Heliyon 10 (2024) e38204

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A vaccination-based COVID-19 model: Analysis and prediction using Hamiltonian Monte Carlo

Touria Jdid [a,*], Mohammed Benbrahim [a], Mohammed Nabil Kabbaj [a], Mohamed Naji [b]

[a] Laboratory of Engineering, Modeling and Systems Analysis (LIMAS), Faculty of Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco
[b] Laboratory of Applied Physics Informatics and Statistics (LPAIS), Faculty of Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco

ARTICLE INFO

Dataset link: https://github.com/CSSEGISandData/COVID-19

Dataset link: https://github.com/touria2000/COV19Sim

Keywords:
Epidemic models
COVID-19
Vaccination
Bayesian inference
Hamiltonian Monte Carlo
Prediction

ABSTRACT

Compartmental models have emerged as robust computational frameworks and have yielded remarkable success in the fight against COVID-19. This study proposes a vaccination-based compartmental model for COVID-19 transmission dynamics. The model reflects the specific stages of COVID-19 infection and integrates a vaccination strategy, allowing for a comprehensive analysis of how vaccination rates influence the disease spread. We fit this model to daily confirmed COVID-19 cases in Tennessee, United States of America (USA), from June 4 to November 26, 2021, in a Bayesian inference approach using the Hamiltonian Monte Carlo (HMC) algorithm. First, excluding vaccination dynamics from the model, we estimated key epidemiological parameters like infection, recovery, and disease-induced death rates. This analysis yielded a basic reproduction number (R_0) of 1.5. Second, we incorporated vaccination dynamics and estimated the vaccination rate for three vaccines: 0.0051 per day for both Pfizer and Moderna and 0.0059 per day for Janssen. The fitted curves show reductions in the epidemic peak for all three vaccines. Pfizer and Moderna vaccines bring the peak down from 8,029 infected cases to 5,616 infected cases, while the Janssen vaccine reduces it, to 6,493 infected cases. Simulations of the model by varying the vaccination rate and vaccine efficacy were performed. A highly effective vaccine (95% efficacy) with a daily vaccination rate of 0.006 halved COVID-19 infections, reducing cases from 8,029 to around 4,000. The results also show that the model's prediction accuracy for new observations improves with the number of observed data used to train the model.

1. Introduction

Approximately four years after its first identification in Wuhan, China in late December 2019 [1], COVID-19 remains an ongoing global crisis affecting different sectors of society. Although the incidence of COVID-19 is decreasing, there is still the threat of emerging new waves [2], especially with the parallel spread of several strains of COVID-19 [3]. Therefore, continuing to combat the disease that has caused 6,972,152 deaths and 771,407,825 confirmed cases worldwide as of October 19, 2023 [4], is crucial to prevent further loss of life and mitigate the impact of the current pandemic.

Compartmental models have emerged as a robust computational framework and have demonstrated remarkable success in the fight against COVID-19 disease. These models have been used to understand the dynamics of COVID-19 [5–7], assess intervention strategies

* Corresponding author.
E-mail address: touria.jdid@usmba.ac.ma (T. Jdid).

https://doi.org/10.1016/j.heliyon.2024.e38204
Received 30 October 2023; Received in revised form 18 September 2024; Accepted 19 September 2024

Available online 23 September 2024

[8–10], predict disease outcomes [11–13], estimate resource needs [14–16], evaluate vaccine impact [17–19], allow data analysis and surveillance [20–22], and scenario planning [19,23,24]. Despite their particularities, most of these models have adopted the Susceptible-Infected-Recovered (SIR) [25] and Susceptible Exposed Infectious Recovered (SEIR) [26] frameworks as their foundation, expanding upon them through integrating additional compartments.

In the context of enhancing the understanding of COVID-19 dynamics, various approaches have been explored. For instance, a fractional-order compartmental model incorporating asymptomatic and hospitalized populations has improved accuracy in capturing the disease's dynamics [27]. Another significant contribution to the field is a model that considers both human-to-human and environment-to-human transmission, underlining the environment's critical role in virus spread [28]. The interaction between SARS-CoV-2 and other infections, such as HBV, has been explored using a co-dynamical model, emphasizing the importance of time-variant controls in mitigating dual infections [29].

Understanding the impact of control measures on COVID-19 transmission is crucial for effective outbreak management. A study demonstrated the effectiveness of various control measures in preventing disease transmission [30]. Additionally, models focusing on COVID-19 dynamics in the presence of comorbidities, like diabetes mellitus, have been developed to identify optimal control strategies for disease prevention [31]. Innovative approaches have also been applied to specific regions, such as Hermosillo, Mexico, where population movement patterns and local demographics were incorporated into a model using the HMC algorithm to address parameter estimation challenges [32].

Modeling studies have extensively investigated the impact of vaccination on controlling COVID-19 outbreaks, especially focusing on developing optimal vaccination strategies. Kiem et al. [24] examined vaccine prioritization and the relaxation of control measures, stressing the importance of targeting at-risk populations to reduce morbidity and mortality, and pointed out that very high vaccination coverage may be required to relax control measures. Additionally, an analysis in [33] assessed multiple vaccination policies using different vaccines, achieving significant reductions in severe COVID-19 cases through diverse approaches.

Research by [17] developed policies based on varying vaccine efficacies and immunity periods to minimize the disease burden and maintain hospital occupancy below critical levels. Similarly, a dynamic model projected the epidemic trajectory of COVID-19 in major US states under different vaccine effectiveness and social distancing scenarios, underscoring the need for high vaccine coverage and continued non-pharmaceutical interventions (NPIs) to suppress the epidemic [19].

Herd immunity is identified as fundamental in informing the required immunization coverage to control disease spread and protect the population effectively [34]. In their study, [18] assessed the impact of an imperfect vaccine on transmission dynamics of COVID-19 in the US, highlighting that combining vaccination with interventions such as face masks and social distancing is crucial for achieving herd immunity. Likewise, the study in [35] focused on achieving herd immunity under limited and full vaccine supply scenarios, indicating that mass vaccination, efficient post-exposure prophylaxis, and rapid distribution are essential for effective outbreak control.

Further studies have explored the evolution of the COVID-19 pandemic through multiple waves, the emergence of new variant strains, and the impact of vaccination efforts. Study [3] described the progression through three distinct waves, noting variations in disease severity, symptoms, and public attitudes, and emphasized the challenges posed by highly transmissible variants like Delta and Omicron, raising concerns about vaccine efficacy and the potential need for annual vaccinations. Parolini et al. [36] also highlighted the impact of vaccination campaigns and the dominance of new virus variants in shaping the dynamics of the COVID-19 pandemic in Italy, offering insights into effective control strategies.

In building epidemic models, the values of the parameters play a determining role in the reliability of the results provided by the model. Thus, the numerical technique for inferring model parameters from infectious disease count data is essential and must be chosen carefully to formulate a mathematical model usable in practice. In this context, Bayesian statistical inference [37] has received increasing attention from the scientific community because of its capability to accurately estimate numerous parameters, even though they are highly correlated.

Over recent years, the Hamiltonian Monte Carlo (HMC) algorithm [38], a robust Markov Chain Monte Carlo (MCMC) method, has proven efficient in searching high-dimensional parameter spaces [39]. This method has become a vital tool for fitting epidemic models, addressing challenges in parameter estimation and uncertainty quantification [10,21,32,39,40]. During the 2020 COVID-19 outbreak in Hermosillo, Mexico, HMC was leveraged in a multi-patch epidemic model to tackle these challenges effectively [32]. Similarly, Iyaniwura et al. [40] employed HMC to fit an age-structured model to COVID-19 data in British Columbia, demonstrating the utility of time-varying contact rates for precise analysis of transmission dynamics.

Furthermore, HMC has been applied in innovative ways to enhance the accuracy of epidemiological parameter estimation. One study used HMC to estimate delay distributions in COVID-19 surveillance data, highlighting the necessity of high-resolution data [21]. Another study integrated HMC within a semiparametric framework using the rstan package in R to model COVID-19, resulting in accurate short-term forecasts for healthcare indicators in Slovenia [10].

Efforts in this field also include the development of tutorials and methodological enhancements to support the effective use of HMC in epidemic modeling. A comprehensive tutorial provides an in-depth guide on building, fitting, and diagnosing disease transmission models in Stan, underscoring Bayesian modeling's strength in quantifying uncertainty and incorporating prior knowledge [41]. Moreover, a detailed investigation into the application of HMC within Stan software emphasized the trade-off between statistical efficiency and computational speed when fitting epidemic models to real outbreak data [42].

In this study, we propose a vaccination-based compartmental model to capture, analyze, and predict the transmission of COVID-19. The model reflects individuals' specific stages of COVID-19 infection and presents a vaccination strategy, allowing for a comprehensive analysis of how vaccination rates influence the disease spread. To estimate key epidemiological parameters including infection rates and R_0, we fit our model to COVID-19 incidence data for Tennessee, USA, using a Bayesian inference approach with the HMC

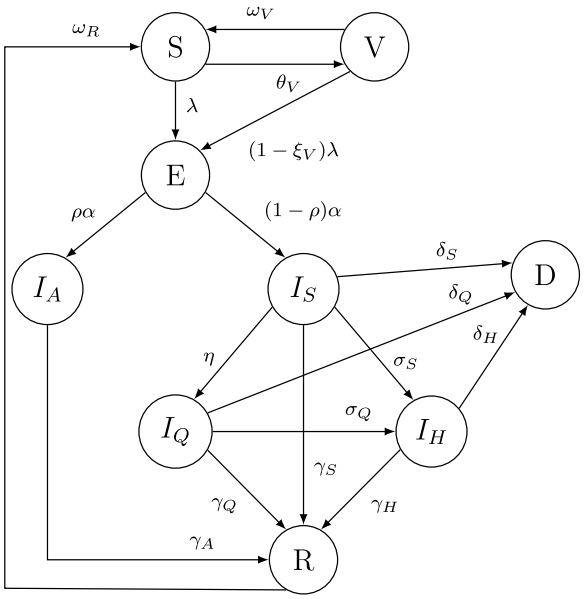

Fig. 1. Schematic illustration of the proposed epidemic model. The total population is partitioned into nine classes: Susceptible S, Exposed E, Symptomatic Infectious I_S, Asymptomatic Infectious I_A, Quarantined I_Q, Hospitalized I_H, Recovered R, Deceased D, and Vaccinated V. The four classes I_S, I_A, I_Q, and I_H contribute to COVID-19 infection.

algorithm. Our model involves estimating vaccination rates for Pfizer, Moderna, and Janssen vaccines. It analyzes how vaccination impacts and potentially controls COVID-19 incidence. Furthermore, the model's predictive ability for unseen data (new observations) increases with the amount of training data (observed days).

Our paper is structured as follows. Section 2 provides the groundwork by formulating an epidemiological model. We then present the COVID-19 data used in our analysis (Section 3). Next, Section 4 delves into the model by deriving the basic reproduction number (R_0). Section 5 outlines the Bayesian inference approach, which allows us to estimate unknown model parameters and generate predictions. Section 6 details the calibration process using COVID-19 data and subsequent evaluation of its predictive capabilities. Finally, Section 7 summarizes key findings and future directions of our research.

2. Model formulation

To conceive a relatively basic epidemiological model for capturing the dynamics of COVID-19 in a human community, we partitioned the overall size of the population of interest at a given time t defined by N(t) into nine classes or compartments. Based on the infection state of individuals, these classes are: Susceptible $S(t)$, Exposed $E(t)$, Symptomatic Infectious $I_S(t)$, Asymptomatic Infectious $I_A(t)$, Quarantined $I_Q(t)$, Hospitalized $I_H(t)$, Recovered $R(t)$, Deceased $D(t)$, and Vaccinated $V(t)$. The diagram in Fig. 1 provides a schematic representation of how individuals transition between different classes.

The infection process for COVID-19 is as follows: At the beginning of the outbreak, all individuals are assumed to be equally susceptible, and they contract the virus upon contact with an asymptomatic, symptomatic, quarantined, or hospitalized individual. The transmission coefficients were $\beta_A I_A(t)$, $\beta_S I_S(t)$, $m_Q\beta_S I_Q(t)$, and $m_H\beta_S I_H(t)$. β_A and β_S indicate the rates of asymptomatic and symptomatic infections, respectively. m_Q and m_H represent the reduction factors of infectivity by a quarantined and hospitalized individual, compared with the symptomatic infectious class, and are constrained by $0 < m_Q < 1$ and $0 < m_H < 1$. Individuals infected with the SARS-CoV-2 virus typically experience a time delay before the onset of symptoms. This time delay, referred to as the incubation period [43], has an average duration ranging between 5 and 5.2 days, as reported in [43–45]. During this period $\left(\frac{1}{\alpha}\right)$, we assume that exposed individuals cannot transmit the disease to others. At the end of the incubation period, a fraction ρ of the exposed individuals displayed no clinical symptoms and transitioned to the I_A class. The remaining fraction $(1-\rho)$ develops clinical signs of COVID-19 and moves to the I_S class. Asymptomatic carriers, a major factor in COVID-19 transmission dynamics [46–48], exhibit viral loads similar to symptomatic individuals [49]. Compartment I_A in our model represents these asymptomatic infectious individuals who go unreported by authorities. They naturally recover from the illness at a rate of γ_A. Class I_S encompasses symptomatic infectious individuals diagnosed and officially reported by the authorities. Diagnosed cases of COVID-19 may be admitted to the hospital (moved to class I_H) for treatment and isolation, enter home quarantine (moved to class I_Q) to avoid contact with others, naturally recover from the disease, or, unfortunately, die from it. Intra-hospital transmission of COVID-19 from patients to visitors or healthcare workers is a critical issue in the COVID-19 outbreak owing to imperfect isolation. Thus, we assumed that hospitalized patients are isolated and treated but can still transmit the disease to others, with a reduction factor of infectivity m_H. They recovered from COVID-19 at a rate of γ_H or suffered disease-related mortality at a rate of δ_H. Infectious individuals quarantined at home can still spread the disease with a reduction factor of infectivity m_Q. These individuals are monitored, and if their disease condition progresses, they are promptly

Table 1
Meanings of the parameters in the epidemic model.

