Valuation and comparison of the actual and optimal control strategy in an emerging infectious disease: Implication from a COVID-19 transmission model

Lili Liu [a], Xi Wang [a, b], Ou Liu [b], Yazhi Li [c], Zhen Jin [a], Sanyi Tang [d], Xia Wang [d, *]

[a] Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Complex Systems Research Center, Shanxi University, Taiyuan, 030006, China
[b] School of Mathematical Sciences, Shanxi University, Taiyuan, 030006, China
[c] School of Mathematics and Statistics, Qiannan Normal University for Nationalities, Guizhou, Duyun, 558000, China
[d] School of Mathematics and Statistics, Shaanxi Normal University, Xi'an, 710119, China

ARTICLE INFO

ABSTRACT

To effectively combat emerging infectious diseases like COVID-19, it is crucial to adopt strict prevention and control measures promptly to effectively contain the spread of the epidemic. In this paper, we propose a transmission model to investigate the influence of two control strategies: reducing contact numbers and improving medical resources. We examine these strategies in terms of constant control and time-varying control. Through sensitivity analysis on two reproduction numbers of the model with constant control, we demonstrate that reducing contact numbers is more effective than improving medical resources. Furthermore, these two constant controls significantly influence the peak values and timing of infections. Specifically, intensifying control measures can reduce peak values, albeit at the expense of delaying the peak time. In the model with time-varying control, we initially explore the corresponding optimal control problem and derive the characteristic expression of optimal control. Subsequently, we utilize real data from January 10th to April 12th, 2020, in Wuhan city as a case study to perform parameter estimation by using our proposed improved algorithm. Our findings illustrate that implementing optimal control measures can effectively reduce infections and deaths, and shorten the duration of the epidemic. Then, we numerically explore that implementing control measures promptly and increasing intensity to reduce contact numbers can make actual control be more closer to optimized control. Finally, we utilize the real data from October 31st to November 18th, 2021, in Hebei province as a second case study to validate the feasibility of our proposed suggestions.

© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Emerging infectious diseases are characterized by their occurrence in the human population for the first time or rapidly increasing incidence (Jones et al., 2008). There have been 335 reported emergences, including multi-drug-resistant

---

tuberculosis and chloroquine-resistant malaria caused by newly evolved pathogen strains, as well as SARS, COVID-19, Lyme disease, and Ebola, caused by novel or historically existing pathogens (Jones et al., 2008). The continuous emergence and spread of these diseases impose a significant burden on global economies and public health. Therefore, it is imperative to develop suitable theories to deepen our understanding of transmission mechanism and devise effective control strategies to mitigate their prevalence.

When faced with an emerging infectious disease, a critical question for decision-makers is how to formulate and implement effective control measures to contain its prevalence. Early adoption of stringent prevention and control measures is typically crucial in curbing the epidemic's spread. However, due to the lack of effective treatments or vaccines against emerging infectious diseases, combined with their high infectiousness, there is often a shortage of medical resources, potentially leading to their collapse. Consequently, incomplete isolation of infected individuals or delayed medical assistance leads the disease to continue spreading in the society, hindering efforts to control its widespread prevalence. The shortage of medical resources, particularly hospital beds, has been described by Zhou et al. using a piecewise smooth model applied to the actual problem of Wuhan, where it was demonstrated that the rapid construction of emergency hospitals prevented 22,786 infections and saved 6524 lives (Zhou et al., 2020). Wang et al. utilized multiple data sources and cross-validation of a COVID-19 epidemic model to assess the impact of improved medical resources in several countries, revealing that a 50% reduction in resources would result in over 590,000 confirmed cases and 60,000 deaths in mainland China by March 27, 2020 (Wang et al., 2021). Further research on medical resources can be found in references (Gupta et al., 2021; KhudaBukhsh et al., 2023; Shreffler et al., 2020; Sun et al., 2020; Wang et al., 2021) and related references. Notably, the aforementioned studies employed piecewise smooth models to describe the shortage of medical resources. However, understanding the dynamic behavior of these models using related theories can be challenging. Thus, if medical resources are described using a smooth function, it is of interest to investigate the resulting model changes and their implications for actual control strategies.

Mathematical epidemiology plays a crucial role in describing and modeling the transmission dynamics of epidemics, providing valuable guidance for policymakers, a fact acknowledged by the World Health Organization (WHO) (Egger et al., 2017). COVID-19, as a classical example of an emerging infectious disease, has garnered global attention, and the numerous studies employ mathematical models to investigate various control strategies (Hao et al., 2020; Maier & Brockmann, 2020; Taboe et al., 2023; Tang, Wang, et al., 2020; Wu et al., 2020). Wu et al. formulated an SEIR-type model to estimate the infection size in Wuhan city and forecast the global spread of COVID-19 (Wu et al., 2020). Tang et al. developed a mathematical model to estimate the transmission risk of Wuhan city in 2020 year, revealing a high disease's transmissibility and offering insights for public health interventions (Tang, Wang, et al., 2020). Aforementioned research plays a significant role in guiding the prevention and control of the epidemic's spread.

To better control the epidemic's situation, policymakers and public health authorities need to comprehensively consider how to carry out optimized control strategies to achieve the best control effects. The most popular way to model epidemic control strategies with optimal control is the so-called Lenhart's approach (Lenhart & Workman, 2007), which provides the objective function(al) and constraint conditions. Usually, the goal of optimal control is to minimize the infections or deaths with the lowest intervention cost, where the widely used theory for ordinary differential equations is Pontryagins Maximum Principle (Pontryagin, 1987). Manuel et al. formulated an optimal control problem with mixed constraints to compare the different vaccination schedules and answered the importance of how and when to vaccinate to reduce the infection size. They showed that it is extremely important for the response regarding vaccine-induced immunity and reinfection periods in alleviating the epidemic (Acuña-Zegarra et al., 2021). Another control theory is the bang-bang control, that is, control measures are switched, either off or maximal, at a given constant level in time (Hansen & Day, 2011). Plank used a highly idealized branching process model and bang-bang control theory to address the optimal way to maintain an elimination state. The results showed that the optimal threshold of introducing controls was negatively correlated with the effective reproduction number and positively correlated with overdispersion of the offspring distribution and the effectiveness of control measures (Plank, 2022). It should be noted that bang-bang control concerns the optimal trigger condition of introducing control to eliminate the break, while Lenhart's optimal approach focuses on the optimal way to reach a herd immunity state. Based on the transmission characteristics of emerging infectious diseases, when it comes, policymakers are more concerned about how to carry out actual control to achieve optimal effects. That is, here, we are concerned with the optimal way in practice. In addition, since the insufficient experience in coping with emerging infectious diseases, there usually exists a significant gap between actual control and optimal control (Penn & Donnelly, 2023). Therefore, this paper will utilize Lenhart's approach to investigate methods of reducing the disparity between actual and optimal control.

This paper's main contributions can be summarized in three aspects. Firstly, we propose a new function to characterize the shortage of medical resources, which can cover bilinear, saturated incidence, and B-D functional response forms. The bilinear incidence function signifies ample medical resources, the saturated incidence function represents an upper bound of medical resources, while the B-D function reflects the interaction between symptomatic and hospitalized infections. Secondly, we propose a improved algorithm to deal with the parameter estimation. The traditional estimation method is the least squares, Markov chain Monte Carlo et al., which typically require strict initial value selection. Hence, the improved algorithm will strive to solve this difficulty. Finally, we focus on how to minimize the gap between actual and optimal control. Specifically, we investigate how constant control measures influence peak infection values and timing, characterize optimal control for time-varying measures, and compare actual and optimal control strategies. The aim of this paper is to determine if actual control can be adjusted and how to optimize it in practice to approach the optimal control level.

The paper is structured as follows: In Section 2, we provide an overview of the methods used, including the formulation of two basic models with constant and time-varying control measures, respectively, as well as details on the data and the method for parameter estimation employed. For the model with constant control, we derive the control reproduction number and perform a corresponding sensitivity analysis. For the model with time-varying control, we analyze optimal control and introduce two actual control indexes. Section 3 presents the results, including the estimation results for two cases, a discussion on sensitivity analysis and the effects of constant control measures, an illustration of the effects of time-varying controls, and an examination of how actual control approaches the optimal control level. Finally, the paper concludes with a discussion and a summary of the findings.

2. Methods

2.1. Model with constant control

Based on the basic SEIR-type compartmental structure, we formulate a mathematical transmission model describing the early stage of COVID-19 transmission, which incorporates the two infection statuses (asymptomatic and symptomatic) and further consider two constant control measures (reducing contact numbers and improving medical resources).

