Infectious Disease Modelling 9 (2024) 527–556

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Assessing the dynamics and impact of COVID-19 vaccination on disease spread: A data-driven approach

Farhad Waseel [a, e, *], George Streftaris [b, d], Bhuvendhraa Rudrusamy [c], Sarat C. Dass [a]

[a] School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, Putrajaya, Malaysia
[b] School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, United Kingdom
[c] School of Engineering and Physical Sciences, Heriot-Watt University Malaysia, Putrajaya, Malaysia
[d] Maxwell Institute for Mathematical Sciences, United Kingdom
[e] Faculty of Mathematics, Kabul University, Kabul, Afghanistan

ARTICLE INFO

Article history:
Received 26 November 2023
Received in revised form 23 February 2024
Accepted 23 February 2024
Available online 12 March 2024
Handling Editor: Dr Daihai He

Keywords:
Bayesian inference
SEIRV
COVID-19
Vaccination policy
Importance sampling
Malaysia
Data-driven

ABSTRACT

The COVID-19 pandemic has significantly impacted global health, social, and economic situations since its emergence in December 2019. The primary focus of this study is to propose a distinct vaccination policy and assess its impact on controlling COVID-19 transmission in Malaysia using a Bayesian data-driven approach, concentrating on the year 2021. We employ a compartmental Susceptible-Exposed-Infected-Recovered-Vaccinated (SEIRV) model, incorporating a time-varying transmission rate and a data-driven method for its estimation through an Exploratory Data Analysis (EDA) approach. While no vaccine guarantees total immunity against the disease, and vaccine immunity wanes over time, it is critical to include and accurately estimate vaccine efficacy, as well as a constant vaccine immunity decay or wane factor, to better simulate the dynamics of vaccine-induced protection over time. Based on the distribution and effectiveness of vaccines, we integrated a data-driven estimation of vaccine efficacy, calculated at 75% for Malaysia, underscoring the model's realism and relevance to the specific context of the country. The Bayesian inference framework is used to assimilate various data sources and account for underlying uncertainties in model parameters. The model is fitted to real-world data from Malaysia to analyze disease spread trends and evaluate the effectiveness of our proposed vaccination policy. Our findings reveal that this distinct vaccination policy, which emphasizes an accelerated vaccination rate during the initial stages of the program, is highly effective in mitigating the spread of COVID-19 and substantially reducing the pandemic peak and new infections. The study found that vaccinating 57–66% of the population (as opposed to 76% in the real data) with a better vaccination policy such as proposed here is able to significantly reduce the number of new infections and ultimately reduce the costs associated with new infections. The study contributes to the development of a robust and informative representation of COVID-19 transmission and vaccination, offering valuable insights for policymakers on the potential benefits and limitations of different vaccination policies, particularly highlighting the importance of a well-planned and efficient vaccination rollout strategy. While the methodology used in this study is specifically applied to national data from Malaysia, its successful application to local regions within Malaysia, such as Selangor and Johor, indicates its adaptability and

* Corresponding author. School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, Putrajaya, Malaysia.
E-mail addresses: farhadwaseel@ku.edu.af (F. Waseel), g.streftaris@hw.ac.uk (G. Streftaris), b.rudrusamy@hw.ac.uk (B. Rudrusamy), s.dass@hw.ac.uk (S.C. Dass).

Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2024.02.010

potential for broader application. This demonstrates the model's adaptability for policy assessment and improvement across various demographic and epidemiological landscapes, implying its usefulness for similar datasets from various geographical regions.

1. Introduction

The discovery of the SARS-CoV-2 virus (COVID-19) in December 2019 marked the beginning of a pandemic that significantly altered global norms in 2020 and the subsequent years (He et al., 2023a, 2023b). Malaysia detected its first cases on January 25, 2020, and there were many waves of COVID-19 in Malaysia subsequently, leading to the widespread transmission of the disease across the nation. Malaysia adopted various movement control orders (MCO) to curtail the virus spread. While these measures initially controlled transmission, they also posed many socio-economic challenges (Dass et al., 2021; Ge et al., 2022; Olivera Mesa et al., 2022; Stokes, Turner, Anselmi, Morciano, & Hone, 2022). Early in 2020, several companies began developing the first COVID-19 vaccines. On February 24, 2021, Malaysia approved the first vaccine developed by Pfizer-BioNTech and began vaccinating citizens. With the introduction of vaccination in 2021, case numbers seemed to decline (Yu, Li, Chen, & He, 2023). However, easing of the MCOs saw a resurgence in cases, prompting questions about vaccine efficacy and the feasibility of relying solely on vaccinations for controlling disease spread (Giordano et al., 2021; He, Cowling, et al., 2023; Moore, Hill, Tildesley, Dyson, & Keeling, 2021; Schwartz, Kaufman, Hu, & Bianco, 2020).

Studies such as the ones reported in (Polack et al., 2020; Saeedi, Waseel, & Jalal, 2022; Thompson et al., 2021; Tian, Yang, & Chen, 2022; Yu et al., 2023) provide insights into the broader implications of vaccination campaigns and examine the impact of different vaccination rates. However, they do not delve into the operational specifics of vaccine distribution and administration strategies. Similarly, (Yamana et al., 2023) offers a valuable perspective on the overall impact of COVID-19 vaccination in reducing disease burden in the United States, highlighting the critical role of vaccines in the public health response. However, this study primarily focuses on the broader effects of vaccination, without detailing the specific strategies for vaccine rollout and administration. In addressing this gap, our study uniquely focuses on specifying vaccine rollout strategies in the context of Malaysia, an aspect not extensively explored in previous research. We consider factors such as the varying efficacy rates and vaccine induced waning rates of different vaccine types, such as reported in (Polack et al., 2020; Suah et al., 2021; Zeng et al., 2023). This consideration is vital for COVID-19 vaccination, where different types of vaccines have been administered. Furthermore, we integrate insights from studies such as (Amin & BN, 2023; Chen, Zhao, Jin, He, & Li, 2023; Ferdinands et al., 2022; He, Chen, Zhao, & Stone, 2023; Lodz et al., 2024; Suah et al., 2022; Suleyman et al., 2023; Waseel & Yusof, 2019), which emphasize the varying transmissibility and immune escape capabilities of different COVID-19 variants, into our analysis. These variants' dynamics are crucial in formulating a responsive and adaptive vaccination strategy.

Additionally, our research incorporates a formal, in-depth analysis of COVID-19 parameters using an Exploratory Data Analysis (EDA) approach. This enhances the data-driven rigor in compartmental modeling, an aspect that has been somewhat under looked in previous studies. For instance, studies such as (Chen et al., 2023; Fei et al., 2022; Saeedi & Waseel, 2022, pp. 1–25; Watson et al., 2020; Wen et al., 2022) focusing on various aspects like transmission estimation methods, but do not fully explore the data-driven methodologies for estimating disease parameters in the context of COVID-19 epidemiology. Our approach, therefore, contributes to this area by providing detailed, model-based formulations for data-driven estimates, which are crucial for a comprehensive understanding and effective parameters of epidemiological modeling.

Therefore, we have the following aims in this study: (1) to develop an epidemic model which incorporates the observed vaccination trend for analyzing vaccination impact on the dynamics of COVID-19 transmission in Malaysia, (2) to offer a distinctive data-driven approach understanding COVID-19 parameters, providing formal in-depth approximation formulas for factors such as transmission rate, exposed individuals, initial conditions, and vaccine efficacy, and (3) to assess the impact of a distinct vaccination rollout policy on disease transmission based on observed disease characteristics.

Mathematical modeling of the epidemiological impact of vaccination has a broad and extensive history. In the context of vaccination, the dynamics of disease transmission and immunity in a population are often modeled using either network or compartmental models (Wagner, Saad-Roy, & Grenfell, 2022; Yamana et al., 2023). Compartmental models provide a good theoretical framework for studying the spread of infectious diseases and vaccination effects, and have been widely used and reported in the scientific literature (Dass et al., 2021; MacIntyre, Costantino, & Trent, 2022; Nuraini et al., 2021). In this paper, a time-varying SEIRV (Susceptible-Exposed-Infected-Recovered-Vaccinated) compartmental epidemic model is developed that incorporates Malaysia's daily vaccination rates. After capturing observed disease spread characteristics based on COVID-19 data in Malaysia, a practical vaccination policy is hypothesized and incorporated into the mathematical model. The aim is to analyze the impact of different vaccination strategies on the spread of infectious diseases by using real-world data from Malaysia. Our modeling provides valuable insights and information for policymakers on how a better vaccination rollout policy could effectively control epidemic outbreaks promptly while adopting an economical approach. Thus, we aim to highlight essential information regarding the potential benefits and limitations of a different vaccination policy to the relevant authorities. This information will enable researchers and policymakers to make well-informed decisions regarding

vaccination rollout strategies, resource allocation, and the implementation of effective measures to control the spread of infectious diseases.

The time frame for our study, from February 24, 2021, to the end of October 2021, has been chosen with precision based on three prevailing conditions and the specific dynamics of COVID-19 variants in Malaysia. First, the issue of reinfection during this period forms a pivotal aspect of our analysis. Vaccines were notably effective against reinfection, with studies reporting rates ranging from a minimal 0.0001%–1.5% (Menegale et al., 2023; Pecoraro, Pirotti, & Trenti, 2022; Stein et al., 2023; Suleyman et al., 2023). These rates were influenced by factors such as the emergence of new variants (Chen et al., 2023; He, Cowling, et al., 2023), He, Cowling, et al., 2023ge demographics, pre-existing health conditions, and the severity of initial infections. This period provides a stable baseline for understanding vaccine efficacy before the complexity of higher reinfection rates associated with later variants. Several studies, including (Feng, Luo, et al., 2022; Liao et al., 2021; Mahmud et al., 2022; Parolini, Dede', Ardenghi, & Quarteroni, 2022; Webb, 2021a; Zuo et al., 2022), relied on negligible reinfection rates within the same timeframe as our study. These researchers have also decided not to include reinfection dynamics in their models, which is consistent with the approach used in our study. Second, the timeframe encapsulates the peak of the pandemic driven by the Delta variant. Detected in June 2021, Delta quickly became the dominant strain, causing a substantial increase in cases (He et al., 2023b, 2023c). The government's interventions, such as stringent lockdowns and restrictions on public gatherings, played a crucial role in curbing the spread. However, despite these efforts, Delta remained a significant threat. Our analysis during this peak period is essential to understand the dynamics of COVID-19 under the significant strain of Delta before the introduction of additional complexities by subsequent variants (He, Chen, et al., 2023; Polack et al., 2020; Tian et al., 2022). Third and finally, our study deliberately avoids the period marked by the Omicron variant. Omicron, first detected in Malaysia towards the end of November 2021, exhibited characteristics distinct from previous strains, including higher transmissibility (Chen et al., 2023; Polack et al., 2020), an increased rate of reinfection (Deng et al., 2022), and notably lower vaccine efficacy (Thompson et al., 2021). A study in (Chen et al., 2023) highlighted Omicron's higher reproduction number, suggesting its transmissibility advantage. This underpins the decision to focus on the pre-Omicron period as the main period to study to avoid altered infection fatality rates and immune escape capabilities (Chen et al., 2023) which necessitates a different modeling approach. The findings of (Zeng et al., 2023) on the reduced effectiveness of booster vaccines against Omicron further emphasize the variant's distinct impact on vaccine strategies. Hence, our study's cutoff at the end of October 2021 provides a focused analysis of earlier phases of the pandemic, particularly the Delta variant period, allowing for a clearer assessment of vaccination strategies before the emergence of these complex Omicron dynamics.

The present study makes two distinct contributions to the field of modeling the COVID-19 epidemic. First, it develops a data-driven exploratory data analysis (EDA) approach for estimating the unknown model parameters of the SEIRV model, and thereby, is able to demonstrate the crucial role of time-varying transmission rates in modeling the spread of the disease accurately. The resulting model is flexible and capable of explaining the broad features of the epidemic with the inclusion of uncertainty intervals obtained via a Bayesian inferential framework. To illustrate the adaptability and flexibility of our model, we applied the approach to two local regions within Malaysia, indicating its potential global implications across different geographical contexts. Second, the study presents a hypothetical vaccination rollout policy, distinct from the current policy employed in Malaysia. We can assess its impact given the observed disease spread characteristics. This alternative policy demonstrates the potential to effectively curb the spread of COVID-19 b y reducing the number of infections achieved through optimal vaccine deployment. The results highlight the significance of having an early, well-planned, and efficient vaccination rollout policy in controlling the pandemic and public health.

The paper is structured into several sections to provide a clear and concise understanding of the methodology, results, and conclusions. In Section 2, we provide an in-depth explanation of the data used for analysis and the methods employed, including Exploratory Data Analysis, Bayesian inference, and the hypothetical vaccination policy. Section 3 showcases the results of the SEIRV model fit, highlighting the positive impact of the vaccination policy on reducing the number of new infections. Finally, the last section summarizes the findings and presents a conclusion and recommendations based on the results of the study. Additional details and results are provided in Appendices in the last section.

2. Materials and methodology

We evaluate disease transmission dynamics using a range of data sources. This can be done, for instance, by model fitting, parameter inference, or estimating model parameters from data. Fitting a model to data is often difficult due to the model's complexity, data restrictions, and the requirement to assimilate many data sources. The subsequent section begins by describing the real-world data used in the analysis. This is followed by the development of the Susceptible-Exposed-Infected-Recovered-Vaccinated (SEIRV) compartmental epidemic model here. The details of the model and its associated parameters are explained in detail. Next, we conduct an Exploratory Data Analysis (EDA) on the observed data to estimate the parameters of the model, including vaccine efficacy and analyze the data-driven disease spread characteristics. To quantify the uncertainties involved, we introduce the statistical inference framework that utilizes Bayesian tools and techniques. Subsequently, we introduce and describe a hypothetical vaccination policy and analyze the effects of vaccination policy with associated prediction intervals on controlling disease spread and new infections.

2.1. Data

COVID-19 data on various types of cases are provided by the Ministry of Health (MoH) Malaysia's open-access data portal (github.com/MoH-Malaysia/covid19-public). These data sets are augmented with daily COVID-specific information including daily new cases, daily total active cases, daily recovered cases, and daily fully vaccinated cases. Fully vaccinated here counts the individuals who received the full two doses of vaccination, which started in February 2021 in Malaysia. Hence, we will use COVID-19 observational data during the specified period of our study for model-fitting purposes and analyzing the vaccination effect. Each group of observed daily cases is denoted as indicated below. These notations will be used throughout this paper.

- O_N (t): to denote the observed number of new cases on day t.
- O_I(t): to denote the observed number of active (infectious) cases on day t.
- O_V(t): to denote the observed number of new fully vaccinated cases on day t.
- O_R(t): to denote the observed number of newly recovered cases on day t.

where $t = 1, 2, \ldots, T$ for days in the specified study period (February 24, 2021 to October 31, 2021), and T being the end time point corresponds to 250 (October 31, 2021).

2.2. The SEIRV model with time-varying transmission rate

To examine the transmission dynamics of COVID-19 and the vaccination effect, the well-known SEIRV epidemic compartmental model divides the population into five compartments: susceptible (S), exposed (E), infected (I), recovered (R) and vaccinated (V). These compartments illustrate COVID-19's development stages in a population (more generally, of an infectious disease with vaccination) (Brauer, Castillo-Chavez, & Feng, 2019). When susceptible individuals come into contact with infectious (i.e., able to transmit the disease) individuals, they are considered to be exposed to the virus (exposed but not yet infectious). During this period, called the incubation period, the virus incubates in these individuals. At the end of the incubation period, the exposed individual becomes infectious, and these individuals are able to spread the disease to other susceptibles. The word “infectious period or recovery period” relates to how long an individual is in the infectious state. A person will either recover, die, or develop immunity to the disease after the infectious period.

In our model, only a proportion of vaccinated individuals, who were previously susceptible to the virus but have now received two doses of vaccination, and are considered to be immune for the specified period of our study. This proportion is chosen based on the prevailing vaccine efficacy during our study period, taking into account the strains of the virus and the corresponding efficacy against those variants. Furthermore, over time, the vaccine's efficacy wanes, leading individuals to revert to a susceptible state. This dynamic, known as the vaccine-induced immunity waning rate, is crucial for understanding the temporal aspects of immunity and disease spread. This approach is reasonable and consistent with the literature on vaccine effectiveness against the existing variants during our study period, reinforcing our primary focus on understanding whether the vaccination rollout strategy would have sufficed to control disease spread in Malaysia.

The compartmental SEIRV model is given by a set of nonlinear ordinary differential equations (ODEs):

$$\frac{dS}{dt} = -\beta(t)S(t)I(t) - \nu_e\,(1-\lambda)\,\nu(t) \tag{1}$$

$$\frac{dE}{dt} = \beta(t)S(t)I(t) - \delta E(t) \tag{2}$$

$$\frac{dI}{dt} = \delta E(t) - \gamma I(t) \tag{3}$$

$$\frac{dR}{dt} = \gamma I(t) \tag{4}$$

$$\frac{dV}{dt} = \nu_e\,(1-\lambda)\,\nu(t) \tag{5}$$

where $S(t)$, $E(t)$, $I(t)$, $R(t)$, and $V(t)$ represent numbers of individuals in each of the (susceptible, exposed, infected, recovered, and vaccinated) compartments at time t, respectively. We consider the re-normalized version of the SEIRV model, i.e., $S(t) + E(t) + I(t) + R(t) + V(t) = 1$ where each compartment represents the proportion of individuals at time t rather than the total number of individuals, i.e., $S(t) = S(t)/N$, $E(t) = E(t)/N$, $I(t) = I(t)/N$, $R(t) = R(t)/N$, and $V(t)/N$, where N is the total population size. The unknown parameters in the SEIRV are denoted by $\Theta \equiv (\underline{\beta}, \delta, \gamma, i_0, e_0)$, which are the time-varying transmission rates $\underline{\beta} \equiv \{\beta(t), t = 1, 2, \ldots, T\}$ denoting the collection of $\beta(t)$ for every day t in the year 2021, the incubation rate (δ), the recovery rate

(γ), and the initial numbers of infectious and exposed individuals i_0 and e_0, respectively. The components of Θ are essential for determining the trajectories of the individuals in each compartment. The vaccination rate $v(t)$ is taken from the daily vaccination rates reported for Malaysia and is considered known for the purposes of this model. The proportion of fully vaccinated individuals who are totally immune to disease spread is denoted as v_e and λ represents vaccine-induced immunity waning rate. These parameters are considered constants in our study; however, their values are carefully estimated to account for a wide range of factors. These include the variety of vaccine types used, the number of doses given, and the specific efficacy of each vaccine type against common COVID-19 variants during the study period. Furthermore, our estimation process incorporates the various rates of immunity decay associated with each vaccine type. This comprehensive approach ensures that the overall values of v_e and λ, taking into account the numerous aspects of vaccination efficacy and immunity waning rates, thereby offering a more precise and broadly applicable analysis of vaccine-induced protection over time. The detailed explanation and justification for the choice of v_e and λ are provided in sections 2.4 and 2.5. We point out that several works in the literature have used constant vaccination waning rates in SIR-type models in their studies; see, for example, (El Khalifi & Britton, 2023; Feng, Obolski, Stone, & He, 2022; Han et al., 2022; Tran et al., 2021). The parameters of the above SEIRV model are listed in Table 1.