Parameter	Interpretation
N	Population size
β_S	Rate of infection from symptomatic to susceptible people
β_A	Rate of infection from asymptomatic to susceptible people
α	Incubation rate
ρ	Proportion of asymptomatic individuals
γ_S	Recovery rate for symptomatic individuals
γ_A	Recovery rate for asymptomatic individuals
γ_H	Recovery rate for hospitalized individuals
γ_Q	Recovery rate for quarantined individuals
σ_S	Rate of transition from symptomatic to hospitalized class
σ_Q	Rate of transition from quarantined to hospitalized class
η	Rate of transition from symptomatic to quarantined class
δ_S	Disease-induced death rate due to symptomatic class
δ_Q	Disease-induced death rate due to quarantined class
δ_H	Disease-induced death rate due to hospitalized class
ω_R	Rate of waning immunity among recovered individuals
ω_V	Rate of waning immunity among vaccinated individuals
θ_V	Vaccination rate
ξ_V	Vaccine efficacy

hospitalized at the rate of σ_Q. They recovered from COVID-19 and died because of it at the rates of γ_Q and δ_Q, respectively. A report suggests that natural immunity to SARS-CoV-2 may decline over time [50], increasing the risk of reinfection. Consequently, recovered individuals might experience waning immunity and return to a susceptible state at a rate of ω_R. Compartment D incorporates those who have died from the disease. Because the simulation was limited to a short period, the model did not include the potential effects of births and natural deaths on disease transmission.

When formulating the vaccination strategy, the following assumptions are taken into account:

- The only variant under consideration is COVID-19.
- The vaccine is only administered to the susceptible population.
- Given the vaccine's profile, an individual eligible for vaccination is expected to get one or two doses.
- Only susceptible individuals exhibit the effects of the vaccine.
- The COVID-19 vaccine is administered to the susceptible population at a rate of θ_V.
- The vaccine's efficacy, denoted by ξ_V, is assumed to be imperfect, where $0 < \xi_V < 1$.
- The likelihood of vaccinated individuals being infected with COVID-19 is possible, with a probability of $(1 - \xi_V)$.
- Vaccinated individuals will experience a reduction in immunity at a rate of ω_V.
- The immunity acquired through vaccination is identical to that achieved through natural infection ($\omega_V = \omega_R$).

Table 1 summarizes the interpretations of the epidemiological parameters applied in the model. The system of ordinary differential equations (ODEs) that controls the dynamics of the disease is expressed by Equation (1)

$$\begin{cases} \frac{dS(t)}{dt} = \omega_R R(t) + \omega_V V(t) - (\lambda + \theta_V) S(t), \\ \frac{dE(t)}{dt} = \lambda S(t) + (1 - \xi_V)\lambda V(t) - \alpha E(t), \\ \frac{dI_A(t)}{dt} = \rho\alpha E(t) - \gamma_A I_A(t), \\ \frac{dI_S(t)}{dt} = (1 - \rho)\alpha E(t) - (\delta_S + \sigma_S + \gamma_S + \eta) I_S(t), \\ \frac{dI_Q(t)}{dt} = \eta I_S(t) - (\delta_Q + \sigma_Q + \gamma_Q) I_Q(t), \\ \frac{dI_H(t)}{dt} = \sigma_S I_S(t) + \sigma_Q I_Q(t) - (\delta_H + \gamma_H) I_H(t), \\ \frac{dR(t)}{dt} = \gamma_A I_A(t) + \gamma_S I_S(t) + \gamma_Q I_Q(t) + \gamma_H I_H(t) - \omega_R R(t), \\ \frac{dD(t)}{dt} = \delta_S I_S(t) + \delta_Q I_Q(t) + \delta_H I_H(t), \\ \frac{dV(t)}{dt} = \theta_V S(t) - (\omega_V + (1 - \xi_V)\lambda) V(t), \end{cases} \tag{1}$$

where λ refers to the force driving infection among susceptible individuals, as defined in Equation (2):

$$\lambda = \frac{\beta_A I_A(t) + \beta_S I_S(t) + m_Q \beta_S I_Q(t) + m_H \beta_S I_H(t)}{N(t)}, \tag{2}$$

and N(t), representing the total population at time t, is given in Equation (3) by:

$$N(t) = S(t) + E(t) + I_A(t) + I_S(t) + I_Q(t) + I_H(t) + R(t) + D(t) + V(t). \tag{3}$$

2.1. Non-negativity and boundedness of the solutions

We assume the model described by Equation (1) starts with the given initial conditions $(S(0), E(0), I_A(0), I_S(0), I_Q(0), I_H(0), R(0), D(0), V(0)) \in \mathbb{R}^9_+$, where $\mathbb{R}^9_+ = \{(X_1, \ldots, X_9) \,|\, X_i \geq 0, i = 1, \ldots, 9\}$ denotes the non-negative cone of $\mathbb{R}^9$. We demonstrate that all solutions to model (1) remain within the non-negative orthant $\mathbb{R}^9_+$, starting from any point within this region [51].

Theorem 1. *The model system (1) is bounded in the region* $\mathbb{R}^9_+$.

Proof. See Appendix A. □

3. Data

This study used state-level time series for daily confirmed COVID-19 cases data publicly accessible via the repository [52] maintained by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Tennessee State in the USA is chosen as a case study, with its total population estimated at 6,910,840 inhabitants based on the 2020 US Decennial Census data [53]. The model was parameterized on these data over six months, from June 4 to November 26, 2021, with Tennessee's total population size being the key parameter N of the model.

4. Basic reproduction number

The basic reproduction number, R_0, serves as a fundamental epidemiological metric representing the expected number of secondary infections generated by a single infected individual during the infection period in a population entirely susceptible to the disease [54]. R_0 determines whether a pathogen can persist within a population. If R_0 is greater than one, it indicates that each infectious individual, on average, is infecting more than one other individual, which suggests that the disease has the potential to spread in the population. Conversely, if R_0 is less than one, it means that each infected individual, on average, is infecting less than one other individual, indicating that the disease is likely to die out in the population [55]. In the computation of R_0, we excluded vaccination dynamics from our model (Equation (1)), as R_0 represents a scenario of complete susceptibility to the disease, assuming no immunity exists in the population [34,56]. The Disease-Free Equilibrium (DFE) state, $X_0 = (S^0, E^0, I_A^0, I_S^0, I_Q^0, I_H^0, R^0, D^0)$, is obtained by setting the right-hand sides of the equations in the system (1) to zero. When no disease is present in the population $(I_A = I_S = I_Q = I_H = 0)$, the system leads to the DFE state, as detailed in Equation (4):

$$X_0 = (N(0), 0, 0, 0, 0, 0, 0, 0), \tag{4}$$

where $N(0)$ is the initial total population size.

The analytical expression of R_0 is determined using the next-generation matrix method, following the notation of Van den Driessche and Watmough in [55]. Let $X = (E, I_A, I_S, I_Q, I_H)^T$ be the vector of infected compartments. Then, for each compartment i representing infected individuals, the dynamics are given by Equation (5):

$$\dot{X}_i = \mathcal{F}_i(X) - \mathcal{V}_i(X), \tag{5}$$

where $\mathcal{F}_i$ represents the rate at which new infections arise in compartment i and $\mathcal{V}_i$ is the rate at which individuals enter or leave infected compartments by all other means. From system (1), we derive $\mathcal{F}$ and $\mathcal{V}$, such that $\mathcal{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_5)^T$ and $\mathcal{V} = (\mathcal{V}_1, \ldots, \mathcal{V}_5)^T$, respectively, as

$$\mathcal{F}(X) = \begin{pmatrix} (\beta_A \frac{I_A}{N} + \beta_S \frac{I_S}{N} + m_Q \beta_S \frac{I_Q}{N} + m_H \beta_S \frac{I_H}{N})S \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and,

$$\mathcal{V}(X) = \begin{pmatrix} \alpha E \\ \gamma_A I_A - \rho \alpha E \\ (\delta_S + \sigma_S + \gamma_S + \eta) I_S - (1-\rho)\alpha E \\ (\delta_Q + \sigma_Q + \gamma_Q) I_Q - \eta I_S \\ (\delta_H + \gamma_H) I_H - \sigma_S I_S - \sigma_Q I_Q \end{pmatrix}.$$

The Jacobians F and V of $\mathcal{F}$ and $\mathcal{V}$ evaluated at the DFE are defined such that

$F = \left[\frac{\partial \mathcal{F}_i}{\partial X_j}\right]_{X=X_0}$, $V = \left[\frac{\partial \mathcal{V}_i}{\partial X_j}\right]_{X=X_0}$, with $1 \le i, j \le 5$. That is,

$$F = \begin{pmatrix} 0 & \beta_A & \beta_S & m_Q\beta_S & m_H\beta_S \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and,

$$V = \begin{pmatrix} \alpha & 0 & 0 & 0 & 0 \\ -\rho\alpha & \gamma_A & 0 & 0 & 0 \\ -(1-\rho)\alpha & 0 & \varepsilon_1 & 0 & 0 \\ 0 & 0 & -\eta & \varepsilon_2 & 0 \\ 0 & 0 & -\sigma_S & -\sigma_Q & \varepsilon_3 \end{pmatrix},$$

where $\varepsilon_1 = \delta_S + \sigma_S + \gamma_S + \eta$, $\varepsilon_2 = \delta_Q + \sigma_Q + \gamma_Q$ and $\varepsilon_3 = \delta_H + \gamma_H$.

The basic reproduction number R_0 is determined as the spectral radius of the next-generation matrix FV^{-1}, as defined in Equation (6):

$$R_0 = \rho(FV^{-1}). \tag{6}$$

By performing a straightforward calculation, we find that

$$V^{-1} = \begin{pmatrix} \frac{1}{\alpha} & 0 & 0 & 0 & 0 \\ \frac{\rho}{\gamma_A} & \frac{1}{\gamma_A} & 0 & 0 & 0 \\ \frac{1-\rho}{\varepsilon_1} & 0 & \frac{1}{\varepsilon_1} & 0 & 0 \\ \frac{(1-\rho)\eta}{\varepsilon_1\varepsilon_2} & 0 & \frac{\eta}{\varepsilon_1\varepsilon_2} & \frac{1}{\varepsilon_2} & 0 \\ \frac{(1-\rho)(\sigma_S\varepsilon_2+\eta\sigma_Q)}{\varepsilon_1\varepsilon_2\varepsilon_3} & 0 & \frac{\sigma_S\varepsilon_2+\eta\sigma_Q}{\varepsilon_1\varepsilon_2\varepsilon_3} & \frac{\sigma_Q}{\varepsilon_2\varepsilon_3} & \frac{1}{\varepsilon_3} \end{pmatrix},$$

and,

$$FV^{-1} = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where,

$$b_{11} = \frac{\rho\beta_A}{\gamma_A} + \frac{(1-\rho)\beta_S}{\varepsilon_1} + \frac{(1-\rho)\eta m_Q\beta_S}{\varepsilon_1\varepsilon_2} + \frac{(1-\rho)(\sigma_S\varepsilon_2+\eta\sigma_Q)m_H\beta_S}{\varepsilon_1\varepsilon_2\varepsilon_3},$$
$$b_{12} = \frac{\beta_A}{\gamma_A},$$
$$b_{13} = \frac{\beta_S}{\varepsilon_1} + \frac{\eta m_Q\beta_S}{\varepsilon_1\varepsilon_2} + \frac{(\sigma_S\varepsilon_2+\eta\sigma_Q)m_H\beta_S}{\varepsilon_1\varepsilon_2\varepsilon_3},$$
$$b_{14} = \frac{m_Q\beta_S}{\varepsilon_2} + \frac{m_H\beta_S\sigma_Q}{\varepsilon_2\varepsilon_3},$$
$$b_{15} = \frac{m_H\beta_S}{\varepsilon_3}.$$

Solving Equation (6) gives Equation (7)

$$R_0 = \frac{\rho\beta_A}{\gamma_A} + \frac{(1-\rho)\beta_S}{\varepsilon_1} + \frac{(1-\rho)\eta m_Q\beta_S}{\varepsilon_1\varepsilon_2} + \left(\frac{(1-\rho)\sigma_S}{\varepsilon_1} + \frac{(1-\rho)\eta\sigma_Q}{\varepsilon_1\varepsilon_2}\right)\frac{m_H\beta_S}{\varepsilon_3}. \tag{7}$$

The Basic reproduction number depicted in Equation (7) can be expressed as the sum of the reproduction numbers associated with each infected compartment: R_{I_A}, R_{I_S}, R_{I_Q}, and R_{I_H}. The reproduction number R_{I_A} is the product of the proportion of exposed individuals that transition to the asymptomatic class (ρ), the infection rate of the asymptomatic individuals (β_A), and the average duration of infection in the asymptomatic individuals $\left(\frac{1}{\gamma_A}\right)$. Similarly, the reproduction number R_{I_S} is the product of the fraction of exposed people who survived the incubation time and transitioned to the symptomatic class ($1-\rho$), the infection rate of the symptomatic individuals (β_S), and the average duration of infection in the symptomatic clinical status $\left(\frac{1}{\varepsilon_1}\right)$. The reproduction number

R_{I_Q} is derived as the product of the proportion of symptomatic individuals that move to the quarantined class $\left(\frac{(1-\rho)\eta}{\varepsilon_1}\right)$, the rate at which quarantined individuals transmit the disease $(m_Q\beta_S)$, and the average infectious period of the quarantined category $\left(\frac{1}{\varepsilon_2}\right)$. Finally, the reproduction number R_{I_H} is given by the sum of the proportions of symptomatic and quarantined individuals who move to the hospitalized class $\left(\frac{(1-\rho)\sigma_S}{\varepsilon_1}+\frac{(1-\rho)\eta\sigma_Q}{\varepsilon_1\varepsilon_2}\right)$ multiplied by the rate at which hospitalized individuals transmit the disease $(m_H\beta_S)$ and the average infectious period of the hospitalized people $\left(\frac{1}{\varepsilon_3}\right)$.

