Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 207 (2023) 533–555
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

Optimal strategies for coordinating infection control and socio-economic activities

Tangjuan Li, Yanni Xiao*

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, 710049, PR China

Received 30 September 2022; received in revised form 10 January 2023; accepted 13 January 2023
Available online 20 January 2023

Abstract

It becomes challenging to identify feasible control strategies for simultaneously relaxing the countermeasures and containing the Covid-19 pandemic, given China's huge population size, high susceptibility, persist vaccination waning, and relatively weak strength of health systems. We propose a novel mathematical model with waning of immunity and solve the optimal control problem, in order to provide an insight on how much detecting and social distancing are required to coordinate socio-economic activities and epidemic control. We obtain the optimal intensity of countermeasures, i.e., the dynamic nucleic acid screening and social distancing, under which the health system is functioning normally and people can engage in a certain level of socio-economic activities. We find that it is the isolation capacity or the restriction of the case fatality rate (CFR) rather than the hospital capacity that mainly determines the optimal strategies. And the solved optimal controls under quarterly CFR restrictions exhibit oscillations. It is worth noticing that, if without considering booster or very low booster rate, the optimal strategy is a "on–off" mode, alternating between lock down and opening with certain social distancing, which reflects the importance and necessity of China's static management on a certain area during Covid-19 outbreak. The findings suggest some feasible paths to smoothly transit from the Covid-19 pandemic to an endemic phase.

Keywords: COVID-19; Epidemiological modeling; Optimal control; Nucleic acid screening; Social distancing

1. Introduction

Since the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), many countries have experienced the multiple outbreaks of infection, accompanied by lockdown, opening and re-lockdown and re-opening and etc. [12,14]. A series of strong prevention and control measures have been taken to effectively keep transmission suppressed for two more years in China [37,42]. However, enforcing these countermeasures for long periods of time can have substantial negative socio-economic impacts. Recently, many developed western countries have almost completely canceled off travel restrictions and other interventions given high coverage of vaccination and natural transmission [26]. Faced with the pressure of imported cases and spreading domestic outbreaks, China's ticking to the dynamic zero-COVID policy becomes great challenging. It is urgent to propose a feasible and novel

* Corresponding author.
E-mail address: yxiao@mail.xjtu.edu.cn (Y. Xiao).

https://doi.org/10.1016/j.matcom.2023.01.017

regime to coordinate the control of SARS-CoV-2 pandemic and socio-economic activities for mainland China, given the waning of immunity and little natural transmission and massive populations.

Mathematical modeling has played a key role in examining the effectiveness of intervention measures and informing the key parameters that affect the disease infection. In response to the SARS-CoV-2 pandemic, some researchers considered the various scenarios with different intensities of nonpharmaceutical measures under which a stable equilibrium with low case numbers was attained [13]. Lauro et al. examined the optimal timing of one-short interventions for epidemic control by observing total attack rate, peak prevalence and average time of infection [17]. Some researchers have investigated optimal control strategies based on Pontryagin's maximum principle [11,21,24,28,33,38]. Bin et al. proposed an on–off (also called bang–bang) social distancing policy for mitigating a second wave [7]. Kumar Das et al. framed an optimal control problem to analyze the effectiveness of diagnosis based on traced contacts made by a confirmed COVID patient [16]. Lu et al. proposed a novel two-stage epidemic model with a dynamic control strategy to seek appropriate control strategies to minimize the control cost and ensure the normal operation of society [30]. The optimal control questions usually involve the minimization of the number of infected individuals as well as the economic cost of the countermeasures by looking for the optimal intensification of nonpharmaceutical interventions (NPIs) [7,24,33]. The optimal vaccination scheme has also been considered by many researchers in order to answer who should be vaccinated first. Matrajt et al. presented optimal age-targeted vaccination policies, and obtained that when minimizing deaths, it is optimal to vaccinate the elderly (younger age groups) for low (high) vaccine efficacy [31]. Grundel et al. coordinated both social distance and vaccination to mitigate SARS-CoV-2 outbreaks by formulating an age-differentiated compartmental model in an attempt to answer how much social distancing measures can be relaxed for various coverage of vaccination [22]. However, as immunity is waning and the lack of clinically effective drugs, simultaneous optimal control of cost and social distancing as well as avoiding run of medical resources becomes challenging and falls within the scope of this study.