Furthermore, only susceptible individuals are presumed to get vaccinated, and not individuals from other compartments. The diagram in Fig. 1 illustrates the flow of disease status of individuals from S to R or V compartments. In (3), the term $\delta E(t)$ corresponds to the proportion of new cases at time t, whereas $\beta(t)S(t)I(t)$ are the proportion of new exposed individuals at time t. It is also assumed that a recovered individual cannot be reinfected; therefore, no new individuals enter the Susceptible compartment (Refer to Section 1 for further details and confer related studies [36, 37, 38, 39, 40, 41] for similar studies regarding the absence of reinfection in recovered individuals analyzed the same time period as our study.). This assumption aligns with the extensive literature on reinfection rates for SARS-CoV-2, where various studies have consistently shown low reinfection rates based on the population and variant involved (Section 1). The minimal occurrence of reinfection, particularly during the study period and within the geographic context and time period of our study, justifies the exclusion of reinfection as a significant factor in our model.

To investigate the Θ values of the SEIRV that are best supported by the observed data, we use a data-driven approach and analyze the observed data to confirm the ranges of the parameters. This process helps us to develop reasonable insights into the values of the unknown parameters and to elicit explicit parametric trends for the time-varying components of Θ. The data-driven approach explained subsequently suggests the following parametric trend for $\beta(t)$:

$$\beta(t) = e^{(a_0+a_1t+a_2t^2+a_3t^3)}, \quad 1 \le t \le T \tag{6}$$

where a_0, a_1, a_2 and a_3 are unknown coefficients. For δ and γ we assume that these parameters remain the same across time since these parameters are intrinsic to the disease and the underlying population (Dass et al., 2021; Watson et al., 2020; Wen et al., 2022). The use of the exponential form for $\beta(t)$ as specified in equation (6) is a direct result of insights gained from the data-driven approach. This representation, while appropriate for the dataset in our study, may not be universally applicable to data from other regions due to variations in epidemiological dynamics and socio-environmental factors. For instance, we use an exponential form for $\beta(t)$ with a quadratic exponent, when applying our model to Malaysia's Johor region (Appendix Appendix B.2). Researchers should conduct a careful data-driven approach (explained in subsequent sections), specified to their specific country's data, to identify the most accurate time-varying trend for $\beta(t)$. This localized approach ensures that the model form is consistent with the unique epidemic characteristics and data patterns observed in various geographical settings.

The subsequent section will explain the data-driven approach (i.e., the exploratory data approach or EDA), performed on the SEIRV model and Malaysian COVID-19 data for this purpose.

2.3. Exploratory data analysis (EDA)

Using a trial and error approach, we found that utilizing a constant value (as opposed to time-varying) for $\beta(t)$ of the SEIRV model does not adequately fit the observed data. To better comprehend the trend and trajectories of observed data in terms of underlying disease parameters, we use a data-driven approach to determine what parameter value ranges (possibly time-

Table 1
SEIRV model parameters.

Parameter	Description	Assumption
$\beta(t)$	disease transmission at day t	unknown
δ	incubation rate	unknown
γ	recovery rate	unknown
v_e	vaccine efficacy	75% (Section 2.4)
λ	vaccine-induced immunity waning rate	0.0026 (Section 2.5)
$v(t)$	vaccination rate at day t	real data (Section 2.1)
i_0	initial infectious individuals	unknown
e_0	initial exposed individuals	unknown

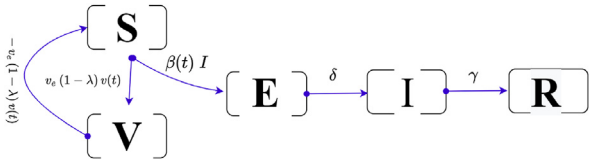

Fig. 1. Flow chart and compartments of the SEIRV model, and their associated parameters.

varying) are better supported by the observed as a suitable range of values for $\beta(t)$, δ, and γ. This method to parameterize the transmission rate using daily reported cases data was used in (Webb, 2021b). Many studies have used similar approaches (see (Webb, 2021c) for more details) to relate reported case data to explain the dynamics of a model. EDA also allows us to understand the daily variation of δ and γ as deviations from common mean values over time. In the subsequent subsections, we will determine the values for the parameters $\underline{\beta}, \delta, \gamma$ and initial values of the SEIRV model using the EDA approach.

2.3.1. Estimation of E(t) and γ

The proportion of exposed individuals ($E(t)$) for each t is a latent compartment of the SEIRV model; no observations are made on these values. Thus, we resort to Exploratory Data Analysis (EDA) to explore possible values supported by observed data. To initiate the EDA for unobserved values of the disease parameters, we assume $\delta = 1/z$, where z represents the incubation period. Recent literature on the COVID-19 models shows that the incubation period ($1/\delta$) is around 4–6 days (see, for example (Dass et al., 2021; Estrada, 2020; Lauer et al., 2020; MacIntyre et al., 2022; Mahmud & Lim, 2020; Zamri et al., 2021),). Thus, to start the EDA for unobserved values of the disease parameters, we assume $\delta = 1/5$ or $z = 5$ days.

To estimate $E(t)$, we match the observed number of new cases at time point $t+1$ to the corresponding theoretical value of new cases, which is $\delta E(t)$. Thus, we have $O_N(t+1) \approx \delta E(t)$ from which we get the estimate of $E(t)$ as

$$\hat{E}(t) = \frac{O_N(t+1)}{\delta} \tag{7}$$

As mentioned earlier, to understand the daily variation of γ, we model $\gamma \equiv \gamma(t)$ to represent its daily time-varying pattern. To estimate $\gamma(t)$, we approximate equation (4) of the SEIRV model as $R(t+1) - R(t) \approx \gamma(t)I(t)$ and replace the theoretical values by their corresponding observed quantities. Thus, $R(t+1) - R(t)$ is approximated by $O_R(t+1)$, the observed daily number of recovered cases at time point $t+1$, and $I(t)$ is approximated by $O_I(t)$, the number of active cases on day t, from which we derive the estimate of $\gamma(t)$ as

$$\hat{\gamma}(t) = \frac{O_R(t+1)}{O_I(t)} \tag{8}$$

2.3.2. Estimation of daily new exposures (ψ(t))

When an individual becomes exposed to the disease, it takes some time (incubation period) for the individual to become infectious, which is later reported as a new case. The new exposed individuals denoted by $\psi(t)$ is defined by the term $\beta(t)S(t)I(t)$ in the SEIRV model ((1)–(2)), i.e.:

$$\psi(t) = \beta(t)S(t)I(t) \tag{9}$$

With the incubation period being z days, the new cases are those who were exposed to the disease z days earlier. However, since the incubation period may vary among individuals, we employ a smoothing window to estimate the average number of daily new exposures.

The use of a smoothing window is a well-established technique in time series analysis, used to smooth out short-term fluctuations and highlight longer-term trends or cycles. This approach is particularly useful when dealing with real-world data, which can often be noisy or contain missing values. By smoothing the data, we can reduce the impact of these issues and obtain a more accurate estimate of the number of new exposures. Moreover, the use of a smoothing window aligns with the biological reality of the disease spread, effectively considering a range of possible incubation periods, which is a more realistic representation of the disease spread (Chatfield & Xing, 2019; Schwartz, 2001). In addition, the use of a smoothing window has been validated in various studies. For instance, a study by (Kandasamy, Baret, Verger, Neveux, & Weiss, 2013) compared different methods for smoothing and gap-filling time series of remote sensing observations and found that the use of a smoothing window provided accurate results. Similarly, a study by (Fu, Pedrini, & Osten, 2007) applied Fourier analysis and windowed Fourier analysis to digital hologram sequences and found that smoothing in time provided accurate measurements of displacement, velocity, and acceleration.

Let $\hat{\psi}(t)$ be an approximation of $\psi(t)$, then for a z being an odd incubation period:

$$\hat{\psi}(t) = \delta \sum_{k=\frac{z+1}{2}}^{k=\frac{3z-1}{2}} O_N(t+k), \ \ for \ \ z = 2n+1, \ \ n \in \mathbb{N} \tag{10}$$

where, z represents the window size or incubation period and $O_N(t+k)$ is observed new cases at day $t+k$.

When the incubation period is an even number ($z = 2n, \ n \in \mathbb{N}$), the formula calculates two estimates, one potentially overestimating ($\hat{\psi_1}(t)$) and the other potentially underestimating ($\hat{\psi_2}(t)$) the number of new exposures. By averaging these two, the formula effectively balances out potential biases, leading to a more accurate approximation. This approach is similar to the method of bounds in numerical analysis, where an overestimate and underestimate are used to find a range in which the true value lies. These two estimations are defined as:

$$\hat{\psi_1}(t) = \delta \sum_{k=\frac{z}{2}}^{k=\frac{3z}{2}} O_N(t+k) \tag{11}$$

$$\hat{\psi_2}(t) = \delta \sum_{k=\frac{z}{2}+1}^{k=\frac{3z}{2}-1} O_N(t+k) \tag{12}$$

By averaging the overestimation and underestimation, we obtain a more accurate approximation for the number of new exposures at day t given as

$$\hat{\psi}(t) = \frac{1}{2}(\hat{\psi_1}(t) + \hat{\psi_2}(t)). \tag{13}$$

For our study purpose, we use the following formula to approximate newly exposed individuals:

$$\hat{\psi}(t) = \frac{1}{5} \cdot \left[\sum_{k=3}^{7} O_N(t+k) \right] \tag{14}$$

which is obtained for values of $t = 0, 1, 2, \ldots, T$, by substituting $z = 5$ in (10), where T is end time point of our study period. The last few values of $\hat{\psi}(t)$ are obtained using the first few observed data from November 2021.

Using $\hat{\psi}(t)$ obtained for all t as above, we next obtain the values of $S(t)$ by approximating equation (1) in the SEIRV model as $S(t+1) - S(t) \approx \psi(t) - v_e\,(1-\lambda)\,v(t)$. Starting from $S(0) = N$, where N is the total number of population, $\hat{\psi}(t)$ is as described previously. v_e, $(1-\lambda)$, and $v(t)$ is vaccine efficacy, vaccine-induced immunity waning rate, and the daily observed vaccination rate respectively. $S(t+1)$ is estimated recursively as

$$\hat{S}(t+1) = \hat{S}(t) + \hat{\psi}(t) - v_e\,(1-\lambda)\,v(t) \tag{15}$$

for $t = 0, 1, 2, \ldots, T$.

2.3.3. Estimation of the time-varying infectivity rate $\beta(t)$

Since $\psi(t) = \beta(t)\,S(t)\,I(t)$, we estimate $\beta(t)$ as:

$$\hat{\beta}(t) = \frac{\hat{\psi}(t)}{\hat{S}(t)\,O_I(t)}, \tag{16}$$

since $O_I(t) \approx I(t)$ based on the arguments presented earlier.

2.3.4. Estimating initial values of E(t) and I(t) based on the EDA approach

To estimate the initial values of $E(t)$ and $I(t)$ for the SEIRV model based on observed daily cases, we consider using the equation $\dot{R}(t) = dR(t)/dt = \gamma I(t)$ from (4). The initial values at $t = 0$, satisfies $\dot{R}(0) = \gamma\ I(0)$ which implies $i_0 \equiv I(0) = \dot{R}(0)/\ \gamma$. Since $O_R(t)$ is a good estimate of R(t), we fit an exponential function using the least squares to the trajectory of O_R(t) for a small window consisting of w initial days ($t \leq w$). The first derivative of the fitted function is then taken as the estimate of $\dot{R}(0)$ and used to approximate i_0 from $i_0 = \frac{\dot{R}(0)}{\gamma}$, where γ is obtained earlier.

To estimate $E(0)$, we consider equation (7) and fit an exponential function to the trajectory of $O_N(t+1)$ for $t \leq w$ using the least squares. The fitted function at $t = 0$ is then taken as the estimate of $E(0)$ and used to approximate e_0 from $e_0 = \frac{E(0)}{\delta}$.

After applying various window sizes $w = 10, 30, 60$ to fit the exponential function, we found that a window size of $t \leq 60(w = 60)$ returned the most accurate results, which were closely aligned with the observed data, demonstrating its superiority over the other tested window sizes.

2.4. Estimating vaccine efficacy v_e

We analyzed the distribution of vaccines in Malaysia for the study period from February to October 2021. From the data, we found that almost 98% of total distributed vaccines belong to Pfizer-BioNTech, Sinovac, and AstraZeneca (See Table 2).

The determination of the vaccine efficacy proportion for our SEIRV model required a comprehensive approach that considers multiple factors. A rigorous review of the literature was conducted, involving a wide range of systematic review papers, including findings from sources [58, 17, 11, 12, 13, 59, 60, 61, 62, 63]. The range of efficacies varied widely for Pfizer-BioNTech (72%–95% for the first dose and 89%–95% for the second dose), Sinovac (17%–65.7% for the first dose and 51%–97% for the second dose), AstraZeneca (18%–83.5% for the first dose and 61%–90% for the second dose), and other brands (50.7%–98.1%). This reflects the complexity and diversity of the global vaccination landscape.

The selected efficacies for each brand and dose were derived by taking the average and potential value of the reported efficacies, thereby providing a representative efficacy value for each vaccine type (Table 3).

Given the significance of the Pfizer-BioNTech vaccine, especially during the initial phase of vaccination in Malaysia, and its known high efficacy, the inclusion of the first vaccine dose in the weighted calculation was essential. This decision aligns with epidemiological best practices, considering the dynamics of the real-world vaccination process.

To derive the overall vaccine efficacy proportion (v_e), a weighted mean calculation was employed, which involved weighting the chosen efficacies by the proportion of individuals vaccinated with each brand and each dose in Malaysia. The efficacy values were then multiplied by the corresponding percentage of the population vaccinated with each vaccine and divided by the total number of vaccinated cases for both the first and second doses. By employing a weighted mean approach, the calculation inherently accounts for the distribution and effectiveness of each vaccine in the real world. By integrating both specific vaccine efficacy and real-world vaccine distribution, the chosen proportion aligns closely with the actual conditions during the study period, further enhancing the model's applicability to the specific context of Malaysia.

We found that the vaccine efficacy for Malaysia is 75% based on our method and study period. The resulting value, $v_e = 0.75$, serves as a robust and contextually relevant parameter for the SEIRV model. It reflects the actual vaccination landscape in Malaysia, ensuring that the model is well-aligned with the observed epidemiological trends and that the assumptions underlying the model are grounded by empirical evidence. We arrived at this estimate in a data-driven way from the various daily case numbers presented in the dataset for Malaysia. Further, extensive literature (such as (Ghazy et al., 2022; Jara et al., 2021; Jin, Li, Zhang, Li, & Zhu, 2022; Polack et al., 2020; Suah et al., 2021; Tanriover et al., 2021; Thompson et al., 2021; Tian et al., 2022; Zeng, Gao, Zhou, Yu, & Sun, 2022; Zheng et al., 2022)) demonstrates high initial vaccine efficacy against COVID-19 transmission for vaccines like Pfizer-BioNTech, AstraZeneca, and CoronaVac, exceeding 92%, 83% and 74%, respectively. Another study which provides a report from various other studies concluded a pooled mean vaccine efficacy of 76% against SARS-CoV-2 based on data on 7 million individuals for all ages and vaccines. The study also reported a sharp decline after 100 days following full vaccination of around 83% (Ssentongo et al., 2022). These studies support our assumption of a high level of initial protection and the incorporation of a vaccine waning rate which is described in the subsequent section.

2.5. Vaccine induced immunity waning rate (λ)

Several studies, including (Amin & BN, 2023; Ferdinands et al., 2022; Husin et al., 2022; Levin et al., 2021; Lodz et al., 2024), and (Townsend, Hassler, Sah, Galvani, & Dornburg, 2022), have shown a decline in antibody levels after vaccination. (Lodz et al., 2024) and (Amin & BN, 2023) emphasize the reduction in anti-S antibody levels, starting from the third month post-vaccination, with Pfizer-BioNTech recipients showing initially higher median levels, indicating a stronger or long-lasting immune response. This trend is confirmed in (Ferdinands et al., 2022), which shows a nuanced reduction in vaccine effectiveness against severe outcomes like hospital admissions, becoming increasingly apparent in the subsequent months post-vaccination. The findings of (Levin et al., 2021) support the claim that immunity declines over time. Furthermore, a systematic review ((Notarte et al., 2022)) provides insight into the trajectory of antibody levels. It identifies a gradual decrease that begins 4–6 months after vaccination after peak levels are reached between 21 and 28 days after the second dose. This pattern indicates a decline in antibody levels, which coincides with a decrease in vaccine efficacy.

Table 2
Vaccination brands distributed by October 31, 2021.

Vaccine brand	Dose one (%)	Dose two (%)
Pfizer-BioNTech	52.50	49.80
Sinovac	39.44	38.81
AstraZeneca	7.99	7.80
Other	0.07	3.59

Table 3
Vaccination efficacy for different brands based on literature for the specified study period.

Vaccine brand	Dose one (%)	Dose two (%)
Pfizer-BioNTech	80	92
Sinovac	35	74
AstraZeneca	36	83

While the exact rate and duration of waning vaccine efficacy vary and are not uniformly available across the literature, it is evident from the study (Suah et al., 2022) that within a 3–5 month timeframe, there is a significant decrease in vaccine efficacy, particularly for Pfizer-BioNTech (from 90.8% to 79.9%) and CoronaVac (from 74.5% to 30.4%). Furthermore, (Townsend et al., 2022) adds details to our understanding by quantifying vaccine efficacy with a 5% infection probability, indicating the average time frame for this risk threshold after vaccination. Another study, (Menni et al., 2022) incorporates a broad analysis of 620,793 participants, revealing that vaccine effectiveness wanes post the second dose. Specifically, at the 5-month mark, the effectiveness of vaccines was decreased from 89-97%–82.1%, and from 92% to 75.7% for Pfizer-BioNTech (BNT162b2) and AstraZeneca (ChAdOx1 nCoV-19) respectively. For other mRNA vaccines, the waning over 5 months decreased from 94% to 84.3%.