5. Parameter estimation and predictions

This section addresses a procedure to estimate unknown epidemiological parameters in the ODEs system (1) from COVID-19 incidence data and generate predictions in a Bayesian inference approach.

5.1. Inference method

We are interested in estimating infection rates β_S and β_A, recovery rates γ_S, γ_A, γ_H, and γ_Q, disease-induced death rates δ_S, δ_Q, and δ_H, and transition rates σ_S, σ_Q, and η. Therefore, given the vector $\Theta = (\beta_S, \beta_A, \gamma_S, \gamma_A, \gamma_H, \gamma_Q, \delta_S, \delta_Q, \delta_H, \sigma_S, \sigma_Q, \eta)$ of unknown parameters, the first step is to define for these parameters prior distributions that capture our prior beliefs about them before observing any data. Then, given the observed data $y^{obs} = \{y_i^{obs}\}$ of $y_1^{obs}, ..., y_N^{obs}$ points, we apply the Bayes rule [57] using a likelihood function to update the priors. The result is a posterior distribution defined by Equation (8)

$$P(\Theta \mid y^{obs}) = \frac{P(y^{obs} \mid \Theta)P(\Theta)}{P(y^{obs})}, \tag{8}$$

where $P(y^{obs} \mid \Theta)$ stands for the likelihood function, $P(\Theta)$ is the prior distribution, $P(y^{obs})$ denotes the evidence, and $P(\Theta \mid y^{obs})$ the posterior distribution, also known as target distribution.

In Bayesian inference, the evidence function $P(y^{obs})$ in Equation (8) is considered fixed and independent of the vector Θ. So, it acts like a normalizing constant to ensure that the target/posterior distribution is a valid distribution. Since the normalizing constant (also known as the marginal distribution of the data) is independent of the unknown parameters, it can be omitted when writing the posterior distribution in Equation (8). In effect, the target distribution becomes proportional to the product of the likelihood function and the priors as expressed in Equation (9).

$$P(\Theta \mid y^{obs}) \propto P(y^{obs} \mid \Theta)P(\Theta). \tag{9}$$

Here, to compute the posterior distribution, we use the HMC algorithm, whose theoretical foundation relies on Hamiltonian dynamics [58]. HMC uses gradient calculations to generate efficient samples from the posterior distribution. This property allows HMC to be fit for high-dimensional and complex models even if there is a strong correlation among the parameters [42]. So, the HMC algorithm converges quickly to a given target distribution, taking a few iterations. However, the HMC's performance can be influenced by the manual tuning of some of its parameters. To overcome this problem, [59] proposed the No-UTurn Sampler (NUTS) algorithm. The NUTS algorithm, a type of HMC, automatically tunes its parameters to maximize efficiency and is highly effective in exploring complex posterior distributions. It is adopted by Stan software [60], a statistical modeling platform that can be used to model compartmental epidemic models [42,41], as the default algorithm for Bayesian inference based on MCMC sampling with the HMC algorithm.

5.2. Observation model

Because observed COVID-19 data do not directly correspond to a specific quantity in the deterministic model (1), a likelihood function is necessary to link these observations to changes in model compartments. We achieved this by assuming that the daily reported incidence of symptomatic infected cases is related to the reported incidence in the model (1) by a negative binomial distribution defined in Equation (10). The negative binomial distribution is an alternative to the Poisson distribution when modeling count data with over-dispersion, where the variance exceeds the mean [7,10,17,41,40]. This flexibility allows us to account for the inherent variability and uncertainty in epidemiological data, which often exhibits over-dispersion. While the Poisson distribution is typically used for modeling events that occur within a given period, such as infection count data, the negative binomial distribution is better suited when there is significant variability beyond what the Poisson distribution can accommodate [61].

$$incidence_i^{obs} \sim NegBinomial(incidence_i^{ODE}(\Theta), \phi), \tag{10}$$

with $incidence_i^{ODE}(\Theta) = (1-\rho)\alpha E_i$.

In Equation (10), $incidence_i^{obs}$ is the observed incidence of infected cases reported on day i. The corresponding symptomatic infection incidence, denoted as $incidence_i^{ODE}(\Theta)$, is produced by the ODEs (1), where Θ is the vector of parameters with values yet to be determined. NegBinomial(μ, ϕ) represents the Negative Binomial distribution in which μ indicates the mean and ϕ is the dispersion parameter. ϕ is unknown and needs to be estimated.

We adopted the parameterization outlined in Gelman et al. [62] for the Negative Binomial distribution and gave its expression by Equation (11)

$$\text{NegBinomial}(n|\mu,\phi) = \binom{n+\mu-1}{\mu-1}\left(\frac{\phi}{\phi+1}\right)^{\mu}\left(\frac{1}{\phi+1}\right)^{n}, \tag{11}$$

where $\mu \in \mathbb{R}^+$, $\phi \in \mathbb{R}^+$, and $n \in \mathbb{N}$ represents the observed number of new infections.

5.3. Prior specification

After determining the likelihood function that models the observed data, we must specify the prior distribution, $P(\Theta)$, of the set of unknown parameters in the vector, $\Theta = (\beta_S, \beta_A, \gamma_S, \gamma_A, \gamma_H, \gamma_Q, \delta_S, \delta_Q, \delta_H, \sigma_S, \sigma_Q, \eta, \phi)$ which is viewed as random. This allowed us to include our pre-existing beliefs about the epidemiological parameters in the model by establishing a prior distribution for each quantity in vector Θ, as provided in Equation (12)

$$P(\Theta) = P(\beta_S)P(\beta_A)P(\gamma_S)P(\gamma_A)P(\gamma_H)P(\gamma_Q)P(\delta_S)P(\delta_Q)P(\delta_H)P(\sigma_S)P(\sigma_Q)P(\eta)P(1/\phi). \tag{12}$$

Previous studies estimated the infection rate from 0.5 to 1.5 day^{-1} [63–65] and found similar viral loads for asymptomatic and symptomatic individuals [49]. To ensure positive and reliable estimates for β_A and β_S, we choose an informative normal prior distribution with a mean of 1 and a standard deviation of 0.2. This prior reflects our belief that these parameters will likely be around 1, with limited variability. Our model considers recovery for individuals in various states: symptomatic and asymptomatic infectious, quarantined, and hospitalized. As reported in [18,66,67], recovery rates for symptomatic and asymptomatic infectious individuals are observed to be faster (0.1429 day^{-1}) compared to hospitalized and quarantined individuals (0.0714 day^{-1}). To reflect this finding, the recovery rates (γ_S and γ_A) for the symptomatic/asymptomatic subpopulations are modeled as normally distributed with a mean of 0.1429 and a standard deviation of 0.1. Similarly, the recovery rates (γ_H and γ_Q) for hospitalized and quarantined individuals are set to follow a normal distribution with a mean of 0.0714 and a standard deviation of 0.01. In this way, we constrain the model to generate reliable estimates of recovery rates, all of which will be positive. The disease-induced death rate is estimated to be 0.0016 day^{-1} in symptomatic individuals, 0.0025 day^{-1} in quarantined individuals [8], and ranges from 0.001 to 0.1 day^{-1} in hospitalized patients [63,68]. We then draw δ_S and δ_Q from a uniform distribution on the interval (0, 0.004) and δ_H from a uniform distribution on the interval (0, 0.04). The transition rate from the quarantined to hospitalized class (σ_Q) is estimated to be 0.14 day^{-1} [68–70]. Similarly, the transition rate from symptomatic to hospitalized class (σ_S) ranges from 0.0514 day^{-1} to 0.14 day^{-1} [18,37,69]. For the transition from symptomatic to quarantined, the rate (η) is estimated to be 0.1327 day^{-1} [66]. We chose uniform distributions for these rates. σ_Q is drawn from a uniform distribution of (0.125, 0.166), σ_S from a uniform distribution of (0.06, 0.14), and η from a uniform distribution of (0.1, 0.25). This selection of uniform distributions reflects the limited prior information available from the literature on the specific distribution of these parameters. This aligns with the established scientific understanding that disease-induced death rates (δ_S, δ_Q, and δ_H) and the rates of transition (σ_Q, σ_S, and η) are inherently positive. As recommended in [41], we specify the prior distribution $P(1/\phi)$ as an exponential distribution with a scale parameter value of 5 to prevent an excessive concentration of prior mass on models exhibiting significant over-dispersion.

5.4. Assessing the model

We conduct posterior predictive checks [41] to evaluate the model's fit to the observed data. After fitting the model, we leverage the estimated posterior distribution $P(\Theta \mid y^{obs})$, of the model parameters Θ, to generate simulated predictions (simulated datasets), denoted as y_{sim}. These simulated datasets represent the outcomes the model would anticipate if it were responsible for generating the data. Then, we compare these datasets with the observed data to assess the consistency between model predictions and reality. We draw simulated datasets, y_{sim}, by following this procedure:

$$\Theta_{post} \sim P(\Theta \mid y^{obs}),$$

$$y_{sim} \sim P(y^{obs} \mid \Theta_{post}).$$

This procedure samples simulated predictions based on parameter values drawn from the posterior distribution, incorporating information from observed data into the parameter estimates.

5.5. Predictions

An essential step in developing epidemic models is model validation, which involves assessing the accuracy and the performance of the model to predict future observations. The ability of the model to make predictions is related to the accurate and reliable estimates of its parameters. Hence, after estimating the unknown quantity Θ in the model (1), we evaluated the model's predictive capacity by comparing model predictions with the observed data. To generate predictions, we divided the observed data y^{obs} into a training set y^{train} and test set y^{test} such that $y^{train} = (y_1^{obs}, y_2^{obs}, ..., y_n^{obs})$, and $y^{test} = (y_{n+1}^{obs}, y_{n+2}^{obs}, ..., y_N^{obs})$, with n representing the size of the training set and N the number of observations. Next, we apply a fit-and-predict approach built-in Stan [71]. In this approach, the predictive model is encapsulated within the generated quantities block of the Stan program. This block is where we specify the quantities of

Table 2
Predetermined parameter values in the system of equation (1).

Parameter	Mean value	95% Confidence Interval (CI)	Source
α	0.1923 day^{-1}	[0.1429, 0.2439]	[43–45]
ρ	0.1790	[0.1550, 0.2020]	[46–48]
ω_R	0.0050 day^{-1}	-	Assumed
m_H	0.01	-	Assumed
m_Q	0.1	-	Assumed

interest generated based on the parameters estimated by the model. By placing the predictive model within this block, we ensure that predictions are made based on the posterior distribution of parameters obtained from the fitted model. When new predictions are required, the entire model-fitting process is repeated. This involves re-estimating model parameters using the updated dataset, which includes any new observations that have become available. This iterative process ensures that the model remains current and that predictions are based on the most recent data.

6. MCMC simulations and results

This section focuses on fitting the proposed epidemic model (Equation (1)) to incidence data under two scenarios. The first scenario explores the model's behavior without vaccination, while the second scenario incorporates vaccination dynamics. We further evaluate the model's ability to generate predictions for new observations.

6.1. Model fitting without vaccination

We fit the vaccination-free version of our deterministic compartmental model (1) $\left(\theta_V = 0 \; and \; V(0) = 0\right)$ to the daily confirmed cases of COVID-19 in the State of Tennessee from June 4 to November 26, 2021, to estimate its unknown parameters Θ. The other parameters in the model (1) are obtained from existing literature on COVID-19. The mean incubation period of COVID-19 ranges between 5 and 5.2 days [43–45]. We take the base value as $\alpha = 0.1923 \; day^{-1}$ (95% confidence interval (CI), 0.1429 to 0.2439). The estimated proportion of asymptomatic individuals in all infected cases with SARS-CoV-2 ranges between 0.1572 and 0.308 [46–48]. We set the base value as $\rho = 0.1790$ (95% credible interval (CrI): 0.155 to 0.202). As reported in [50], reinfections are prompting inquiries regarding long-term immunity to COVID-19 and its implications for vaccine development. We assumed waning immunity among recovered individuals at 183 days ($\omega_R^{-1} = 183$). Given that the reduction factors of infectivity, m_H and m_Q, must be strictly between 0 and 1, we have chosen $m_H = 0.01$ and $m_Q = 0.1$ to illustrate a specific scenario in our simulations. These values represent the significant decrease in transmission from hospitalization (m_H) and the moderate reduction from quarantine (m_Q). Table 2 lists the fixed and assumed parameters of our model. We also set initial conditions for the state variables as follows: $E(0) = 35,870$, $I_A(0) = 2,701$, $I_S(0) = 15,890$, $I_Q(0) = 18,905$, $I_H(0) = 28,686$, $R(0) = 30,417$, $D(0) = 11,372$, and $S(0) = (N - E(0) - I_A(0) - I_S(0) - I_Q(0) - I_H(0) - R(0) - D(0))$. Here, N refers to the Tennessee population size as is given in Section 3.