2.1.1. Model formulation

The total population (N) is divided into several epidemiological statuses, namely, susceptible (S), exposed (E), asymptomatic (A), symptomatic (I), hospitalized (H), and recovered (R). Given the brevity of the early stage of natural transmission, we do not take into account the birth or immigration of susceptible classes, nor the natural death rate of all classes. It is assumed that the susceptible individuals can be infected by not only the symptomatic infections at a rate β, but also the asymptotic infections at a rate $\beta\sigma$, where σ denotes the modification factor of the transmissibility of asymptomatic infections. After infection, it takes a duration of the exposed period, $1/\alpha$, to become infectious, where infected individuals with a proportion p enter into the asymptomatic infected class and with a proportion $1 - p$ become symptomatic infections. Both asymptomatic and symptomatic infections can recover at rates r_A and r_I, respectively. Due to the assumption that asymptomatic infections own stronger immunity, we omit the disease-induced death rate of asymptomatic infections and only consider the disease-induced death of symptomatic infections with the rate d_I. Along with disease progress, asymptomatic infections can convert to symptomatic infections at a rate δ. The symptomatic infections need to be diagnosed and treated in the hospital at a baseline rate of m. To describe the effect of medical resources' shortage on the transmission of COVID-19, we take a new type function $f(I, H) = mI/(k + aI^x + bH^y)$, where k, a and b denote the saturated parameters of medical resources, and x and y denote the degree of constraint on medical resources. It should be noted that this type $f(I, H)$ can cover several different incidence forms, such as mI, $mI/(k + aI)$, $mI/(k + aI^2)$, $mI/(k + bH)$, $mI/(k + bH^2)$, $mI/(k + aI + bH)$ et al. Hence, this function is more general. After being treated in a hospital, the infections can recover or die at rates r_H or d_H. All recovered individuals enter the recovered class and are assumed to have full immunization and can not be infected again for a short time. It is assumed that both asymptomatic infections (A) and symptomatic infections (I) are infectious and can cause the transmission of emerging infectious diseases in society. In contrast, due to the good isolation measures, the hospitalized infection (H) can not be transmitted to others. The schematic diagram of the transmission model is shown in Fig. 1.

To model two control measures: reducing contact numbers and improving medical resources, we first introduce two controls, denoted by u_1 and u_2, where $0 \leq u_1, u_2 \leq 1$. Note that in the subsequent sections, we will discuss the effects of two controls from two aspects, constant and time-varying. Since reducing contact numbers can effectively reduce the transmission probability, the transmission rate β is modified as $(1 - u_1)\beta$. Improving medical resources includes increasing the diagnosis rate and the number of hospital beds (ICU and isolated hospital beds, etc.), thus can admit more infected individuals into hospital to be treated. The improving hospitalization rate is modeled as $(1 + u_2)f(I, H)$. All parameters are considered non-negative constants, and their biological meanings are summarized in Table 1. Based on the above assumptions, we formulate the following transmission dynamic model with two control measures, which is described by the following ordinary differential equations
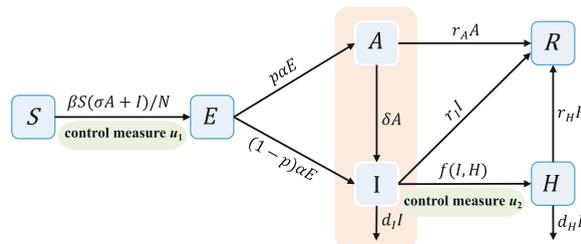

Fig. 1. The schematic diagram of the proposed model.

Table 1
Descriptions of parameters in system (2.1).

Variables	Descriptions
β	Transmission rate
σ	Modification factor of transmission rate
p	Proportion of asymptomatic infection
α	Progression rate of exposed individuals to infections
δ	Transition rate from asymptomatic to infected class
r_A	Recovery rate of asymptomatic infections
r_I	Recovery rate of symptomatic infections
r_H	Recovery rate of hospitalized infections
d_I	Disease-induced death rate of infected individuals
d_H	Disease-induced death rate of hospitalized infections
m	Hospitalization rate
k, a, b, x, y	Dimensionless parameters

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{(1-u_1)\beta S(\sigma A + I)}{N}, \\[2mm] \dfrac{dE}{dt} = \dfrac{(1-u_1)\beta S(\sigma A + I)}{N} - \alpha E, \\[2mm] \dfrac{dA}{dt} = p\alpha E - (\delta + r_A)A, \\[2mm] \dfrac{dI}{dt} = (1-p)\alpha E + \delta A - (r_I + d_I)I - (1+u_2)f(I,H), \\[2mm] \dfrac{dH}{dt} = (1+u_2)f(I,H) - (r_H + d_H)H, \\[2mm] \dfrac{dR}{dt} = r_A A + r_I I + r_H H. \end{cases} \tag{2.1}$$

2.1.2. The control reproduction number

It is clear that system (2.1) always has a disease-free equilibrium $E_0 = (N^0, 0, 0, 0, 0, 0)$, where N^0 denotes the initial number of population. Following the method given by Diekmann and Dreessche (Diekmann et al., 1990; Dreessche & Watmough, 2002), we denote two matrixes $\mathscr{F}$ and $\mathscr{V}$ as follows

$$\mathscr{F} = \begin{pmatrix} \dfrac{(1-u_1)\beta S(\sigma A + I)}{N} \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathscr{V} = \begin{pmatrix} \alpha E \\ -p\alpha E + (\delta + r_A)A \\ -(1-p)\alpha E - \delta A + (r_I + d_I)I + (1+u_2)f(I,H) \end{pmatrix}.$$

Linearizing system (2.1) at E_0 yields two sensitivity matrixes of $\mathscr{F}$ and $\mathscr{V}$, that is

$$F = \begin{pmatrix} 0 & (1-u_1)\sigma\beta & (1-u_1)\beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} \alpha & 0 & 0 \\ -p\alpha & \delta + r_A & 0 \\ -(1-p)\alpha & -\delta & D_1 \end{pmatrix},$$

where $D_1 = r_I + d_I + (1+u_2)m$. Thus, we can derive the conrol reproduction number $\mathfrak{R}_c(u_1, u_2)$ as the spectrum radius of the next generation matrix (FV^{-1}), denoted by

$$\mathfrak{R}_c(u_1, u_2) = \rho(FV^{-1}) = (1-u_1)\left[\frac{p\sigma\beta}{\delta + r_A} + \frac{(1-p)\beta}{D_1} + \frac{p\delta\beta}{D_1(\delta + r_A)}\right]. \tag{2.2}$$

It should be noted that if we discard the control measures, i.e., $u_1 = 0$ and $u_2 = 0$, the control reproduction number $\mathfrak{R}_c(u_1, u_2)$ will reduce to be the basic reproduction number with the following form

$$\mathfrak{R}_0 = \mathfrak{R}_c(0,0) = \frac{p\sigma\beta}{\delta + r_A} + \frac{(1-p)\beta}{r_I + d_I + m} + \frac{p\delta\beta}{(\delta + r_A)(r_I + d_I + m)}, \tag{2.3}$$

which implies the number of new infected individuals produced by one infected individual entering an entire susceptible population. The biological meaning of $\mathfrak{R}_0$ is clear. There are three items in the expression of $\mathfrak{R}_0$, which means that three

routines can spread disease in the population. The first item denotes the new infection caused by asymptomatic infections. The second item is the new infection caused by a asymptomatic infections, which are directly from the exposed individuals, while the third item is the new infection produced by a symptomatic infection, which first undergo asymptomatic infected $(p\delta\beta/(\delta + r_A))$ and then survive into symptomatic infected class $(1/(r_I + d_I + m))$.

2.1.3. Sensitivity analysis

Sensitivity analysis is performed to explore the impact of constant controls and other parameters on two reproduction numbers. Here, we use the normalized forward sensitivity index, also called the local sensitivity index (Chitnis et al., 2008) and has the form $\gamma_\theta^{\Re_j} = \partial\Re_j/\partial\theta \times \theta/\Re_j$ $(j = 0, c)$ judging the sensitivity of $\Re_j$ to parameter θ. If $\gamma_\theta^{\Re_j}$ is larger than zero, then $\Re_j$ will increase with the increase of parameter θ, and otherwise it will decrease with the increase of parameter θ. It should be noted that local sensitivity index is obtained by changing one parameter θ at a time, while all the other parameters are fixed at their baseline values. This text explores the dependence of the basic reproduction number $\Re_0$ in (2.3) on other parameters and verifies the reasonableness of control measures. To do this, we need to compute the analytical expression for sensitivity indices of $\Re_0$ with respect to each parameter shown in the following

$$\gamma_\beta^{\mathscr{R}_0} = 1, \quad \gamma_\sigma^{\mathscr{R}_0} = \frac{p\sigma Q_1}{Q_3}, \quad \gamma_p^{\mathscr{R}_0} = \frac{p(\sigma Q_1 - r_A)}{Q_3}, \quad \gamma_\delta^{\mathscr{R}_0} = \frac{p\delta(-\sigma Q_2 + r_A)}{Q_2 Q_3}, \quad \gamma_{r_A}^{\mathscr{R}_0} = -\frac{p r_A(\sigma Q_1 + \delta)}{Q_2 Q_3},$$

$$\gamma_{r_I}^{\mathscr{R}_0} = -\frac{r_I[(1-p)Q_2 + p\delta]}{Q_1 Q_3}, \quad \gamma_{d_I}^{\mathscr{R}_0} = -\frac{d_I[(1-p)Q_2 + p\delta]}{Q_1 Q_3}, \quad \gamma_m^{\mathscr{R}_0} = -\frac{m[(1-p)Q_2 + p\delta]}{Q_2 Q_3},$$

where $Q_1 = r_I + d_I + m$, $Q_2 = \delta + r_A$ and $Q_3 = p\sigma Q_1 + (1-p)Q_2 + p\delta$.

Furthermore, we show the sensitivity results of $\Re_c$ to constant controls u_1 and u_2. From the expression of $\Re_c$ in (2.2), we can easily have that

$$\gamma_{u_1}^{\Re_c} = -\frac{u_1}{1-u_1}, \quad \gamma_{u_2}^{\Re_c} = -\frac{m u_2 D_2}{D_1(p\sigma D_1 + D_2)},$$

where $D_2 = (1-p)(\delta + r_A) + p\delta$. Clearly, $\gamma_{u_1}^{\Re_c} < 0$ and $\gamma_{u_2}^{\Re_c} < 0$, which implies that increasing the constant controls can decrease the value of $\Re_c$, thus contributes to the remission of COVID-19. Because $\partial|\gamma_{u_1}^{\Re_c}|/\partial u_1 > 0$ and $\partial|\gamma_{u_2}^{\Re_c}|/\partial u_2 > 0$, the larger u_1 and u_2 are, the stronger the correlation between $\Re_c$ and them is.