Using these distinct but converging insights, the evidence suggests that vaccine efficacy declines around the 4–5 month mark. This trend is highlighted by the observed reduction in efficacy, particularly in the Pfizer-BioNTech vaccine, where efficacy is significantly waned by 10.9–14.9%, 16.3%, and 44.1% for Pfizer-BioNTech, AstraZeneca, and CoronaVac, respectively.

We assume that vaccine efficacy (VE) declines exponentially, which is consistent with established models in epidemiological research and supported various studies (for example [44, 45, 46]). These studies demonstrate the exponential model's ability to accurately represent the gradual loss of vaccine-induced immunity, which is an important factor in understanding vaccine performance over time. Notably, the contrast with the classic epidemic model, particularly in terms of gradual waning via exponential decay functions, emphasizes the model's relevance. Therefore, we can describe the rate of decline in efficacy over time using an exponential decay model (Menegale et al., 2023). This model is commonly used to describe processes where the quantity of interest decreases at a rate proportional to its current value. The efficacy of the vaccine, $V_e(t_{end})$ at time, t_{end} can be described as:

$$V_e(t_{end}) = V_e(t_{init}) \cdot e^{-\lambda t} \tag{17}$$

where, $V_e(t_{end})$ is vaccine efficacy at the time t_{end}, $V_e(t_{init})$ is the initial vaccine efficacy (right after receiving the second dose), λ is the constant decay rate (which we want to determine), t is the time duration post-vaccination in days, e is the base of the natural logarithm.

Given the data from Table 4 and assuming the efficacy reduction at 120 or 150 days (4 or 5 months) as given, we can rearrange the formula to solve the constant decay rate λ for each vaccine:

$$\lambda = -\frac{1}{t} \cdot \ln\left(\frac{V_e(t_{end})}{V_e(t_{init})}\right) \tag{18}$$

The decay rate λ provides a measure of how quickly the efficacy decreases for each type of vaccine. It's important to note that this model assumes a constant rate of decay, which may be a simplification depending on the actual biological processes at play. Given (18) and vaccine efficacy wane in literature, the average approximated waning rates for each type of vaccine are shown in Table 5.

To compute the overall waning rate of vaccine efficacy, a method similar to that used to calculate the overall vaccine efficacy proportion was used. This involved using a weighted mean approach, in which the waning rates for each vaccine were weighted based on the proportion of the population inoculated with each vaccine type. Specifically, the waning rate for each vaccine was multiplied by the proportion of the population who received that vaccine. The resulting values were then aggregated and divided by the total number of vaccinated individuals, taking into account second doses, to calculate the overall waning rate across all vaccines. We found that the overall rate is $\lambda = 0.0026$. This value is close to several studies: For

Table 4
The waning of vaccine efficacy (VE) within a 4–5 month period post-vaccination, for different types of vaccines.

Reference	Vaccine Type	Vaccine Efficacy at vaccination	Vaccine Efficacy post-vaccination
[22, 68]	Pfizer-BioNTech	90.8%	79.9% (4 Months)
	CoronaVac	74.5%	30.4% (4 Months)
	AstraZeneca	–	–
[68, 70]	Pfizer-BioNTech	89–97%	82.1% (5 Months)
	CoronaVac	–	–
	AstraZeneca	92%	75.7% (5 Months)

Table 5
Waning rate of vaccines per day.

Vaccine Type	λ
Pfizer-BioNTech	0.000905168
CoronaVac	0.006054338
AstraZeneca	001300069

example, (Han et al., 2022) reports the waning rate to be between 0.002 and 0.005, (Feng, Obolski, et al., 2022) chose waning rates as 0.00027 and 0.0021, (Zhou et al., 2022) assumes 0.002739, and (Fisher, Xu, He, & Wang, 2023) utilized a waning rate 0.0027 and 0.005. This rate will be used for additional analysis in our model to account for the waning effect of vaccine efficacy over time, resulting in a more accurate and dynamic representation of vaccine performance in the population.

2.6. Verifying EDA results

To investigate the adequacy of the estimated parameters based on our EDA approach, these parameters are reused in the SEIRV model ((1)–(5)) with time-varying estimated parameters $\hat{\beta}(t), \hat{\gamma}(t), O_V(t)$, and v_e obtained from EDA based on the assumption of $\delta = 1/5$. The ODE system is solved and the resulting trajectories are shown in Fig. 2. The trajectories in Fig. 2 show that the model and observed data match closely, which assures that the EDA approach works well for choosing suitable values for $\beta(t)$ and $\gamma(t)$ consistent with the observed data.

The estimated $\hat{\beta}(t)$ and $\hat{\gamma}(t)$ trajectories based on our EDA approach are shown in Fig. 3. Fig. 3 shows that $\hat{\beta}(t)$ is increasing exponentially. Thus, based on the $\hat{\beta}(t)$ trajectory, a time-dependent exponential function for the transmission rate ($\beta(t)$) is postulated to describe the dynamics of the disease. To estimate the model parameters, we used regression analysis with time as an independent variable and estimated transmission rates as the dependent variable. The regression coefficients, standard errors, and significance levels show a strong fit, supporting the inclusion of temporal components in $\beta(t)$. Although the model captures the dominant trend, there may be some deviations between the estimated and fitted ($\beta(t)$) values, due to factors such as changes in contact patterns, public health interventions, or varying levels of immunity within the population. The $\beta(t)$ exponential form is explained earlier in (6). The $\hat{\gamma}(t)$ trajectory looks dispersed around the mean with no significant trend. We take the mean of $\hat{\gamma}(t)$ as the estimated value for a common γ.

2.7. Uncertainty quantification

The insights derived from EDA on the Malaysian COVID-19 data formed the basis for the following stages. Using the findings from the EDA, we selected priors on the components of Θ and iteratively adjusted hyperparameters of the priors by

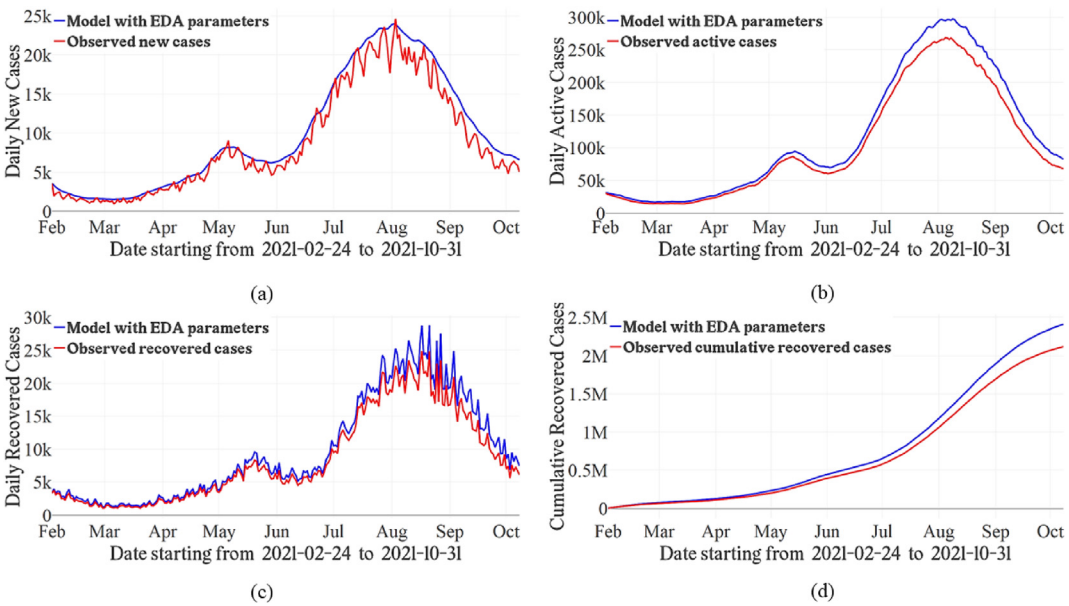

Fig. 2. Malaysia's observed data trajectories (red line) and trajectories based on the SEIRV model with estimated parameters using the EDA approach (blue line): Figure panels show the following types of daily cases: (a) new cases, (b) active cases, (c) daily recovered cases and (d) cumulative recovered cases.

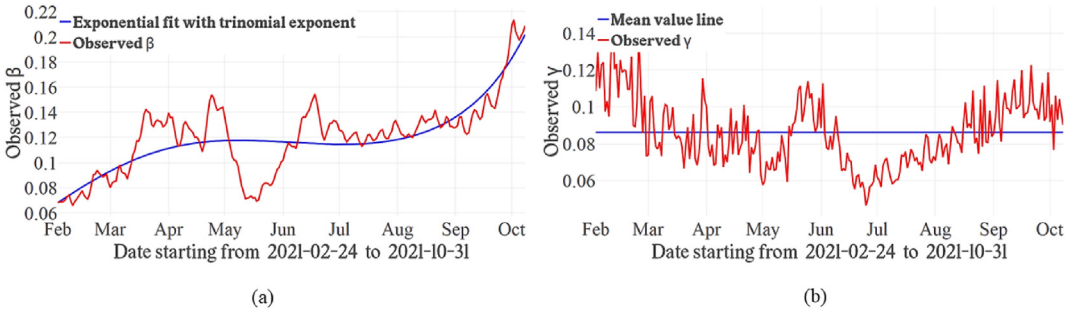

Fig. 3. The estimated trajectories of (a) $\hat{\beta}(t)$ and (b) $\hat{\gamma}(t)$ using the outlined EDA approach.

step-by-step improvement of the model fit to the observed data. Thus, our iterative procedure is similar to the selection of priors in the empirical Bayes approach (Watson et al., 2020; Wen et al., 2022). After establishing the best model fit, we extended the framework proposed in (Dass et al., 2021) to apply Bayesian methods to infer unknown parameters of the SEIRV models. The application of the Bayesian framework allows us to generate point estimates and quantify uncertainties via the posterior distributions of the unknown parameters.

Our Bayesian data-driven methodology is explained in detail in the following subsections.

2.7.1. The likelihood

The negative binomial (NB) distribution is widely used in epidemiological studies because it effectively handles over-dispersed count data. The NB distribution used in our study is parameterized in terms of its mean (μ) and dispersion parameter (α) with probability mass function (PMF):

$$Pr(X = x|\mu, \alpha) = \frac{\Gamma(\alpha + x)}{x!\Gamma(\alpha)} \left(\frac{\mu}{\mu + \alpha}\right)^{x} \left(1 + \frac{\mu}{\alpha}\right)^{-\alpha}, \quad \mu > 0, \alpha > 0 \tag{19}$$

where x is a non-negative integer taking values in $\{0, 1, 2, \cdots\}$. The mean and variance of X is μ and $\mu + \frac{\mu^2}{\alpha}$. In the statistics literature, the quantity $\tau = \frac{1}{\alpha}$ is commonly used due to its preferable properties for inference. Using the τ parameterization, the variance on NB distribution becomes $\mu + \mu^2\tau$ (Lloyd-Smith, 2007). Both α and τ are called dispersion parameters and the term 'shape parameter', 'size', or 'clustering coefficient' are also referred to as α. In this study, all calculations are conducted using τ. The observed total number of daily cases O_I(t), O_N(t), and O_R(t) (from I and R compartments in the SEIRV model) are assumed to be distributed according to an NB distribution given the underlying trajectories of the SEIRV ODE model:

$$O_N(t) \sim NB(\bullet\ ;\ \delta E(t), \tau) \tag{20}$$

$$O_R(t) \sim NB(\bullet\ ;\ \gamma I(t), \tau) \tag{21}$$

where $NB(\bullet; \mu, \tau)$ denotes the negative binomial probability density function in (19).

2.7.2. Data-driven prior assignment

As mentioned earlier, the hyperparameters corresponding to the priors on Θ components are selected in a data-driven way to obtain the best model fit in an approach similar to empirical Bayes. The hyperparameters of the priors for the

Table 6
Lower(A) and upper(B) bounds for the uniform priors, as chosen based on existing literature and the aforementioned EDA approach.

Parameter	A	B
a_0	-2.801	-2.800
a_1	0.01175	0.01185
a_2	$-6.8 \times e^{-5}$	$-6.6 \times e^{-5}$
a_3	$1.24 \times e^{-7}$	$1.26 \times e^{-7}$
x	-1.6054	-1.6034
y	-2.479	-2.478
τ	0.001	0.01
i_0	$\zeta_{i_0} - 100$	$\zeta_{i_0} + 100$
e_0	$\zeta_{e_0} - 100$	$\zeta_{e_0} + 100$

SEIRV model are shown in Table 6. More details on choosing the ranges for these hyperparameter values are provided in subsequent paragraphs.

Let Θ denote the collection of all unknown parameters of the model for the entire time period of our study period: $\Theta \equiv \{a_0, a_1, a_2, a_3, \delta, \gamma, \tau, i_0, e_0\}$. The prior is assumed to be uniformly distributed on each individual component of Θ, whose lower and upper bounds are determined using the EDA approach described earlier. The complete joint prior on Θ is considered as the product of the individual prior assignments, assuming that the components of Θ are independent apriori.

For every $\epsilon \in \{a_0, a_1, a_2, a_3\}$ (parameters associated with the transmission rate $\beta(t)$), the priors are assumed to be independent and uniform with lower and upper bounds A_ϵ and B_ϵ, respectively. The values for A_ϵ and B_ϵ are selected based on the EDA and are given in Table 6. The range of values for the parameters δ and γ were determined based on reported literature and the EDA from Section 2.3. For the prior on δ, we consider $\delta = e^x$ and choose $x \sim U(A_x, B_x)$ with A_x and B_x as given in Table 6. Similarly, γ is parameterized as $\gamma = e^y$ where $y \sim U(A_y, B_y)$ with A_y, B_y values reported in Table 6.

The priors on the initial number of infectious and exposed individuals i_0 and e_0, respectively, are taken as $i_0 \sim U(\zeta_{i_0} - \Delta_{i_0}, \zeta_{i_0} + \Delta_{i_0})$ and $e_0 \sim U(\zeta_{e_0} - \Delta_{e_0}, \zeta_{e_0} + \Delta_{e_0})$, where the choices of the mid-points, ζ_{i_0} and ζ_{e_0} and half-widths Δ_{i_0} and Δ_{e_0} are made based on the EDA estimations explained in Section 2.3.4 with setting based on the exponential fit in the initial window size of 60 days. The half-widths Δ_{i_0} and Δ_{e_0} are also adjusted to be large enough to represent the maximum extent of prior uncertainty surrounding ζ_{i_0} and ζ_{e_0}, respectively. Table 6 shows the hyperparameter settings assigned to model the SEIRV model.

2.7.3. The posterior distribution

Based on the negative binomial likelihood and prior elicitation, the posterior of Θ can be derived using the Bayes theorem as

$$\pi(\Theta|D) \propto \pi(\Theta) \times \prod NB(O_N(t);\ \delta E(t), \tau)\ \times\ \prod NB(O_R(t); \gamma I(t), \tau)\ =\ \pi(\Theta) \cdot L(\Theta) \tag{22}$$

where $\delta E(t)$ and $\gamma I(t)$ are the mean values of $O_N(t)$ and $O_R(t)$, respectively, and τ is the dispersion parameter under negative binomial likelihoods, D is the collection of all observed daily cases of $O_N(t)$ and $O_R(t)$. For the entire year of 2021, $L(\Theta)$ is the full likelihood, and $\pi(\Theta)$ is the full prior joint specification on Θ as explained in Section 2.7.2.

2.8. Importance sampling

The Monte Carlo method of importance sampling is developed to conduct Bayesian inference on Θ by obtaining sample and weight pairs that can approximate the posterior distribution of Θ, empirically. Based on specified priors in Table 6, a total of M samples, Θ_i for $i = 1, 2, \cdots, M$, are generated for large M. The likelihood, $L(\Theta_i)$, is calculated based on each sample Θ_i and weights w_i are obtained as:

$$w_i = \frac{L(\Theta_i)}{\sum_{i=1}^{M} L(\Theta_i)} \tag{23}$$

where, $w_i \geq 0$ and $\sum_{i=1}^{M} w_i = 1$.

Subsequently, importance resampling is used to obtain new samples from the weighted samples $\{w_i, \Theta_i\}_{i=1}^{M}$ with selection probabilities given by w_i. Denote $\{w_i^*, \Theta_i^*\}_{i=1}^{M}$ to be the resampled Θ_i values with new weights $w_i^* = 1/M$. The resampled Θ_i^* values give an unbiased estimate of $E(g(\Theta)|D)$ and its corresponding variance based on the usual sample mean and variance of $\{\Theta_i^*\}_{i=1}^{M}$:

$$E(g(\Theta)|D) \hat{=} \frac{1}{M} \sum_{i=1}^{M} g(\Theta_i^*)$$

and

$$Var(g(\Theta)|D) = \frac{1}{M} \sum_{i=1}^{M} (g(\Theta_i^*) - \bar{g})^2$$

An approximation to the Maximum-a-Posteriori (MAP) estimator of Θ can also be found based on Θ_i samples, since

$$\Theta_{MAP} \equiv\ \arg\max_{\Theta} \pi(\Theta|D)\ \Longrightarrow\ \hat{\Theta}_{\text{MAP}} \approx \Theta_{\text{MAP}} \tag{24}$$

for large M, where $\hat{\Theta}_{MAP} \equiv \arg\max_{\Theta} \pi(\Theta_i|D)$. The validation of the above approximation is explained in (Dass et al., 2021; Tokdar & Kass, 2010).

The resampled values $\{\Theta_i^*\}_{i=1}^M$ can be used to quantify the uncertainty present in the inference of the unknown parameters. Since the importance sampling method enables the computation of sample variance based on $\{g(\Theta_i^*)\}_{i=1}^M$ for any function g, credible bands around the sample mean $\frac{1}{M}\sum_{i=1}^{M} g(\Theta_i^*)$ can be obtained for a given confidence level. The resampled values $\{\Theta_i^*\}_{i=1}^M$ can also be used to quantify the uncertainty present in the predictive distributions corresponding to the observed quantities which are the daily new cases, the active cases and the recovered cases. Each sample in $\{\Theta_i^*\}_{i=1}^M$ is incorporated into the SEIRV model, resulting in predictive distributions that summarize the estimated future states of the epidemic. Following this, a negative binomial distribution is employed to simulate the observed data. For each time point t, we compute the 2.5% and 97.5% percentiles of the simulated data, forming the bounds of the credible interval. Alongside these, the median (50% percentile) is calculated to provide a measure of central tendency. This approach enables a more accurate evaluation of the model's performance, offering a better understanding of the uncertainty surrounding the model's predictions.