We employ the NUTS algorithm based on the RStan package [71] in R software version 4.3.1 to draw samples. In the settings of NUTS, we set the number of parallel chains to four, the number of iterations per chain to 2000, and the number of iterations to discard per chain (warm-up) to 1000. For each estimated parameter and the associated function, Basic reproduction number R_0, 4000 samples were obtained (1000 sample draws per chain). These samples are utilized in computing summary statistics for the fitted parameters, including posterior mean, standard deviation, quantiles, effective sample size (n_{eff}), and $\hat{R}$ statistics to assess the MCMC chains.

To gauge if our inference is reliable, we first provide in Table 3 a summary of statistical metrics of the fitted quantities. Next, we examine the central posterior uncertainty intervals (Figs. 2 and 3), marginal posterior densities (Fig. 4), and trace plots (Fig. 5) for parameters of interest. As shown in Table 3, the $\hat{R}$ value is close to 1 for all parameters, indicating good convergence of the MCMC chains. In addition, the effective sample sizes for all sampled parameters were high, with the parameter η having the lowest number of effective samples at 3168, a value still regarded as high. This indicates that the MCMC chains have successfully explored the posterior distribution. The NUTS algorithm uses the $\hat{R}$ statistics and n_{eff} as diagnostic tools to assess the quality of the MCMC samples generated by the algorithm [41,72]. $\hat{R}$ with values close to 1 indicates the convergence of parallel chains, and higher values of n_{eff} (over 100) signify more informative samples [73]. Figs. 2 and 3 show that the parameters have narrow intervals for central posterior uncertainty with $\hat{R}$ near 1 ($\hat{R} \leqslant 1.05$). In Fig. 4, we observe that the four parallel Markov chains for each parameter distribution, ignoring the influence of other parameters, are overlapped, confirming that the MCMC chains converge. The mixing and agreement of the Markov chains can also be seen in the traceplots, as presented in Fig. 5.

After evaluating the inference, we assess model fit using posterior predictive checks (Section 5.4), and we provide the result in Fig. 6. This assessment involves generating simulated predictions for daily confirmed cases and plotting the constructed curve with uncertainty measures, 50% and 97.5% credible intervals (CrI) (Fig. 6(a)). Using the estimated parameters, we present in Fig. 6(b) the fitted curve for confirmed cases from model (1) with 97.5% CrI.

Fig. 6(a) shows a good fit between the observed incidence data (black dots) and the posterior median of the simulated predictions (solid blue line). The model uncertainty, visualized by the credible intervals (CrIs), captures the structure of the variability in the data. The dark blue shaded area represents the 50% CrI and encompasses the most likely range of predictions. The wider light blue shaded

Table 3
Estimates of the model parameters in the absence of vaccination dynamics.

Parameter	Posterior mean	Standard deviation	Quantile 2.5%	Quantile 97.5%	n_{eff}	$\hat{R}$	Source
β_S	0.7161 day^{-1}	0.1138	0.5016	0.9389	6128	0.9996	Fitted
β_A	1.2446 day^{-1}	0.2104	0.8341	1.6524	5166	0.9995	Fitted
γ_S	0.3712 day^{-1}	0.0850	0.2084	0.5403	3807	0.9993	Fitted
γ_A	0.4157 day^{-1}	0.0633	0.3040	0.5504	6378	1.0002	Fitted
γ_H	0.0724 day^{-1}	0.0102	0.0518	0.0929	6458	0.9995	Fitted
γ_Q	0.0725 day^{-1}	0.0100	0.0525	0.0921	6396	0.9995	Fitted
δ_S	0.0020 day^{-1}	0.0012	0.0001	0.0039	5336	0.9991	Fitted
δ_Q	0.0020 day^{-1}	0.0012	0.0001	0.0039	5135	0.9994	Fitted
δ_H	0.0125 day^{-1}	0.0103	0.0004	0.0370	4376	0.9995	Fitted
σ_S	0.1140 day^{-1}	0.0207	0.0714	0.1419	4930	1.0008	Fitted
σ_Q	0.1475 day^{-1}	0.0120	0.1265	0.1659	5297	0.9993	Fitted
η	0.2013 day^{-1}	0.0374	0.1150	0.2483	3168	1.0001	Fitted
$1/\phi$	0.2559	0.0262	0.2097	0.3113	5265	0.9996	Fitted
ϕ	3.9489	0.4012	3.2123	4.7676	6143	0.9995	Fitted

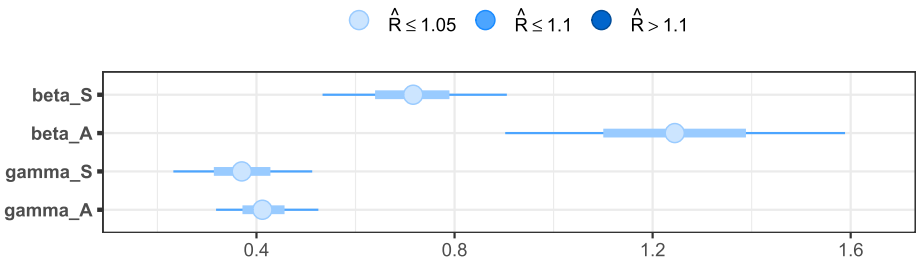

Fig. 2. The plot depicts the median estimates and their credible intervals for parameters β_S (beta_S), β_A (beta_A), γ_S (gamma_S), and γ_A (gamma_A). The color scheme corresponds to the $\hat{R}$ value for each parameter. A long, thin blue line signifies the 90% credible interval, while the short, thick blue line represents the 50% credible interval. The large, shaded blue circle denotes the median estimate of the parameter.

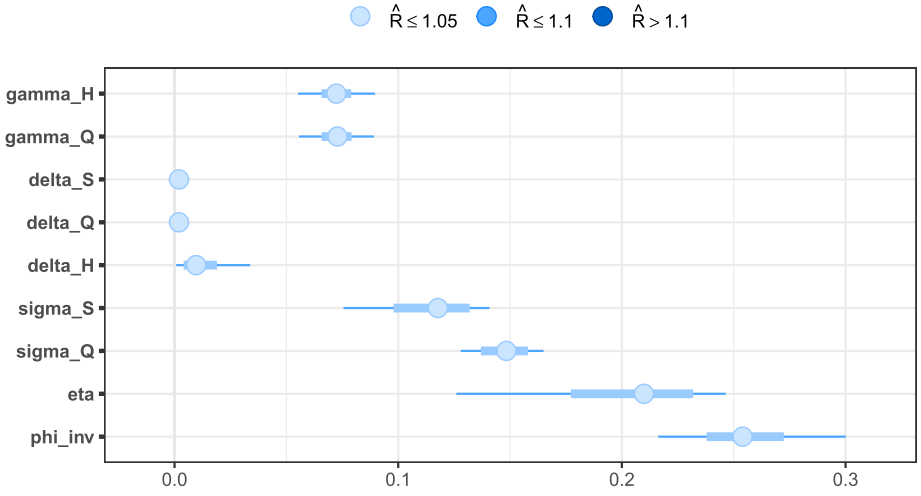

Fig. 3. The plot depicts the median estimates and their credible intervals for parameters γ_H (gamma_H), γ_Q (gamma_Q), δ_S (delta_S), δ_Q (delta_Q), δ_H (delta_H), σ_S (sigma_S), σ_Q (sigma_Q), η (eta), and $1/\phi$ (phi_inv). The color scheme corresponds to the $\hat{R}$ value for each parameter. A long, thin blue line signifies the 90% credible interval, while the short, thick blue line represents the 50% credible interval. The large, shaded blue circle denotes the median estimate of the parameter.

area represents the 97.5% CrI, highlighting the broader range of possible outcomes considering the model's uncertainty. Similarly, Fig. 6(b) depicts a satisfactory fit between the model estimation (solid blue line) and the observed data points. Likewise, the model uncertainty is represented by the 97.5% CrI (light blue shaded area), accounting for the variability observed in the data. Hence, using the HMC algorithm, our epidemiological model produces the COVID-19 incidence in Tennessee State.

Table 3 shows that the estimated infection rates, β_S and β_A, are 0.716 and 1.244 $days^{-1}$, respectively. These values align with the range of 0.5–1.5 $days^{-1}$ supported by [63–65]. Furthermore, the estimated values for β_S and β_A suggest that the effective contact rate for asymptomatic individuals (β_A) is higher than that of symptomatic individuals (β_S). This result emphasizes the importance of asymptomatic carriers as the main contributors to the COVID-19 pandemic in Tennessee. However, a study [49] estimated the

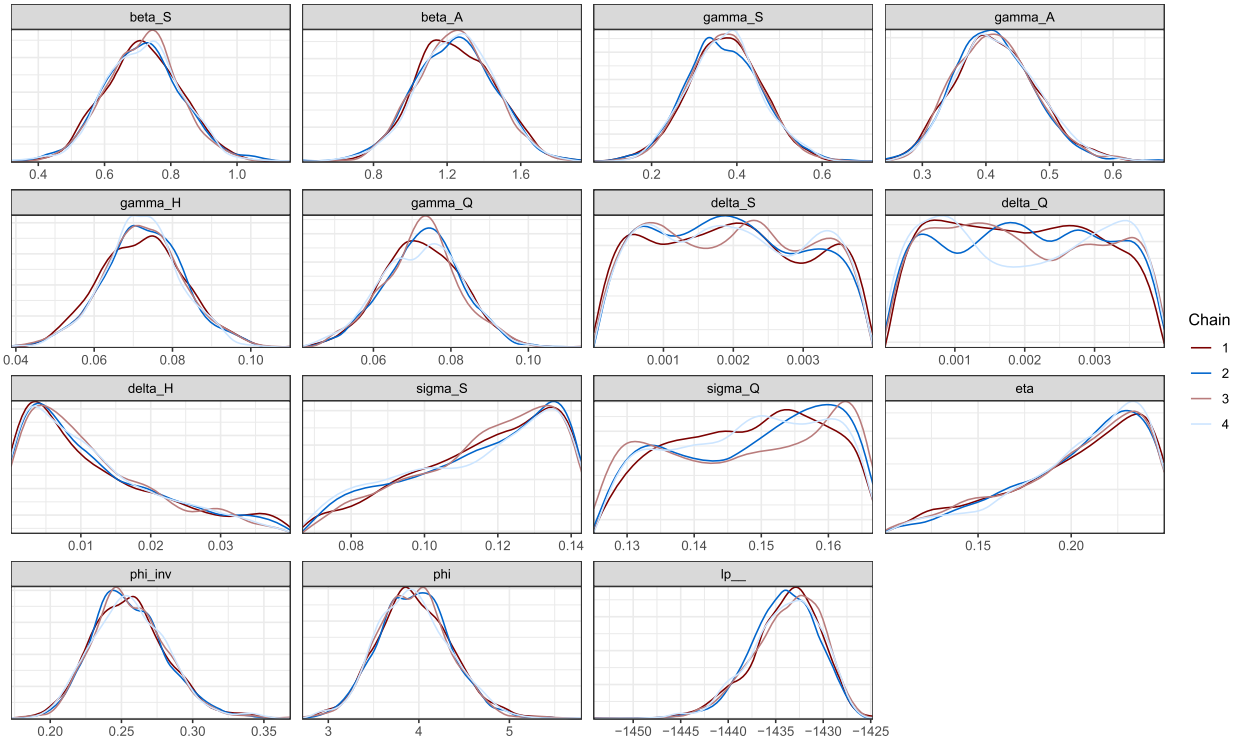

Fig. 4. Marginal posterior densities plot for parameters β_S, β_A, γ_S, γ_A, γ_H, γ_Q, δ_S, δ_Q, δ_H, σ_S, σ_Q, η, $1/\phi$, ϕ, and `lp__` (log-posterior). A line depicts a Markov chain.

