **Recovery Rate (γ_1)**:

$$\frac{dA(t)}{dt} = -\gamma_1 A(t)$$

where γ_1 is calculated as:

$$\gamma_1 = \frac{C}{A + C + D}$$

- **Death Rate (γ_2)**:

$$\frac{dA(t)}{dt} = -\gamma_2 A(t)$$

where γ_2 $is\ calculated\ as$:

$$\gamma_2 = \frac{D}{A + C + D}$$

These equations provide a baseline for understanding long-term cumulative outcomes at each facility without considering vaccination data. Solving these first-order differential equations yields exponential decay functions that model the decline in active cases over time due to recovery and death:

- **Recovery Solution**

$$A_{\text{recovery}}(t) = A_0 \cdot e^{-\gamma_1 \cdot t}$$

- **Death Solution**

$$A_{\text{death}}(t) = A_0 \cdot e^{-\gamma_2 \cdot t}$$

where A_0 is the initial active cases count (we used $A_0 = 10.03$), t = days.

These solutions model the decline in active cases over time, offering insights into long-term cumulative public health outcomes. Such insights can guide resource planning and forecasting in public health settings. Part B builds upon the baseline model by incorporating vaccination data to assess its impact on recovery and death rates.

2.8. Mathematical Modeling: Part B

Part B extends the model by incorporating cumulative vaccination data, focusing on individuals who received Pfizer/Moderna or Janssen, representing fully vaccinated individuals. The adjusted recovery and death rates are calculated as:

- **Vaccination-Adjusted Recovery Rate**:

$$\gamma_1\left(Adjusted, V_{\text{fully vaccinated}}\right) = \gamma_1^0 \times \left(1 + \alpha \frac{V_{\text{fully vaccinated}}}{A + C + D}\right)$$

- **Vaccination-Adjusted Death Rate**:

$$\gamma_2\left(A, V_{\text{fully vaccinated}}\right) = \gamma_2^0 \times \left(1 - \beta \frac{V_{\text{fully vaccinated}}}{A + C + D}\right)$$

Here, γ_1^0 and γ_2^0 denote baseline cumulative recovery and death rates, respectively; $V_{fully\ vaccinated}$ represents the cumulative number of fully vaccinated individuals both Pfizer/Moderna (Dose 2 of 2) and Janssen (Dose 1 of 1).Parameters α=0.1 and β=0.05 reflect conservative estimates for vaccination impact, validated in prior studies [1,11]. These values can be adapted for specific populations or interventions. The adjustment factors, α and β, are used to model the proportional effects of vaccination on recovery and death rates, an approach commonly applied in models that account for modifying factors such as vaccination or co-infection [2,4].

As in Part A, the Solution for Decline in Active Cases approach may also apply here, adjusted, as needed, for use with the first-order differential equations to yield solutions for recovery or death trends. This consistent framework models case resolution dynamics across both parts.

For a detailed step-by-step calculation of Part A and Part B, please refer to Appendix B where numerical examples demonstrate how the equations are applied.

2.9. Consistency Check-a Sensitivity Analysis: Plan

To evaluate the stability and robustness of the model, a Consistency Check was conducted by applying the same modeling procedures to a subsequent cumulative dataset of the same type but from a later date (File C; dated November 7, 2024; with data from 2020 to Present). The consistency check, a form of sensitivity analysis, evaluates the model's robustness and stability. This step confirms that the model produces reliable results with updated data, supporting long-term applicability.

A "stable result" in this context is defined by three key criteria:

1. **Relative Consistency**: Model outputs (e.g., recovery and death rates) should remain within a 5-8% tolerance of the initial cumulative results, indicating consistent predictions.
2. **No Significant Change in Outcomes**: Similar trends in cumulative outcomes indicate model alignment with long-term case resolution dynamics.
3. **Key Model Parameters Consistency**: Cumulative recovery and death rates should show minimal deviation, supporting confidence in model reliability. This includes consistent integration of fully vaccinated data counts, with Dose 2 of 2 and Dose 1 of 1 used as the standard measure of full vaccination.