The computational algorithm is developed in R. The likelihood value of $L(\Theta_i)$ for each Θ_i is computed numerically based on the ODE system in (1)-(5) using 'stats', 'bayestestR', and 'deSolve' packages in R.

2.9. Vaccination policy

The impact of a distinct vaccination rollout program on the disease burden of COVID-19 in Malaysia can be studied based on our developed SEIRV model and Bayesian inference methodology. We introduce a hypothetical but practical vaccine rollout policy while keeping Malaysia's vaccine supply in mind and aiming for a minimum number of vaccinations to considerably reduce the disease burden. Our hypothetical vaccination rollout policy is shown in Fig. 4, which has a trapezoidal shape. The function of proposed vaccination policy $v(t, \theta)$ is defined as:

$$v(t,\theta) = \begin{cases} \dfrac{e^{\theta}}{(T_1 - T_0)}\, t - \dfrac{e^{\theta}}{(T_1 - T_0)}\, T_0, & \text{if } T_0 \le t \le T_1 \\ e^{\theta}, \quad \text{if } T_1 \le t \le T_2 - \dfrac{e^{\theta}}{(T_3 - T_2)}\, t + \dfrac{e^{\theta}}{(T_3 - T_2)}\, T_3, & \\ & \text{if } T_2 \le t \le T_3 \\ 0, & \text{if otherwise} \end{cases} \tag{25}$$

where θ represents the natural logarithm of the total daily vaccination rate achieved. This vaccination policy is defined by four key time points: T_0, T_1, T_2, and T_3: T_0 represents the start point of the vaccination program where the vaccination rate increases linearly until the maximum daily number of vaccinations is achieved at time T_1, the maximum vaccination rate continues until T_2 and then decreases linearly until T_3 which marks the end of the vaccination program. This policy lets the vaccination rate grow until it reaches the maximum number of vaccination rates required to control new infections, and then linearly decreases as the number of new infections decreases. More details on the vaccine policy function and its associated parameters are available in Appendix A. The salient feature of the policy is the time point T_0, where T_0 is the vaccination start time, and in our study, it is determined by Malaysia's vaccination start date.

To study the impact of the hypothesized vaccination policy on COVID-19 incidence using our fitted models, we formulate a new SEIRV model coupled with the hypothetical vaccination policy and a time-varying disease transmission rate $\beta(t)$ to explore the disease spread within the population, where the empirical $v(t)$ is replaced by $v(t, \theta)$ for the unknown parameter θ. The resulting SEIRV formulation is:

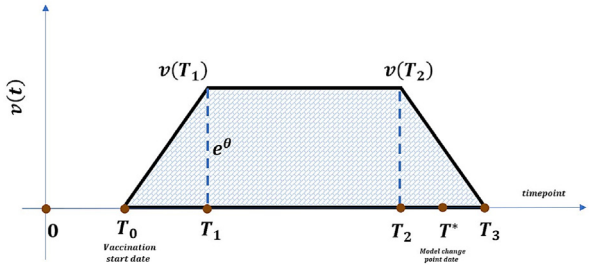

Fig. 4. Hypothetical vaccination policy.

$$\frac{dS}{dt} = -\beta(t)S(t,\theta)I(t,\theta) - \nu_e\,(1-\lambda)\,\nu(t,\theta) \tag{26}$$

$$\frac{dE}{dt} = \beta(t)S(t,\theta)I(t,\theta) - \delta E(t,\theta) \tag{27}$$

$$\frac{dI}{dt} = \delta E(t,\theta) - \gamma I(t,\theta) \tag{28}$$

$$\frac{dR}{dt} = \gamma I(t,\theta) \tag{29}$$

$$\frac{dV}{dt} = \nu_e\,(1-\lambda)\,\nu(t,\theta) \tag{30}$$

It would be ideal to vaccinate all populations against COVID-19 as early as possible; however, socio-economic factors, supply shortages, and the price of the vaccine make this impractical. To study the optimal value for this strategy, we construct a loss function to estimate the optimal value of θ at which the number of new infections will drop after reaching a specified time point with a particular vaccination rate and percentage. The loss function governing our policy is predicated upon two key metrics: the total number of individuals vaccinated and the disease burden. The function is designed to account for both of these factors, providing a comprehensive assessment of the efficacy of our policy. The function is defined as

$$J(\theta) = e^{\theta}\frac{(T_2 - T_1 + T_3 - T_0)}{2}\cdot C + H\int_0^{T_3}\delta\,E(t,\theta)\,dt \tag{31}$$

where $e^{\theta}\frac{(T_2-T_1+T_3-T_0)}{2}$ is the area of trapezium representing the total number of people vaccinated, the parameters C and H are vaccination costs and hospitalization costs, respectively, and $\int_0^{T_3}\delta E(t,\theta)$ is the disease burden representing the total number of individuals infected by COVID-19. The disease burden can be estimated numerically as

$$J(\theta) = e^{\theta}\frac{(T_2 - T_1 + T_3 - T_0)}{2}\cdot C + H\delta\sum_{i=0}^{T_3-1}E(t_i,\theta) \tag{32}$$

leading to an approximated loss function of (31).

The optimal value of θ is found by minimizing $J(\theta)$ with respect to θ using the gradient descent algorithm. More details on the loss function, its associated parameters, and assigned values are provided in Appendix A.3. Identifying the optimal θ can help public health policymakers create a vaccination program that reduces disease spread while remaining economically viable. The θ parameter helps public health officials set a clear target for disease control by determining the optimal peak vaccination rate. This priority supports efficient resource allocation for vaccines, personnel, and logistics rather than aiming for broad maximum-rate campaigns that may be unsustainable or less effective. It emphasizes the cost-effectiveness of these strategies. Resources saved by optimizing vaccination targets may be redirected to other essential health services or future pandemic preparedness. Finally, this parameter serves as the link between vaccination rates, epidemiological impact, and resource management, providing an evidence-based foundation for decision-making in public health.

3. Results and discussions

Our study's time period is from February 24, 2021, to October 31, 2021. In subsequent sections, the Bayesian inference approach as described in Section 2.7 is applied to the SEIRV model with the fitted exponential time-varying transmission rate and vaccine efficacy proportion. Parameter estimates and uncertainty bands are provided based on the Bayesian analysis described earlier. Finally, we apply our proposed vaccination rollout policy on the fitted SEIRV model to the Malaysia data and explain the vaccination rollout policy's impact on controlling the pandemic compared to the actual rollout in Malaysia during 2021.

3.1. SEIRV MAP estimates and uncertainty bands

The Monte Carlo importance sampling method is applied with a total of $M = 50,000$ simulations to obtain the map estimates of Θ, $\hat{\Theta}_{MAP}$. Also, samples $\{\Theta_i^*\}_{i=1}^{M}$ are obtained from the posterior of Θ in (22). The trajectories for $\delta E(t)$, $I(t)$, and $R(t)$ compartments are shown in Fig. 5 for each day t along with the observed daily cases $O_N(t)$, $O_I(t)$ and $O_R(t)$, respectively. The numerical MAP estimates are shown in Table 7.

These curves indicate that the model closely fits the observed data and performs better with a time-varying disease transmission rate. According to the MAP estimates, the incubation period for Malaysia is roughly 4.7 days, while the

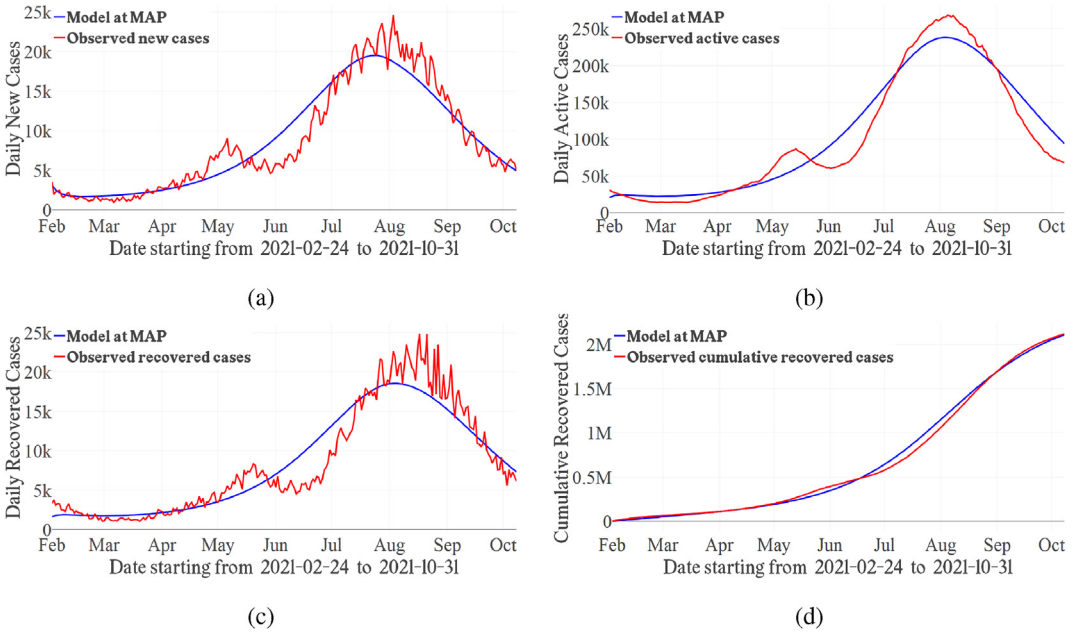

Fig. 5. Plots based on $\hat{\Theta}_{MAP}$ of SEIRV model (blue line) and Malaysia's observed daily reported cases (red line). Figure panels show the following types of daily cases: (a) new cases, (b) active cases, (c) daily recovered cases and (d) cumulative recovered cases.

infectiousness period is roughly two weeks (12.8 days). Although we point out that our model does not account for the local peak that occurs around June, it is possible to do so by modifying the exponential functional form of the disease transmission rate with further flexible forms (Fei et al., 2022) as suggested by the EDA. Since we are interested in capturing overall trends of the disease (which includes a global peak in 2021), we found that the exponential form for $\beta(t)$ is simple enough and yet sufficient for our purpose. This approach keeps our model focused on broader epidemiological insights, which are critical for public health planning and response strategies.

The uncertainty estimates are obtained for all unknown parameters in Θ based on the ensemble $\{\Theta_i^*\}_{i=1}^M$. The predictive variability of the fitted model is shown in Fig. 6 in comparison with the observed trajectories. In Fig. 6, the shaded purple regions show the 95% prediction interval based on the predictive distributions of O_N, O_I, and O_R. The uncertainty area in Fig. 6 shows that the SEIRV model is able to capture broad features of observed cases, such as global peak and overall trend. In Fig. 6, the 50th percentile (blue curves) of the predictive distribution, also known as the median, is less sensitive to extreme values in the data compared to the MAP estimate. By using the median curve, we provide a more robust central tendency measure and further provide the 95% credible interval around the median as a measure of uncertainty.

The SEIRV model shows flexibility, notably in its incorporation of time-varying transmission rate $\beta(t)$, which allows a variety of time trends. These time-varying forms of $\beta(t)$ can be interpreted as distinct parameters, each estimable from observed data, and serve as the benchmark for model fitting and precise parameter estimation. The incorporation of time-varying transmission rates enables the model to capture complex time-based dynamics and changing patterns of disease transmission, further enhancing the adaptability of the model. This functionality is particularly critical in epidemiological studies, where the disease transmission rate often fluctuates due to changes in environmental factors, interventions, or population behavior. Inherently, the flexibility to utilize different forms of the time-varying transmission rate empowers the SEIRV model to efficiently adapt to and accurately represent a diverse array of epidemic scenarios and trajectories. This feature, coupled with a robust statistical framework, enables the model to provide reliable insights into the spread and control of infectious diseases, thereby strengthening its value as a robust tool for epidemiological forecasting and public health decision-making.

Having established the robustness and effectiveness of the SEIRV model in capturing the broad features of observed cases, we now discuss the implications of a hypothetical implemented vaccination policy. This policy has three main goals: to reduce the number of infections, to minimize the required total number of vaccinations, and concurrently, to minimize the total cost associated with new infections. The evaluation of this policy hinges on the foundation laid out by our SEIRV model. The vaccination policy, thus, serves as an extension of our primary model, introducing another parameter and offering insights into potential public health policies. In the following section, we will explore the results yielded by this proposed vaccination policy. We will investigate its effectiveness, explore its impact on the trajectories of the SEIRV model, and ultimately, understand how it can shape the course of the pandemic in Malaysia.

Table 7
MAP parameter estimates based on Malaysia daily reported cases.

Parameter	MAP-SEIRV
a_0	−2.800367
a_1	0.01179586
a_2	$-6699985 \times e^{-5}$
a_3	$1.251495 \times e^{-7}$
τ	0.009990192
$1/\delta$	4.680573
$1/\gamma$	12.82215
i_0	20624.81
e_0	14595.44

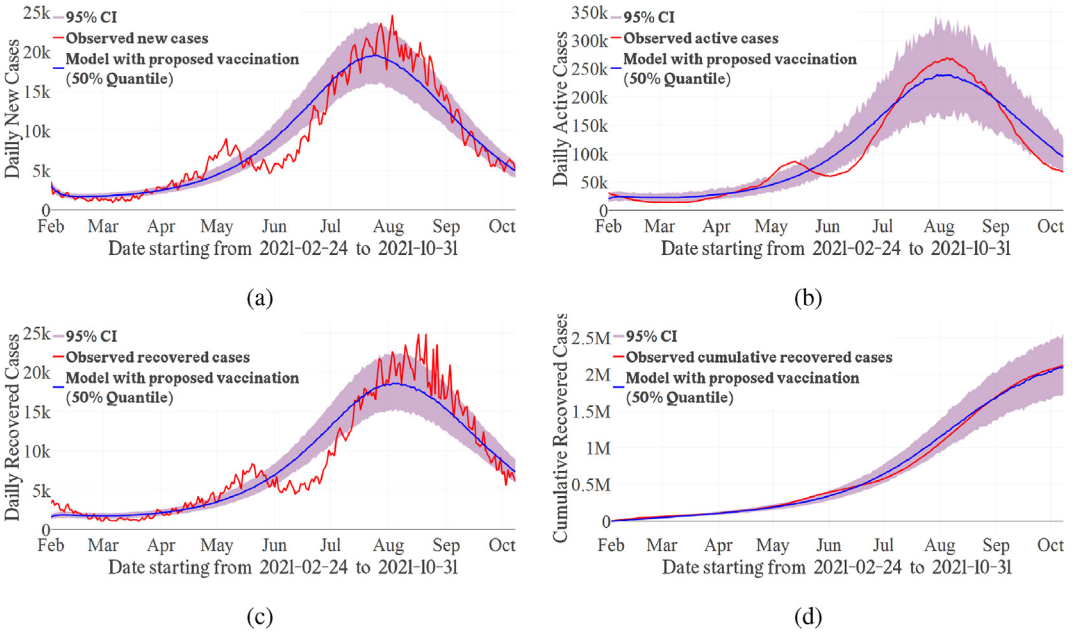

Fig. 6. MAP and variability estimate of daily cases in the SEIRV model: Malaysia's observed daily reported cases (red line), estimated daily cases based on MAP (blue line) and posterior prediction bands of daily cases (purple shaded). Figure panels show the following types of daily cases: (a) new cases, (b) active cases, (c) daily recovered cases and (d) cumulative recovered cases.

3.2. Proposed vaccination policy on SEIRV model

In Section 2.9, we described the vaccination rollout policy, the corresponding parameters, the loss function, and the gradient descent algorithm for the SEIRV model. To provide further clarity, the parameters were carefully chosen to consider the real-world (observed) data and literature. For instance, the cost of vaccination per individual ($C = 77.35$) was determined based on the Malaysian government's estimates, and the cost of hospitalization ($H = 846.10$) was derived by taking median cost informed by several media and government reports about daily hospitalization costs. The time points ($T_0 = 21.5$, $T_1 = 115.5$, $T_2 = 175.5$, and $T_3 = 245.5$) were selected to align with the actual and anticipated vaccination rollout trajectories, aiming to capture the dynamics of the vaccination program and its impact on new infections. More detailed explanations and rationales behind these parameter choices are provided in Appendix A.2. The fixed parameter values are provided in Table 8.

With these parameters in place, we initiate the gradient descent algorithm with an initial value of $\theta = 14$ and a learning rate of $\eta = 0.001$. The algorithm is designed to converge when the change in θ is small enough (specifically, when $\eta\nabla(\theta) \leq 10^{-8}$), ensuring the stability and accuracy of our results. After executing the algorithm, the derivative converges to zero after 141 iterations (≈ 0.00001087324) and the minimal loss function is obtained at $\theta = 11.85390$. The trajectories of the SEIRV model with this optimal θ for the vaccination rollout policy are presented in Fig. 7.