2.2. Model with time-varying control

In practice, the intensity of control measures will change with the current epidemic situation. That is to say, the control parameters change over time. Hence, in this subsection, we focus on the model with time-varying control, referring to optimal control and actual control.

2.2.1. Model formulation

Changing the constant controls u_1 and u_2 in system (2.1) to time-varying forms, denoted by $u_1(t)$ and $u_2(t)$, we obtain the following time-varying system

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{(1-u_1(t))\beta S(\sigma A + I)}{N}, \\[2mm] \dfrac{dE}{dt} = \dfrac{(1-u_1(t))\beta S(\sigma A + I)}{N} - \alpha E, \\[2mm] \dfrac{dA}{dt} = p\alpha E - (\delta + r_A)A, \\[2mm] \dfrac{dI}{dt} = (1-p)\alpha E + \delta A - (r_I + d_I)I - (1 + u_2(t))f(I, H), \\[2mm] \dfrac{dH}{dt} = (1 + u_2(t))f(I, H) - (r_H + d_H)H, \\[2mm] \dfrac{dR}{dt} = r_A A + r_I I + r_H H. \end{cases} \tag{2.4}$$

2.2.2. Optimal control

Denote two control measures as $U(t) = (u_1(t), u_2(t))$ and the control set as $\Theta = \{U \in (L^\infty(0, T), L^\infty(0, T)) | 0 \le u_i(t) \le 1, i = 1, 2\}$. Then, we define the objective functional as

$$J(U) = \int_0^{t_f} L(A(t), I(t), u_1(t), u_2(t))dt,$$

where t_f denotes the end time of control and the integrand function is given by

$$L(A(t), I(t), u_1(t), u_2(t)) = A_1 A(t) + A_2 I(t) + \frac{1}{2}(B_1 u_1^2(t) + B_2 u_2^2(t)).$$

The objective of optimal control is to find the optimal control u_1^* and u_2^* such that the number of nonhospitalized infections (asymptomatic and symptomatic) is minimized with the lowest cost of the control measures. Here, A_1 and A_2 represent the weight coefficients of the asymptomatic and symptomatic infected individuals, respectively. B_1 and B_2 denote the weight coefficients for the costs associated with the control variables $u_1(t)$ and $u_2(t)$, respectively.

The existence and characteristics expression of optimal control is shown in the following two theorems, respectively, and the proofs are shown in Appendix B and Appendix C.

Theorem 2.1. There exists an optimal control $U^* = (u_1^*, u_2^*) \in \Theta$ such that $J(U^*) = \min_{U \in \Theta} J(U)$.

Theorem 2.2. For any optimal control U* $\in \Theta$ and solution (S, E, A, I, H, R) of system (2.4), the optimal control solution of optimal control problem can be obtained as

$$u_1^* = \min\{1, \max\{0, u_1^c\}\}, \quad u_2^* = \min\{1, \max\{0, u_2^c\}\},$$

where

$$u_1^c = \frac{\beta S(\sigma A + I)}{N} \frac{\lambda_2 - \lambda_1}{B_1}, \quad u_2^c = \frac{mI(\lambda_4 - \lambda_5)}{B_2(k + aI^x + bH^y)} \tag{2.5}$$

and $\lambda_i (i = 1, \ldots, 6)$ denotes the adjoint funtions which are defined in Appendix C.

2.2.3. Actual control

In this section, we concern on the actual control, described by two time-varying control indexes. Once the disease outbreak occurs, strict measures, such as wearing masks, reducing aggregation and lockdown, etc., are implemented immediately to lessen the contact numbers. At the same time, the capability of nucleic acid detecting is quickly improved, and the number of hospital beds is drastically increased along with the occupation of specialized hospitals. Those actual controls effectively inhibited the epidemic's spread. Based on the research of Zhou (Zhou et al., 2020), we define two time-varying control indexes, reflecting the control intensity of reducing contact numbers and improving medical resources, with the following form.

$$U_1(t) = 1 - \frac{c(t)}{c_0}, \quad U_2(t) = 1 - \frac{q_0}{q(t)}, \tag{2.6}$$

where c_0 is the basic contact numbers without any control measures, and q_0 is the basic diagnosis and hospitalization rate of infected individuals at the initial stage of the disease. $c(t)$ denotes the time-varying contact numbers, reflecting the control measure of the restriction on social distance. $q(t) \in [0, 1]$ reflects the improvement rate of medical resources. Under different epidemic stages, the forms of the two non-negative functions $c(t)$ and $q(t)$ are also different. We will provide specific forms of $c(t)$ and $q(t)$ based on the actual data collected in Section 2.3. Clearly, two control indexes satisfy $U_1, U_2 \in [0, 1]$ so that we can compare the control effects with the optimal control.

2.3. Data and parameter estimation

2.3.1. Data collection

The data includes mainly two parts: the COVID-19 epidemic in Wuhan city from January 10 to April 12, 2020, and in Hebei province from October 31 to November 18, 2021 (See Fig. 2. Data sources (Hebei Provincial Health Commission, 2022; Hubei Provincial Health Commission, 2022):). It should be noted that there is no death related to COVID-19 during the considered period in Hebei province.

Based on the reported data of Wuhan city, the specific forms of $c(t)$ and $q(t)$ in (2.6) are taken as follows

$$c(t) = \begin{cases} c_0, & t < t_1, \\ (c_0 - c_1)e^{-r_c(t-t_1)} + c_1, & t \geq t_1, \end{cases} \quad \frac{1}{q(t)} = \begin{cases} \dfrac{1}{q_0}, & t < t_1, \\ \left(\dfrac{1}{q_0} - \dfrac{1}{q_1}\right)e^{-r_q(t-t_1)} + \dfrac{1}{q_1}, & t_1 \leq t < t_2, \\ \dfrac{1}{q_1}, & t \geq t_2, \end{cases}$$

where c_1 is the minimum contact number with the awareness of the disease and reducing contact numbers, r_c is the decreasing rate of the contact numbers. The control measure of reducing contact numbers was carried out from 23 January 2020 (Day 13), thus, taking $t_1 = 13$. Before 23 January, the infected individuals can be diagnosed and hospitalized at a basic rate q_0 at the initial stage; along with improving detecting capability and hospital beds, the waiting time from being diagnosed to hospitalized is decreased at a rate r_q. The maximum rate of improving medical resources is q_1. Due to the sharp increase of confirmed cases on 12 February 2020 (Day 33), the capacity of diagnosed and hospitalized reaches its maximum at $t_2 = 33$ and keeps it going.

Based on the reported data of Hebei province, we can observe that the epidemic sustained about 16 days from the beginning to the end. This implies the controls are taken timely and effectively. Thus, we assume that two controls, reducing contact numbers and improving medical resources, were implemented on the first day of the disease outbreak. Furthermore, the medical resources can reach the maximum level, which is reasonable since the capability of nucleic acid detection, diagnosis, mobile cabin hospital, and centralized isolation places in most districts is significantly promoted after the 2020 year. Hence, the specific forms of $c(t)$ and $q(t)$ in (2.6) are defined as

$$c(t) = (c_0 - c_1)e^{-r_c(t-t_1)} + c_1, \quad q(t) = q_1, \quad t \geq t_1.$$

Here, $t_1 = 1$ represents the first day to implement two control measures.

2.3.2. Parameter estimation

Considering the influence of the data randomness on parameter estimation, we assume that the daily numbers of new confirmed and new deaths follow a Poisson distribution with a mean value of the actual data. Here, we take 1000 iterations to randomly generate samples of datasets for fitting. According to computations, the mean values, standard deviations, and 95% Let $\hat{X}(t_i)$ and $X(t_i, \psi)$ be the reported numbers and simulation numbers at date t_i, respectively, and ψ is the vector of parameters to be estimated. Based on the classical nonlinear least-square (LS) method, we propose the improved Particle Swarm Optimization (IPSO) algorithm in Matlab to find the parameter value to minimize the objective function with the following form

$$f(\psi) = \kappa \sum_{i=1}^{n} (\hat{X}(t_i) - X(t_i, \psi))^2, \tag{2.7}$$

where κ is the weight coefficient to stress the importance of fitting data, n is the size of the epidemic data period. By implementing the IPSO in Matlab, we obtain the values of unknown parameters. It should be noted that the IPSO, which integrates two operations, crossover and acceptance operations, into the classical PSO, is first proposed and can better improve the speed for searching the optimal solution. More details on IPSO algorithm are referred to Appendix A.
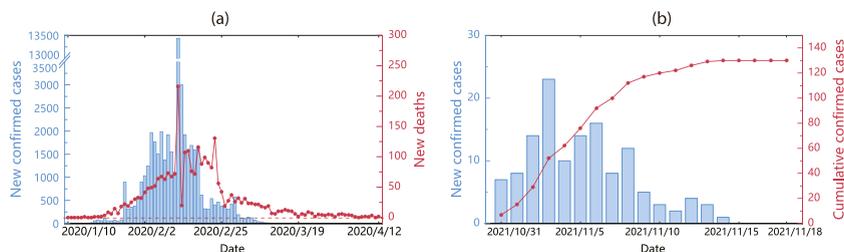

Fig. 2. The reported data. (a) Wuhan City from 10 January to 12 April 2020. (b) Hebei Province from 31 October to 18 November 2021.