2.10. Limitations

Part A does not account for vaccination status, which may affect the accuracy of recovery and death rate estimates in highly vaccinated facilities. This limitation is mitigated by a flexible model structure, adaptable with additional vaccination data. In Part B, Dose 1 of 1 (Janssen) and

Dose 2 of 2 (Pfizer/Moderna) data were included to represent fully vaccinated individuals. The decision to include both full vaccination types reflects the most complete vaccination information available, providing more accurate estimates for fully vaccination populations. The model does not consider the time between doses, account for differences between vaccine types, or other facility-level factors. Additionally, the model assumes constant recovery and death rates, simplifying real-time application but may reduce precision over time. Future updates could incorporate these details for enhanced precision without sacrificing the model's real-time application. For example, incorporating time-varying vaccination effects or regional demographic variations could improve the model's accuracy.

Our model assumes homogeneous mixing, meaning each individual is equally likely to interact with others, regardless of demographic or social differences. This simplifies calculations, making them practical for health system decision-making, which relies on aggregated trends rather than individual-level clinical details [9]. Although this assumption is standard in epidemiological modeling to keep the model practical, it permits actionable insights for resource allocations and strategic planning. While this assumption facilitates timely decision-making, it limits the ability to capture heterogeneity in populations with varied demographic or social structures.

2.11. Ethical Considerations

This project relied on publicly available data (or public-use data) posted by the Agency on the public-facing website, with compliance with HIPAA and other data privacy regulations. The use of publicly available data ensures compliance with ethical and privacy considerations but may introduce biases due to reporting inaccuracies. Since this model focuses on system-level decision-making, future applications could integrate demographic and socioeconomic data to address health equity challenges and ensure interventions reach underserved populations. Such limitations are inherent in using real-world public health data. However, the structured data management activities described in Appendix A aim to mitigate their impact and ensure reliability for decision-making. No individuals were contacted to obtain the data used in this project. Ethical guidelines were followed in the use of GenAI tools to ensure transparency and accountability.

3. SUMMARY OF FINDINGS

Using a structured differential equation model and public-use datasets, this analysis establishes baseline recovery and death rates for COVID-19. In Part A, baseline recovery and death rates were modeled without vaccination, establishing a control for subsequent analysis. In Part B, the prior estimates were adjusted by incorporating cumulative vaccination data. This distinction helps assess COVID-19 impact and supports strategic health planning and public health decision-making.

Part A Results: Baseline Model

The baseline model offers foundational insights into cumulative COVID-19 outcomes without vaccination influence, using active cases, recoveries, and deaths from across healthcare facilities. This model serves as a reference to quantify the vaccination effect in Part B, ensuring a clear basis for comparative evaluation.

Estimated Recovery and Death Rates:

The baseline model defines recovery (γ_1) and death (γ_2) rates follow:

$$\gamma_1 = \frac{970{,}474}{998{,}748} \approx 0.9717 \approx 0.972$$

$$\gamma_2 = \frac{26{,}619}{998{,}748} \approx 0.0267$$

First-order differential equations:
The first-order differential equations for recovery and death rates are defined as:

$$\frac{dA(t)}{dt} = -\gamma_1 A(t)$$

$$\frac{dA(t)}{dt} = -\gamma_2 A(t)$$

Solution for Decline in Active Cases:
The first-order differential equations yield solutions for recovery and death trends:

$$A_{\text{recovery}}(t) = A_0 \cdot e^{-\gamma_1 \cdot t}$$

$$A_{\text{death}}(t) = A_0 \cdot e^{-\gamma_2 \cdot t}$$

where A_0 represents the initial count of active cases. See Appendix B for application.

Interpretation: Collectively, the high recovery rate and the relatively low death rate indicate favorable outcomes for COVID-19 cases in the healthcare system. This suggests a resilient healthcare system, capable of managing surges during a communicable disease outbreak. These baseline rates enable us to assess the impact of vaccination on recovery and death rates, as interpreted in the following section.

Part B Results: Baseline Model with Cumulative Vaccination Data

Part B refines recovery and death rates by incorporating cumulative vaccination data. The vaccination-adjusted rates underscore a measurable improvement in recovery rates and reduction in death rates resulting from public health action. This adjustment highlights the measurable effects of vaccination on recovery and death trends in a healthcare population.