Our results, as illustrated in Fig. 7, show a significant decline in daily reported cases upon the implementation of our proposed vaccination policy. This policy predicts vaccinating a total of 20,022,019 individuals, a figure remarkably lower than the observed data, which stands at 24,801,906. Notably, the distinctive structure of our loss function is specifically designed to find an optimal balance between reducing the number of vaccinations and optimizing the cost of the disease burden. This

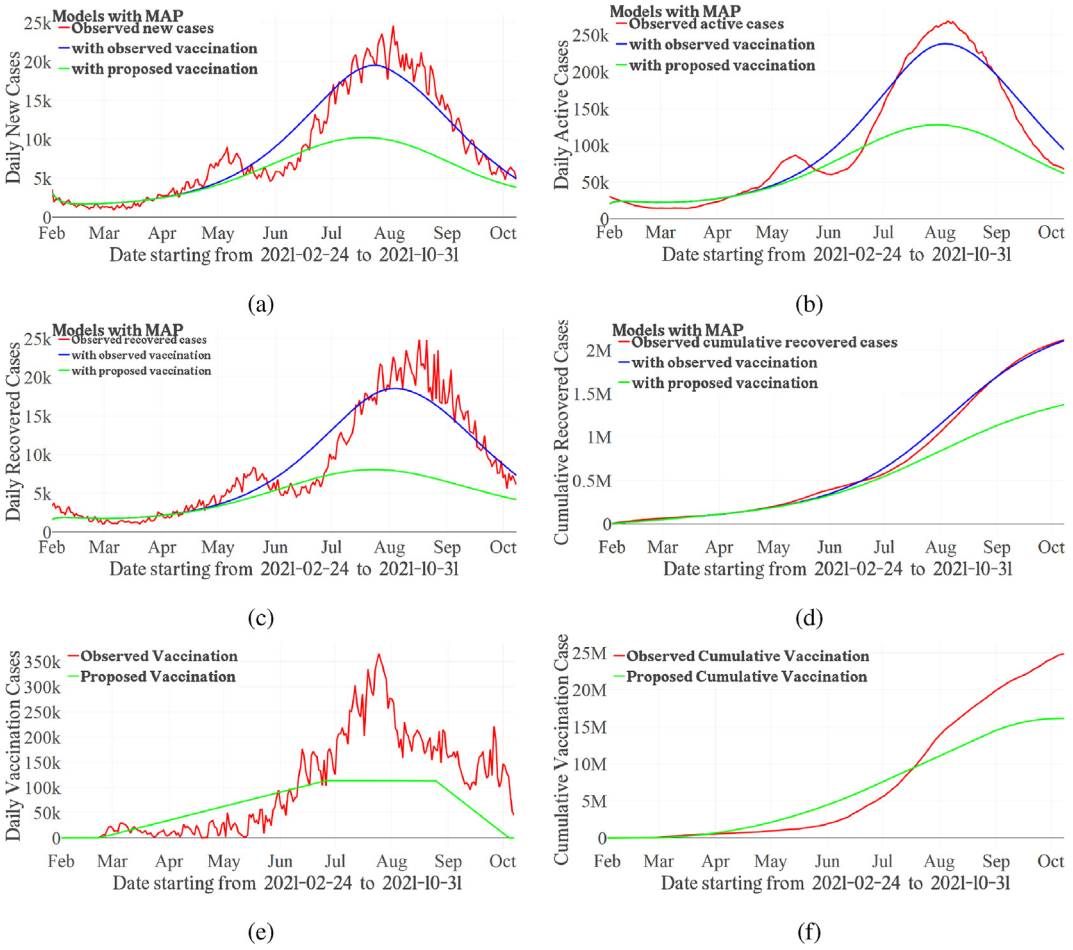

Fig. 7. Plots based on $\hat{\Theta}_{MAP}$ for the observed vaccination (blue line), proposed vaccination rate with $\theta = 11.85390$ (green line), and Malaysia's observed daily reported cases (red line). Figure panels show the following types of daily cases: (a) New cases, (b) Active cases, (c) Daily recovered cases, (d) Cumulative recovered cases, (e) Daily vaccinated cases, and (f) Cumulative vaccinated cases.

strategic balance is achieved through the regularization rule, which ensures that both the vaccination numbers and disease burden costs are considered. The observed vaccination trajectory, shown in panel (e), reveals an initial slow vaccination rate that subsequently experiences an exponential surge. However, we propose an alternative approach that promotes a higher initial rate of vaccination for the first six months. This is followed by a steady, constant rate that does not require further increment. Eventually, as the vaccination program gains control over the disease spread, the rate should gradually decrease. Fundamentally, our proposed vaccination model emphasizes an efficient rollout policy that could potentially lead to a substantial reduction in new infections. Interestingly, this can be achieved even with a lower total number of people vaccinated — a figure hovering around 61.26% (20.02 million) as opposed to the observed 76.08% (24.83 million). Moreover, the structure of loss function and minimization technique not only aims to reduce the number of vaccinations but also minimize the cost associated with the disease burden. This balanced approach, in turn, has considerable economic implications, allowing for better management of resources during the health crisis. The results from our study underline the significant influence of a well-designed vaccination policy in effectively managing disease outbreaks. This highlights the need for strategic policy and implementation in public health measures, particularly in our ongoing efforts against the COVID-19 pandemic. The uncertainty analysis is based on MAP estimates and θ resampling is shown in Appendix A.

Having evaluated the effectiveness of our proposed vaccination rollout policy, we now focus on the robustness of the results under different conditions. Specifically, we aim to analyze how the selection of the key time points T_1 and T_2 impacts the model outcomes. While T_0 and T_3 are well-defined based on real-world data—reflecting the start and end of the vaccination program, respectively—the time points T_1 and T_2 represent an interval where the vaccination rate remains constant. Their selection is not strictly determined by data but guided by it, hence introducing a degree of uncertainty. However, θ, another key parameter in the SEIRV model, depends on these time points and directly influences the vaccination policy and the resulting number of new cases. Given these considerations, it is crucial to assess the sensitivity of our results to changes in T_1, T_2, and optimal θ. A sensitivity analysis can provide valuable insights into the robustness of our findings and their

dependence on these parameters. In the upcoming section, we explore how various values for T_1 and T_2 affect the model's outcomes while accounting for optimal θ in the SEIRV model.

3.2.1. Sensitivity analysis

In the previous section, we evaluated the effectiveness of our hypothetical vaccination rollout policy. We found that the policy was effective in reducing the number of new infections. However, the results were based on a specific set of assumptions, including the values of the key time points T_1 and T_2. In this section, we conduct a sensitivity analysis to assess the robustness of our results to changes in these parameters. We vary the values of T_1 and T_2 and re-run the model to see how these new values influence the number of new infections. By conducting this analysis, we aim to assess the robustness of our model and the reasonableness of our chosen values for T_1 and T_2.

Let n be an integer such that $n \in [-30, 30]$. We define T_1^* and T_2^* as new values for T_1 and T_2, respectively, where $T_1^* = T_1 + n$ and $T_2^* = T_2 + n$. We run the model with a one-day variation to each T_1^* and T_2^*, respectively, while all other parameters are fixed as described in the previous section. The sample is selected such that the distance between T_1^* and T_2^* is also kept constant. Table B10 and Fig. 8 show the influence of T_1^* and T_2^* on the total number of vaccinations and new cases.

The detailed outcomes of the sensitivity analysis are listed in the table (See Appendix A). The results, underline the significant impact of the timing of the maximum vaccination rate (T_1) on both the cost of the disease and the total vaccination requirement to suppress new cases. Interestingly, it is found that a delay in achieving the maximum vaccination rate (each day delayed in reaching T_1) can lead to a considerable escalation in the disease's overall cost. This delay also concurrently increases the total number of required vaccinations to mitigate the spread of new cases. On the other hand, reaching the maximum vaccination rate earlier yields the opposite effect, reducing both the disease's cost and the total number of vaccinations required. To quantitatively examine this relationship, we employed the least squares method to fit a linear function to the total cost and total vaccinations derived from the sensitivity analysis (as presented in Table B10). The fitted results are graphically represented in Fig. 9. According to our fitted model, each day of delay in reaching T_1 would result in an additional expenditure of approximately 9,286,394 MYR and necessitate an additional 36,759 vaccinations to optimally minimize new cases. This finding further underscores the importance of rapid mobilization and the cost-effectiveness of achieving maximum vaccination rates as quickly as possible. Rapid action proves to be highly beneficial, not only in mitigating the number of new cases but also in significantly reducing the economic burden associated with the disease. The findings emphasize the significance of a well-planned and efficiently executed vaccination strategy, which plays a crucial role in managing disease outbreaks. The linear relationship observed between T_1 and both total cost and total vaccinations provides predictability and serves as a valuable tool for resource allocation and strategic planning in the face of disease outbreaks. The results also align with the study in (Yamana et al., 2023) where analysis shows that COVID-19 vaccination reduced the burden of disease. In conclusion, our analysis highlights the critical role of timely and effective vaccination strategies in managing the health and economic impacts of disease outbreaks. By emphasizing the importance of rapid action and the benefits of achieving peak vaccination rates earlier, these findings underscore the significance of proactive measures in curbing the spread of diseases and minimizing their associated costs.

4. Conclusion

In our analysis of the dynamics of COVID-19 transmission and the impact of vaccination in Malaysia, we developed a model that captures the multifaceted nature of the disease's spread during the year 2021. This research represents a significant advancement in modeling disease transmission, particularly concerning COVID-19 in Malaysia. We integrated a time-varying disease transmission rate into our model, allowing for a more accurate representation of real-world disease dynamics and providing a data-driven method to insight model parameters. The Bayesian inference approach employed is a notable methodological contribution, offering a nuanced method of parameter inference while accounting for inherent uncertainties. Additionally, our proposed vaccination policy offers a distinct perspective, emphasizing efficiency and cost-effectiveness in

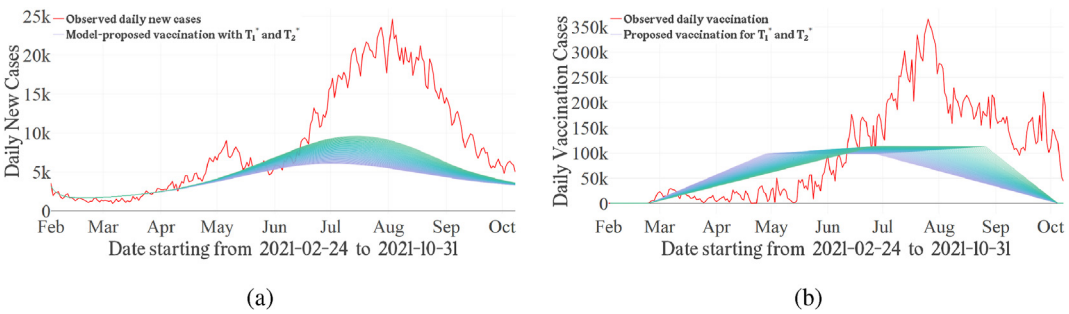

Fig. 8. Plots based on $\hat{\Theta}_{MAP}$, hypothetical vaccination policy and different time points for T_1 and T_2 (colored lines), and Malaysia's observed daily reported cases (red line). Figure panels show the following types of daily cases: (a) new cases and (b) daily vaccinated cases.

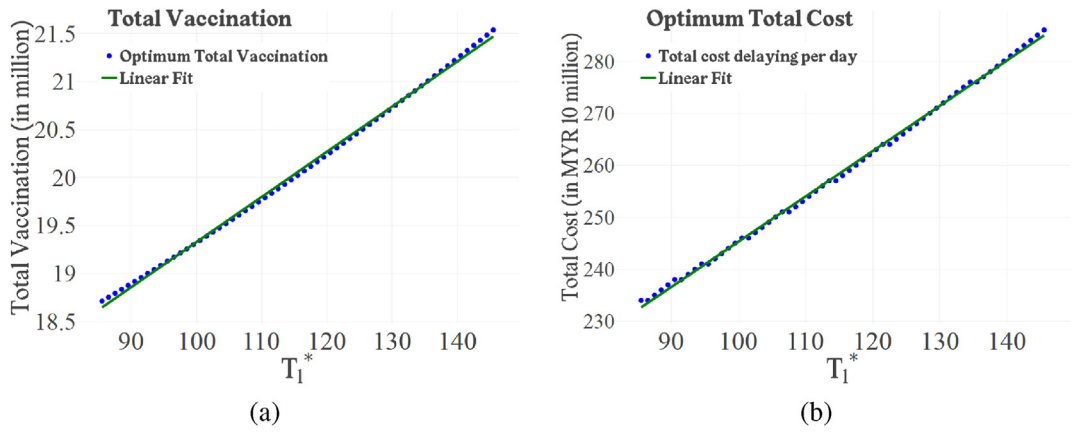

Fig. 9. Linear fit plot based on Table D10 (green line), and actual values (blue line). Figure panels show the following types of information: (a) optimum total vaccination and (b) optimum total cost.

controlling the pandemic. These findings not only contribute to the academic literature but also provide practical insights for policymakers and public health officials.

Our findings, which estimate the incubation period of COVID-19 in Malaysia to be approximately 4.6 days and the infectious period to be around 13 days, are in concordance with the results of global studies conducted in (Chen et al., 2023; He et al., 2023a, 2023b), underscoring a global consistency in the temporal dynamics of the disease. Considering both the costs associated with vaccination and the disease burden, our study found that a well-planned policy could substantially reduce the pandemic peak and curb the emergence of new infections. The effectiveness of an increased vaccination rate during the first few months after the program's launch is particularly highlighted and recommended by this study. The proposed vaccination policy suggests that, given Malaysia's observed disease transmission profile, a smaller percentage of the vaccinated population could still effectively contain the pandemic. Furthermore, our study suggests that a decelerated vaccination rate substantially escalates the overall disease burden and associated economic costs. While this phenomenon aligns with the observations made by (Watson et al., 2020; Yamana et al., 2023), our research distinctively outlines the optimal balance between the costs incurred due to disease burden and total vaccination cost, thereby contributing a nuanced perspective to the discourse on pandemic response optimization.

However, there are some limitations. The model does not account for the local peak observed around June, which suggests that other external factors or patterns in transmission might not have been adequately captured. Our model was sensitive to the time-varying disease transmission rate, which requires further improvement. Approaches such as MCMC and deep learning could be employed to explore deeper into the characteristics of such parameters. By reducing dimensions and refining the accuracy, deep learning models can further enhance the predictive capabilities of our existing framework.

Comparatively, our proposed vaccination policy appears more efficient, with a recommended vaccination of 57–66% of the population by the end of October 2021 to effectively control the pandemic. This contrasts with the empirical estimate of 76%, indicating potential over-vaccination, given our model's predictions. Our focus was the period from February to October 2021, thus a compelling path for future research would be the integration of reinfection dynamics, particularly in the context of newer strains like the Omicron variant. To gain a comprehensive understanding of Malaysia's entire COVID-19 trajectory of 2021 and later on, a new model should incorporate these dynamics. Furthermore, by analyzing the impact of vaccination against these later strains, researchers can extend the current models. This would be important for those interested in understanding the evolving nature of the virus, including the effects and implications of strains like Omicron and its vaccination impact and rollout strategies.

In addition to the national analysis, the expansion of our model to the regions of Selangor and Johor demonstrates the adaptability and robustness of our approach. The tailored parameterization for disease transmission rate, $\beta(t)$, in these regions, with a trinomial exponent for Selangor and a quadratic exponent for Johor, demonstrates the model's flexibility and precision in capturing the unique epidemiological characteristics of diverse demographic settings. The findings from these regions not only align with national trends but also provide more detailed, localized insights, particularly into disease transmission dynamics and the efficacy of vaccination strategies. This regional analysis confirms the model's ability to aid targeted public health interventions and emphasizes the importance of adaptable, data-driven models in managing public health crises.

The use of time-dependent parameters is pivotal for our models to represent real-time pandemic dynamics accurately (i.e., mimic the observed trajectories of case numbers and capture their general trends). Our estimation of time-dependent parameters is informed by the detailed exploratory data analysis (EDA) tailored to the given SEIRV model. The EDA methodology is thus flexible for capturing trends of time-dependent parameters in a data-driven way for different regions and different time periods, both for short and long terms. We have illustrated this in the revised manuscript by including the analysis of

case numbers for two additional regions within Malaysia with different trajectories of disease spread, which were estimated using the same EDA methodology and SEIRV model. In the short term, the EDA and fitted SEIRV time-varying parameter model capture rapid shifts in transmission dynamics. In the long term, the same methodology is able to estimate stable transmission and evolutionary trends when the disease achieves endemicity. However, for the long term when the disease achieves endemicity, the approach necessitates additional analytical investigation into time-series forecasting tools and techniques to understand the impact of trends, seasonality and structural changes on transmission dynamics over extended periods. Additionally, for future research it would be beneficial to expand the model's adaptability to other regions and diseases, utilizing other Bayesian inference approaches to account for uncertainties inherent in epidemiological data. Our study sets a foundation for understanding disease transmission and the profound impact of vaccination policies. Future research should aim at refining model parameters, employing other Bayesian (Wen et al., 2022) and deep learning methods (Watson et al., 2020), and expanding the scope to include other factors such as varying public health measures (Chen et al., 2023), and evaluating the long-term impacts of such vaccination policies (Zeng et al., 2023).

In summary, our research underscores the need for a proactive, data-driven approach in public health decision-making. By leveraging advanced methods and focusing on understanding and modeling the disease transmission and vaccination rollout strategies, we could have more informed, effective, and adaptive health policies. This study also emphasizes the need for adaptability, continuous learning, and the integration of emerging methodologies.

Funding

This work was funded by the James Watt PhD Scholarship program supported by Heriot-Watt University. The funders had no role in the design of the study, the collection, analysis, and interpretation of data, or the writing of the manuscript.

Data availability

The data supporting our study is publicly available by the Ministry of Health, Malaysia in the GitHub repository at the following URL: https://github.com/MoHMalaysia/covid19-public.

Authors contribution

FW: Conceptualization; Data curation; Formal analysis; Investigation; Methodology development and assessment; Resources; Software; Visualization; Writing - original draft; and writing — review and editing. SCD: Conceptualization; Methodology development and assessment; Project administration; Supervision; Validation; review and editing GS and BR: Investigation; Methodology development and assessment; Validation; Visualization; review and editing. All authors critically assessed the data analysis, interpretation, and revision of the manuscript. All authors commented on previous versions of the manuscript, and read and approved the final version.

CRediT authorship contribution statement

Farhad Waseel: Writing − review & editing, Writing − original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **George Streftaris:** Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Bhuvendhraa Rudrusamy:** Validation, Methodology, Investigation, Formal analysis. **Sarat C. Dass:** Validation, Supervision, Project administration, Methodology, Conceptualization.

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.

Appendix A. Vaccination policy function

Appendix A1. Vaccination policy function

In this section, we develop the mathematical form of $v(t, \theta)$ consistent with the vaccination policy shown in trapezium shape in the main article. This vaccination policy is defined by four key time points: T_0, T_1, T_2, and T_3: T_0 represents the start point of the vaccination program where the vaccination rate increases linearly until the maximum daily number of vaccinations is achieved at time T_1. The maximum vaccination rate continues until T_2 and then decreases linearly until, T_3 which marks the end of the vaccination program.