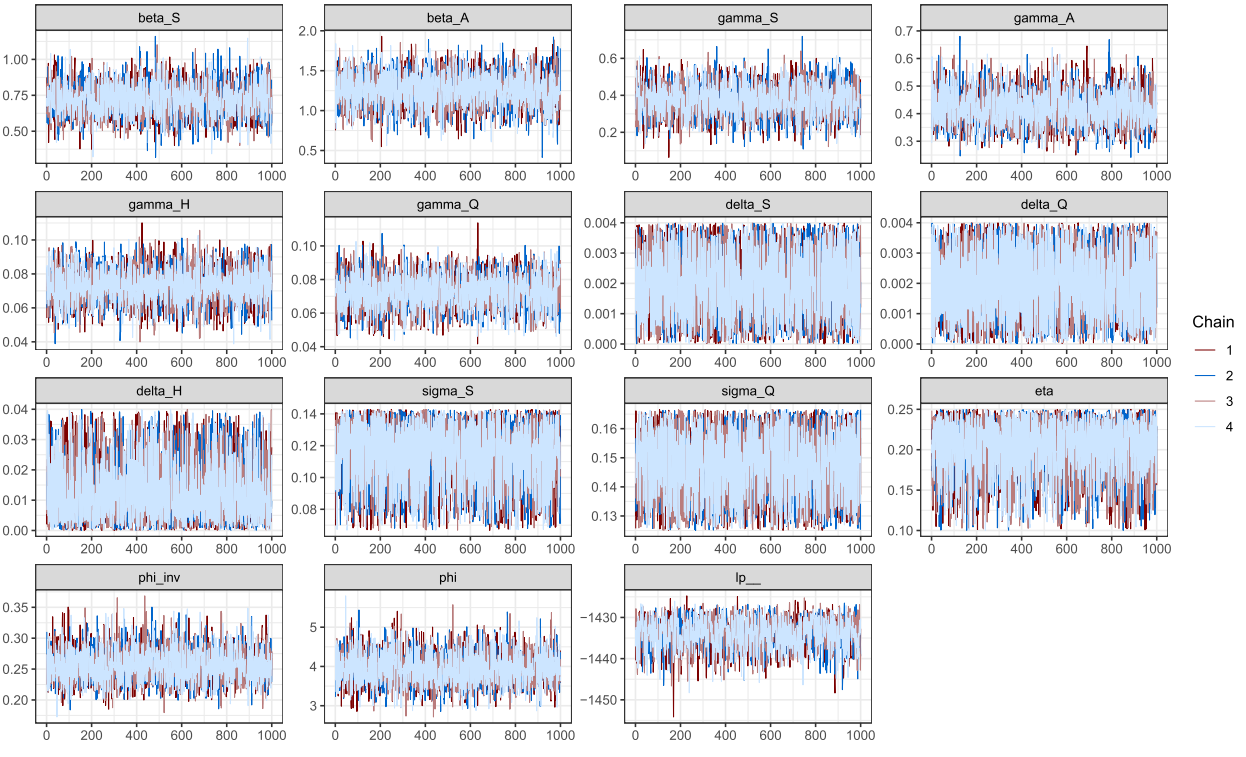

Fig. 5. Trace plots for parameters β_S, β_A, γ_S, γ_A, γ_H, γ_Q, δ_S, δ_Q, δ_H, σ_S, σ_Q, η, $1/\phi$, ϕ, and `lp__` (log-posterior) indicating the value for each MCMC chain at every iteration during the stationary phase (excluding the warm-up period). A line depicts a Markov chain.

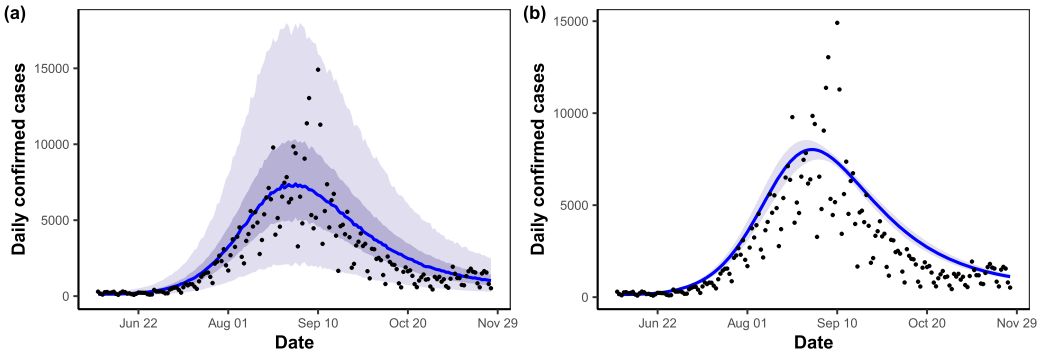

Fig. 6. Model (1) fitted in the absence of vaccination to daily COVID-19 confirmed cases (black dots) from June 4 to November 26, 2021. (a) Posterior predictive check for COVID-19 incidence. The solid blue line represents the median posterior, the dark blue shaded area represents the 50% credible interval (CrI), and the light blue shaded area indicates the 97.5% CrI. (b) The solid blue line shows the model estimation and the light blue shaded area illustrates the corresponding 97.5% CrI.

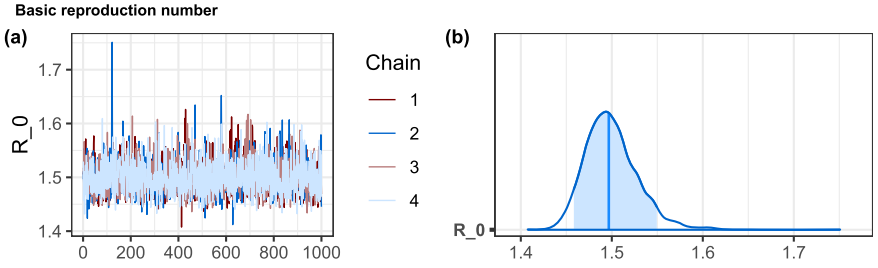

Fig. 7. Trace plot and posterior distribution of the Basic reproduction number. (a) Trace plot indicating the value of each MCMC chain at each iteration of the stationary phase. (b) Posterior distribution with median and 95% interval.

viral load of COVID-19 to be similar for both asymptomatic and symptomatic persons. Thus, we can explain the high probability of COVID-19 transmission by asymptomatic individuals as follows. First, because asymptomatic carriers do not show clinical signs, they may be unaware that they have the virus and are infectious. Therefore, they inadvertently transmit the virus. Second, asymptomatic individuals are likely to continue their normal activities, including social interactions, with no precautions taken. This behavior can lead to increased contact with others and a greater likelihood of transmission. Finally, asymptomatic individuals are less likely to be diagnosed unless they have had direct contact with a confirmed case or through targeted screening. This delay in diagnosis allows asymptomatic carriers more time to spread the virus unknowingly. The average recovery rates for hospitalized (γ_H) and quarantined (γ_Q) individuals are 0.0724 and 0.0725 $days^{-1}$, respectively, consistent with the findings of [18,66,67]. The average recovery rates for symptomatic (γ_S) and asymptomatic (γ_Q) individuals are estimated to be around 0.37 $days^{-1}$, which is faster than those reported in [18,66,67]. Disease-induced death rates are estimated to be relatively low, ranging from 0.002 day^{-1} for symptomatic infectious and quarantined individuals to 0.0125 day^{-1} for hospitalized individuals. These estimates support the findings reported in [8] for symptomatic infectious and quarantined individuals, and in [63,68] for hospitalized individuals. Our estimates for transition rates align with previous studies: 0.1475 day^{-1} from quarantined to hospitalized and 0.1140 day^{-1} from symptomatic to hospitalized [18,37,69,68,70]. However, the estimated transition rate from symptomatic to quarantined (0.2013 day^{-1}) is higher than the value reported in [66].

We estimated the basic reproduction number (R_0) by incorporating Equation (7) into the block of generated quantities within the Stan program. The mean of R_0, as estimated from the observed data, was 1.5 (97.5% credible interval (CrI): 1.46–1.56). Fig. 7 displays both the trace plot and the posterior distribution of R_0. Specifically, Fig. 7(a) demonstrates the mixing and convergence of the MCMC chains, whereas Fig. 7(b) illustrates the posterior distribution of R_0 with a median estimate of 1.5. Our estimated value of R_0 (approximately 1.5) aligns closely with early estimates from other studies conducted in Tennessee. For instance, [27] reported an R_0 of 1.0343, while [28] estimated R_0 at 1.791 in Hamilton County, Tennessee. The observed discrepancies may be due to the difference in model assumptions and structures.

6.2. Model fitting with vaccination

Throughout this section, we fit the model in Equation (1) to COVID-19 confirmed cases on incidence, from June 4 to November 26, 2021, by accounting for vaccination $\left(\theta_V \neq 0 \; and \; V(0) \neq 0\right)$. To implement a vaccination strategy, we examined three COVID-19 vaccines: Pfizer-BioNTech with 95% efficacy [74], Moderna with 94.5% efficacy [74], and Janssen with 67% efficacy [75]. These vaccines have received emergency use authorization from the Food and Drug Administration (FDA) in Tennessee State [76]. Our primary objective is to estimate the vaccination rate from the data for each of the three COVID-19 vaccines. Following the same procedure described above, the model was calibrated separately for each vaccine efficacy. We inform the prior distribution for θ_V

Table 4
Mean vaccination rate estimates for each COVID-19 vaccine.

θ_V	mean	Standard deviation	Quantile 2.5%	median	Quantile 97.5%	n_{eff}	$\hat{R}$	Source
Pfizer	0.0051	0.0009	0.0034	0.0051	0.0068	3129	0.9993	Fitted
Moderna	0.0051	0.0009	0.0034	0.0051	0.0069	3492	0.9999	Fitted
Janssen	0.0059	0.0018	0.0025	0.0059	0.0095	3356	0.9997	Fitted

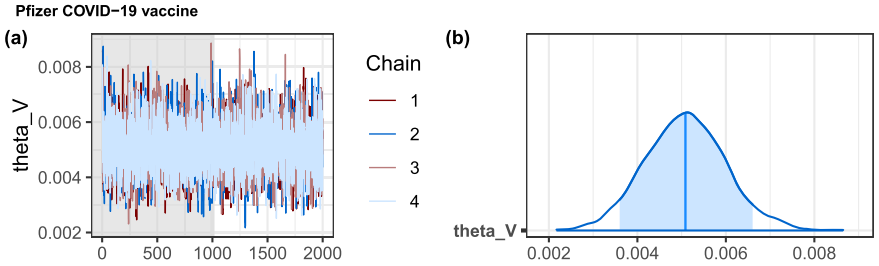

Fig. 8. Trace plot and posterior distribution of the parameter θ_V (or theta_V) when we use the Pfizer COVID-19 Vaccine. (a) Trace plot indicating the value of each MCMC chain at each iteration with the grey area showing the warm-up phase. (b) Posterior distribution with median and 95% interval.

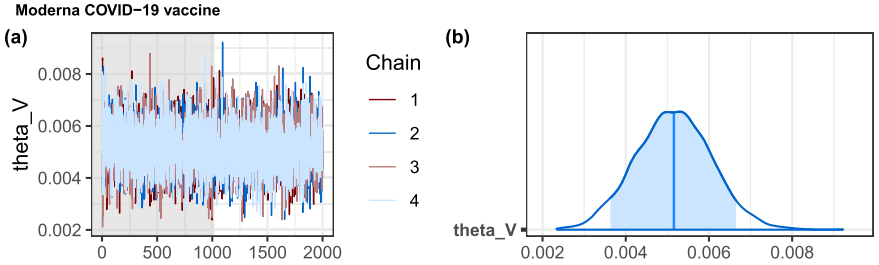

Fig. 9. Trace plot and posterior distribution of the parameter θ_V (or theta_V) when we use the Moderna COVID-19 Vaccine. (a) Trace plot indicating the value of each MCMC chain at each iteration with the grey area showing the warm-up phase. (b) Posterior distribution with median and 95% interval.

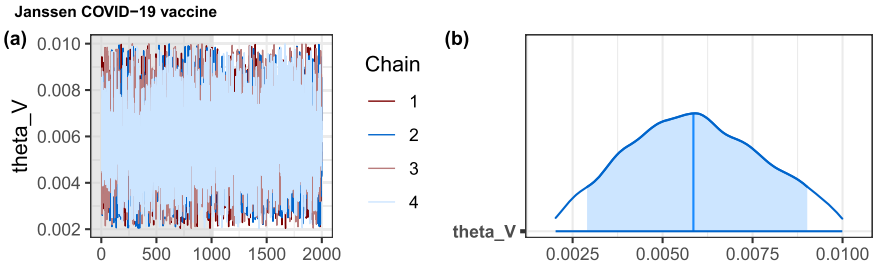

Fig. 10. Trace plot and posterior distribution of the parameter θ_V (or theta_V) when we use the Janssen COVID-19 Vaccine. (a) Trace plot indicating the value of each MCMC chain at each iteration with the grey area showing the warm-up phase. (b) Posterior distribution with median and 95% interval.

(vaccination rate) from the literature. A previous study [27] estimated the Pfizer, Moderna, and Janssen vaccine vaccination rates to be 0.0059 day^{-1}, 0.0042 day^{-1}, and 0.0059 day^{-1}, respectively. Another study [17] suggested a wider range of 0 to 0.009 day^{-1} for the vaccination rate. Given this information, we choose a uniform distribution for θ_V, with a lower bound of 0.002 day^{-1} and an upper bound of 0.01 day^{-1} $\left(\theta_V \sim U(0.002, 0.01)\right)$. The initial number of vaccinated individuals was set to 25,460. That is, $V(0) = 25,460$.