Our main purpose is not to eliminate the infection, but rather to demonstrate how to coordinate disease control and social distancing, and to provide some guiding principles of not only benefiting public health and psychological well-being but also profiting the economy, which will apply to our national policy on mitigating COVID-19 pandemic. To this end, we propose a novel mathematical model with waning of immunity and examine the optimal contact rate and detection rate such that cost is minimized, and as well as the number of hospitalized/isolated individuals no greater than the hospital/isolation capacity. In order to transit from COVID-19 pandemic to endemic phase, we further include the case fatality rate (CFR) as a restriction. Whereafter, the effect of vaccine on controlling the spread of COVID-19 is also analyzed. Our results have essential implications for ongoing discussion of "dynamics zero" policy and provide insights into preparedness for future pandemics.

2. A compartmental model and optimal control problems

In this section, we establish a dynamic model tailored to COVID-19 by incorporating social distancing, nucleic acid screening and vaccination. Considering fairly high vaccination rate and the fact that immunity improves after vaccination and/or natural infection but wanes as antibodies decay [1], we divided the population into those with high immunity provided by booster or prior infection and low immunity. In fact, there is evidence showing that individuals with high immunity have relatively low risk of infection, and they are less likely to get serious illness if they do get it [15,20]. We consider a SEIS-type model, then, the population is divided into susceptible individuals (S_i), exposed individuals (E_i), infectious individuals with either mild symptoms (I_{im}) who may only need to admit to the isolated compartment (Q_i), or severe symptoms (I_{is}) who need to be hospitalized for treatment, i.e., admitted to hospitalized compartment (H_i), and death compartment (D_i). Here we use $i = 1$ or 2 to represent individuals with low or high immunity. We assume that mild cases that go undetected eventually recover on their own, and those who are detected are immediately isolated and then recovered after a period of isolation, but severe cases are hospitalized once detected and cannot cause further transmission, in which they may either die or recover. All infected individuals become susceptible with high immunity as soon as they recover. The model flow diagram is

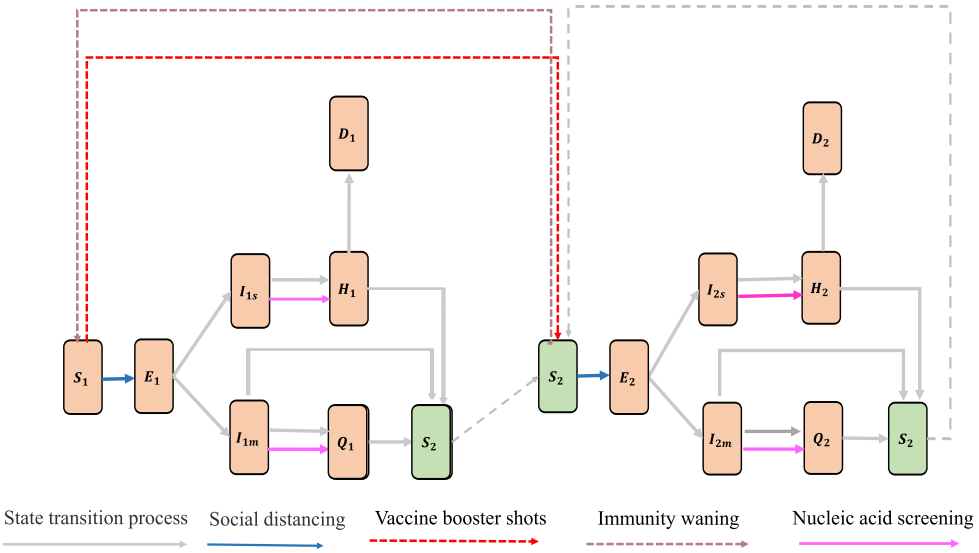

Fig. 1. Flow diagram for a dynamic model tailored to COVID-19 with social distancing, nucleic acid screening, and vaccination.