3. Results

3.1. Parameter estimation

For the data of Wuhan city, according to the method in Section 2.3.2, we set $\hat{X}_1(t_i)$, $\hat{X}_2(t_i)$ be the reported cumulative numbers of confirmed cases and deaths at date t_i, and $X_1(t_i, \psi)$, $X_2(t_i, \psi)$ be the simulation cumulative numbers of confirmed cases and deaths at date t_i, respectively. Furthermore, we define the objective function with the following form

$$f(\psi) = \kappa_1 \sum_{i=1}^{n} (\hat{X}_1(t_i) - X_1(t_i, \psi))^2 + \kappa_2 \sum_{i=1}^{n} (\hat{X}_2(t_i) - X_2(t_i, \psi))^2, \tag{3.1}$$

where $\kappa_1 = 3/5$, $\kappa_2 = 2/5$ and $n = 94$. It should be pointed out that the weight values of κ_1 and κ_2 represent the importance of fitting data. The estimation values of parameters are summarized in Table 2. Moreover, the fitting curve with 95% confidence intervals is shown in Fig. 3.

Furthermore, using parameter values in Table 2., we gain the estimated cumulative numbers of hospitalized cases and deaths with 50869 and 2471, respectively, while the reported cumulative numbers of hospitalized cases and deaths are 50008 and 2580, which are shown in Fig. 3. It can be seen that the margin of error between the estimated values and the actual values is only 1.72% and 4.22%. Therefore, model (2.1) is credible.

For the data of Hebei province, we set $\hat{X}(t_i)$, $\hat{Y}(t_i)$ be the reported new confirmed number and its cumulative numbers at date t_i, and $X(t_i, \psi)$, $Y(t_i, \psi)$ be the simulation new confirmed numbers and its cumulative numbers at date t_i, respectively. It should be noted that there is no death related to COVID-19 during the considered period in Hebei province. Hence, we define the objective function with the following form

$$f(\psi) = \kappa_1 \sum_{i=1}^{n} (\hat{X}(t_i) - X(t_i, \psi))^2 + \kappa_2 \sum_{i=1}^{n} (\hat{Y}(t_i) - Y(t_i, \psi))^2, \tag{3.2}$$

Table 2
Initial values of state variables and parameters in system (2.1).

Variables		Initial value (Std.)		Resource
		Wuhan city	Hebei province	
S		11081000	74610235	Data
E		87(2)	47(9)	Estimated
A		48(3)	45(3)	Estimated
I		57(1)	4.8(1)	Estimated
H		38	1	Data
R		2	0	Data

| Parameters		Mean value (Std.)		Resource
		Wuhan city	Hebei province	
β		1.7746(0.0156)	1.6746(0.5684)	Estimated
σ		0.3123(0.0284)	0.8990(0.0060)	Estimated
p		0.13166 (Tang, Bragazzi, et al., 2020)	0.3001(0.0005)	Estimated
α		1/5 (Tang, Bragazzi, et al., 2020)	1/3 (Zhang et al., 2021)	
δ		0.5348(0.0210)	0.4122(0.0326)	Estimated
r_A		0.2605(0.0183)	0.2851(0.1194)	Estimated
r_I		0.2745(0.0054)	0.0737(0.0117)	Estimated
r_H		0.2480(0.0112)	0.5000(0.0000)	Estimated
d_I		0.0003(0.0003)	0 (Data)	Estimated
d_H		0.0126(0.0007)	0 (Data)	Estimated
m		0.3956(0.0096)	1.9972(0.0275)	Estimated
k		1.3226(0.0184)	1	Estimated
a		0.1438(0.0084)	0.2084(0.0624)	Estimated
b		0.1268(0.0063)	0.3500(0.000)	Estimated
x		0.1530(0.0075)	0.8731(0.0569)	Estimated
y		0.0172(0.0049)	0.6000(0.0000)	Estimated
$c(t)$	c_0	23.8325(0.3320)	30(0.0000)	Estimated
$q(t)$	c_1	0.5688(0.0797)	0.1000(0.0000)	Estimated
	r_c	0.0810(0.0031)	0.4997(Assumed)	Estimated
	q_0	0.0510(0.0024)	—	Estimated
	q_1	0.8313(0.0101)	0.9956(0.0313)	Estimated
	r_q	0.0114(0.0018)	—	Estimated
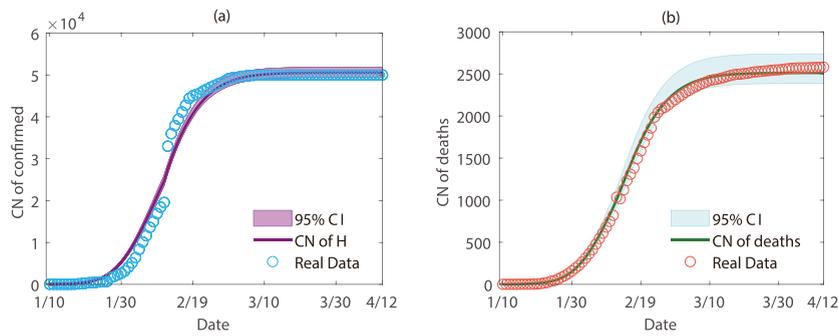
Fig. 3. Fitting result for the number of cumulative confirmed cases and deaths from 10 January to 12 April 2020 in Wuhan city. Note: CI—confidence interval; CN—cumulative numbers; CC—confirmed cases.

where $\kappa_1 = 9/10$, $\kappa_2 = 1/10$ and $n = 19$. Here, κ_1 is larger than κ_2, which shows that we pay more attention to fitting daily data. Thus, the estimation values of parameters are summarized in Table 2, and the fitting curve with 95% confidence intervals is shown in Fig. 4.

3.2. The effect of constant control

Based on the parameter values in Table 2, and based on Section 2.1.3, we can get the absolute sensitivity of the basic reproduction number $\Re_0$ to each parameter which is shown in Fig. 5(a), where a clear observation is that the transmission probability (β) has the most important effect on $\Re_0$ and the hospitalized rate of symptomatic infections (m) takes the second place. To better control the spread of COVID-19, the most suitable control measures are reducing contact numbers (u_1) and improving medical resources (u_2), which should be adopted. From Fig. 5(b), one can see that constant controls u_1 and u_2 are negatively correlated with the control reproduction number $\Re_c$. Comparing the influence of u_1 with u_2 on $\Re_c$, one can find that u_1 is more sensitive than u_2. Moreover, single improving u_2 can not reduce the value of $\Re_c$ to below 1, which implies that reducing contact numbers and providing ample medical resources are both vital.

To understand how the constant controls influence the peak values and peak time of symptomatic infections ($I(t)$) and hospitalized infections ($H(t)$), we simulate the contour plots of the peak value and peak time of $I(t)$ and $H(t)$ with respect to constant controls u_1 and u_2 (see Fig. 6). From Fig. 6(a)–(b), one can observe that increasing the intensity of two control measures contributes to a decrease in the peak values of $I(t)$ and $H(t)$, and the larger u_2 is, the better the control effect on the peak value is, but when the intensity of u_1 is increased to 0.6, two controls will have little or no effect on the peak value of $I(t)$ and $H(t)$. For instance, if u_2 is fixed, increasing u_1 will initially favor reducing the peak value of infections, but if u_1 exceeds about 0.6, there have little effects. For the peak time of $I(t)$ and $H(t)$, If u_1 exceeds 0.7, u_2 has little effect on shortening the length of the epidemic, while it takes nearly 200 days to reach the peak value for $I(t)$ and $H(t)$, which will further make people more burdensome. Thus, only two controls simultaneously being adopted could effectively mitigate the epidemic's spread.

3.3. The effect of time-varying control

3.3.1. Case1: Wuhan city

In this section, we will verify the proposed model's reliability, compare the results of the optimal control with the actual case, and offer related suggestions so that the actual control can better approach the optimal control level.
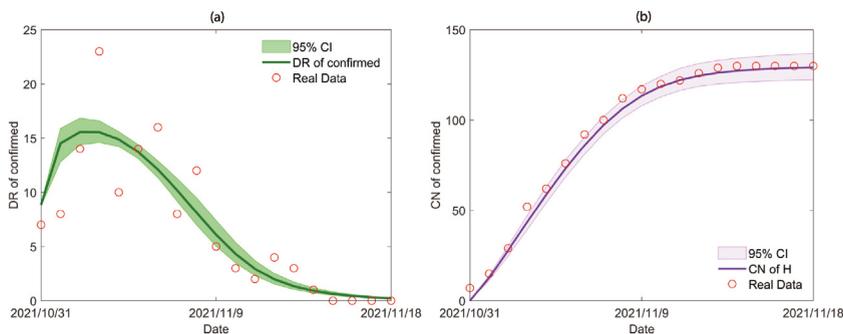