Our estimates reveal a mean vaccination rate of 0.0051 day^{-1} for both the Pfizer and Moderna vaccines, while the Janssen vaccine showed a slightly higher rate of 0.0059 day^{-1} (see Table 4). Figs. 8, 9, and 10 present the trace plots and posterior distributions with median values and 95% credible intervals for the vaccination rates of the three vaccines. From Figs. 8(a), 9(a), and 10(a), we observe the mixing and convergence of the MCMC chains, even during the warm-up phase, for the vaccination rates of all three vaccines. Figs. 8(b), 9(b), and 10(b) illustrate the posterior distributions of the vaccination rates, with a median estimate of 0.0051 for both the Pfizer and Moderna vaccines and 0.0059 for the Janssen vaccine. The estimated values make scientific sense. First, because the Moderna vaccine has efficacy in the same order as the Pfizer vaccine, we obtain a similar estimate of the vaccination rate for both vaccines. Second, as expected [17,35], the vaccination rate for Janssen's vaccine is higher than for Pfizer's and Moderna's vaccines. Model fitting for all three vaccines yields approximately the same estimated values for the other parameters.

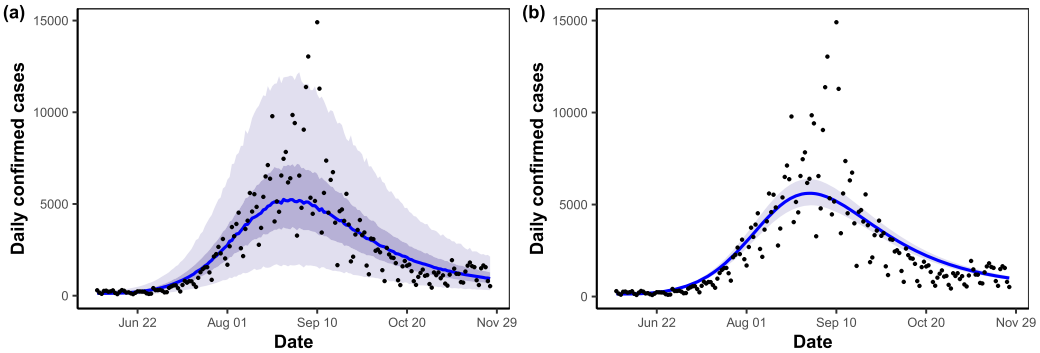

Fig. 11. Model (1) with vaccination dynamics fitted to daily COVID-19 confirmed cases (black dots) from June 4 to November 26, 2021, using the Pfizer-BioNTech COVID-19 Vaccine. (a) Posterior predictive check for COVID-19 incidence. The solid blue line represents the median posterior, the dark blue shaded area represents the 50% credible interval (CrI), and the light blue shaded area indicates the 97.5% CrI. (b) The solid blue line shows the model estimation and the light blue shaded area illustrates the corresponding 97.5% CrI.

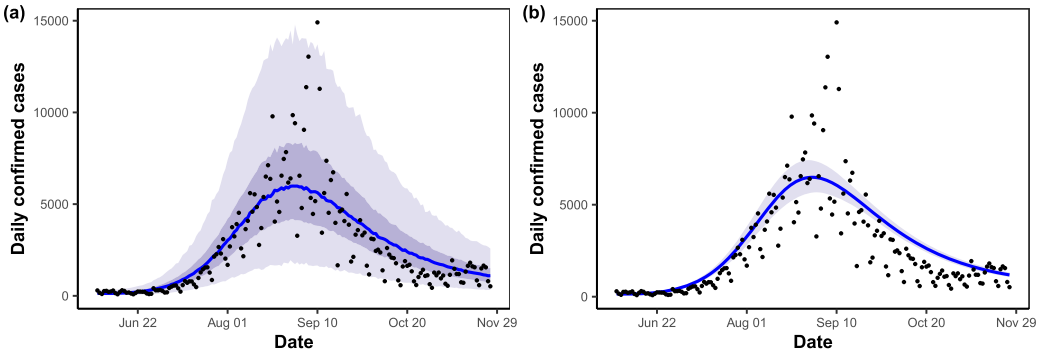

Fig. 12. Model (1) with vaccination dynamics fitted to daily COVID-19 confirmed cases (black dots) from June 4 to November 26, 2021, using the Janssen COVID-19 Vaccine. (a) Posterior predictive check for COVID-19 incidence. The solid blue line represents the median posterior, the dark blue shaded area represents the 50% credible interval (CrI), and the light blue shaded area indicates the 97.5% CrI. (b) The solid blue line shows the model estimation and the light blue shaded area illustrates the corresponding 97.5% CrI.

Calibration of the model using the efficacy of the Pfizer and Moderna vaccines produced a similar fit curve, so we present only the fit for Pfizer in Fig. 11. Fig. 12 shows the fit for the Janssen vaccine. Figs. 11(a) and 12(a) display a good fit between the observed incidence data (black dots) and the posterior median of the simulated predictions (solid blue line). The model's uncertainty is captured by the credible intervals (CrIs), with the dark blue shaded area representing the 50% CrI and the light blue shaded area illustrating the 97.5% CrI, reflecting the range of likely outcomes based on the model. By comparing Figs. 6(b), 11(b), and 12(b), there was a reduction in the epidemic peak, from 8,029 infected cases on August 30 (Fig. 6(b)) to 5,616 on August 30 (Fig. 11(b)) and 6,493 on August 30 (Fig. 12(b)). Thus, a decrease in the number of daily cases of infection is observed when a vaccination strategy is implemented.

To explain the effects of vaccination on the incidence of COVID-19, we simulated our model (1) using the estimated parameter values in Table 3, the fixed parameter values in Table 2, and the initial conditions mentioned previously from June 4 to November 26, 2021. To do this, we first set the vaccine efficacy to that of the Pfizer vaccine ($\xi_V = 0.95$) and varied the vaccination rate. Next, we kept the vaccination rate at 0.0051 day^{-1} and varied the vaccine efficacy. Fig. 13 shows the numerical simulations. Fig. 13(a) shows that increasing the vaccination rate by maintaining vaccine effectiveness at 0.95 reduces the number of daily infections and flattens the epidemic curve. Likewise, Fig. 13(b) shows that adopting highly effective vaccines decreases the incidence but not as much as when increasing vaccination rate. Therefore, the effectiveness of a vaccination strategy in reducing COVID-19 incidence depends on the vaccine uptake or vaccination rate and the vaccine efficacy. It also depends on other factors, such as vaccination coverage [17–19] and the characteristics of the disease itself [3,24].

Overall, a vaccination strategy intends to reduce disease incidence in a population by increasing the proportion of individuals who are immune to the disease. Vaccinating a large proportion of the population reduces the spread of the disease and, ultimately, the incidence of new cases. The success of a vaccination strategy can be assessed by monitoring changes in incidence over time.

6.3. Model prediction

Considering the vaccination-free version of our model and the data of Tennessee State, from June 4 to November 26, 2021, we explore four distinct scenarios with the size of training set $n \in \{52, 88, 123, 158\}$. So, the training set includes 30%, 50%, 70%, and 90% of the observations y^{obs}, respectively. In Fig. 14, we present the prediction results for the four scenarios.

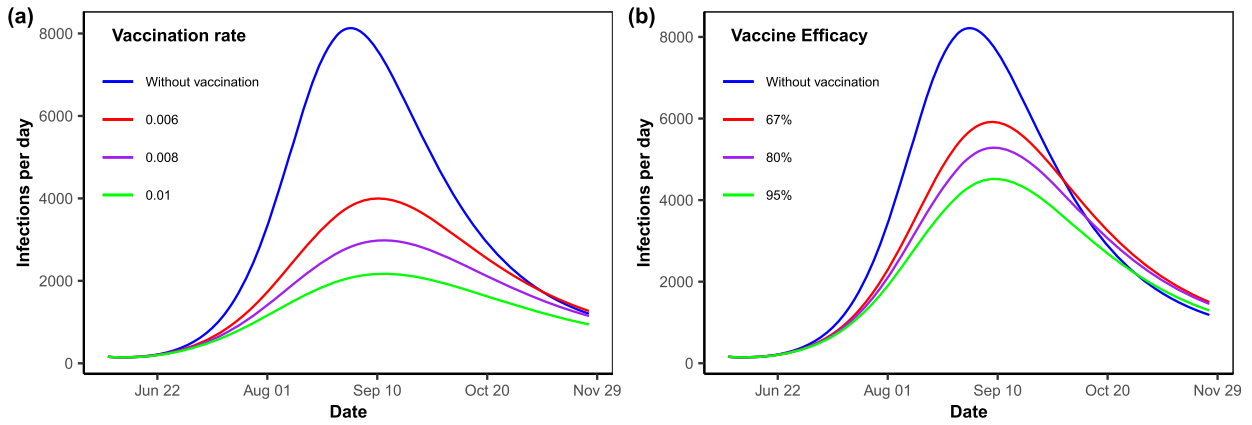

Fig. 13. Effect of vaccination on the incidence behavior of COVID-19. (a) Evolution of daily confirmed cases with the vaccination rate by keeping constant the vaccine efficacy to 95%. (b) progression of daily confirmed cases with vaccine efficacy by holding the vaccination rate at 0.0051 day^{-1}.

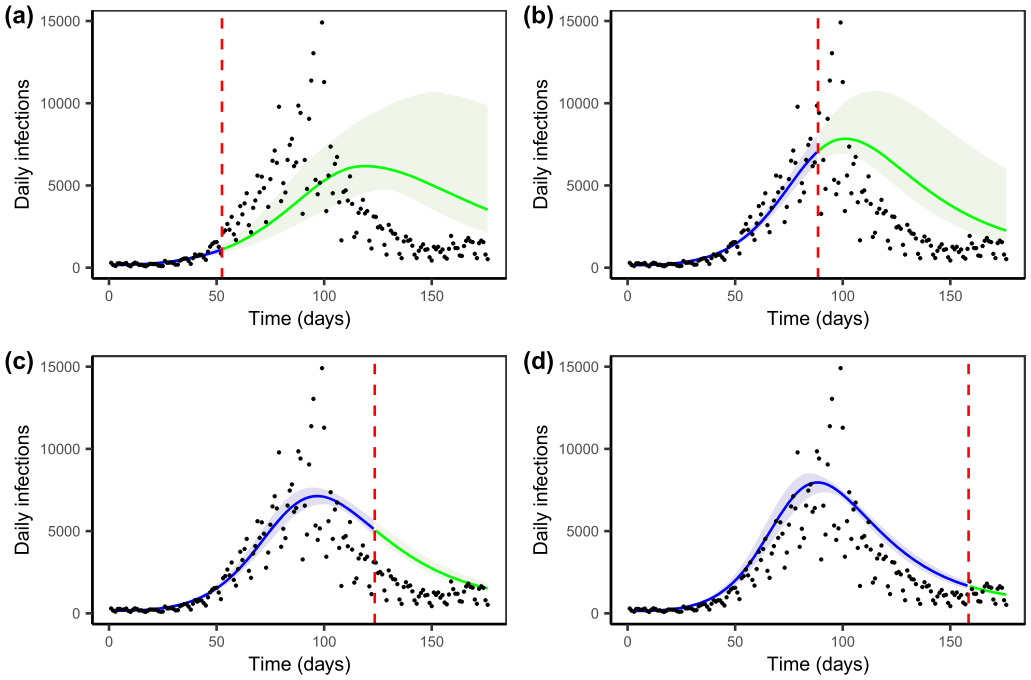

Fig. 14. Predictions for Tennessee are generated using training data of 52, 88, 123, or 158 days. The solid blue curve represents the model trained to real data (black dots). The solid green curve depicts the predicted daily confirmed COVID-19 cases. The light blue and green shaded areas illustrate the 97.5% credible interval (CrI) for model training and prediction. The vertical dashed red line indicates the starting prediction date.

When we train the model with only 30% of observations (52 days, June 4 to July 25, 2021), we see from Fig. 14(a) that the prediction uncertainty (green area) is wide and doesn't match the shape of the data before and after observing the inflection point of the outbreak. In addition, the predicted curve (solid green line) underestimates the observed data from the July 26-August 31 period and overestimates the observed data for the latter period. However, from Fig. 14(b), increasing the training data from 30% to 50% of the observations (88 days, June 4 to August 30, 2021) improves the model's predictive performance. Indeed, from this Figure, we observe that the predicted curve roughly predicts the epidemic peak, and the prediction uncertainty decreases and captures the shape of the new data after the inflection point. From Fig. 14(c), by augmenting training data to 70% of observations (123 days, June 4 to October 4, 2021), we find that the prediction uncertainty becomes narrow, and the model predictions improve considerably. In Fig. 14(d), training the model on 90% of the observed data (158 days, June 4 to November 8, 2021) shows that the model accurately generates predictions in new data. So, the model prediction line is approximately close to the fitted curve in Fig. 6(b).

In summary, predictions for new observations are generated using Stan's fit-predict approach. Our analysis demonstrates that the quality of these predictions is closely related to the number of observed days used to train the model. As the size of the training dataset increases, the prediction accuracy improves, evidenced by a reduction in bias and more precise estimates. This is illustrated in Fig. 14(c, d), where larger training sets lead to better alignment between predicted and observed data. Additionally, our posterior

predictive checks confirm that the credible intervals realistically capture the variability in the data, maintaining that the reduction in uncertainty remains well-calibrated and not attributed to overfitting. This approach can guide public health decision-making by providing insights into the trajectory of COVID-19 transmission. Besides, the model can be leveraged for further analyses to assess the impact of various non-pharmaceutical interventions (NPIs) such as social distancing, mask-wearing, and lockdowns. This capability can also be extended to evaluate the effectiveness of different vaccination coverage levels. By combining these NPIs with vaccination programs, the model can inform the development of optimal control strategies for COVID-19.