shown in Fig. 1 and the model equations are as follows.

$$\begin{cases} \frac{dS_1(t)}{dt} = \omega_2 S_2 - \nu_1 S_1 - \sum_{j=1}^{2} c(t)\hat{c}\beta_1 S_1(\varphi I_{jm} + I_{js}), \\ \frac{dS_2(t)}{dt} = -\omega_2 S_2 + \nu_1 S_1 - \sum_{j=1}^{2} c(t)\hat{c}\beta_2 S_2(\varphi I_{jm} + I_{js}) \\ \qquad + \sum_{j=1}^{2}(\gamma_Q Q_j + \gamma_{H_j} H_j + \gamma_{I_m} I_{jm}), \\ \frac{dE_i(t)}{dt} = \sum_{j=1}^{2} c(t)\hat{c}\beta_i S_i(\varphi I_{jm} + I_{js}) - \sigma E_i, \\ \frac{dI_{im}(t)}{dt} = (1-\rho_i)\sigma E_i - \delta(t) I_{im} - \xi_{I_m} I_{im} - \gamma_{I_m} I_{im}, \\ \frac{dI_{is}(t)}{dt} = \rho_i \sigma E_i - \delta(t) I_{is} - \xi_{I_s} I_{is}, \\ \frac{dQ_i(t)}{dt} = \delta(t) I_{im} + \xi_{I_m} I_{im} - \gamma_Q Q_i, \\ \frac{dH_i(t)}{dt} = \delta(t) I_{is} + \xi_{I_s} I_{is} - \gamma_{H_i} H_i - d_i H_i, \\ \frac{dD_i(t)}{dt} = d_i H_i, \end{cases} \tag{2.1}$$

where S_i, E_i, I_{im}, I_{is}, Q_i, H_i and D_i $(i = 1, 2)$ are considered as the corresponding population proportions, ν_1 represents the rate of booster vaccination, ω_2 represents immunity waning rate, $\hat{c}$ is the baseline number of daily contacts before the COVID-19 outbreak, β_i is the infection risk given a contact between a susceptible individual S_i and a serve case, ρ_i denotes the probability of infected individuals who are with severe symptoms, ξ_{I_m} or ξ_{I_s} is the symptom-driven testing rate for mild or serve patients. $\delta(t) : [0, \infty] \longrightarrow [0, 1]$ is detection rate due to nucleic acid screening and $c(t) : [0, \infty] \longrightarrow [0, 1]$ represents social distancing strategy, here $c(t) = 1$ means no social distancing, $c(t) = 0$ means complete lockdown. Functions $c(t)$ and $\delta(t)$ represent the intensities of control measures we need to solve, and we assume that those two countermeasures are kept constant over one week, which are realistic and feasible in practice.

When no NPIs are taken, i.e., $c(t) = 1$ and $\delta(t) = 0$, the effective reproduction number gives

$$R_t(t) = \left((1-\rho_1)\frac{\varphi\hat{c}\beta_1}{\gamma_{I_m} + \xi_{I_m}} + \rho_1\frac{\hat{c}\beta_1}{\xi_{I_s}}\right) S_1(t) + \left((1-\rho_2)\frac{\varphi\hat{c}\beta_2}{\gamma_{I_m} + \xi_{I_m}} + \rho_2\frac{\hat{c}\beta_2}{\xi_{I_s}}\right) S_2(t). \tag{2.2}$$