Fig. 4. Fitting result for the number of daily confirmed and cumulative confirmed cases from 31 October to 18 November 2021 in Hebei province. Note: CI—confidence interval; DR—daily reported; CN—cumulative numbers.
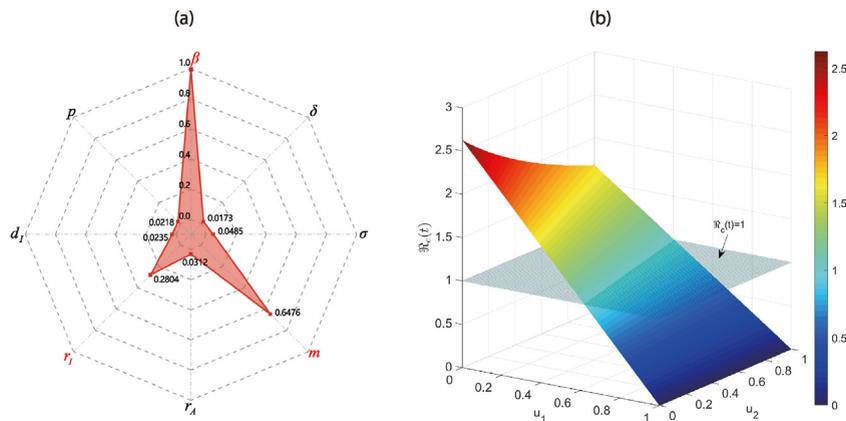
Fig. 5. The influence of parameters and constant controls on reproduction number. (a) Absolute sensitivity indices of $\Re_0$ in (2.3); (b) The influence of u_1 and u_2 on $\Re_c$ in (2.2).
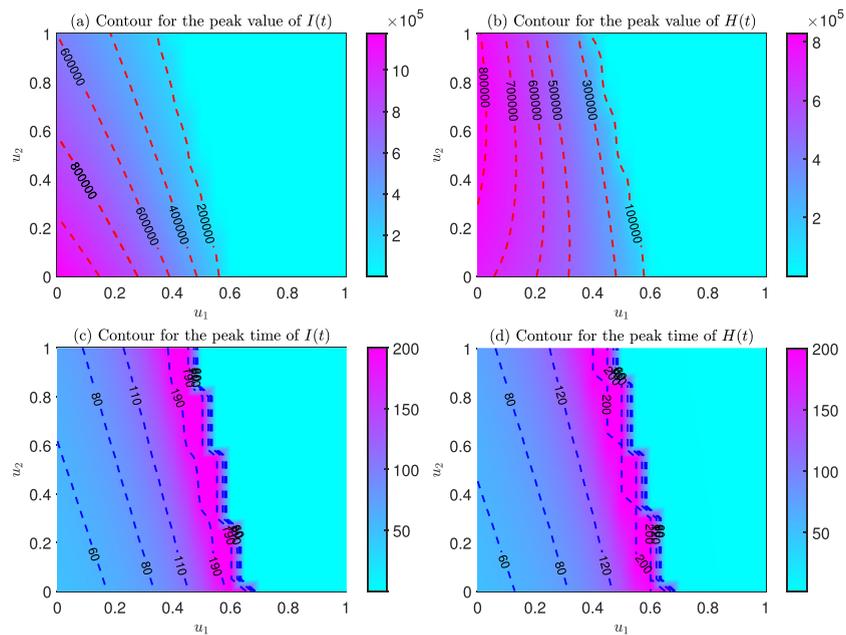

Fig. 6. Contour plots on the peak value and peak time of $I(t)$ and $H(t)$.

Simulations are used to compare the effects of two different control strategies, optimal control, and actual control, on the spread of COVID-19 from 10 January to 12 April 2020 in Wuhan city. Fig. 7 shows the time series of two different time-varying control indexes. Following Fig. 7(a), one can observe that the way of optimal control is that controls are immediately implemented once the epidemic occurs. The control intensity reaches the maximum and maintains about 34 and 8 days for reducing contact numbers and improving medical resources, respectively, and then gradually decreases to zero at day 50. Following Fig. 7(b), it takes 13 days to realize the severity of the epidemic, and then two controls are implemented. Whereas, for optimal control case, one can observe from Fig. 7(c) that the duration of epidemic can be shortened to 21 days, we can calculate the cumulative numbers of hospitalized infections and deaths re 3595 and 64 cases. This can prevent 46,413 people from being infected and save 2516 lives, compared with the actual values, 50008 and 2580. Compare Fig. 7(c) with Fig. 7(d), we can find that the peak values of $I(t)$ and $H(t)$ will be reduced to 60 and 50 cases, respectively, which is different from the situation of actual control with 7028 and 7621 cases. Moreover, the peak time can also be sharply shortened compared with the actual control. Optimal control is an effective strategy for combating emerging infectious disease.

Next, we discuss how the actual control can approach the optimal control level. It follows from the above comparison that implementing optimal control can significantly reduce the size and duration of an epidemic compared with the actual control. We focus on how to achieve the possible effect of actual control can approach the optimal control level. From two different
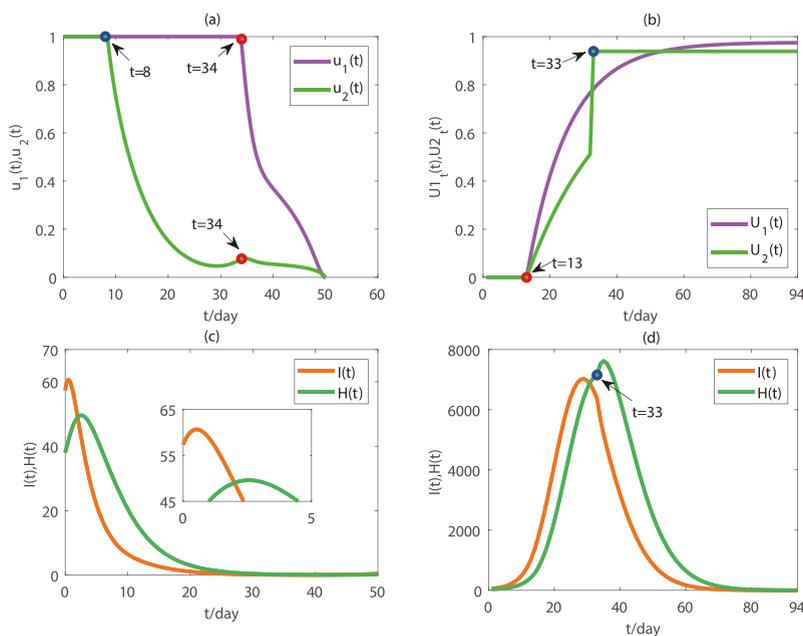
Fig. 7. Time series diagram of time-varing control indexes and infections. (a) Control indexes of optimal control case; (b) Control indexes of actual control case for Wuhan city; (c) Infections of optimal control case; (d) Infections of actual control case for Wuhan city.

time-varying control curves in Fig. 7(a)–(b), we can realize that the beginning time of controls has a direct impact on the spread of the epidemic, including the peak value and peak time of $I(t)$ and $H(t)$, even the duration of the epidemic. To explore the influence of controls' start time, we set three comparison groups for the combinations of two controls' start time, whose impacts on the spread of COVID-19 are shown in Fig. 8 and Table 3. Fig. 8(a)–(b) shows the impact of the control start time on the control indexes $U_1(t)$ and $U_2(t)$. From Fig. 8(c)–(d), a precise observation is that the peak value of infections is decreased, and the peak time is advanced along with the advance of control implementation time. Specifically, it follows from Table 3 that if the implementation time of two controls is brought forward to day 10 (20 January) and day 20 (31 January), respectively, then the peak value of $I(t)$ and $H(t)$ will decrease by 63.85% and 55.51%, respectively. If the implementation time of two controls can be further advanced to the first day, then there will be 98.14% and 97.33% decrease in the peak value. At the same time, the peak time also can be advanced, and the epidemic can be ended at day 54, which is closer to the optimal control level, though there is still a particular gap.

How to adjust the actual control to narrow the gap with the optimal control level? To deal with this question, we try to change the decreasing rate of contact numbers (r_c) under the case that $t_1 = t_2 = 1$. Here two scenarios $r_c = 0.1810$ and $r_c = 0.4997$ are considered. Fig. 8(e)–(f) and Table 3 shows the impacts of r_c on the spread of COVID-19, and it is clear that the peak value of $I(t)$ and $H(t)$ decreases with the increase of r_c. Primarily, following Table 3, we can obtain that when $r_c = 0.1810$, the peak value of $I(t)$ and $H(t)$ can be fallen by 55.64% and 62.56%, respectively. When r_c is promoted to 0.4997, the results of actual control, including the peak value, peak time of infections, and the length of the epidemic, are closer to the optimal control level.

Actually, after two years development, medical resources, including nucleic acid detection capability, hospital beds and establishment of Fangcang shelters, have been greatly improved. As a result, medical resources were able to reach a maximum improvement level of 0.8313, which is close to 1 at the beginning of the COVID-19 epidemic, which is consistent with the actual situation. Consequently, we can conclude that the ideal actual control in practice is that controls are implemented from the beginning time and the decreasing rate of contact numbers can reach 0.4997 shown in Table 2.

3.3.2. Case 2: Hebei province

Whether the adjustment in Section 3.3.1 can be adopted in practice? To verify the suggestion on the start time and intensity of controls, we take the reported data from 31 October to 18 November 2021 in Hebei province as a second case.