First, for $t \in [T_0, T_1]$, $v(t, \theta)$ is linear with a positive slope. Thus, $v(t, \theta) = at + b = f(t)$, say. We know that $f(T_0) = a\,T_0 + b = C_0$, and $f(T_1) = a\,T_1 + b = e^{\theta}$, where C_0 is a constant indicating the total number of people already vaccinated at the commencement of the vaccination date (T_0). We can calculate a and b by solving this system of equations, obtaining:

$$v(t,\theta) = \frac{e^{\theta}}{(T_1 - T_0)}\ t + C_0 - \frac{e^{\theta} - C_0}{(T_1 - T_0)}\ T_0. \tag{A.1}$$

for $t \in [T_0, T_1]$. Second, we define a constant function for the next stage by assuming that daily vaccination numbers should be kept at a constant rate, thus for $t \in [T_1, T_2]$,

$$v(t,\theta) = e^{\theta}. \tag{A.2}$$

In the third partition, we define a linear function with a negative slope as $v(t, \theta) = -at + b = g(t)$, say, as the vaccination rate decreases after time T_2. We know that $g(T_3) = -a\,T_3 + b = C_1$ and $g(T_2) = -a\,T_2 + b = e^{\theta}$, where C_1 is a constant indicating the number of persons remaining for daily vaccination at the time T_3. Thus, for $t \in [T_2, T_3]$

$$v(t,\theta) = -\frac{e^{\theta} - C_1}{(T_3 - T_2)}\ t + C_1 + \frac{e^{\theta} - C_1}{(T_3 - T_2)}\ T_3. \tag{A.3}$$

combining all time periods, we obtain the equation of $v(t, \theta)$ as

$$v(t,\theta) = \begin{cases} \dfrac{e^{\theta}}{(T_1 - T_0)}\ t + C_0 - \dfrac{e^{\theta} - C_0}{(T_1 - T_0)}\ T_0, & \text{if } T_0 \le t \le T_1 \\ e^{\theta}, \quad \text{if } T_1 \le t \le T_2 - \dfrac{e^{\theta} - C_1}{(T_3 - T_2)}\ t + C_1 + \dfrac{e^{\theta} - C_1}{(T_3 - T_2)}\ T_3, & \\ & \text{if } T_2 \le t \le T_3 \\ 0, & \text{if } t < T_0 \text{ or } t > T_3. \end{cases} \tag{A.4}$$

We assume that the number of vaccinations at time T_0 and T_3 is zero because T_0 is the vaccination program's start time point and T_3 is the time point where we expect the vaccination program to reach zero. Thus, the final form of the function $v(t, \theta)$ is

$$v(t,\theta) = \begin{cases} \dfrac{e^{\theta}}{(T_1 - T_0)}\ t - \dfrac{e^{\theta}}{(T_1 - T_0)}\ T_0, & \text{if } T_0 \le t \le T_1 \\ e^{\theta}, \quad \text{if } T_1 \le t \le T_2 - \dfrac{e^{\theta}}{(T_3 - T_2)}\ t + \dfrac{e^{\theta}}{(T_3 - T_2)}\ T_3, & \\ & \text{if } T_2 \le t \le T_3 \\ 0, & \text{if } t < T_0 \text{ or } t > T_3. \end{cases} \tag{A.5}$$

Appendix A2. Parameters of hypothetical vaccination policy

Numerous factors of our vaccination policy reflect the SEIRV model's trajectories. For the parameters of the SEIRV model ($\beta(t)$, δ and γ), we employ the values of MAP estimates obtained using importance sampling earlier (Section 2.2). In this section, we will explain the values of vaccination function parameters T_0, T_1, T_2, T_3 (Time points in trapezium), C (cost of two doses of vaccination per individual), and H (Cost of hospitalization per individual). The proposed vaccination policy's time points and trends have been thoughtfully constructed in light of the observed vaccination data.

The time point T_0 is chosen as the start date of Malaysia's vaccination program. For the Pfizer-BioNTech vaccine, which is Malaysia's first authorized vaccination, the recommended time between doses is three weeks (21 Days) [75]. We also assumed that each individual would receive two doses of vaccines, Thus, we chose March 16, 2021 (The initial date was February 24, 2021) as the vaccination start date ($T_0 = 21$).

During the initial phase, spanning from day 21–115, the vaccination rate displays a linear growth pattern. This consistent, yet modest trajectory, is represented by $y(t) = 3947.1 + 188.3 \times t$. It might encompass the logistical learning curve or initial public hesitancy. A strong start, maximizing on the momentum of the early days, is essential for setting this stage which is slowing the spread of the disease, and establishing a precedent for the subsequent phases. Therefore, a linear trend is proposed at this stage and the point is chosen based on the observed trajectory.

The two, time points of T_1 and T_2 are more significant because they represent a period where the vaccination is constant. These points should also be practical and ideal to keep vaccinations to a minimum while considering Malaysia's vaccination rollout supply availability and healthcare system capacity. We examined the current Malaysia rollout trajectories and discovered that Malaysia already has the total capacity and sufficient supply from mid-June 2021 to August 2021. The data shows an exponential growth in the vaccination rate, captured by $y(t) = e^{(7.19+0.03t)}$. This upsurge suggests that the vaccination program is progressing, possibly due to improved infrastructure, increased public urgency, or increased vaccine availability.

However, the policy's perspective on this trend is informative. Instead of following this exponential trajectory, it advocates for a more consistent, steady rate. This strategy is based on two main reasons. First, by definition, a constant rate ensures logistical predictability and operational simplicity. Second, and perhaps more importantly, the cautious strategy taken in the initial phase — achieving a higher rate prior to T_1, eliminates the need for an exponential surge later on. This balanced approach aims to keep the overall vaccination trajectory in balance. The maximum vaccination is at time point 175 based on observed data. Therefore, it is practical that the time points T_1 and T_2 can be located between these dates ($T_1, T_2 \in [80, 180]$). The values for the time points are chosen as $T_1 = 115$ and $T_2 = 175$. Additionally, we then perform the sensitivity analysis, which will be explained later, to check the sensitivity of our vaccination policy with these time points.

The final phase, which runs from day 175–250, defines an obvious slowdown in the vaccination rate. Thus, the next time point is T_3, and we expect that the vaccination program will reach zero at this point. According to Malaysian vaccination data, second-dose vaccination rates will remain constant after October 10, 2021, with a daily vaccination rate of fewer than 10,000 people. This is a reasonable choice of $T_3 = 245$, which is October 25, 2021, while we expect the second vaccination dose to be sufficient to reduce new infections. To summarize, the proposed vaccination policy parameters are more than just numbers. They are detailed patterns of strategic decisions, informed by observed trends, and constructed with a single goal in mind: an effective, feasible, and strategically sound vaccination rollout.

On the other hand, according to the reports, the Malaysian government estimated that two vaccination doses would cost 77.35 MYR (Malaysian Ringgit) per individual. Even though the rates for each type of vaccination may be different, the cost is considered to be average. As a result, the cost of vaccination, $C = 77.35$, is chosen for further investigation. Various reports (For example [76] and [77]), on the other hand, estimate that each day of hospitalization in Malaysia costs $\$427 (1750 - 1794$ MYR) at the pediatric intensive care unit (ICU) and $\$1324 (5428 - 5560$ MYR) at the general ICU. The work [78] estimated that the cost of hospital-based case management activities is $\$206.38 (825 - 870$ MYR) per day in the hospital for severe cases and $\$2011.43 (8045 - 8448$ MYR) for critical cases. However, in Malaysia, critical cases with ICU needs are much lower, so we chose $H = 846.10$ as a reasonable value for the cost of disease-burden hospitalization.

Based on the above explanations, the parameter values are shown in Table 8.

Appendix A3. Loss function and gradient descent algorithm

Consider the loss function (32), where $e^{\theta}(T_2 - T_1 + T_3 - T_0)/2$ is the area of the trapezium, representing the total number of people vaccinated. The parameters C and H are vaccination costs and hospitalization costs, respectively, and $H\delta\sum_{i=0}^{T_3-1} E(t_i, \theta)$ is the disease burden. This function represents a trade-off between the total number of vaccinated individuals (first term) and the total number of new infections (second term) at time t. This balance aligns with public health policy, where increasing the number of vaccinations can lead to a decrease in new infections, and vice versa, and can also be viewed as a form of regularization. Regularization techniques are often used in statistical and machine learning models to prevent overfitting and to balance competing objectives [79, 80]. In this case, the loss function performs a similar role, providing a balance between two competing public health objectives: minimizing the number of vaccinations (to save resources and because every vaccination carries some risk, however minimal) and minimizing the number of new infections (to protect public health). The C and H parameters further emphasize this balance and serve to weigh the two terms in the loss function. This allows the model to prioritize either vaccinations or infections depending on the specific values of these parameters, representing the trade-offs involved in public health policy decisions and capturing the inverse relationship between vaccinations and new infections.

As a result, there exists an optimal solution for θ such that it reduces the number of new infections and overall costs. The conditions for applying a gradient descent algorithm are met, so we use the gradient descent algorithm to minimize the function for θ and find the optimal value.

Gradient descent is a recursive first-order optimization process that locates the local minimum/maximum of a given function. Given the function $J(\theta)$ in (32), we optimize the function with respect to θ:

$$\theta_{n+1} = \theta_n - \eta \ \nabla J(\theta) \tag{C.1}$$

where, n represents the number of steps (next point), η is the learning rate (adaption rate) which scales the gradient and thus controls the step size, and $\nabla J(\theta) = \frac{\partial}{\partial\theta} J(\theta)$ is the gradient. Thus:

Table 8
Fixed and pre-determined parameters for hypothetical vaccination policy.

Parameter	Description	Value	Reference
T_0	Starting day of vaccination	21.5	Appendix A.2
T_1	Day where vaccination rate reach maximum	115.5	Appendix A.2
T_2	Day where vaccination rate is constant at maximum	175.5	Appendix A.2
T_3	Day where vaccination rate decline to zero	245.5	Appendix A.2
C	Cost of vaccination per individual for two doses	77.35 (MYR)	Appendix A.2
H	Hospitalization cost per individual per day	846.10 (MYR)	Appendix A.2

$$\frac{\partial J}{\partial \theta} = e^{\theta} \cdot \frac{(T_2 - T_1 + T_3 - T_0)}{2} \cdot C + H\delta \sum_{i=0}^{T_3-1} \frac{\partial E(t_i, \theta)}{\partial(\theta)} \tag{C.2}$$

This algorithm changes the parameters so that the new values lie some distance down the steepest error gradient from the current values. The algorithm is overly reliant on the learning rate parameter. The slower the learning rate is, the longer the gradient descent takes to converge or reach maximal iteration steps, before reaching the optimum point. If the learning rate is too high, the algorithm may bounce around or diverge. As a result, it is critical to begin the algorithm with a near θ, a suitable learning rate, and the number of steps. We use a graphical method to see the minimum point of loss function by plotting vaccination and infection burden. It is found that $\theta \approx 14$ is a good starting point for choosing $\eta \approx 0.0005$. The calculation process of optimum θ and solving SEIRV differential equations is explained in Table A9.

Table A9
Gradient descent algorithm for optimization of θ and solution for SEIRV differential equations.

Calculation process of optimum θ and solving SEIRV differential equations
1. **Initialization:** •Define Dates: DATE ←{*date*1, *date*2, …, *dateN*} •**Define SEIRV Parameters:** SEIRV Parameters ←{a_0, a_1, a_2, a_3, δ, γ, e_0, i_0} •**Define Vaccination Parameters:** Vaccination Parameters ←{T_0, T_1, T_2, T_3, C, H} •**Define Gradient Descent Parameters:** GD Parameters ←{η, θ, n, *stp*} 2. **Gradient Descent Loop:** •**For** (i *in* 1: n) - Solve the SEIRV differential equations and store the result in OUTPUT. - Extract the gradient of the cost function, $\partial E(\theta)/\partial(\theta)$, from the OUTPUT. - Update θ using the gradient descent update rule: $\theta = \theta - \eta \nabla J(\theta)$. - If the update to θ is less than the stopping threshold, i.e., $\eta^* \nabla J(\theta) < stp$, then stop the loop.

To estimate $\frac{\partial E(t_i,\theta)}{\partial(\theta)}$, we need to take the derivative of equations (26)–(30) with respect to θ. This process involves applying the rules of differentiation to each term in these equations. In this context, we look at how the function $E(t_i, \theta)$ changes as θ changes, while holding the time t_i constant. Thus, the derivative of the SEIRV model is given by:

$$\frac{\partial S(t,\theta)}{\partial t\, \partial \theta} = -\beta(t)\, \frac{\partial S(t,\theta)}{\partial \theta} I(t,\theta) - \beta(t)\, S(t,\theta) \frac{\partial I(t,\theta)}{\partial \theta} - v_e\, (1-\lambda) \frac{\partial v(t,\theta)}{\partial \theta} \tag{C.3}$$

$$\frac{\partial E(t,\theta)}{\partial t\, \partial \theta} = \beta(t)\, \frac{\partial S(t,\theta)}{\partial \theta} I(t,\theta) + \beta(t)\, S(t,\theta) \frac{\partial I(t,\theta)}{\partial \theta} - \delta\, \frac{\partial E(t,\theta)}{\partial \theta} \tag{C.4}$$

$$\frac{\partial I(t,\theta)}{\partial t\, \partial \theta} = \delta\, \frac{\partial E(t,\theta)}{\partial \theta} - \gamma\, \frac{\partial I(t,\theta)}{\partial \theta} \tag{C.5}$$

$$\frac{\partial R(t,\theta)}{\partial t\, \partial \theta} = \gamma\, \frac{\partial I(t,\theta)}{\partial \theta} \tag{C.6}$$

$$\frac{\partial V(t,\theta)}{\partial t\, \partial \theta} = v_e\, (1-\lambda) \frac{\partial v(t,\theta)}{\partial \theta} \tag{C.7}$$

The parameters (T_0, T_1, T_2, T_3) for the partial derivative $\frac{\partial v(t,\theta)}{\partial \theta}$, as referenced in equations (A.8) and (A.12), are selected to ensure the differentiability of the function. To achieve this, we adopt a strategy of selecting a point in the neighborhood of the original time point. For instance, T_1 is 100, and we select 100.5 as our point of interest. This slight shift into the neighborhood of the original time-point ensures that we are always working with a small, smooth segment of the function that can be approximated by a straight line — the very definition of differentiability. This approach ensures the continuity and differentiability of the function across the chosen interval, thereby enhancing the robustness and reliability of our mathematical model.

Appendix B. Additional Results

Appendix B1. Vaccination uncertainty

The uncertainty analysis is based on MAP estimates and θ resampling is shown in Figure B10.

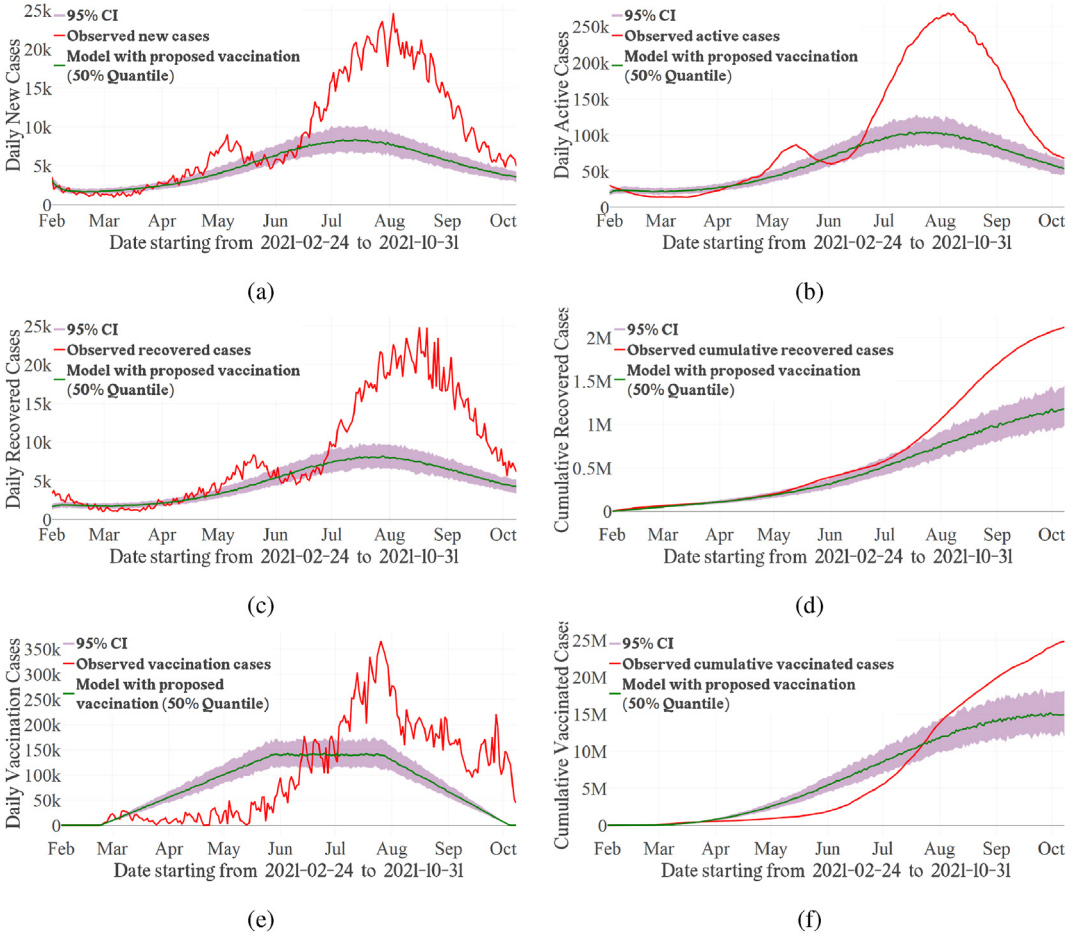

Fig. B.10. Plots based on $\hat{\Theta}_{MAP}$ with proposed vaccination rate (green line), and Malaysia's observed daily reported cases (red line). Figure panels show the following types of daily cases: (a) new cases, (b) active cases, (c) recovered cases, (d) cumulative recovered cases, (e) vaccination cases, and (f) cumulative vaccination cases.

Appendix B2. Vaccination sensitivity analysis

The results of sensitivity analysis are shown in Table B10.

Table B10
Sensitivity analysis results based on T_1^*.