Vaccination-Adjusted Recovery and Death Rates: With the inclusion of cumulative vaccination data, the vaccination-adjusted rates are calculated as follows for recovery ($\gamma_{1,\text{adjusted, fully vaccinated}}$) and death ($\gamma_{2,\text{adjusted, fully vaccinated}}$):

$$\gamma_1\left(Adjusted, V_{\text{fully vaccinated}}\right) = 0.9717 * \left(1 + 0.1 * \frac{5{,}497{,}748}{998{,}748}\right) \approx 1.5066$$
$$\gamma_2\left(A, V_{\text{fully vaccinated}}\right) = 0.0267 * \left(1 - 0.05 * \frac{5{,}497{,}748}{998{,}748}\right) \approx 0.0193$$

First-order differential equations for Adjusted Model:
For the vaccination-adjusted model, the first-order differential equations are:

$$\frac{dA_{\text{recovery}}(t)}{dt} = -\gamma_{1,\text{adjusted}} \cdot A(t)$$

$$\frac{dA_{\text{death}}(t)}{dt} = -\gamma_{2,\text{adjusted}} \cdot A(t)$$

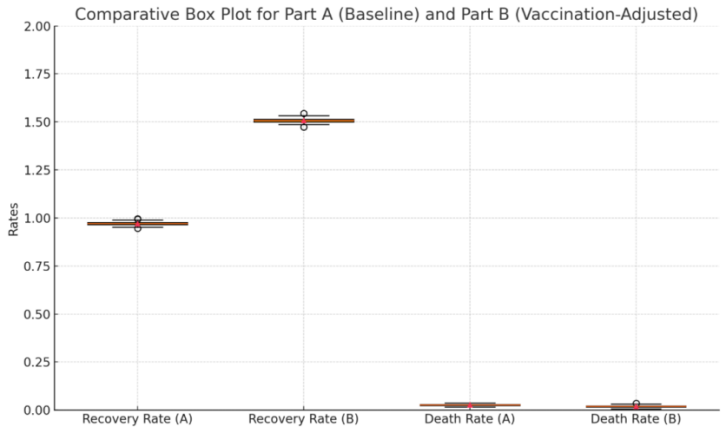
Solution for Decline in Active Cases:

Use the same mathematical forms as shown in Part A. See Appendix B for application.

Interpretation: The adjustments illustrate that the inclusion of cumulative vaccination data led to a higher recovery rate and reduced death rate in the healthcare system. This aligns with broader public health findings where vaccination in a population contributes to reduced death rates from the communicable disease in a healthcare system with high vaccination rates. These adjusted rates highlight the effectiveness of vaccination as shown in the comparative analysis with baseline results.

Comparative Analysis: Part A and Part B

Comparing baseline (Part A) and vaccination data-adjusted (Part B) models reveals significant differences in recovery and death rates, providing a robust foundation for resource allocation, public health policy planning, and targeted public health intervention strategies (see Figure 1). This analysis highlights the model's utility in supporting public health and healthcare administrators in health planning and optimizing resource distribution.



Source: **COVID-19 National Summary** from the U.S. Department of Veterans Affairs (VA), Veterans Health Administration (VHA), dated October 22, 2024.

Figure 1: Comparative Box Plot for Part A (Baseline) and Part B (Vaccination-Adjusted)

In Part A, the baseline recovery rate (γ_1) of 0.972 and death rate (γ_2) of 0.0267 represent COVID-19 case resolution in the unvaccinated population. In contrast, the adjusted model in Part B shows a recovery rate approximately 1.5066 and a death rate approximately by 0.0193, underscoring the cumulative benefits of vaccination.

Again, the consistency check serves as a form of sensitivity analysis, as it evaluates the model's robustness to variations in input datasets and confirms the stability of recovery and death rates.This comparison not only highlights vaccination effectiveness in enhancing recovery outcomes but also provides a clear basis for public health resource planning and healthcare prioritization in a population. The findings underscore the model's decision-support utility, which is further explored in the next section.