Simulations are performed to verify whether the actual control in Hebei province can approach the optimal control level under the above suggestions. Using the parameter values in Table 2, we plot the time series of two different time-varying control indexes shown in Fig. 9. It follows from Fig. 9(a)–(b) that both the time-varying control indexes are different. Under optimal control (see Fig. 9(a)), the maximum control intensity is reached and lasts for 7 days and 13 days for two controls, respectively, and then gradually decreases to 0 till the end of the epidemic. It is important to note that when the control intensity of reducing contact numbers drops to day 13.5, the control intensity of improving medical resources has a slight rise to ensure that the epidemic does not rebound. This phenomenon is reasonable and identical to the reality. Under
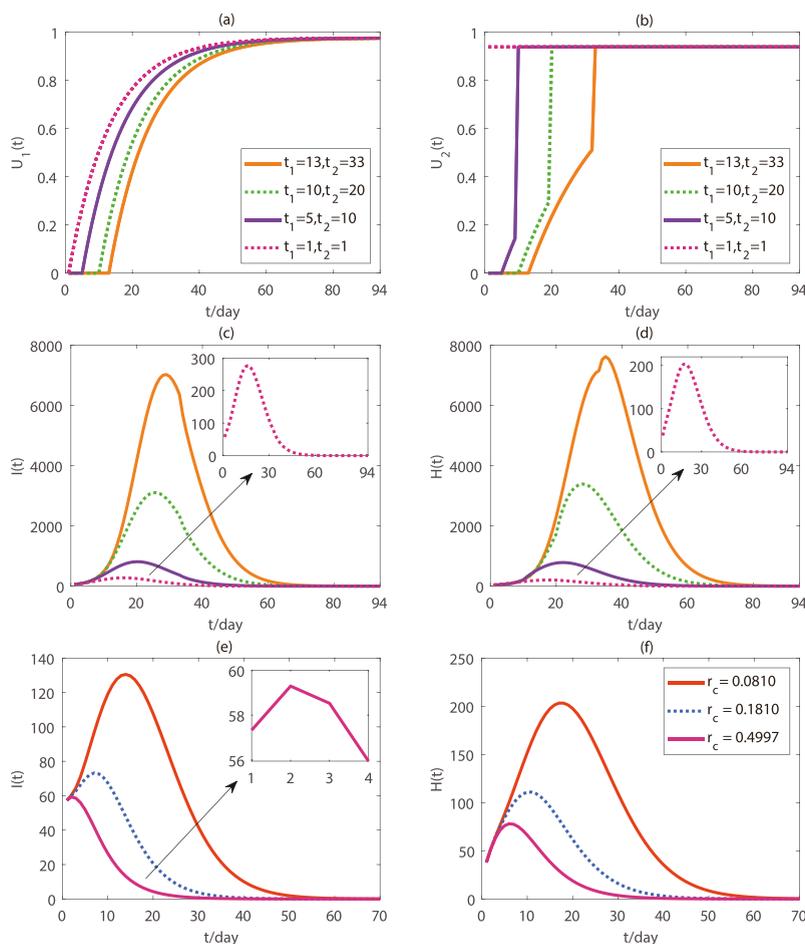
Fig. 8. The impact of the start time and decreasing rate for reducing contact numbers on the spread of COVID-19 in Wuhan city.

Table 3
The impact of controls' start time and intensity on the spread of COVID-19 in Wuhan city.

Control		$I(t)$		$H(t)$		Duration
		PVs	PT	PVs	PT	
Actual control		7027	29	7621	29	89
	$t_1 = 10, t_2 = 20$	2540	20	3391	20	80
$r_c = 0.0810$	$t_1 = 5, t_2 = 10$	528	18	784	18	66
	$t_1 = 1, t_2 = 1$	131	14	203	14	54
$t_1 = t_2 = 1$	$r_c = 0.1810$	73	7	111	11	38
	$r_c = 0.4997$	59	2	76	6	28
Optimal control		60	2	50	4	21

Note: PVs − Peak values; PT − Peak time.

actual control for Hebei province (see Fig. 9(b)), based on the rich experience of resistance to COVID-19, at the beginning of the epidemic, the medical resources can reach the maximal improving level about 0.9956. Concurrently, individuals are aware of the need to reduce contact with others, for example by wearing masks and avoiding large gatherings. When an epidemic occurs and government starts to implement more strict control, people can promptly respond at a rate 0.4997 to decrease the contact numbers.

To verify the suggestion on adjusting the decreasing rate (r_c), we set $r_c = 0.4997$, which is the same as the value of r_c in the last simulation for Wuhan city. From Fig. 9(c)−(d) and Table 4, we can find that the peak value of $I(t)$, the peak time of $H(t)$ and duration of the epidemic under actual control are 28, 6, and 14, respectively, which are very close to the corresponding optimal control level. This illustrates that the actual control taken in 2021 year for Hebei province can better approach the optimal control level. However, comparing the actual control with optimal control in the aspects of peak time of $I(t)$ and peak
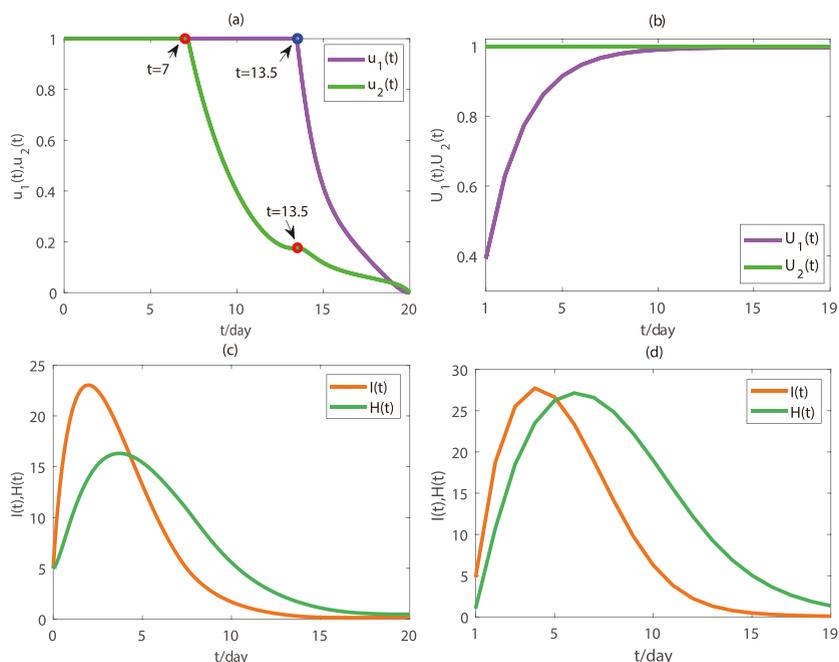
Fig. 9. Time series diagram of time-varying control indexes and infections. (a) Control indexes of optimal control case; (b) control indexes of actual control case for Hebei province; (c) infections of optimal control case; (d) infections of actual control case for Hebei province.

value of $H(t)$, there is still a certain gap to be researched, which deserves us to explore some more unknown mechanisms for controlling epidemic.

4. Discussion and conclusion

The outbreak of COVID-19 quickly spread around on a global scale and caused a "pandemic" announced by the WHO, which brought serious harm and social burden to the world. In the absence of effective vaccines to battle this emerging infectious disease, the timely adoption of strict prevention and control measures was crucial in controlling the spread of disease. For instance, the related departments in China took urgent measures to control the epidemic's prevenlance, including lockdown cities to decrease the contact numbers of individuals, promoting the ability to be diagnosed, constructing designated hospitals and shelters to improve the hospitalized rate, being aided by other provinces and countries, etc. Thus, the epidemic in Wuhan city was controlled effectively, and the adopted control has been highly praised.

4.1. Summary of findings

In this paper, we formulate a COVID-19 transmission model with two control measures, reducing contact numbers and improving medical resources. To model the serious shortage of medical resources at the initial stage of the epidemic, we propose a new type of function $f(I, H) = mI/(k + aI^x + bH^y)$ to describe that not all individuals can be timely diagnosed and treated in hospital. Two different forms (constant control and time-varying control) of two control measures are applied to the proposed model. Theoretically, the control reproduction number and the basic reproduction number are calculated, and the sensitivity of the reproduction number to various parameters is analyzed, as well as the existence and formal characteristics of the optimal control. In terms of numerical simulation, we propose the ISPO algorithm to conduct the parameter estimation. Based on the actual data of Wuhan city from January 10 to April 12 in 2020 year, and the rationality of the model is verified by comparing the difference between the estimated cumulative number of confirmed cases and the actual value.

Table 4
The comparison on the spread of COVID-19 in Hebei province.

Control	I(t)		H(t)		Duration
	PVs	PT	PVs	PT	
Actual control	28	4	27	6	14
Optimal control	23	2	16	4	12

Then, the effects of constant control and time-varying control are discussed separately. In particular, the path to make the actual control as close as possible to the optimal control level is discussed, and the effectiveness of the proposed path is verified with the actual data of Hebei Province from October 31 to November 18 in 2021 year. To sum up, the main results of this article are as follows.

(1) In the model with constant control, the transmission probability (β) has the most important effect on the basic reproduction number ($\Re_0$), and the hospitalized rate of symptomatic infections (m) takes the second place. u_1 and u_2 are negatively correlated with the control reproduction number ($\Re_c$). Furthermore, u_1 is more sensitive than u_2, and a single improvement of u_2 can not reduce the value of $\Re_c$ to below 1.

(2) For the constant control, as the control intensity increases, the peak values of $I(t)$ and $H(t)$ decrease, and the peak time will be delayed. Regardless of the intensity of u_2, if the intensity of u_1 is increased to 0.6, two controls will have little effect on the peak values of $I(t)$ and $H(t)$; if the intensity of u_1 is increased to 0.7, two controls will have little effect on the peak time of $I(t)$ and $H(t)$.

(3) For the time-varying control, we have the following findings:

(3.1) Based on the data of Wuhan city, one can observe that optimal control has the best effect in suppressing COVID-19. Moreover, in practice, as implementation time of the actual control advances, the peak value is reduced and peak time is delayed. Thus, the more timely the implementation of control is, the better the effect is. Finally, when the decreasing rate of the contact numbers (r_c) is promoted to 0.4997, the control results, including the peak value, peak time and the duration of the epidemic, are closer to the optimal control level.

(3.2) Based on the data of Hebei Province, we validate the feasibility of adjusting (r_c) to make the actual control effect closer to the optimal control level.