T_1^*	Total Vaccination (In million)	Cost (In 10 million)	T_1^*	Total Vaccination (In million)	Cost (In 10 million)
85.5	18.71	234	116.5	20.07	259
86.5	18.75	234	117.5	20.12	260
87.5	18.79	235	118.5	20.16	261
88.5	18.84	236	119.5	20.21	262
89.5	18.88	237	120.5	20.26	263
90.5	18.92	238	121.5	20.31	264
91.5	18.96	238	122.5	20.36	264
92.5	19.00	239	123.5	20.41	265
93.5	19.04	240	124.5	20.46	266
94.5	19.09	241	125.5	20.50	267
95.5	19.13	241	126.5	20.55	268
96.5	19.17	242	127.5	20.60	269
97.5	19.21	243	128.5	20.65	270
98.5	19.26	244	129.5	20.70	271
99.5	19.30	245	130.5	20.75	272
100.5	19.34	246	131.5	20.80	273
101.5	19.39	246	132.5	20.85	274

Table B10 (*continued*)

T_1^*	Total Vaccination (In million)	Cost (In 10 million)	T_1^*	Total Vaccination (In million)	Cost (In 10 million)
102.5	19.43	247	133.5	20.91	275
103.5	19.48	248	134.5	20.96	276
104.5	19.52	249	135.5	21.01	276
105.5	19.56	250	136.5	21.06	277
106.5	19.61	251	137.5	21.11	278
107.5	19.65	251	138.5	21.16	279
108.5	19.70	252	139.5	21.22	280
109.5	19.74	253	140.5	21.27	281
110.5	19.79	254	141.5	21.32	282
111.5	19.84	255	142.5	21.38	283
112.5	19.88	256	143.5	21.43	284
113.5	19.93	257	144.5	21.48	285
114.5	19.98	257	145.5	21.54	286
115.5(T_1)	20.02	258			

Appendix D3. Model application on other regions

To verify the robustness and adaptability of our SEIRV model, we expanded our analysis beyond Malaysia's national scale and focused on two distinct regions: Selangor and Johor. According to the Department of Statistics, Malaysia (2022), Selangor is the most populous state on Malaysia's west coast, around the capital Kuala Lumpur, whereas Johor is a state in the south and the second most populous state in Malaysia.

We not only assess the model's accuracy but also demonstrate its applicability across a wide range of demographic settings. Our model's adaptability is notable for its ability to accommodate various forms of the disease transmission parameter, $\beta(t)$. Our model for Selangor used an exponential form with a trinomial exponent, whereas Johor used a quadratic exponent. These specific forms were not chosen randomly, but rather as a result of precise exploratory data analysis (EDA), emphasizing the model's ability to fine-tune its parameters to closely match regional data variations. This parameterization flexibility, particularly for the transmission rate, demonstrates the model's versatility and potential for broader applicability, providing a tailored analysis that corresponds to the unique epidemiological dynamics of each region.

The approach adopted is similar to the one explained in 2.7: we first utilized EDA to extract information about priors and then used the Monte Carlo importance sampling approach with a total of $M = 50,000$ simulations to obtain $\hat{\Theta}_{MAP}$. The MAP estimates for Selangor and Johor are presented in Table B11.

Table B11
MAP estimated parameter results for Selangor and Johor daily reported cases.

Parameter	Selangor	Johor
a_0	-2.859259	-3.090414
a_1	0.01305965	0.01240119
a_2	$-5.49493 \times e^{-5}$	$-3.829866 \times e^{-5}$
a_3	$1.93258 \times e^{-8}$	–
τ	0.009964003	0.09990511
$1/\delta$	4.994595	5.103786
$1/\gamma$	12.19794	13.10336
i_0	9441.642	5840.989
e_0	5248.725	227.1173

The trajectories of these MAP estimates and associated uncertainties are shown in Figures B11 and B.12 for Selangor and Johor, respectively. The MAP estimates indicate that Selangor has a higher transmission rate ($\beta(t)$) than Johor, possibly due to its larger population and closeness to the capital. Furthermore, the model suggests that the incubation period in Selangor is close to the national average of 4.68 days, at 4.99 days, whereas Johor has a slightly longer incubation period of around 5.10 days. Furthermore, the duration of infectiousness or recovery time for individuals in both states ranges from 12 to 13 days. It also indicates a longer recovery period for Johor residents compared to those in Selangor.

The outcomes of implementing a hypothetical vaccination policy in these regions were consistent with the promising results observed in national data. Our findings, as shown in figures B13 and B.14, highlight a significant decrease in daily reported cases following the implementation of our model's vaccination strategy. This strategy suggests vaccinating 65% of the population in Selangor, which is significantly lower than the observed 68% vaccination coverage. In Johor, the model recommends vaccinating 62% of the population, which is significantly lower than the observed 78%. This approach, based on the unique structure of our loss function, has been carefully constructed to find an optimal balance, prioritizing both a reduction in total vaccinations and minimizing disease-related costs.

The observed vaccination trajectories in both regions show different initial rates, which eventually result in higher overall vaccination rates than our model predicts. Our model suggests a more efficient approach, advocating for a strong initial vaccination effort followed by a consistent and steady pace, eliminating the need for rapid increases in vaccination rates. The model predicts that as the vaccination initiative progresses and disease spread is controlled, the necessary vaccination rate will decrease.

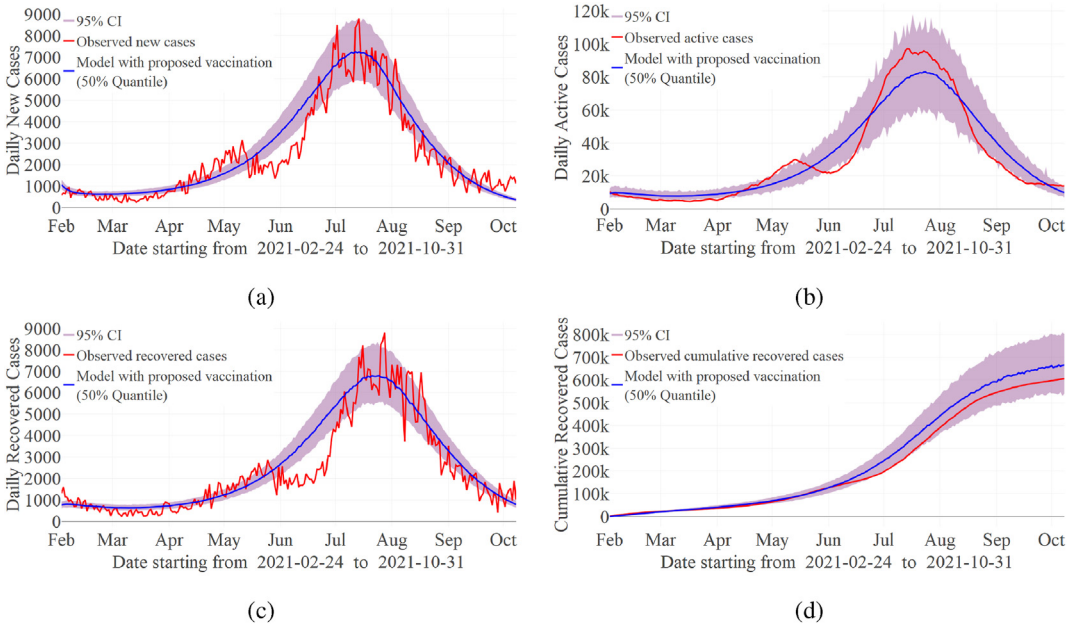

Fig. B11. MAP and variability estimate of daily cases in the SEIRV model: Selangor's observed daily reported cases (red line), estimated daily cases based on MAP (blue line) and posterior prediction bands of daily cases (purple shaded). Figure panels show the following types of daily cases: (a) new cases, (b) active cases, (c) daily recovered cases and (d) cumulative recovered cases.

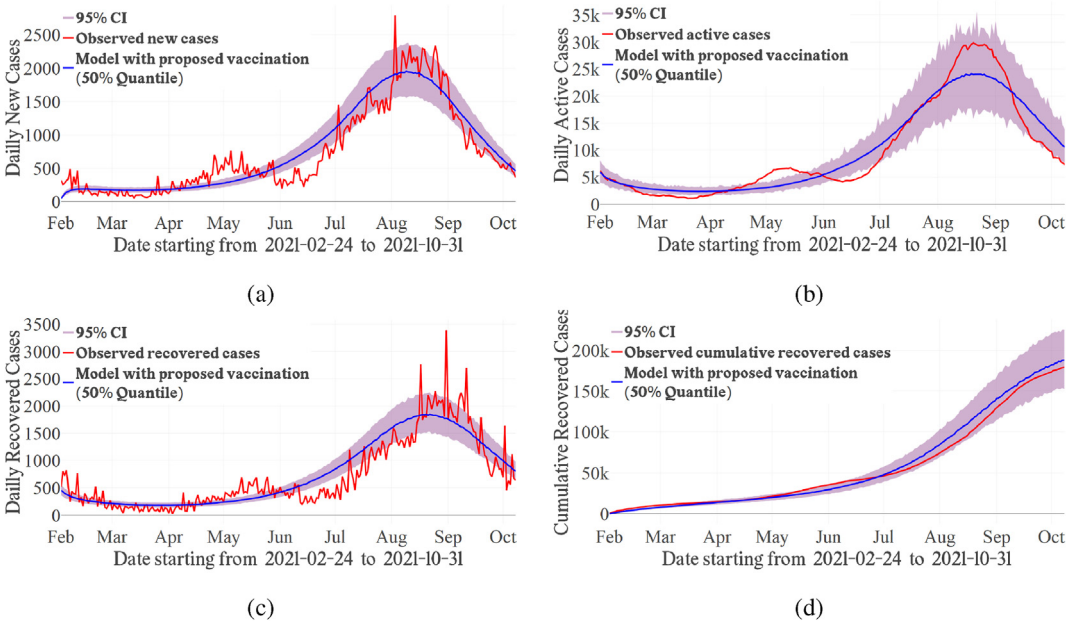

Fig. B12. MAP and variability estimate of daily cases in the SEIRV model: Johor's observed daily reported cases (red line), estimated daily cases based on MAP (blue line) and posterior prediction bands of daily cases (purple shaded). Figure panels show the following types of daily cases: (a) new cases, (b) active cases, (c) daily recovered cases and (d) cumulative recovered cases.

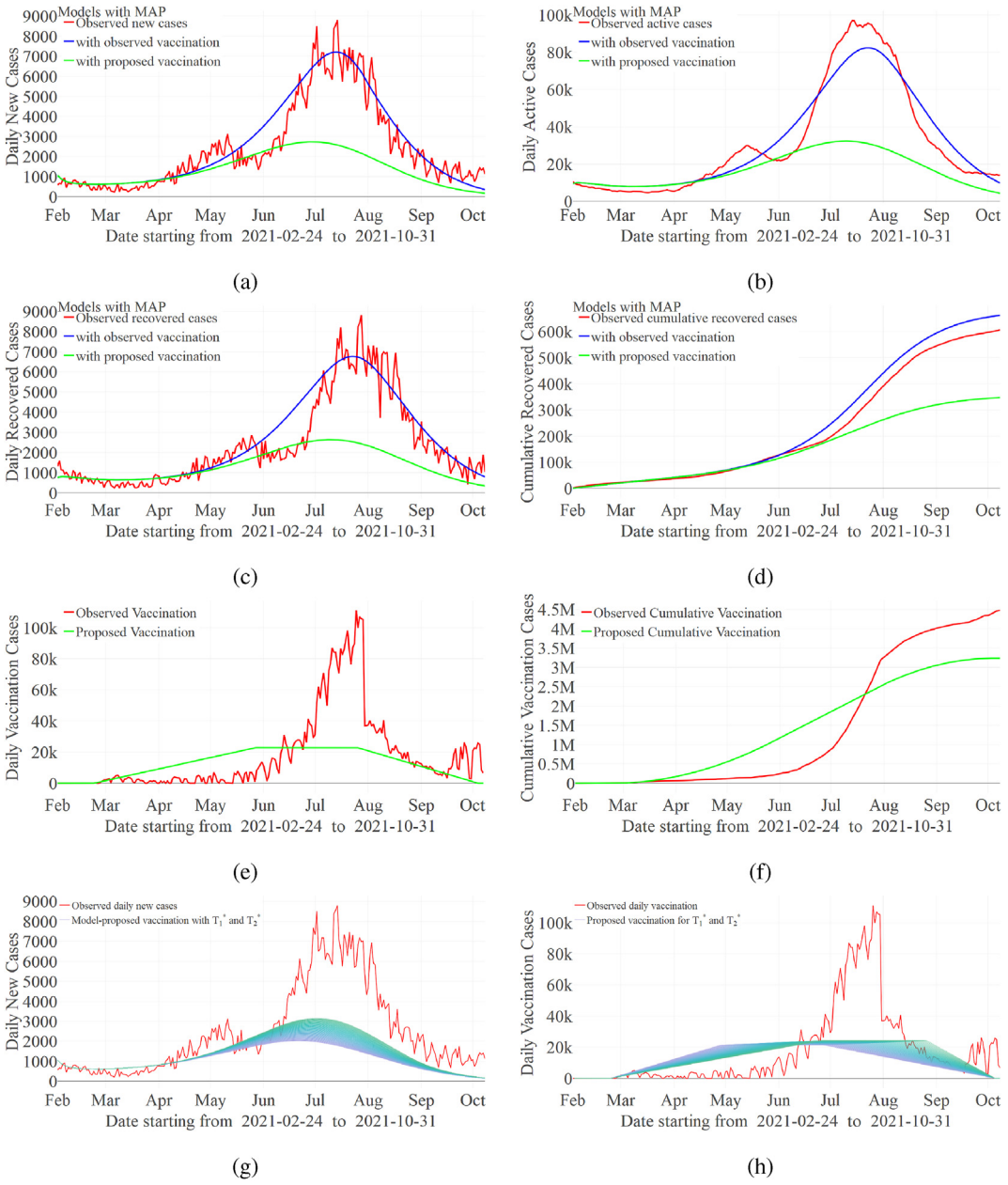

Fig. B13. Plots (a)–(f) based on $\hat{\Theta}_{MAP}$ for the observed vaccination (blue line), proposed vaccination rate with $\theta = 10.32175$ (green line), and Selangor's observed daily reported cases (red line). Figure panels show the following types of daily cases: (a) New cases, (b) Active cases, (c) Daily recovered cases, (d) Cumulative recovered cases, (e) Daily vaccinated cases, and (f) Cumulative vaccinated cases. Plots (g) and (h) are based on $\hat{\Theta}_{MAP}$, hypothetical vaccination policy and different time points for T_1 and T_2 (colored lines), and Selangor's observed daily reported cases (red line). Figure panels show the following types of daily cases: (g) new cases and (h) daily vaccinated cases.

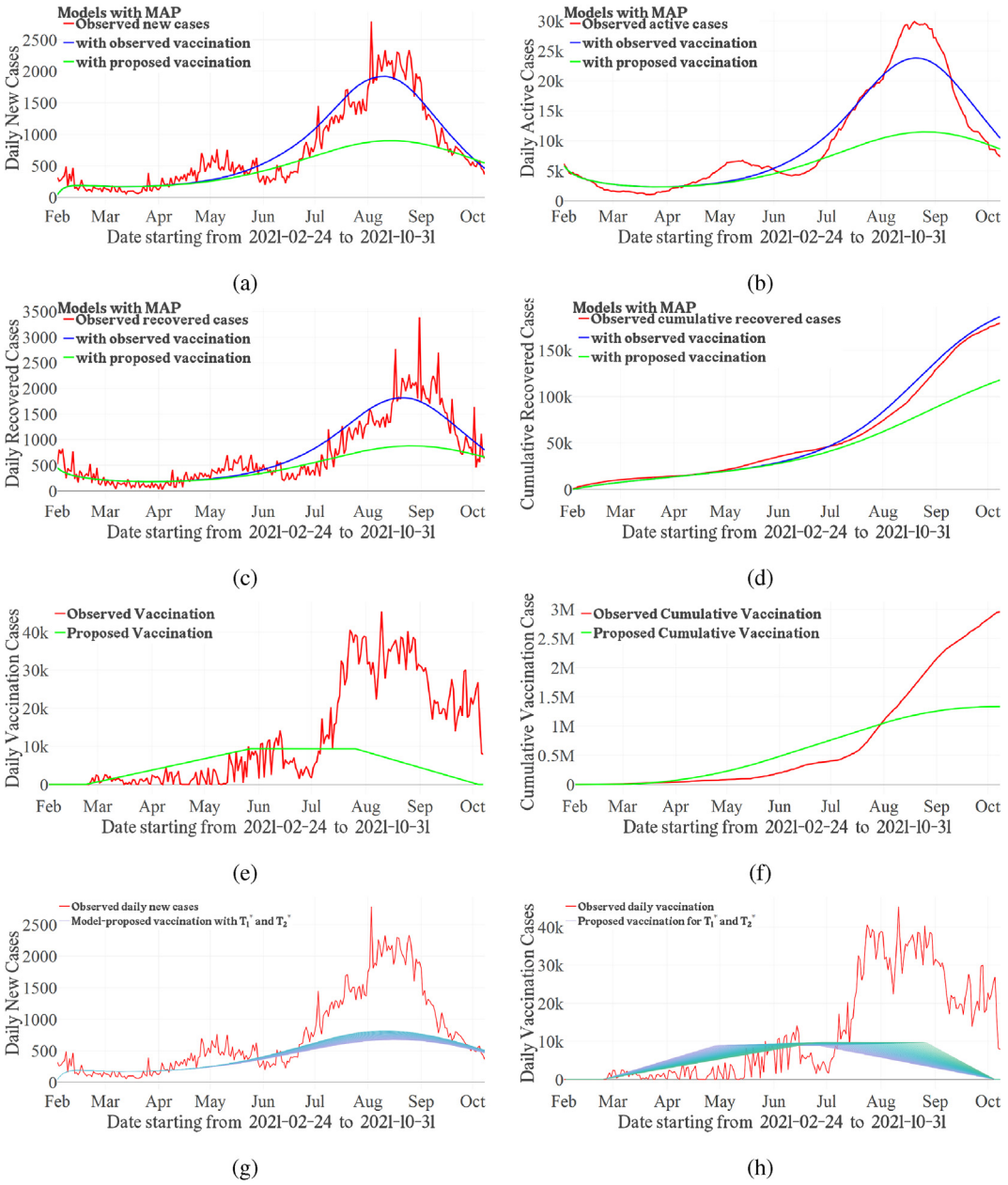

Fig. B14. Plots (a)–(f) based on $\hat{\Theta}_{MAP}$ for the observed vaccination (blue line), proposed vaccination rate with $\theta = 9.428116$ (green line), and Johor's observed daily reported cases (red line). Figure panels show the following types of daily cases: (a) New cases, (b) Active cases, (c) Daily recovered cases, (d) Cumulative recovered cases, (e) Daily vaccinated cases, and (f) Cumulative vaccinated cases. Plots (g) and (h) are based on $\hat{\Theta}_{MAP}$, hypothetical vaccination policy and different time points for T_1 and T_2 (colored lines), and Johor's observed daily reported cases (red line). Figure panels show the following types of daily cases: (g) new cases and (h) daily vaccinated cases.