In order to coordinate socio-economic activities and epidemic prevention and control, we need long-term, sustainable strategies to contain the spread. Then, we will establish different optimal control problems (OCPs) based on model (2.1) to determine the optimal combination of social distancing strategy (contact rate) and detection rate within a given period to minimize the costs of taking countermeasures and maintain strict cap on the hospitalized beds and beds used for isolation. Let Q_{max} and H_{max} represent the capacities of the isolated resources and the

hospital beds that can accept COVID-19 patients respectively. Then the OCP is then given by

$$\begin{aligned}
\min_{c,\delta} \quad & \int_0^T (1-c(t))^2 + \kappa\delta^2(t)\sum_{j=1}^2 (S_j + E_j + I_{jm} + I_{js})\,dt, && (a)\\
\text{subject to} \quad & \text{model (2.1)}, && (b)\\
& 0 \le c(t) \le 1, && (c)\\
& 0 \le \delta(t) \le 1, && (d)\\
& N\cdot(H_1(t)+H_2(t)) \le H_{max}, && (e)\\
& N\cdot(Q_1(t)+Q_2(t)) \le Q_{max}, && (f)\\
& c(t) = c(n\Delta t), t\in[n\Delta t,(n+1)\Delta t], n=0,1,\ldots,K, && (g)\\
& \delta(t) = \delta(n\Delta t), t\in[n\Delta t,(n+1)\Delta t], n=0,1,\ldots,K, && (h)
\end{aligned} \tag{2.3}$$

where (a) is the objective function, which includes the normalized cost of taking social distancing to reduce contacts (the first term) and the relative cost of taking nucleic acid screening to block some transmission (the second term), and constant κ is the cost regulator of taking nucleic acid screening. (c) and (d) represent restrictions on control measures. (e) is to prevent exceeding the maximum number of hospital beds that can accept COVID-19 patients. (f) indicates that the number of quarantined persons does not exceed isolation capacity. Note that N is the total number of individuals and hence the left sides of (e) and (f) represent the numbers of individuals who seek the hospital beds or to be quarantined. (g) and (h) mean that countermeasures are kept constant over a given period (say, Δt may be a week), K (or T) is the total number of weeks (days) considered. And in the objective function, we use $(1-c(t))^2$ and $\delta^2(t)$ to discourage very strict contact restrictions and very frequent testing in order to avoid occurring of bang–bang/singular controls, as Refs. [19,22,35] suggested.

Omicron has demonstrated an increased transmissibility relative to Delta but possibly lower clinical severity than Delta [10], and consequently ones hope to transit from COVID-19 pandemic to endemic phase. If the mortality of COVID-19 infections is similar to that of influenza, we may coexist with COVID-19 through coordinating socioeconomic development and suitable/feasible measures. Then we further include the case fatality rate (CFR) at the end of our considered period as an additional restriction on the basis of the OCP (2.3).

In order to calculate the CFR, let us denote the cumulative number of confirmed cases by $C(t)$, then we have

$$\frac{dC}{dt} = \sum_{j=1}^{2} \left(\delta(t)(I_{im}+I_{is}) + \xi_{I_m} I_{im} + \xi_{I_s} I_{is}\right). \tag{2.4}$$

So the CFR within T days is $\frac{D_1(T)+D_2(T)}{C(T)}$. The corresponding optimal control problem is

$$\begin{aligned}
\min_{c,\delta} \quad & \int_0^T (1-c(t))^2 + \kappa\delta^2(t)\sum_{j=1}^2 (S_j + E_j + I_{jm} + I_{js})\,dt, && (a)\\
\text{subject to} \quad & \text{model (2.1) and (2.4)}, && (b)\\
& 0 \le c(t) \le 1, && (c)\\
& 0 \le \delta(t) \le 1, && (d)\\
& N\cdot(H_1(t)+H_2(t)) \le H_{max}, && (e)\\
& N\cdot(Q_1(t)+Q_2(t)) \le Q_{max}, && (f)\\
& \frac{D_1(T)+D_2(T)}{C(T)} \le r_{max}, && (g)\\
& c(t) = c(n\Delta t), t\in[n\Delta t,(n+1)\Delta t], n=0,1,\ldots,K, && (h)\\
& \delta(t) = \delta(n\Delta t), t\in[n\Delta t,(n+1)\Delta t], n=0,1,\ldots,K, && (i)
\end{aligned} \tag{2.5}$$

where (2.5)(g) indicates that the yearly CFR is less than a given level r_{max}. Note that both OCP (2.3) and (2.5) are optimal control problems with pure state constraints [8,19] since constraints (2.3)(e–f) and (2.5)(e–f) do not include the controls. We use the direct method [32,34,38], which appends to the Hamiltonian a penalty multiplier that directly multiplies the constraints, to handle pure state constraints. See Appendix A for the corresponding necessary conditions [34] and fast direct multiple shooting algorithms [18] to solve our OCPs numerically.