4. DISCUSSION

Considerations: For public health and healthcare system decision-makers

The findings from Part A and Part B underscore the critical role of vaccination in managing COVID-19 outcomes within a public healthcare system. Baseline recovery rate in Part A and vaccination-adjusted rates in Part B show a strong capacity for recovery and reduced mortality within VHA healthcare system. This distinction highlights the values of mathematical modeling for this and future communicable disease outbreaks [12]:

- **Resource Allocation:** Facilities with slower case resolution may require additional resources to manage the burden of disease at that location, while those with higher recovery rates could serve as examples of best-practice.

- **Vaccination Strategy Adjustment:** Observing facility-level recovery rates in vaccinated populations can identify areas needing intensified vaccination outreach, particularly in regions with lower vaccination uptake.

- **Operational Planning:** Projected timelines for case resolution assists public health and healthcare system administrators in forecasting healthcare workforce requirements, maintaining essential operations, and preparing for potential surges.

The comparison (see Figure 1) reminds administrators how many lives can be spared from the burden of disease by having resource savailable to support clinical providers in delivering vaccination and treatment to protect the population from communicable disease while reducing the death rate in the population. While the homogeneous mixing assumption simplifies interactions for modeling purposes, it aligns with the model's focus on providing actionable insights for health system decision-making rather than individual-level outcomes. Further adaptions could incorporate population heterogeneity and regional differences to improve applicability. Although this model relies on cumulative vaccination data, future adaptations could incorporate vaccination timing to improve dynamic modeling of population immunity.

The public health data collected during the population during the public health emergency would be processed and analyzed using both statistical techniques for probabilistic trends and mathematical methods for precise articulation of mathematical values on an ongoing basis (see Appendix B). The enterprise-level model would be solved using public health data from a geographic area, making the results applicable to that region. The findings, combined with public health insights, could help decision-makers fine-tune efforts to support frontline clinical providers. Although beyond the scope of this project, future iterations of the model might incorporate geographic-specific trends to further refine predictions and support localized public health decision-making.

Considerations: For the baseline model

The baseline model assumes constant recovery and death rates, as commonly used in real-time applications to simplify decision-making when using a linear first-order differential equation. Such models are effective for real-time public health surveillance and decision-making due to their adaptability and simplicity [1,11]. While effective for rapid insights, this assumption may limit long-term precision. Future updates could incorporate time-dependent factors to increase accuracy while balancing computational demands [3,9]. The baseline model equations are revisited:

Model Setup: Active cases A(t), cumulative recoveries C, and deaths D are used to model disease progression.

Equations:
These equations model rate of decrease in active cases due to recovery and, separately, for deaths:

$$dA(t)/dt = -\gamma_1 A(t)$$
$$dA(t)/dt = -\gamma_2 A(t)$$

where γ_1 and γ_2 represent the recovery rate and death rate, respectively.

The minus sign is important and represents the decrease in active cases over time due to recovery and death [11]. This sign indicates a negative rate of change, aligning with real-world epidemiological dynamics where active cases are expected to decline as individuals either recover or pass away. Without the minus sign, these equations would incorrectly suggest an increase in active cases as recoveries or deaths occur, misrepresenting disease progression and undermining the utility of the model as a guide for public health decision-making [9].

Parameters:

These constants, $\gamma_1 \; and \; \gamma_2$, are estimated based on the cumulative data:

$$\gamma_1 = \frac{C}{A + C + D}$$
$$\gamma_2 = \frac{D}{A + C + D}$$

The terms C and D represent the cumulative recoveries and deaths, respectively. These provide long-term estimates of recovery and death rates without vaccination.

Appendix A provides a structured data management protocol that includes data cleaning, data validation, and exploratory data analysis (EDA). By validating and cleaning data with this structured approach, the model can utilize accurate case counts, recovery rates, and death rates, which are essential for calculating parameters like $-\gamma_1 \; and \; -\gamma_2$. The appendix highlights the careful handling of real-world data, ensuring that the model's outputs would be reliable and actionable for public health decision-makers.