7. Conclusions

In this work, we introduced a vaccination-based compartmental model to capture, analyze, and predict the transmission of COVID-19. The model represents different stages of COVID-19 infection and includes a vaccination strategy, enabling a thorough analysis of the influence of vaccination rates on the spread of the disease. This model was calibrated using daily confirmed COVID-19 cases in Tennessee, USA, from June 4 to November 26, 2021. A Bayesian inference framework with the HMC algorithm was employed to explore the posterior distribution of model parameters.

We presented results from two scenarios. In the first scenario, we fit the model (Equation (1)) from June 4 to November 26, 2021, excluding vaccination, to estimate the epidemiological parameters such as infection, recovery, and disease-induced death rates. During this period, the Tennessee population was undergoing a vaccination campaign. Although the model in this scenario does not explicitly account for vaccination, the effects of vaccination are inherently reflected in the data. Vaccination efforts reduce the number of susceptible individuals, affecting disease transmission by limiting the number of effective contacts between infected and susceptible individuals [56]. Consequently, the estimates of the underlying epidemiological parameters can be modified, potentially leading to a bias in capturing the true dynamics of an unvaccinated population.

In the second scenario, we incorporated vaccination dynamics into the model (Equation (1)) and estimated the vaccination rate for each vaccine type: Pfizer, Moderna, and Janssen. These estimates are 0.0051 per day for both the Pfizer and Moderna vaccines and 0.0059 per day for the Janssen vaccine. The fitted curves for these vaccines all show reductions in the epidemic peak. Pfizer and Moderna vaccines achieved a greater reduction, bringing the peak down from 8,029 infected cases to 5,616 infected cases, while the Janssen vaccine reduced it, to 6,493 infected cases.

We assessed the impact of vaccination on epidemic control through simulations. Fixing vaccine efficacy at 95% and varying the vaccination rate from 0.006 to 0.01 $days^{-1}$ reduces daily infections and flattens the epidemic curve. Likewise, with the vaccination rate fixed at 0.0051 $days^{-1}$, varying vaccine efficacy from 67% to 95% decreases the incidence, but not as much as when increasing the vaccination rate. These simulations suggest that a highly effective vaccine provides a strong foundation for reducing infections, but achieving a high vaccination rate is also crucial to maximize the impact.

We evaluated our model's ability to predict unseen data (new observations). The amount of training data directly impacts the accuracy of these predictions. As the training data increases, the model's predictive accuracy improves for future cases.

The model's basic reproduction number (R_0) was estimated from the fitting to be 1.5, implying that an infected individual will, on average, transmit the infection to 1.5 other people in a completely susceptible population. This indicates moderate COVID-19 transmissibility, highlighting the need for vigilance and interventions like vaccination to curb the spread and prevent outbreaks. Our findings further emphasize this, showing that Pfizer and Moderna vaccines significantly reduce infections. Combining vaccination with NPIs like social distancing, mask-wearing, and lockdowns can strategically form optimal control strategies for COVID-19.

Although this framework is particularly suitable for capturing COVID-19 spread in homogeneous populations, real-world communities are more complex due to heterogeneous factors like age, disease severity, and health status. In future work, we will extend our model to incorporate age stratification, allowing us to explore COVID-19 dynamics in diverse community settings. This approach will leverage datasets from the Tennessee Department of Health [77], which provide comprehensive data on confirmed cases, testing, deaths, hospitalizations, and vaccinations, as well as trends over time and demographic breakdowns. By utilizing this wealth of information, we aim to tailor strategies that target high-risk groups more effectively and maximize the impact of vaccination campaigns on reducing transmission and mortality rates. This targeted approach will enable us to optimize vaccination programs and improve public health outcomes across various population segments.

In addition to enhancing our model's ability to capture population heterogeneity, future research will also focus on establishing a robust mathematical foundation by conducting a well-posedness analysis. This will involve proving the existence of a unique solution that changes continuously with the variations in the data. Furthermore, we will perform a stability analysis to understand the model's long-term behavior under varying conditions. By addressing these aspects, we aim to strengthen the model's reliability and applicability, ultimately improving its capacity to predict the impact of vaccination strategies on COVID-19 transmission and contributing to more effective public health interventions.

CRediT authorship contribution statement

Touria Jdid: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohammed Benbrahim:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Mohammed Nabil Kabbaj:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis. **Mohamed Naji:** Writing – review & editing, Writing – original draft, Validation, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets used and analyzed during the current study are available from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University: https://github.com/CSSEGISandData/COVID-19. The code necessary to replicate the key findings of this study is available at: https://github.com/touria2000/COV19Sim.

Acknowledgements

The authors are appreciative and thank Sidi Mohamed Ben Abdellah University (USMBA) and the National Center for Scientific and Technical Research (CNRST) of Morocco for their support under grant number (Cov/2020/54). The authors are also grateful to the editor and the anonymous reviewers for their invaluable suggestions and constructive comments, which significantly improved the clarity and focus of the manuscript.

Appendix A

Proof of Theorem 1. By re-writing the model system (1), we obtain

$$\frac{dX}{dt} = F(t, X), \quad X(0) = X_0 \geq 0,$$

where,

$$X(t) = (S(t), E(t), I_A(t), I_S(t), I_Q(t), I_H(t), R(t), D(t), V(t))^T,$$
$$F(t, X) = \left(F_1(t, X), \ldots, F_9(t, X)\right)^T.$$

$F(t, X)$ is defined for all $t \geq 0$, $X \in \mathbb{R}^9_+$, with

$$F_1(t, X) = \omega_R R(t) + \omega_V V(t) - (\lambda + \theta_V) S(t),$$
$$F_2(t, X) = \lambda S(t) + (1 - \xi_V)\lambda V(t) - \alpha E(t),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$F_9(t, X) = \theta_V S(t) - (\omega_V + (1 - \xi_V)\lambda) V(t).$$

We note that

$$\left.\frac{dS}{dt}\right|_{S=0} = \omega_R R(t) + \omega_V V(t) \geq 0,$$
$$\left.\frac{dE}{dt}\right|_{E=0} = \lambda S(t) + (1 - \xi_V)\lambda V(t) \geq 0,$$
$$\left.\frac{dI_A}{dt}\right|_{I_A=0} = \rho\alpha E(t) \geq 0,$$
$$\left.\frac{dI_s}{dt}\right|_{I_S=0} = (1 - \rho)\alpha E(t) \geq 0,$$
$$\left.\frac{dI_Q}{dt}\right|_{I_Q=0} = \eta I_S(t) \geq 0,$$
$$\left.\frac{dI_H}{dt}\right|_{I_H=0} = \sigma_S I_S(t) + \sigma_Q I_Q(t) \geq 0,$$
$$\left.\frac{dR}{dt}\right|_{R=0} = \gamma_A I_A(t) + \gamma_S I_S(t) + \gamma_Q I_Q(t) + \gamma_H I_H(t) \geq 0,$$
$$\left.\frac{dD}{dt}\right|_{D=0} = \delta_S I_S(t) + \delta_Q I_Q(t) + \delta_H I_H(t) \geq 0,$$
$$\left.\frac{dV}{dt}\right|_{V=0} = \theta_V S(t) \geq 0.$$

The previously established inequalities hold throughout the positive orthant $\mathbb{R}^9_+$, including its boundary hyperplanes. Given solutions starting from non-negative initial conditions and non-negative time derivatives, system (1) remains in the region $\mathbb{R}^9_+$. This ensures our model system (1) is biologically realistic and well-defined. □

References

[1] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: emergence, transmission, and characteristics of human coronaviruses, J. Adv. Res. 24 (2020) 91–98, https://doi.org/10.1016/j.jare.2020.03.005.

[2] Y. Ge, W.-B. Zhang, H. Liu, C.W. Ruktanonchai, M. Hu, X. Wu, Y. Song, N.W. Ruktanonchai, W. Yan, E. Cleary, L. Feng, Z. Li, W. Yang, M. Liu, A.J. Tatem, J.-F. Wang, S. Lai, Impacts of worldwide individual non-pharmaceutical interventions on COVID-19 transmission across waves and space, Int. J. Appl. Earth Obs. Geoinf. 106 (2022) 102649, https://doi.org/10.1016/j.jag.2021.102649.

[3] R.M. El-Shabasy, M.A. Nayel, M.M. Taher, R. Abdelmonem, K.R. Shoueir, E.R. Kenawy, Three waves changes, new variant strains, and vaccination effect against COVID-19 pandemic, Int. J. Biol. Macromol. 204 (2022) 161–168, https://doi.org/10.1016/j.ijbiomac.2022.01.118.

[4] WHO coronavirus (COVID-19) dashboard, https://covid19.who.int, May 2023.

[5] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of COVID-19 in China by dynamical modeling, https://doi.org/10.48550/arXiv.2002.06563, 2020.

[6] A.B. Gumel, E.A. Iboi, C.N. Ngonghala, E.H. Elbasha, A primer on using mathematics to understand COVID-19 dynamics: modeling, analysis and simulations, Infect. Dis. Model. 6 (2021) 148–168.

[7] D.G. Whittaker, A.D. Herrera-Reyes, M. Hendrix, M.R. Owen, L.R. Band, G.R. Mirams, K.J. Bolton, S.P. Preston, Uncertainty and error in SARS-CoV-2 epidemiological parameters inferred from population-level epidemic models, J. Theor. Biol. 558 (2023) 111337, https://doi.org/10.1016/j.jtbi.2022.111337.

[8] V.P. Bajiya, S. Bugalia, J.P. Tripathi, Mathematical modeling of COVID-19: Impact of non-pharmaceutical interventions in India, Chaos, Interdiscip. J. Nonlinear Sci. 30 (2020) 113143, https://doi.org/10.1063/5.0021353.

[9] J. Panovska-Griffiths, C.C. Kerr, R.M. Stuart, D. Mistry, D.J. Klein, R.M. Viner, C. Bonell, Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study, Lancet Child Adolesc. Health 4 (2020) 817–827, https://doi.org/10.1016/S2352-4642(20)30250-9.

[10] D. Manevski, N. Ružić Gorenjec, N. Kejžar, R. Blagus, Modeling COVID-19 pandemic using Bayesian analysis with application to Slovene data, Math. Biosci. 329 (2020) 108466, https://doi.org/10.1016/j.mbs.2020.108466.

[11] Y. Ma, S. Xu, Q. An, M. Qin, S. Li, K. Lu, J. Li, L. Lei, L. He, H. Yu, J. Xie, Coronavirus disease 2019 epidemic prediction in Shanghai under the "dynamic zero-COVID policy" using time-dependent SEAIQR model, J. Biosaf. Biosecurity 4 (2022) 105–113, https://doi.org/10.1016/j.jobb.2022.06.002.

[12] R. Li, Y. Song, H. Qu, M. Li, G.-P. Jiang, A data-driven epidemic model with human mobility and vaccination protection for COVID-19 prediction, J. Biomed. Inform. 149 (2024) 104571, https://doi.org/10.1016/j.jbi.2023.104571.

[13] T. Zhou, Y. Ji, Semiparametric Bayesian inference for the transmission dynamics of COVID-19 with a state-space model, Contemp. Clin. Trials 97 (2020) 106146, https://doi.org/10.1016/j.cct.2020.106146.

[14] T. Zhang, Q. Wang, Z. Leng, Y. Yang, J. Yang, F. Chen, M. Jia, X. Zhang, W. Qi, Y. Xu, S. Chen, P. Dai, L. Ma, L. Feng, W. Yang, A scenario-based evaluation of COVID-19-related essential clinical resource demands in China, Engineering 7 (2021) 948–957, https://doi.org/10.1016/j.eng.2021.03.020.

[15] G. Rainisch, E.A. Undurraga, G. Chowell, A dynamic modeling tool for estimating healthcare demand from the COVID19 epidemic and evaluating population-wide interventions, Int. J. Infect. Dis. 96 (2020) 376–383, https://doi.org/10.1016/j.ijid.2020.05.043.

[16] J.M. Garrido, D. Martínez-Rodríguez, F. Rodríguez-Serrano, J.M. Pérez-Villares, A. Ferreiro-Marzal, M.M. Jiménez-Quintana, R.J. Villanueva, Mathematical model optimized for prediction and health care planning for COVID-19, Med. Intensiva (English Ed.) 46 (2022) 248–258, https://doi.org/10.1016/j.medine.2022.02.020.

[17] M.A. Acuña-Zegarra, S. Díaz-Infante, D. Baca-Carrasco, D. Olmos-Liceaga, COVID-19 optimal vaccination policies: a modeling study on efficacy, natural and vaccine-induced immunity responses, Math. Biosci. 337 (2021) 108614, https://doi.org/10.1016/j.mbs.2021.108614.

[18] E.A. Iboi, C.N. Ngonghala, A.B. Gumel, Will an imperfect vaccine curtail the COVID-19 pandemic in the U.S.?, Infect. Dis. Model. 5 (2020) 510–524, https://doi.org/10.1016/j.idm.2020.07.006.

[19] M. Shen, J. Zu, C.K. Fairley, J.A. Pagán, L. An, Z. Du, Y. Guo, L. Rong, Y. Xiao, G. Zhuang, Y. Li, L. Zhang, Projected COVID-19 epidemic in the United States in the context of the effectiveness of a potential vaccine and implications for social distancing and face mask use, Vaccine 39 (2021) 2295–2302, https://doi.org/10.1016/j.vaccine.2021.02.056.