3. Numerical studies

Throughout the simulations, we choose parameters tailored to the COVID-19 outbreak by Omicron variant in Xi 'an, the capital city of Shaanxi Province, China. However, we expect that the conclusions are suitable to other developed countries. We get the permanent population of Xi 'an, i.e., $N = 12952907$, and the total

number of hospital beds (which is 66,506) according to the Statistical yearbook of Xi 'an [2,3], and we obtain $H_{max} = 13,700 = 20\% * 66,506$ by assuming that 20% of beds can be used for COVID-19 patients. We let $Q_{max} = 42,000$ as the baseline quarantined capacity according to the ability during previous outbreak [4], in which more than 42,000 people were quarantined in Xi'an city during 2021 outbreak in Shaanxi province. We hypothesize that people with high immunity will transit to low immunity after six month due to the wane of immunity [1], that is $\omega_2 = 1/180$, and high immunity induce an 80% reduction in being infection. Symptom-driven testing automatic rate of serve cases is $1/3.8$ [45], which means that it takes an average of 3.8 days for a serve case to see a doctor. We can further assume that a mild case takes 5 days to see a doctor considering its mild symptoms, i.e., $\xi_{I_m} = 1/5$.

Here, the parameters describing disease progression meet SARS-CoV-2 Omicron variant's characteristics, detailed meanings and values for parameters are shown in Table 1. We assume that the initial number of people with high immunity is about one-fifth of the population size and the cost regulator κ of nucleic acid screening is equal to 2, that is, the cost of nucleic acid screening is twice as much as the cost of social distancing. There is an evidence that shows the basic reproduction number of Delta variant is 3.2 (95% CI 2.0–4.8) according to outbreak in Guangdong Province, China [46], and the basic reproduction number of Omicron is about 1.5–2.5 times that of Delta [6,25]. Then we can take $\beta_1 = 0.2$ as baseline value to make the effective reproduction number ($R_t(0)$) in the beginning to be 6.55 in a population where the number of people with high immunity is one-fifth of the population size, or 7.55 in a population with low immunity, i.e. a population that has never experienced the COVID-19 or got a booster shot, see details in Table 1. We would like to develop time-varying strategies over one year, so we fix $K = 52$ and $T = 364$.

We numerically solve the optimal problem (2.3) by making use of fast tailored direct optimal control solvers [39]. In particular, we compute the gradients of the left-hand sides of the inequality constraints (2.3)(e) and (2.3)(f) using a continuous forward method. The simulation of all optimal control problems is completed in MATLAB with the help of open OCL-open optimal control library [27]. We will measure the impact of optimal strategies on four quantities of interest: the cumulative numbers of deaths ($D_1(T) + D_2(T)$), confirmed cases ($C(T)$), total costs of taking countermeasures within a year (say T) and the effective reproduction numbers ($R_t(T)$). We also focus on the variation in solved optimal social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) with hospital beds and isolated resources occupancy, the effective reproduction number (R_t) and the evolution of the epidemic under the optimal controls in every case. Note that the term "optimal solutions" ("minimized") in the following refers to the notion that the best possible disease control policy is identified from a set of "local optimal solutions" ("local minimized") that satisfy optimality conditions since iterative schemes often only guarantee local optimality [27,39].

3.1. Optimal control problem with limited hospital and isolation capacity

We initially let $H_{max} = 13,700$ and $Q_{max} = 42,000$ (named as *the benchmark case*) to get the optimal controls and optimal solutions, shown in Fig. 2. It shows that at approximately 230th day, quarantine resources are exhausted up till about the 315th day, $c^*(t)$ and $\delta^*(t)$ stay around 0.7708 and 0.1638 during this period, respectively. And hospital beds are never used up and the proportion of susceptible individuals with high immunity remains at 0.94 for most of the considered period. Note that the disease infection takes off once the controls end at the end of our considered period.