4.2. Revelation gained

Based on the conclusions obtained in this paper, we have the following insights. Firstly, simply expanding medical resources cannot effectively control the diseases that spread quickly in practice. Instead, it is necessary to find ways to reduce the number of contacts as much as possible. Only through the joint implementation of two measures can the disease be better controlled. Secondly, when constant control is adopted, the stronger the control force is at the beginning, the better the control effects are, but this intensity does not need to be increased to the maximum. Because when the contact number is decreased to a certain extent, the speed and intensity of disease transmission will significantly decrease, and with the effective allocation of medical resources, the disease can be completely suppressed. Even if the control intensity is continued to be increased, the control effect will not be further improved. Of course, from the perspective of ease of operation and cost savings in control implementation, relevant departments also do not want to take the maximum control measures. Thirdly, in practice, we always hope to take optimal control measures. However, when implementing them, it is difficult for the actual control to be completely consistent with the optimal control, and we can only try to be as close as possible to it. In situations where various prevention and control measures are relatively complete, the earlier the control is implemented, the smaller the peak value of the epidemic is, the shorter the peak time is, and the shorter the epidemic time lasts, which indicates that timely analysis and prevention and control of the epidemic are crucial. Furthermore, increasing the decreasing rate of the contact numbers (r_c) is another method to make actual control closer to the optimal control. This reminds us to effectively educate the public to wear masks, maintain social distance, and isolate after infection, which is beneficial for reducing the contact numbers.

4.3. Limitations and further work

The results presented in this paper have valuable implications for practical applications, particularly providing insights for other countries in dealing with emerging infectious diseases in the future. However, there are several areas that could be further improved and addressed. Firstly, when an emerging infectious disease occurs, the related policy-makers are hard to promptly identify the infectivity and harmfulness, and whether the related control intensity can reach the maximum is still challenging. Thus, there exists a delay from occurring epidemic to implementing the control measures. Secondly, the current model utilizes ordinary differential equations to capture the transmission dynamics of the epidemic and evaluate the effectiveness of control measures. Actually, the deaths mainly focus on older individuals with some underlying diseases. The infections originate from South China Seafood City in the Jianghan region of Wuhan city and then diffuse to the related regions. Hence, the infections depend on the individuals' age and geographical distribution. The heterogeneity model, taking age and space into consideration, would be more realistic to predict or assess the effectiveness of related control strategies. Finally, along with the use of vaccines, variations of virus strains, and the adjustment of control strategies, more practical models should be developed to describe the complex phenomena, such as multiple waves and scatterer bursts in many countries. Although this paper can show how the actual control approaches the optimal control, unfortunately, it can not absolutely reach, which implies that more potential mechanisms should be explored so as to achieve accurate control by the theoretical optimal control level. These aspects will be the focus of future research endeavors.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

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.

Acknowledgments

Appendix A. IPSO algorithm

Particle swarm optimization (PSO) algorithm is a Swarm Intelligence Algorithm developed by Kennedy and Eberhart based on the social behavior of birds (Kennedy & Eberhart, 1995). This paper integrates two operations, crossover and acceptance, into the classical PSO algorithm to search for better local optimal solutions, and then apply this improve PSO (IPSO) algorithm to implement the parameter estimation of our epidemic model. To better understand the IPSO, we first introduce some parameters and variables in the classical PSO algorithm, which is summerized in Table A1.

Now, we will introduce the IPSO algorithm in detail from the following two parts.

Table A1
The basic variables and parameters in the classical PSO algorithm

Variables	Description	Type
Iteration	Iteration	Integer
Position	The position of all particles	Three-axis matrix
Velocities	Velocities of all particles	Three-axis matrix
IndividualBestPosition	Individual best position of a particle	Martix
GobalBestPosition	Best position of all particles	Vector
Value	Value of the particle in this iteration	Vector
Individual	Individual best value of the particle	Vector
Parameters		
w	A scaling factor influencing old velocity	Constant
c_{Self}	"Cognitive" coefficient	Constant
c_{Social}	"Social" coefficient	Constant
$rand_1$	Random number in the range [0,1]	Scalar
$rand_2$	Random number in the range [0,1]	Scalar
P_c	Probability of crossover	Constant
d	Number of variables	Constant
C	A Parameters of Accpet Function	Constant
MaxIreration	Max Iteration	Constant
random	Random number in the range [0,1]	Scalar

Part 1. Crossover operation.

The idea introducing crossover operation is inspired by genetic algorithms. Specifically, for all *numparticle* particles, they will take place intersection with a probability P_c, that is, there is *numparticle* × P_c particles performing the crossing operation in each iteration. For the selected particles *A* and *B*, system will randomly choose two components *i* and *j* from their *d* variables, which satisfy *i* < *j*, and perform crossover operation. After crossover operation, partial velocities of two random particles can exchanging (Algorithm 1). Thus, the particles can deviate from their original trajectories and explore new positions that would not have been searched otherwise, which allows for an overall search of solution space.

Algorithm 1: Crossover Operation: Cross Function

---

1 $CrossNumber = numparticle * P_c$;

2 $Cross$ = Select the label of $CorssNumber$ from all particles at random;

3 **for** $k \leftarrow 1$ *to CorssNumber* **do**

4 i, j = Select 2 integers at random from 1 to d and $i < j$;

5   For the $Velocities(CorssNumber(k))$, crossover its i-th to j-th components to $Velocities(CorssNumber(k+1))$

6 **end**

7 return $Velocities$

Part 2. Acceptance operation.

The idea of acceptance function is inspired by the simulated annealing algorithm. To avoid situations where convergence is difficult, we adjust the temperature of particles to be a variable denoted as $\Delta(number) = MaxIteration - Iteration + C$, where *MaxIteration* represents the maximum number of iterations, *Iteration* is the current iteration count, and *C* is a factor adjusting the probability. If the new solution is worse than the old, we perform the acceptance function according to the formula $p = 1/(1 + e^{\frac{\Delta}{\Delta(number)}})$, where p represents the acceptance probability, and $\Delta = |NewFval - OldFval|$ denotes the difference between new and old solutions (Algorithm 2), where the worse solution is referred that its fitness value is greater than its value of Individual. The optimization principle of acceptance function that it can influence the search method of particles in the next iteration by accepting a worse solution with a certain probability, thus searching for the solution space more comprehensively.

Algorithm 2: Acceptance Operation: Acceptance function

---

1 **for** $i \leftarrow 1$ *to numparticle* **do**

2   **if** *Value(i) < Individual(i)* **then**

3 $IndividualBestPosition(i) = Position(i)$;

4   **else**

5 $\Delta = |Individual(i) - Value(i)|$;

6 $\Delta(count) = MaxIteration - Iteration + C$;

7 $p = \frac{1}{1 + e^{\frac{\Delta}{\Delta(count)}}}$;

8   **if** $p < random$ **then**

9 $IndividualBestPosition(i) = Position(i)$;

10   **end**

11   **end**

12 **end**

13 return $Individual$

After introducing crossover and acceptance operations into the classical PSO algorithm, we formulate the IPSO algorithm (see Algorithm 3). For parameter estimation, the IPSO algorithm has the advantage in determining initial values: one is that it can simultaneously generate multiple initial values allows for a faster exploration of the solution space. The other is that by optimizing the initial values based on their positions using update formulas, it is more rational compared to blindly searching for initial values. Moreover, the IPSO algorithm also has the advantage in searching for the suboptimal solution, which is ample for the parameter estimations of epidemic models. If we intend to further search for the optimal solution, one can use the suboptimal solution as the initial value and conduct a next estimation by using least squares method.

Algorithm 3: The pseudo-code of IPSO algorithm

---

1 $Iteration = 0$;

2 $Position(Iteration) =$ randomly initialized population;

3 evaluate fitness of each individual in $Position(Iteration)$;

4 while *termination condition = false* **do**

5 $Iteration = Iteration + 1$;

6 $GlobalBestPosition =$ the best position of $IndividualBestPosition$;

7 $Velocities(i, Iteration) = w \times Velocities(i, Iteration - 1) + c_{Self} \times rand_1 \times$
         $(IndividualBestPosition(i) - Position(i, Iteration - 1)) + c_{Social} \times rand_2 \times$
         $(GobalBestPosition - Position(i, Iteration - 1)) \% Velocities(i, Iteration - 1) = \mathbf{0}$ when
         $Iteration = 1$;

8 $Velocities(Iteration) = CrossoverFunction(Velocities(Iteration))$;

9 $Position(i, Iteration) = Position(i, Iteration - 1) + Velocities(Iteration)$;

10 $Value =$ the fitness value of each individual in $Position(Iteration)$;

11 $Individual =$ the fitness value of each individual in $IndividualBestPosition$;

12 $IndividualBestPosition = AcceptanceFunction(Value, Individual, Position(Iteration))$;

13   Verify that the termination condition has been met;

14 end

15 $GlobalBestPosition =$ the best position of $IndividualBestPosition$;

16 $BestValue =$ the fitness value of $GlobalBestPosition$;

17 return $GobalBestPosition, BestValue$

Appendix B. The proof of Theorem 2.1

Proof. We apply the results in (Lukes, 1982) to prove this theorem. Note that (i) The state variables and control variables are non-negative. (ii) The control set Θ is closed and convex. (iii) The optimal system is bounded, which implies the compactness of the optimal control. (iv) The integrand of the objective functional $J(U)$ is convex on Θ. (v) There exist constants $a_1 > 0$, $a_2 > 0$ and $q > 1$ such that the integrand of the objective functional $J(U)$ satisfies

$$J(U) \geq a_1(|u_1|^2 + |u_2|^2)^q + a_2.$$

Therefore, there exists an optimal control $U^* = (u_1^*, u_2^*) \in \Theta$ such that $J(U^*) = \min_{U \in \Theta} J(U)$.