[20] N.G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, R.M. Eggo, Age-dependent effects in the transmission and control of COVID-19 epidemics, Nat. Med. 26 (2020) 1205–1211, https://doi.org/10.1038/s41591-020-0962-9.

[21] R.P.D. Inward, F. Jackson, A. Dasgupta, G. Lee, A.L. Battle, K.V. Parag, M.U.G. Kraemer, Impact of spatiotemporal heterogeneity in COVID-19 disease surveillance on epidemiological parameters and case growth rates, Epidemics 41 (2022) 100627, https://doi.org/10.1016/j.epidem.2022.100627.

[22] W. Yang, D. Zhang, L. Peng, C. Zhuge, L. Hong, Rational evaluation of various epidemic models based on the COVID-19 data of China, Epidemics 37 (2021) 100501, https://doi.org/10.1016/j.epidem.2021.100501.

[23] H. Hu, S. Xiong, X. Zhang, S. Liu, L. Gu, Y. Zhu, D. Xiang, M. Skitmore, The COVID-19 pandemic in various restriction policy scenarios based on the dynamic social contact rate, Heliyon 9 (2023) e14533, https://doi.org/10.1016/j.heliyon.2023.e14533.

[24] C.T. Kiem, C.R. Massonnaud, D. Levy-Bruhl, C. Poletto, V. Colizza, P. Bosetti, A. Fontanet, A. Gabet, V. Olié, L. Zanetti, P.-Y. Boëlle, P. Crépey, S. Cauchemez, A modelling study investigating short and medium-term challenges for COVID-19 vaccination: from prioritisation to the relaxation of measures, eClinicalMedicine 38 (2021), https://doi.org/10.1016/j.eclinm.2021.101001.

[25] W. Kermack, A. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. Ser. A 115 (772) (1927) 700–721.

[26] M.Y. Li, J.R. Graef, L. Wang, J. Karsai, Global dynamics of a SEIR model with varying total population size, Math. Biosci. 160 (1999) 191–213, https://doi.org/10.1016/S0025-5564(99)00030-9.

[27] T.A. Biala, A.Q.M. Khaliq, A fractional-order compartmental model for the spread of the COVID-19 pandemic, Commun. Nonlinear Sci. Numer. Simul. 98 (2021) 105764, https://doi.org/10.1016/j.cnsns.2021.105764.

[28] C. Yang, J. Wang, Modeling the transmission of COVID-19 in the US – a case study, Infect. Dis. Model. 6 (2021) 195–211, https://doi.org/10.1016/j.idm.2020.12.006.

[29] A. Omame, M. Abbas, Modeling SARS-CoV-2 and HBV co-dynamics with optimal control, Phys. A, Stat. Mech. Appl. 615 (2023) 128607, https://doi.org/10.1016/j.physa.2023.128607.

[30] D. Okuonghae, A. Omame, Analysis of a mathematical model for COVID-19 population dynamics in Lagos, Nigeria, Chaos Solitons Fractals 139 (2020) 110032, https://doi.org/10.1016/j.chaos.2020.110032.

[31] A. Omame, N. Sene, I. Nometa, C.I. Nwakanma, E.U. Nwafor, N.O. Iheonu, D. Okuonghae, Analysis of COVID-19 and comorbidity co-infection model with optimal control, Optim. Control Appl. Methods 42 (2021) 1568–1590, https://doi.org/10.1002/oca.2748.

[32] A.O. Akuno, L.L. Ramírez-Ramírez, J.F. Espinoza, Inference on a multi-patch epidemic model with partial mobility, residency, and demography: case of the 2020 COVID-19 outbreak in Hermosillo, Mexico, Entropy 25 (2023) 968, https://doi.org/10.3390/e25070968.

[33] A. Omame, D. Okuonghae, U.K. Nwajeri, C.P. Onyenegecha, A fractional-order multi-vaccination model for COVID-19 with non-singular kernel, Alex. Eng. J. 61 (2022) 6089–6104, https://doi.org/10.1016/j.aej.2021.11.037.
[34] H.E. Randolph, L.B. Barreiro, Herd immunity: understanding COVID-19, Immunity 52 (2020) 737–741, https://doi.org/10.1016/j.immuni.2020.04.012.
[35] C.R. MacIntyre, V. Costantino, M. Trent, Modelling of COVID-19 vaccination strategies and herd immunity, in scenarios of limited and full vaccine supply in NSW, Australia, Vaccine 40 (2022) 2506–2513, https://doi.org/10.1016/j.vaccine.2021.04.042.
[36] N. Parolini, L. Dede', G. Ardenghi, A. Quarteroni, Modelling the COVID-19 epidemic and the vaccination campaign in Italy by the SUIHTER model, Infect. Dis. Model. 7 (2022) 45–63, https://doi.org/10.1016/j.idm.2022.03.002.
[37] P. Chen, K. Wu, O. Ghattas, Bayesian inference of heterogeneous epidemic models: application to COVID-19 spread accounting for long-term care facilities, Comput. Methods Appl. Mech. Eng. 385 (2021) 114020, https://doi.org/10.1016/j.cma.2021.114020.
[38] M. Betancourt, A conceptual introduction to Hamiltonian Monte Carlo, https://doi.org/10.48550/arXiv.1701.02434, 2018.
[39] J. Andrade, J. Duggan, An evaluation of Hamiltonian Monte Carlo performance to calibrate age-structured compartmental SEIR models to incidence data, Epidemics 33 (2020) 100415, https://doi.org/10.1016/j.epidem.2020.100415.
[40] S.A. Iyaniwura, R.C. Falcão, N. Ringa, P.A. Adu, M. Spencer, M. Taylor, C. Colijn, D. Coombs, N.Z. Janjua, M.A. Irvine, M. Otterstatter, Mathematical modeling of COVID-19 in British Columbia: an age-structured model with time-dependent contact rates, Epidemics 39 (2022) 100559, https://doi.org/10.1016/j.epidem.2022.100559.
[41] L. Grinsztajn, E. Semenova, C.C. Margossian, J. Riou, Bayesian workflow for disease transmission modeling in Stan, Stat. Med. 40 (2021) 6209–6234, https://doi.org/10.1002/sim.9164.
[42] A. Chatzilena, E. van Leeuwen, O. Ratmann, M. Baguelin, N. Demiris, Contemporary statistical inference for infectious disease models using Stan, Epidemics 29 (2019) 100367, https://doi.org/10.1016/j.epidem.2019.100367.
[43] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S. Leung, E.H. Lau, J.Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T.T. Lam, J.T. Wu, G.F. Gao, B.J. Cowling, B. Yang, G.M. Leung, Z. Feng, Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia, N. Engl. J. Med. 382 (2020) 1199–1207, https://doi.org/10.1056/NEJMoa2001316.
[44] S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application, Ann. Intern. Med. 172 (2020) 577–582, https://doi.org/10.7326/M20-0504.
[45] N.M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A.R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data, J. Clin. Med. 9 (2020) 538, https://doi.org/10.3390/jcm9020538.
[46] K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020, Euro Surveill. 25 (2020), https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180.
[47] K. Ravindra, V.S. Malik, B.K. Padhi, S. Goel, M. Gupta, Asymptomatic infection and transmission of COVID-19 among clusters: systematic review and meta-analysis, Publ. Health 203 (2022) 100–109, https://doi.org/10.1016/j.puhe.2021.12.003.
[48] H. Nishiura, T. Kobayashi, T. Miyama, A. Suzuki, S.-m. Jung, K. Hayashi, R. Kinoshita, Y. Yang, B. Yuan, A.R. Akhmetzhanov, N.M. Linton, Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19), Int. J. Infect. Dis. 94 (2020) 154–155, https://doi.org/10.1016/j.ijid.2020.03.020.
[49] L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, Q. Guo, T. Song, J. He, H.-L. Yen, M. Peiris, J. Wu, SARS-CoV-2 viral load in upper respiratory specimens of infected patients, N. Engl. J. Med. 382 (2020) 1177–1179, https://doi.org/10.1056/NEJMc2001737.
[50] H. Ledford, Coronavirus reinfections: three questions scientists are asking, Nature 585 (2020) 168–169, https://doi.org/10.1038/d41586-020-02506-y.
[51] H.R. Thieme, Mathematics in Population Biology, 2018.
[52] CSSEGISandData, COVID-19 data repository by the center for systems science and engineering (CSSE) at Johns Hopkins University, https://github.com/CSSEGISandData/COVID-19, Jun. 2023.
[53] U.C. Bureau, Redistricting data program, https://www.census.gov/rdo, Nov. 2022.
[54] R.M. Anderson, R.M. May, Vaccination and herd immunity to infectious diseases, Nature 318 (1985) 323–329, https://doi.org/10.1038/318323a0.
[55] P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, Math. Biosci. 180 (2002) 29–48, https://doi.org/10.1016/S0025-5564(02)00108-6.
[56] P.L. Delamater, E.J. Street, T.F. Leslie, Y.T. Yang, K.H. Jacobsen, Complexity of the basic reproduction number (R0), Emerg. Infect. Dis. 25 (2019) 1–4, https://doi.org/10.3201/eid2501.171901.
[57] A. Lye, A. Cicirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: a tutorial, Mech. Syst. Signal Process. 159 (2021), https://doi.org/10.1016/j.ymssp.2021.107760.
[58] M. Betancourt, L.C. Stein, The geometry of Hamiltonian Monte Carlo, https://doi.org/10.48550/arXiv.1112.4118, 2011.
[59] M.D. Hoffman, A. Gelman, The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, J. Mach. Learn. Res. 15 (2014) 1593–1623.
[60] B. Carpenter, A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, A. Riddell, Stan: a probabilistic programming language, J. Stat. Softw. 76 (2017), https://doi.org/10.18637/jss.v076.i01.
[61] F. Mutiso, J.L. Pearce, S.E. Benjamin-Neelon, N.T. Mueller, H. Li, B. Neelon, Bayesian negative binomial regression with spatially varying dispersion: modeling COVID-19 incidence in Georgia, Spat. Stat. 52 (2022) 100703, https://doi.org/10.1016/j.spasta.2022.100703.
[62] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 1995.
[63] S.E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, A.B. Gumel, To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic, Infect. Dis. Model. 5 (2020) 293–308, https://doi.org/10.1016/j.idm.2020.04.001.
[64] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), Science 368 (2020) 489–493, https://doi.org/10.1126/science.abb3221.
[65] J.M. Read, J.R.E. Bridgen, D.A.T. Cummings, A. Ho, C.P. Jewell, Novel coronavirus 2019-nCoV (COVID-19): early estimation of epidemiological parameters and epidemic size estimates, Philos. Trans. R. Soc. Lond. B, Biol. Sci. 376 (2021) 20200265, https://doi.org/10.1098/rstb.2020.0265.
[66] B. Tang, X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao, J. Wu, Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions, J. Clin. Med. 9 (2020) 462, https://doi.org/10.3390/jcm9020462.
[67] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, B. Cao, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, Lancet 395 (2020) 1054–1062, https://doi.org/10.1016/S0140-6736(20)30566-3.
[68] A. Senapati, S. Rana, T. Das, J. Chattopadhyay, Impact of intervention on the spread of COVID-19 in India: a model based study, J. Theor. Biol. 523 (2021) 110711, https://doi.org/10.1016/j.jtbi.2021.110711.
[69] S. Sharma, V. Volpert, M. Banerjee, Extended SEIQR type model for COVID-19 epidemic and data analysis, https://doi.org/10.1101/2020.08.10.20171439, Aug. 2020.
[70] L.B. Puglisi, G.S.P. Malloy, T.D. Harvey, M.L. Brandeau, E.A. Wang, Estimation of COVID-19 basic reproduction ratio in a large urban jail in the United States, Ann. Epidemiol. 53 (2021) 103–105, https://doi.org/10.1016/j.annepidem.2020.09.002.
[71] S.D. Team, Stan user's guide, https://mc-stan.org/docs/stan-users-guide/, 2023.
[72] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P.-C. Bürkner, Rank-normalization, folding, and localization: an improved $\hat{R}$ for assessing convergence of MCMC (with discussion), Bayesian Anal. 16 (2021) 667–718, https://doi.org/10.1214/20-BA1221.

[73] L. Grinsztajn, E. Semenova, C.C. Margossian, J. Riou, Bayesian workflow for disease transmission modeling in Stan, https://mc-stan.org/users/documentation/case-studies/boarding_school_case_study.html#fnref7.
[74] J.H. Kim, F. Marks, J.D. Clemens, Looking beyond COVID-19 vaccine phase 3 trials, Nat. Med. 27 (2021) 205–211, https://doi.org/10.1038/s41591-021-01230-y.
[75] U.S. Food, D. Administration, FDA issues emergency use authorization for third COVID-19 vaccine, https://www.fda.gov/news-events/press-announcements/fda-issues-emergency-use-authorization-third-covid-19-vaccine, Feb. 2021.
[76] COVID-19 vaccine EUA recipient/caregiver fact sheets | CDC, https://www.cdc.gov/vaccines/covid-19/eua/index.html, Jun. 2022.
[77] T.D. of Health, Downloadable datasets, https://www.tn.gov/health/cedep/ncov/data/downloadable-datasets.html.