If we simply consider a constrain of the shortage of hospital beds, i.e., without constraint (2.3)(f), it follows from Fig. 2 that at approximately 90th day, hospital beds are exhausted up till the end, $c^*(t)$ and $\delta^*(t)$ stay around 0.94 and 0.07 during considered period, respectively, which means a single constrain may allow the relatively loose strategies, compared to the optimal controls under two constrains. Note that in such scenario relatively large infections may be caused but with no run of hospital resources. Further, we notice that the obtained optimal controls $c^*(t)$ and $\delta^*(t)$ are dynamic, any combination of constant controls are not indeed as well as our solved ones in terms of contribution to our objectives. See Fig. B.1 in Appendix B for details. And small changes in the probability of severe symptoms (ρ_1, ρ_2) have little impact on the required intensity of prevention and control measures, but have a significant impact on the number of hospitalizations and deaths. See Fig. B.2 and Table B.1 in Appendix B for details.

To examine the variation in the optimal intensifies of interventions with the capacities of isolation resources and hospital beds, we conducted the following scenarios: cases with isolation resources expanded by half (*Case I*: $H_{max} = 13,700$, $Q_{max} = 63,000$), doubled (*Case II*: $H_{max} = 13,700$, $Q_{max} = 84,000$) and a case with the capacity of hospital beds doubled (*Case III*: $H_{max} = 13,700 \times 2$, $Q_{max} = 42,000$), shown in Fig. 3. It can be seen

Table 1
Detailed meanings and values for parameters describing disease progression meet SARS-CoV-2 Omicron variant's characteristics.

Parameters	Description	Value	Resource
ν_1	The rate of booster vaccination	0.1	Assumed
ω_2	The rate of waning of immunity	$1/180$	[1]
$\hat{c}$	The baseline number of daily contacts before the outbreak	14.6	[44]
β_i	Risk of infection per contact between a susceptible population S_i and a severe case	$\beta_1 = 0.2$ $\beta_2 = 0.2\beta_1$	Assume
φ	Modification factor of the risk of infection per contact between a susceptible population S_i and a mild case	1	Assume
$1/\sigma$	Incubation/latent period	3	[9]
ξ_{I_m}	Symptom-driven testing automatic rate of mild cases	$1/5$	Assumed
ξ_{I_s}	Symptom-driven testing automatic rate of serve cases	$1/3.8$	[45]
ρ_i	The probability of severe symptoms	$\rho_1 = 0.01$ $\rho_2 = 0.007$	[15]
γ_{I_m}	Recovery rate of I_{im}	$1/5.68$	[15]
γ_Q	Recovery rate of Q_i	$1/7$	Assumed
γ_{H_i}	Recovery rate of H_i	$\gamma_{H_1} = 0.1673$ $\gamma_{H_2} = 0.1636$	[15,23]
d_i	Mortality rate of H_i	$d_1 = 0.0145$ $d_2 = 0.0182$	[15,23]