Appendix C. The proof of Theorem 2.2

Proof. We will apply the Pontryagin's Maximum Principle (Pontryagin, 1987) to find the characteristic express of optimal control. To do so, we define the Hamiltonian function as follows

$$G(S, E, A, I, H, R, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = L + \sum_{i=1}^{6} \lambda_i(t) \frac{dX}{dt}, \tag{C.1}$$

for $i = 1, 2 \ldots, 6$ and $X = S, E, A, I, H, R$. Here $\lambda_i(t)$ are adjoint functions and can be derived by $\dot{\lambda}_i(t) = -\partial G / \partial X$ for $i = 1, 2 \ldots, 6$ and $X = S, E, A, I, H, R$. Combining system (2.4) with Hamiltonian function (C.1), we can obtain the adjoint system satisfying the follows differential equations

$$\dot{\lambda}_1(t) = \lambda_1(t)\left[\left(1-u_1(t))\beta(\sigma A+I)\left(\frac{1}{N}-\frac{S}{N^2}\right)\right] - \lambda_2(t)\left[\left(1-u_1(t))\beta(\sigma A+I)\left(\frac{1}{N}-\frac{S}{N^2}\right)\right],$$

$$\dot{\lambda}_2(t) = -\lambda_1(t)(1-u_1(t))\frac{\beta S(\sigma A+I)}{N^2} + \lambda_2(t)\left[\left(1-u_1(t))\frac{\beta S(\sigma A+I)}{N^2}+\alpha\right] - \lambda_3(t)p\alpha$$
$$-\lambda_4(t)(1-p)\alpha,$$

$$\dot{\lambda}_3(t) = -A_1 + \lambda_1(t)(1-u_1(t))\beta S\left(\frac{\sigma}{N}-\frac{\sigma A+I}{N^2}\right) - \lambda_2(t)(1-u_1(t))\beta S\left(\frac{\sigma}{N}-\frac{\sigma A+I}{N^2}\right)$$
$$+\lambda_3(t)(\delta+r_A) - \lambda_4(t)\delta - \lambda_6(t)r_A,$$

$$\dot{\lambda}_4(t) = -A_2 + \lambda_1(t)(1-u_1(t))\beta S\left(\frac{1}{N}-\frac{\sigma A+I}{N^2}\right) - \lambda_2(t)\beta S(1-u_1(t))\left(\frac{1}{N}-\frac{\sigma A+I}{N^2}\right)$$
$$+\lambda_4(t)\left[r_I+d_I+(1+u_2(t))\frac{m(k+aI^x+bH^x)-maxI^x}{(k+aI^x+bH^y)^2}\right]$$
$$-\lambda_5(t)(1+u_2(t))\frac{m(k+aI^x+bH^x)-maxI^x}{(k+aI^x+bH^y)^2} - \lambda_6(t)r_I,$$

$$\dot{\lambda}_5(t) = -\lambda_1(t)(1-u_1(t))\frac{\beta S(\sigma A+I)}{N^2} + \lambda_2(t)(1-u_1(t))\frac{\beta S(\sigma A+I)}{N^2}$$
$$-\lambda_4(t)(1+u_2(t))\frac{mbyIH^{y-1}}{(k+aI^x+bH^y)^2} + \lambda_5(t)\left[\left(1+u_2(t))\frac{mbyIH^{y-1}}{(k+aI^x+bH^y)^2}+r_H+d_H\right]\right.$$
$$-\lambda_6(t)r_H,$$

$$\dot{\lambda}_6(t) = -\lambda_1(t)(1-u_1(t))\frac{\beta S(\sigma A+I)}{N^2} + \lambda_2(t)(1-u_1(t))\frac{\beta S(\sigma A+I)}{N^2}.$$

$$(C.2)$$

The transversality conditions are

$$\lambda_i(t_f) = 0, i = 1,...,6.\qquad\qquad (C.3)$$

State system (2.4), adjoint system (C.2) and transversality conditions (C.3) forms an optimal control problem. Thus, one can have the following theorem. The optimal control can be obtain by solving the following equations

$$\frac{\partial G}{\partial u_1} = B_1 u_1(t) + \lambda_1\frac{\beta S(\sigma A+I)}{N} - \lambda_2\frac{\beta S(\sigma A+I)}{N} = 0,$$
$$\frac{\partial G}{\partial u_2} = B_2 u_2(t) - \lambda_4\frac{mI}{k+aI^x+bH^y} + \lambda_5\frac{mI}{k+aI^x+bH^y} = 0.$$

Thus, one has

$$u_1^c = \frac{\beta S(\sigma A+I)}{N}\frac{\lambda_2-\lambda_1}{B_1}, \quad u_2^c = \frac{mI(\lambda_4-\lambda_5)}{(k+aI^x+bH^y)B_2},$$

which together with the upper and the lower bounds of $u_1(t)$ and $u_2(t)$, derive the characteristic express of optimal control.

References

Acuña-Zegarra, M. A., Díaz-Infante, S., Baca-Carrasco, D., & Olmos-Liceaga, D. (2021). COVID-19 optimal vaccination policies: A modeling study on efficacy, natural and vaccine-induced immunity responses. *Mathematical Biosciences, 337*, Article 108614.

Chitnis, N., Hyman, J. M., & Cushing, J. M. (2008). Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. *Bulletin of Mathematical Biology, 70*(5), 1272–1296.

Diekmann, O., Heesterbeek, J., & Metz, J. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology, 28*(4), 365–382.

Dreessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences, 180*(1–2), 29–48.

Egger, M., Johnson, L., Althaus, C., Schoni, A., Salanti, G., Low, N., & Norris, S. L. (2017). Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies. *Fl000Res, 29*(6), 1586.

Gupta, N., Dhamija, S., Patil, J., & Chaudhari, B. (2021). Impact of COVID-19 pandemic on healthcare workers. *Industrial Psychiatry Journal, 30*(Suppl 1), S282–S284.

Hansen, E., & Day, T. (2011). Optimal control of epidemics with limited resources. *Journal of Mathematical Biology, 62*, 423—451.

Hao, X. J., Cheng, S. S., Wu, D. G., Wu, T. C., Lin, X. H., & Wang, C. L. (2020). Reconstruction of the full transmission dynamics of COVID-19 in wuhan. *Nature, 584*(7821), 420—424.

Hebei Provincial Health Commission. (2022). http://wsjkw.hebei.gov.cn/. (Accessed 19 October 2022).

Hubei Provincial Health Commission. (2022). http://wjw.hubei.gov.cn/. (Accessed 19 October 2022).

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature, 451*(7181), 990—993.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942—1948). IEEE.

KhudaBukhsh, W. R., Bastian, C. D., Wascher, M., Klaus, C., Sahai, S. Y., Weir, M. H., Kenah, E., Root, E., Tien, J. H., & Rempała, G. A. (2023). Projecting COVID-19 cases and hospital burden in Ohio. *Journal of Theoretical Biology, 561*, Article 111404.

Lenhart, S., & Workman, J. T. (2007). *Optimal control applied to biological models*. CRC press.

Lukes, D. L. (1982). *Differential equations: Classical to controlled* (Vol. 162). New York: Academic press.

Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science, 368*(6492), 742—746.

Penn, M. J., & Donnelly, C. A. (2023). Optimality of maximal—effort vaccination. *Bulletin of Mathematical Biology, 85*(8), 73.

Plank, M. J. (2022). Minimising the use of costly control measures in an epidemic elimination strategy: A simple mathematical model. *Mathematical Biosciences, 351*, Article 108885.

Pontryagin, L. S. (1987). *Mathematical theory of optimal processes*. CRC press.

Shreffler, J., Petrey, J., & Huecker, M. (2020). The impact of COVID-19 on healthcare worker wellness: A scoping review. *Western Journal of Emergency Medicine, 21*(5), 1059—1066.

Sun, G.-Q., Wang, S.-F., Li, M.-T., Li, L., Zhang, J. Z., Zhang, W., Jin, Z., & Feng, G.-L. (2020). Transmission dynamics of COVID-19 in Wuhan, China: Effects of lockdown and medical resources. *Nonlinear Dynamics, 101*, 1981—1993.

Taboe, H. B., Asare-Baah, M., Iboi, E. A., & Ngonghala, C. N. (2023). Critical assessment of the impact of vaccine-type and immunity on the burden of COVID-19. *Mathematical Biosciences, 360*, Article 108981.

Tang, B., Bragazzi, N. L., Li, Q., Tang, S., Xiao, Y., & Wu, J. (2020). An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infect. Dis. Model., 5*, 248—255.

Tang, B., Wang, X., Li, Q., Bragazzi, N., Tang, S., Xiao, Y., & Wu, J. (2020). Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine, 9*, 462.

Wang, X., Li, Q., Sun, X., He, S., Xia, F., Song, P., Shao, Y., Wu, J., Cheke, R. A., Tang, S., & Xiao, Y. (2021). Effects of medical resource capacities and intensities of public mitigation measures on outcomes of COVID-19 outbreaks. *BMC Public Health, 21*(1), 605.

Wu, J., Leung, K., & Leung, G. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in wuhan, China: A modelling study. *Lancet, 9*(10225), 689—697.

Zhang, M., Xiao, J., Deng, A., Zhang, Y., Zhuang, Y., Hu, T., Li, J., Tu, H., Li, B., Zhou, Y., et al. (2021). Transmission dynamics of an outbreak of the COVID-19 delta variant B. 1.617. 2—Guangdong province, China, May—June 2021. *China CDC Weekly, 3*(27), 584.

Zhou, W., Wang, A., Wang, X., Cheke, R. A., Xiao, Y., & Tang, S. (2020). Impact of hospital bed shortages on the containment of COVID-19 in Wuhan. *International Journal of Environmental Research and Public Health, 17*(22), 8560.