Considerations: Consistency Check-a sensitivity analysis

The model findings show that vaccination improves COVID-19 recovery rates and reduces mortality among those using the VHA system. Baseline recovery and death rates were previously reported; corresponding rates adjusted for vaccination were also documented earlier.

To validate these findings, a consistency check was conducted using new, unseen data (see Methods section for the identification of File C). These data were cleaned following the same actions described in the Methods section. This check revealed the following descriptive summary of the new data file:

Total Confirmed Cases: 1,000,174
Active Cases: 1,528
Recovered Cases (Convalescent): 971,976
Known Deaths: 26,670

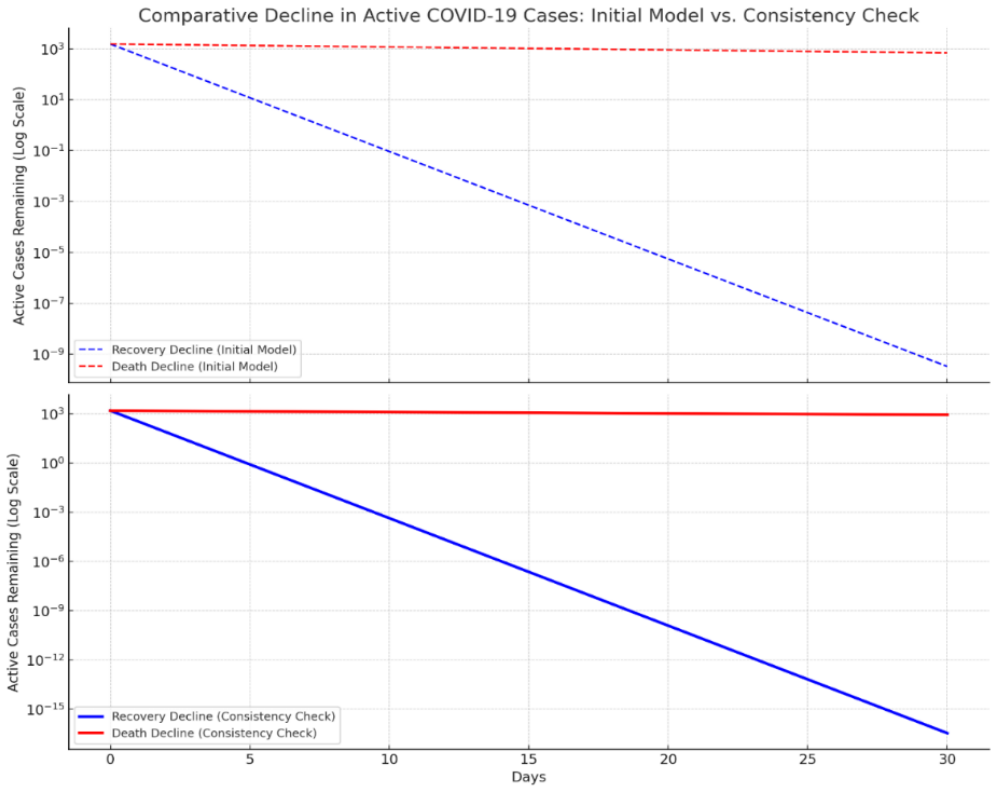
In Part A, the baseline recovery rate (γ_1) = 0.9718 and the baseline death rate (γ_2) = 0.0267. Furthermore, in Part B, the vaccination vaccination-adjusted rates were recovery rate (γ_1) = 1.5078 and death rate (γ_2) = 0.0193. The comparison of rates is in the following (see Table):

Table. Comparison of Rates: Baseline Model vs. Consistency Check

	recovery rate (γ_1)	**death rate (γ_2)**
Part A		
File B: Initial Baseline Model	0.9717	0.0267
File C: Consistency Check	0.9718	0.0267
Part B		
File B: Initial Baseline Model	1.5066	0.0193
File C: Consistency Check	1.5078	0.0193

Note: The differences are within the pre-established tolerance.