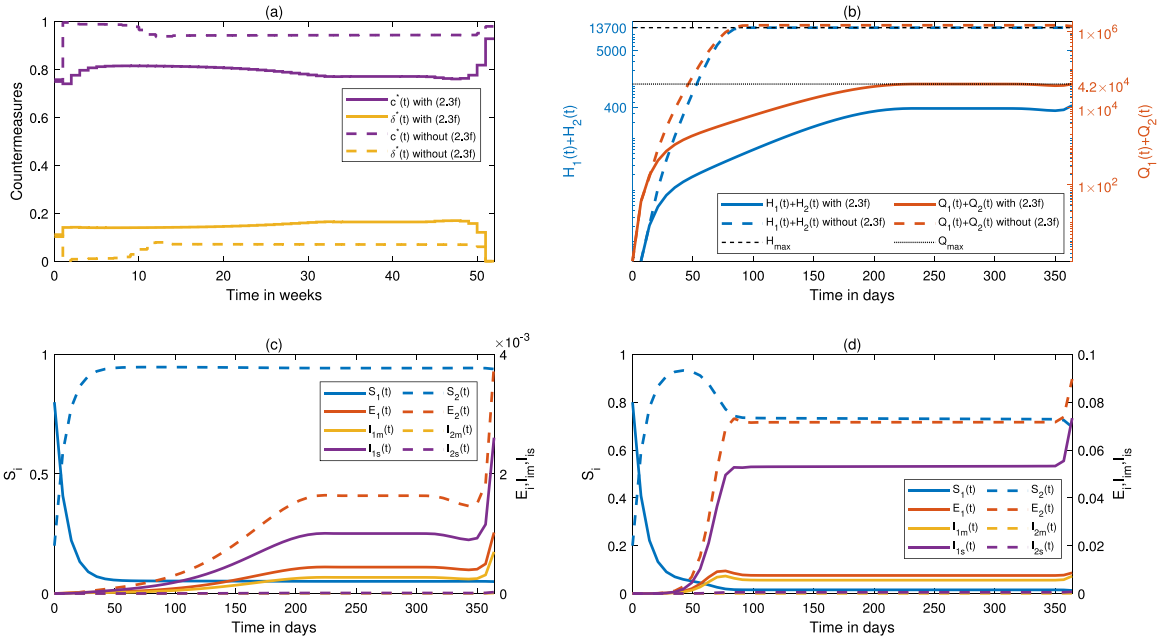

Fig. 2. Results for the OCP (2.3) of *The benchmark case* with or without constraint (f). The optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) is shown in (a), where the solid line corresponds to the case considering constraint (2.3)(f), and the dotted line represents the case where constraint (2.3)(f) is not considered. The variation in hospital beds and isolated resources occupancy with a logarithmic scale on the vertical axis is shown in (b), where the solid line corresponds to the case considering constraint (2.3)(f), and the dotted line represents the case where constraint (2.3)(f) is not considered. The evolution of the epidemic under the optimal controls with or without constraint (2.3)(f) is shown in (c) or (d) respectively, where the left vertical axis corresponds to evolution of susceptible individuals (S_i), and the right vertical axis corresponds to evolution of latent (E_i), mild (I_{im}), and severe (I_{is}) patients ($i = 1, 2$).