References

Amin, A. M., & Bn, C. Z. L. (2023). Humoral response to SARS-CoV-2 vaccines among healthcare workers in a tertiary hospital in Malaysia. *Medical Journal of Malaysia, 78*(1), 21.

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical models in epidemiology* (Vol. 32). Springer.

Chatfield, C., & Xing, H. (2019). *The analysis of time series: An introduction with R*. CRC Press.

Chen, B., Zhao, Y., Jin, Z., He, D., & Li, H. (2023). Twice evasions of Omicron variants explain the temporal patterns in six Asian and Oceanic countries. *BMC Infectious Diseases, 23*(1), 1–12.

Dass, S. C., Kwok, W. M., Gibson, G. J., Gill, B. S., Sundram, B. M., & Singh, S. (2021). A data driven change-point epidemic model for assessing the impact of large gathering and subsequent movement control order on COVID-19 spread in Malaysia. *PLoS One, 16*(5), Article e0252136.

Deng, L., Li, P., Zhang, X., Jiang, Q., Turner, D., Zhou, C., et al. (2022). Risk of SARS-CoV-2 reinfection: A systematic review and meta-analysis. *Scientific Reports, 12*(1), Article 20763.

El Khalifi, M., & Britton, T. (2023). Extending susceptible-infectious-recovered-susceptible epidemics to allow for gradual waning of immunity. *Journal of The Royal Society Interface, 20*(206), Article 20230042.
Estrada, E. (2020). COVID-19 and SARS-CoV-2. modeling the present, looking at the future. *Physics Reports, 869*, 1–51.
Fei, Y., Xu, H., Zhang, X., Musa, S. S., Zhao, S., & He, D. (2022). Seroprevalence and infection attack rate of COVID-19 in Indian cities. *Infectious Disease Modelling, 7*(2), 25–32.
Feng, S., Luo, X.-F., Pei, X., Jin, Z., Lewis, M., & Wang, H. (2022). Modeling the early transmission of covid-19 in new york and san francisco using a pairwise network model. *Infectious Disease Modelling, 7*(1), 212–230. https://doi.org/10.1016/j.idm.2021.12.009. https://www.sciencedirect.com/science/article/pii/S2468042721000907
Feng, A., Obolski, U., Stone, L., & He, D. (2022). Modelling COVID-19 vaccine breakthrough infections in highly vaccinated Israel—the effects of waning immunity and third vaccination dose. *PLOS global public health, 2*(11), Article e0001211.
Ferdinands, J. M., Rao, S., Dixon, B. E., Mitchell, P. K., DeSilva, M. B., Irving, S. A., et al. (2022). Waning of vaccine effectiveness against moderate and severe COVID-19 among adults in the US from the vision network: Test negative, case-control study. *Bmj, 379*.
Fisher, A., Xu, H., He, D., & Wang, X. (2023). Effects of vaccination on mitigating COVID-19 outbreaks: A conceptual modeling approach. *Mathematical Biosciences and Engineering, 20*(3), 4816–4837.
Fu, Y., Pedrini, G., & Osten, W. (2007). Vibration measurement by temporal fourier analyses of a digital hologram sequence. *Applied Optics, 46*(23), 5719–5727.
Ge, Y., Zhang, W.-B., Wu, X., Ruktanonchai, C. W., Liu, H., Wang, J., et al. (2022). Untangling the changing impact of non-pharmaceutical interventions and vaccination on European COVID-19 trajectories. *Nature Communications, 13*(1), 3106.
Ghazy, R. M., Ashmawy, R., Hamdy, N. A., Elhadi, Y. A. M., Reyad, O. A., Elmalawany, D., et al. (2022). Efficacy and effectiveness of SARS-CoV-2 vaccines: A systematic review and meta-analysis. *Vaccines, 10*(3), 350.
Giordano, G., Colaneri, M., Di Filippo, A., Blanchini, F., Bolzern, P., De Nicolao, G., et al. (2021). Modeling vaccination rollouts, SARS-CoV-2 variants and the requirement for non-pharmaceutical interventions in Italy. *Nature Medicine, 27*(6), 993–998.
Han, X., Liu, H., Lin, X., Wei, Y., Ming, M., & Zhang, Y. Z. (2022). Dynamic analysis of a VSEIR model with vaccination efficacy and immune decline. *Advances in Mathematical Physics, 2022*, 1–16.
He, D., Chen, B., Zhao, S., & Stone, L. (2023). The immune evasion ability of Delta variant is comparable to that of Beta variant in South Africa. *BMC Public Health, 23*(1), 1–4.
He, D., Cowling, B. J., Ali, S. T., & Stone, L. (2023). Rapid global spread of variants of concern of SARS-CoV-2. *IJID regions, 7*, 63–65.
He, D., Lin, L., Artzy-Randrup, Y., Demirhan, H., Cowling, B. J., & Stone, L. (2023). Resolving the enigma of iquitos and manaus: A modeling analysis of multiple COVID-19 epidemic waves in two amazonian cities. *Proceedings of the National Academy of Sciences, 120*(10), Article e2211422120.
Husin, M., Tok, P. S. K., Suah, J. L., Thevananthan, T., Tng, B. H., Peariasamy, K. M., et al. (2022). Real-world effectiveness of BNT162b2 vaccine against SARS-CoV-2 infection among adolescents (12 to 17-year-olds) in Malaysia. *International Journal of Infectious Diseases, 121*, 55–57.
Jara, A., Undurraga, E. A., González, C., Paredes, F., Fontecilla, T., Jara, G., et al. (2021). Effectiveness of an inactivated SARS-CoV-2 vaccine in Chile. *New England Journal of Medicine, 385*(10), 875–884.
Jin, L., Li, Z., Zhang, X., Li, J., & Zhu, F. (2022). CoronaVac: A review of efficacy, safety, and immunogenicity of the inactivated vaccine against SARS-CoV-2. *Human Vaccines & Immunotherapeutics, 18*(6), Article 2096970.
Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of methods for smoothing and gap filling time series of remote sensing observations–application to MODIS LAI products. *Biogeosciences, 10*(6), 4055–4071.
Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine, 172*(9), 577–582.
Levin, E. G., Lustig, Y., Cohen, C., Fluss, R., Indenbaum, V., Amit, S., et al. (2021). Waning immune humoral response to BNT162b2 COVID-19 vaccine over 6 months. *New England Journal of Medicine, 385*(24), e84.
Liao, Z., Lan, P., Fan, X., Kelly, B., Innes, A., Liao, Z., et al. (2021). A covid-19 deep learning prediction model based on time-dependent sirvd. *Computers in Biology and Medicine, 138*, Article 104868. https://doi.org/10.1016/j.compbiomed.2021.104868. https://www.sciencedirect.com/science/article/pii/S0010482521006624
Lloyd-Smith, J. O. (2007). Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. *PLoS One, 2*(2), Article e180.
Lodz, N. A., Mat Tamizi, N. F., Abd Mutalip, M. H., Ganapathy, S. S., Lin, C. Z., Ismail, R., et al. (2024). Anti-S antibodies against SARS-CoV-2 infection among four types of vaccines in Malaysia, Asia Pacific. *Journal of Public Health*, Article 10105395231223332.
MacIntyre, C. R., Costantino, V., & Trent, M. (2022). Modelling of COVID-19 vaccination strategies and herd immunity, in scenarios of limited and full vaccine supply in NSW, Australia. *Vaccine, 40*(17), 2506–2513.
Mahmud, M. S., Kamrujjaman, M., Adan, M. M. I. Y., Hossain, M. A., Rahman, M. M., Islam, M. S., et al. (2022). Vaccine efficacy and sars-cov-2 control in California and u.s. during the session 2020–2026: A modeling study. *Infectious Disease Modelling, 7*(1), 62–81. https://doi.org/10.1016/j.idm.2021.11.002. https://www.sciencedirect.com/science/article/pii/S2468042721000701
Mahmud, A., & Lim, P. Y. (2020). *Applying the seir model in forecasting the COVID-19 trend in Malaysia: A preliminary study*. MedRxiv, 2020–04.
Menegale, F., Manica, M., Zardini, A., Guzzetta, G., Marziano, V., d'Andrea, V., et al. (2023). Evaluation of waning of SARS-CoV-2 vaccine–induced immunity: A systematic review and meta-analysis. *JAMA Network Open, 6*(5), Article e2310650.
Menni, C., May, A., Polidori, L., Louca, P., Wolf, J., Capdevila, J., et al. (2022). COVID-19 vaccine waning and effectiveness and side-effects of boosters: A prospective community study from the ZOE coCOVIDvid study. *The Lancet Infectious Diseases, 22*(7), 1002–1010.
Moore, S., Hill, E. M., Tildesley, M. J., Dyson, L., & Keeling, M. J. (2021). *Vaccination and non-pharmaceutical interventions: When can the UK relax about COVID-19?* Preprint: MedRxiv, 2020–12.
Notarte, K. I., Guerrero-Arguero, I., Velasco, J. V., Ver, A. T., Santos de Oliveira, M. H., Catahay, J. A., et al. (2022). Characterization of the significant decline in humoral immune response six months post-SARS-CoV-2 mrna vaccination: A systematic review. *Journal of Medical Virology, 94*(7), 2939–2961.
Nuraini, N., Sukandar, K., Hadisoemarto, P., Susanto, H., Hasan, A., & Sumarti, N. (2021). Mathematical models for assessing vaccination scenarios in several provinces in Indonesia. *Infectious Disease Modelling, 6*, 1236–1258.
Olivera Mesa, D., Hogan, A. B., Watson, O. J., Charles, G. D., Hauck, K., Ghani, A. C., et al. (2022). Modelling the impact of vaccine hesitancy in prolonging the need for non-pharmaceutical interventions to control the COVID-19 pandemic. *Communication and Medicine, 2*(1), 14.
Parolini, N., Dede, L., Ardenghi, G., & Quarteroni, A. (2022). Modelling the covid-19 epidemic and the vaccination campaign in Italy by the suihter model. *Infectious Disease Modelling, 7*(2), 45–63. https://doi.org/10.1016/j.idm.2022.03.002. https://www.sciencedirect.com/science/article/pii/S2468042722000100
Pecoraro, V., Pirotti, T., & Trenti, T. (2022). Evidence of SARS-CoV-2 reinfection: Analysis of 35,000 subjects and overview of systematic reviews. *Clinical and Experimental Medicine*, 1–12.
Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2020). Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *New England Journal of Medicine, 383*(27), 2603–2615.
Saeedi, G., & Waseel, F. (2022). *Existence of solutions for a class of quasilinear elliptic equations involving the p-Laplacian, Complex Variables and Elliptic Equations*.
Saeedi, G., Waseel, F., & Jalal, H. (2022). Existence of solutions for a class of quasilinear elliptic equations with a parameter. *Communications on Pure and Applied Analysis, 21*(9).
Schwartz, J. (2001). Is there harvesting in the association of airborne particles with daily deaths and hospital admissions? *Epidemiology, 12*(1), 55–61.

Schwartz, I. B., Kaufman, J., Hu, K., & Bianco, S. (2020). *Predicting the impact of asymptomatic transmission, non-pharmaceutical intervention and testing on the spread of COVID-19*. Preprint: MedRxiv, 2020–04.
Ssentongo, P., Ssentongo, A. E., Voleti, N., Groff, D., Sun, A., Ba, D. M., et al. (2022). SARS-CoV-2 vaccine effectiveness against infection, symptomatic and severe COVID-19: A systematic review and meta-analysis. *BMC Infectious Diseases, 22*(1), 1–12.
Stein, C., Nassereldine, H., Sorensen, R. J., Amlag, J. O., Bisignano, C., Byrne, S., et al. (2023). Past SARS-CoV-2 infection protection against re-infection: A systematic review and meta-analysis. *The Lancet, 401*(10379), 833–842.
Stokes, J., Turner, A. J., Anselmi, L., Morciano, M., & Hone, T. (2022). The relative effects of non-pharmaceutical interventions on wave one COVID-19 mortality: Natural experiment in 130 countries. *BMC Public Health, 22*(1), 1113.
Suah, J. L., Husin, M., Tok, P. S. K., Tng, B. H., Thevananthan, T., Low, E. V., et al. (2022). Waning COVID-19 vaccine effectiveness for BNT162b2 and CoronaVac in Malaysia: An observational study. *International Journal of Infectious Diseases, 119*, 69–76.
Suah, J. L., Tok, P. S. K., Ong, S. M., Husin, M., Tng, B. H., Sivasampu, S., et al. (2021). PICK-Ing Malaysia's epidemic apart: Effectiveness of a diverse COVID-19 vaccine portfolio. *Vaccines, 9*(12), 1381.
Suleyman, G., Fadel, R., Patel, K., Shadid, A. M., Stuart, H. B. C., Kattula, M., et al. (2023). Outcomes associated with SARS-CoV-2 reinfection in individuals with natural and hybrid immunity. *Journal of Infection and Public Health*.
Tanriover, M. D., Doğanay, H. L., Akova, M., Güner, H. R., Azap, A., Akhan, S., et al. (2021). Efficacy and safety of an inactivated whole-virion SARS-CoV-2 vaccine (CoronaVac): Interim results of a double-blind, randomised, placebo-controlled, phase 3 trial in Turkey. *The Lancet, 398*(10296), 213–222.
Thompson, M. G., Burgess, J. L., Naleway, A. L., Tyner, H., Yoon, S. K., Meece, J., et al. (2021). Prevention and attenuation of COVID-19 with the BNT162b2 and mRNA-1273 vaccines. *New England Journal of Medicine, 385*(4), 320–329.
Tian, F., Yang, R., & Chen, Z. (2022). Safety and efficacy of COVID-19 vaccines in children and adolescents: A systematic review of randomized controlled trials. *Journal of Medical Virology, 94*(10), 4644–4653.
Tokdar, S. T., & Kass, R. E. (2010). Importance sampling: A review. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(1), 54–60.
Townsend, J. P., Hassler, H. B., Sah, P., Galvani, A. P., & Dornburg, A. (2022). The durability of natural infection and vaccine-induced immunity against future infection by SARS-CoV-2. *Proceedings of the National Academy of Sciences, 119*(31), Article e2204336119.
Tran, T. N.-A., Wikle, N. B., Albert, E., Inam, H., Strong, E., Brinda, K., et al. (2021). Optimal SARS-CoV-2 vaccine allocation using real-time attack-rate estimates in Rhode Island and Massachusetts. *BMC Medicine, 19*, 1–14.
Wagner, C. E., Saad-Roy, C. M., & Grenfell, B. T. (2022). Modelling vaccination strategies for COVID-19. *Nature Reviews Immunology, 22*(3), 139–141.
Waseel, F., & Yusof, F. (2019). Attitudes and challenges toward the use of ICT skills: An assessment of ICT skills among university lecturers and students in Afghanistan. *Online Submission, 8*, 314–321.
Watson, G. L., Xiong, D., Zhang, L., Zoller, J. A., Shamshoian, J., Sundin, P., et al. (2020). *Fusing a Bayesian case velocity model with random forest for predicting COVID-19 in the US*. medRxiv, 2020–05.
Webb, G. (2021a). A covid-19 epidemic model predicting the effectiveness of vaccination in the us. *Infectious Disease Reports, 13*(3), 654–667. https://doi.org/10.3390/idr13030062. https://www.mdpi.com/2036-7449/13/3/62
Webb, G. (2021b). A COVID-19 epidemic model predicting the effectiveness of vaccination. *Mathematics in Applied Sciences and Engineering, 2*(2), 134–148.
Webb, G. (2021c). A COVID-19 epidemic model predicting the effectiveness of vaccination in the US. *Infectious Disease Reports, 13*(3), 654–667.
Wen, C., Wei, J., Ma, Z. F., He, M., Zhao, S., Ji, J., et al. (2022). Heterogeneous epidemic modelling within an enclosed space and corresponding Bayesian estimation. *Infectious Disease Modelling, 7*(2), 1–24.
Yamana, T. K., Galanti, M., Pei, S., Di Fusco, M., Angulo, F. J., Moran, M. M., et al. (2023). The impact of COVID-19 vaccination in the US: Averted burden of SARS-CoV-2-related cases, hospitalizations and deaths. *PLoS One, 18*(4), Article e0275699.
Yu, B., Li, Q., Chen, J., & He, D. (2023). The impact of COVID-19 vaccination campaign in Hong Kong SAR China and Singapore. *Infectious Disease Modelling, 8*(1), 101–106.
Zamri, A. S. S. M., Singh, S., Ghazali, S. M., Herng, L. C., Dass, S. C., Aris, T., et al. (2021). Effectiveness of the movement control measures during the third wave of COVID-19 in Malaysia. *Epidemiology and Health, 43*.
Zeng, B., Gao, L., Zhou, Q., Yu, K., & Sun, F. (2022). Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern: A systematic review and meta-analysis. *BMC Medicine, 20*(1), 1–15.
Zeng, T., Lu, Y., Zhao, Y., Guo, Z., Sun, S., Teng, Z., et al. (2023). Effectiveness of the booster dose of inactivated COVID-19 vaccine against Omicron BA. 5 infection: A matched cohort study of adult close contacts. *Respiratory Research, 24*(1), 246.
Zheng, C., Shao, W., Chen, X., Zhang, B., Wang, G., & Zhang, W. (2022). Real-world effectiveness of COVID-19 vaccines: A literature review and meta-analysis. *International Journal of Infectious Diseases, 114*, 252–260.
Zhou, W., Tang, B., Bai, Y., Shao, Y., Xiao, Y., & Tang, S. (2022). The resurgence risk of COVID-19 in China in the presence of immunity waning and ade: A mathematical modelling study. *Vaccine, 40*(49), 7141–7150.
Zuo, C., Zhu, F., Meng, Z., Ling, Y., Zheng, Y., & Zhao, X. (2022). Analyzing the covid-19 vaccination behavior based on epidemic model with awareness-information. *Infection, Genetics and Evolution, 98*, Article 105218. https://doi.org/10.1016/j.meegid.2022.105218. https://www.sciencedirect.com/science/article/pii/S1567134822000156