Appendix B illustrates the decline in active cases by solving the first-order differential equations for values of t, time in days. Calculations confirm that the recovery and death rates for both the Baseline Model and Consistency Check Model remain within acceptable tolerance, highlighting stability in numerical results. Figure 2 further illustrates this consistency visually, showing nearly identical trends in active case decline between the Baseline Model and the Consistency Check Model. This alignment in both numerical and graphical results reinforces the model's reliability for real-time public health decision-making.



Source: **COVID-19 National Summary** from the U.S. Department of Veterans Affairs (VA), Veterans Health Administration (VHA), dated: October 22, 2024 and November 7, 2024

Figure 2: Comparative Decline in Active COVID-19 Cases: Initial Model vs. Consistency Check

In the dual line graph with subplots (Figure 2), the comparative decline in active COVID-19 cases for both models is evident. This graph expands on the differential equation approach in Appendix B, where exponential decay function model the reduction of active cases through recovery and death rates. The logarithmic scale on the y-axis effectively captures this exponential decay. The recovery trend, indicated by blue lines, shows a steeper decline compared to the death trend, which is shown in red. This emphasizes the importance of recovery as a factor reducing the number of active cases in the healthcare system, even over a longer time period.

The transparency of the methods and the deliberate use of first-order differential equations in this project facilitate adaptability to other public health and health system needs. A key strength of this approach lies in its reliance on regularly collected, high-quality aggregate data, which is commonly available in many public health settings. By leveraging such data and applying straightforward applied calculus, this model can be adapted to address similar decision-making needs in other public health settings. For instance, a large public health district managing a communicable disease outbreak with vaccination and case resolution data could utilize this model to support resource allocations, operational planning and public health intervention strategies. The combination of accessible methods, cumulative data, and replicable techniques provides a generalizable framework for public health decision-making.

The use of first-order differential equations reflects a deliberate focus on accessibility and practicality. As a topic familiar to many professionals through applied calculus, these equations provide a transparent and effective framework for addressing real-time challenges in public health decision-making. This contrasts with more complex models, which often demand specialized expertise and computational resources, potentially limiting their application in health systems. The linear first-order differential equation approach applied here appears appropriate for public health and healthcare system decision-making. This suggests that the model could provide practical value for public health strategy and resource allocation, as summarized in the final section.

5. CONCLUSION

This project modeled COVID-19 case resolution within the Veterans Health Administration healthcare system using a first-order differential equations, providing a practical tool for public health decision-making. The model's baseline recovery and death rates of 0.9717 and 0.0267, respectively. With vaccination data, these rates adjust to 1.5066 (recovery) and 0.0193 (death), underscoring the impact of vaccination. This model is a reliable tool for real-time public health decision-making, demonstrating that simplified differential equations can track and forecast outcomes in healthcare systems.

For public health epidemiologists, the Methods section provides a structured approach for data management and model application, supporting mathematical modeling in epidemiology. Appendix A enhances transparency and reproducibility, offering a reliable framework tailored to the rigorous demands of public health epidemiology [13,14]. Appendix B offers a clear model of the decline in active cases over time, aiding in public health forecasting and strategic planning. The project underscores the utility of simplified, first-order differential equation models in epidemiology, offering reliable insights with minimal computational demands. Consistency checks confirm model stability across datasets, supporting its use in real-time public health applications.

ACKNOWLEDGEMENTS

I used GenAI/ChatGPT Data Analyst, by OpenAI, in data management and preparation, confirming my calculations, graphing data, brain-storming, and proof-reading the manuscript. Date: November 14, 2024.

REFERENCES

[1] Mac S, Flahault A, Opatowski L. Modeling the COVID-19 pandemic: A comprehensive guide of infectious disease and decision-analytic models. J Clin Epidemiol. 2021;131(1):25–35.
[2] Hye Y, Song H, Choi M. A mathematical model for the transmission of co-infection with COVID-19 and kidney disease. J Infect Public Health. 2024;17(1):32–40.
[3] Bjornstad O, Shea K, Krzywinski M, Altman N. Modeling infectious epidemics. Nat Methods. 2020;17(5):455–56.
[4] Grassly N, Fraser C. Mathematical models of infectious disease transmission. Nat Rev Microbiol. 2008;6(6):477–87.
[5] Mata B, Dourado C. Mathematical modeling applied to epidemics: Challenges and opportunities with complex networks. PLoSComput Biol. 2021;17(9).
[6] Alfaki S, Masih M. Modeling and forecasting using time-series ARIMA models: An application to Naphtha sales in Libya. Energy Econ. 2015;49:570–80.
[7] Nikolaou M. Revisiting SIR modeling for infectious diseases: Improvements with Padé approximations. Math Biosci. 2022;345(3):1052–60.
[8] Johns Hopkins Bloomberg School of Public Health. Infectious Disease Transmission Modeling Primer. 2020.
[9] Brauer F, Castillo-Chavez C, Feng Z. "Mathematical Models in Epidemiology." New York: Springer; 2019.
[10] Fant GV. Mathematical disease modeling for public health epidemiologists: differential equations supporting public health interventions. Applied Mathematics and Sciences: An International Journal (MathSJ). 2024; 11(3):9-21.
[11] Vynnycky E, White RG. "An Introduction to Infectious Disease Modelling." Oxford: Oxford University Press; 2011.
[12] Thomas RK. “Health Services Planning, second edition.” New York: Kluwer Academics/Plenum Publishers, 2003.
[13] Friis RH, Sellers TA. “Epidemiology for Public Health Practice, Sixth Edition.” Burlington, MA: Jones & Bartlett, 2021.
[14] Hopkins RS, Buehler JW. Public Health Surveillance. In: “Modern Epidemiology, Fourth Edition,” eds.TL Lash, TJ VanderWeele, S Haneuse, KJ Rothman. Philadelphia: Wolters Kluwer, 2021.

AUTHOR

Dr GV Fant, DSc, PhD, MACE, Epidemiologist, Population Health Informatics, Digital Health Office, Veterans Health Administration, U.S. Department of Veterans Affairs, Washington, DC.

Disclaimer: The views expressed in the paper are those of the author and do not represent the official position of the U.S. Government nor the U.S. Department of Veterans Affairs.

Key publication dates: Submitted: 16/11/2024; Accepted: 25/01/2025; Revised: 30/01/2025; Approved: 31/01/2025; Published Online : 31/01/2025

Appendix A: Structured Data Management and Analysis Protocol

1. **Define the Objective**

- **Purpose:** Clearly define the **purpose** of the project or analysis.
- **Outcome Goals:** Identify the specific **outcomes** you are aiming to achieve (e.g., understanding disease dynamics, improving resource allocation).

2. **Data Collection and Ingestion**

- **Source Identification:** Ensure that all data sources (e.g., ELR, vaccination reports, case data) are properly identified.
- **Data Ingestion:** Define a **structured process** for data import. Use clear variable names and appropriate data formats.

3. **Data Cleaning and Preprocessing**

- **Data Cleaning Process:** Specify the steps taken to clean and prepare the data.
- **Data Structuring:** Ensure that the data is structured in a logical, analyzable format (e.g., long or wide format).

4. **Exploratory Data Analysis (EDA)**

- **Initial Data Exploration:** Perform initial exploration to understand the distribution of key variables and any correlations.
- **Visualization:** Create simple visualizations (histograms, bar charts, scatter plots) to identify trends and outliers.

5. **Define the Analytical Approach**

- **Method Selection:** Choose appropriate **mathematical models** or **analytical techniques**.
- **Algorithm Transparency:** Clearly document any GenAI models, algorithms, or tools used. Avoid black-box approaches by explaining: How the algorithm processes the data and the rationale behind its selection.

6. **Data Analysis Execution**

- **Step-by-Step Process:** Break down the analysis into clear, understandable steps.
- **First-order differential equations**: Apply first-order differential equations to model changes in active or convalescent cases over time, estimating rates like recovery and transmission based on the data.

7. **Validation and Quality Assurance**

- **Verification of Results:** Ensure the accuracy of the analysis by validating results at each step.
- **Cross-Check with Other Methods:** Use alternative analytical methods to verify that results remain consistent and reliable.

8. **Interpretation of Results**

- **Data-Driven Conclusions:** Translate the analytical outcomes into clear, actionable conclusions.
- **Limitations and Assumptions:** Document any assumptions made during analysis (e.g., constant recovery rate in first-order differential models).

9. Reporting and Visualization

- **Report Generation:** Create structured reports with tables, visualizations, and narratives explaining the analysis.
- **Visualization of Findings:** Ensure that all visualizations used (e.g., charts, graphs) are clear, labeled, and easy to interpret for stakeholders.

Appendix B: Application-Solution for Decline in Active Cases

This appendix applies first-order differential equations from the Methods Section to model the decline in active COVID-19 cases. Step-by-step solutions for both data files demonstrate exponential decay driven by recovery and death rates, supporting the findings in the main text.

Part A: Baseline Model
Recovery Rate: $A_{\text{recovery}}(t) = A_0 \cdot e^{-\gamma_1 \cdot t}$

where A_0 is initial count of cases ($A_0 = 10.03$) ; t is days

$A_{\text{recovery}}(t) = A_0 \cdot e^{-0.972(t)}$

File B: Initial Baseline Model
t = 0: $A_0 \cdot e^{-0.972(0)} = 10.03$
t = 1: $A_0 \cdot e^{-0.972(1)} = 3.79$
t = 7: $A_0 \cdot e^{-0.972(7)} = 0.011$

$A_{\text{recovery}}(t) = A_0 \cdot e^{-0.972(t)}$

File C: Consistency Check Model
t = 0: $A_0 \cdot e^{-0.972(0)} = 10.03$
t = 1: $A_0 \cdot e^{-0.972(1)} = 3.79$
t = 7: $A_0 \cdot e^{-0.972(7)} = 0.011$

Death Rate: $A_{\text{death}}(t) = A_0 \cdot e^{-\gamma_2 \cdot t}$

$A_{\text{death}}(t) = A_0 \cdot e^{-0.027(t)}$
t = 0: $A_0 \cdot e^{-0.027(0)} = 10.03$
t = 1: $A_0 \cdot e^{-0.027(1)} = 9.76$
t = 7: $A_0 \cdot e^{-0.027(7)} = 8.30$

$A_{\text{death}}(t) = A_0 \cdot e^{-0.027(t)}$
t = 0: $A_0 \cdot e^{-0.027(0)} = 10.03$
t = 1: $A_0 \cdot e^{-0.027(1)} = 9.76$
t = 7: $A_0 \cdot e^{-0.027(7)} = 8.30$

Part B: Baseline Model with Cumulative Vaccination Data
Recovery Rate: $A_{\text{recovery}}(t) = A_0 \cdot e^{-\gamma_{1,adj} \cdot t}$

$A_{\text{recovery}}(t) = A_0 \cdot e^{-1.506(t)}$

File B: Initial Baseline Model
t = 0: $A_0 \cdot e^{-1.506(0)} = 10.03$
t = 1: $A_0 \cdot e^{-1.506(1)} = 2.224$
t = 7: $A_0 \cdot e^{-1.506(7)} = 0.00026$

$A_{\text{recovery}}(t) = A_0 \cdot e^{-1.508(t)}$

File C: Consistency Check Model
t = 0: $A_0 \cdot e^{-1.508(0)} = 10.03$
t = 1: $A_0 \cdot e^{-1.508(1)} = 2.220$
t = 7: $A_0 \cdot e^{-1.508(7)} = 0.00026$

Death Rate: $A_{\text{death}}(t) = A_0 \cdot e^{-\gamma_{2,adj} \cdot t}$

$A_{\text{death}}(t) = A_0 \cdot e^{-0.019(t)}$

t = 0: $A_0 \cdot e^{-0.019(0)} = 10.03$
t = 1: $A_0 \cdot e^{-0.019(1)} = 9.84$
t = 7: $A_0 \cdot e^{-0.019(7)} = 8.78$

$A_{\text{death}}(t) = A_0 \cdot e^{-0.019(t)}$

t = 0: $A_0 \cdot e^{-0.019(0)} = 10.03$
t = 1: $A_0 \cdot e^{-0.019(1)} = 9.84$
t = 7: $A_0 \cdot e^{-0.019(7)} = 8.78$