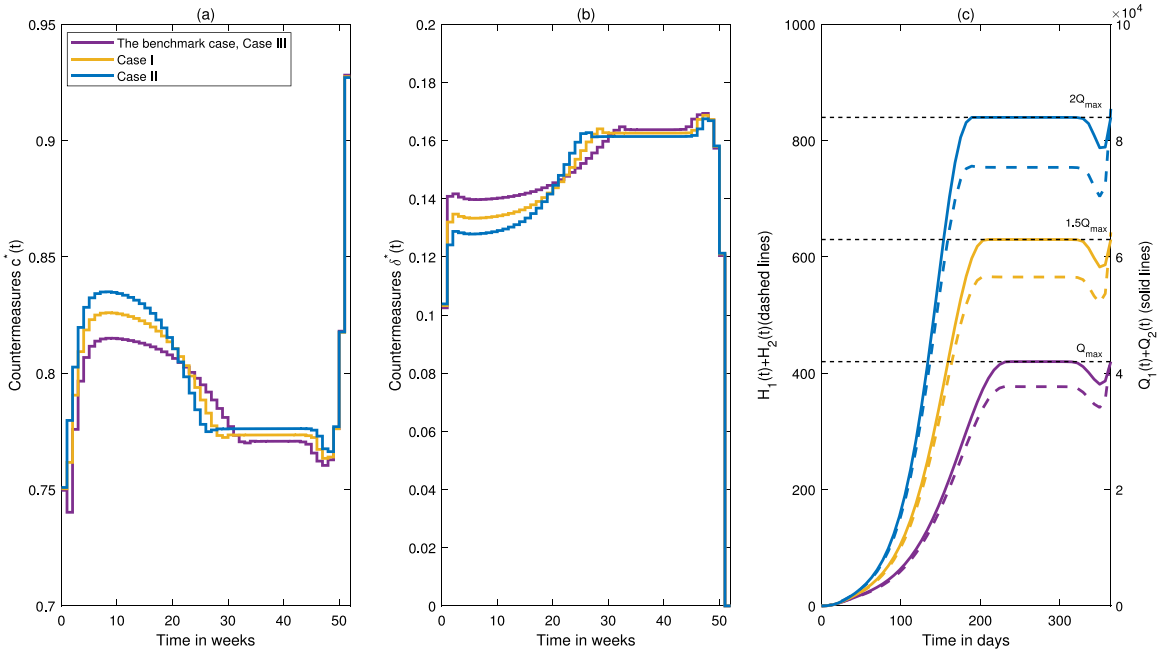

Fig. 3. Results for the OCP (2.3) under *The benchmark case* and *Cases I-III*. (a) and (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) under those cases; (c): The variation in hospital beds and isolated resources occupancy under those cases. Note that the curves of the benchmark case and case III coincide.

that with the increase of isolation resources, the optimal contact rate increases and detection rate decreases during the time when the quarantined resource has been fully occupied, and the time to reach the upper limit of isolation resources advances. This implies that given our control purpose, more isolation resources can allow looser contact restriction and less detection. Further, we list the corresponding quantities of interest for *case I - III* in Table 2. This demonstrates that with the increase of isolation resources, the cumulative number of deaths further increases while the total cost of taking countermeasures decreases (note that the cost of increasing quarantined resources is not considered here). That is because the more quarantine resources, the larger the scale of the infection allowed.

It is interesting to note that for *the benchmark case* and *case III*, the optimal controls coincide and the corresponding quantities of interest are the same. That is to say, when the hospital capacity is increased, the optimal intensity of control measures and the scale of infections remain the same as those with baseline capacity. This means that optimal countermeasures are primarily influenced by the capacity of quarantine resources, that is because only a small number of infected individuals needs to be hospitalized, which corresponds to the fact that individuals with high (low) immunity have a 0.7%(1%) chance of becoming seriously ill and being hospitalized once infected [15].

3.2. Optimal control problem involving influenza-based case-fatality constraints

On the basis of CDC statistics on estimated influenza disease burden for past seasons [5], the annual CFR of influenza for the period of 2015–2019 was approximately 0.09% to 0.14%. We then take this information as the case fatality restrictions when considering the OCP (2.5). For comparison, we still assume that the respective capacities of isolation and hospital beds are kept at baseline levels.

To analyze the influence of additional CFR restriction on optimal controls, we analyze some cases with different CFR restrictions. It follows from Fig. 4 that as the CFR restriction becomes loose (i.e., r_{max} increases), the optimal contact rates increase (decreases) in the early (late) stage of a year, and the optimal detection rates change in exactly the opposite way. This implies that the stricter the CFR restriction, the stricter control measures needed to be taken in the early stage, while the relatively more relaxed control measures allowed in the later stage. It corresponds to the fact that strict control measures in the early stage lead to relatively few new infections and hence the low CFR, the control measures required in late stage will then be relatively relaxed.

Table 2
The cumulative number of deaths ($D_1(T)+D_2(T)$), the cost of taking countermeasures in a year and the effective reproduction number at time T of OCP (2.3) under different cases.

	$D_1(T)+D_2(T)$	The total cost	$R_t(T)$
The benchmark case	1381	32.54	1.85
case I	2208	31.55	1.83
case II	3070	30.72	1.82
case III	1381	32.54	1.85
$v_1 = 0.12$	1189	28.7683	1.8029
$v_1 = 0.15$	964	24.9918	1.7536
$v_1 = 0.12$ and $Q_{max} = 21000$	395	30.0220	1.8180
$v_1 = 0.001$	936	133.5674	\
$v_1 = 0$	1057	153.0742	\

The number of people in this table is rounded to an integer.
The total cost includes the normalized cost of taking social distancing and the relative cost of taking nucleic acid screening in a year.

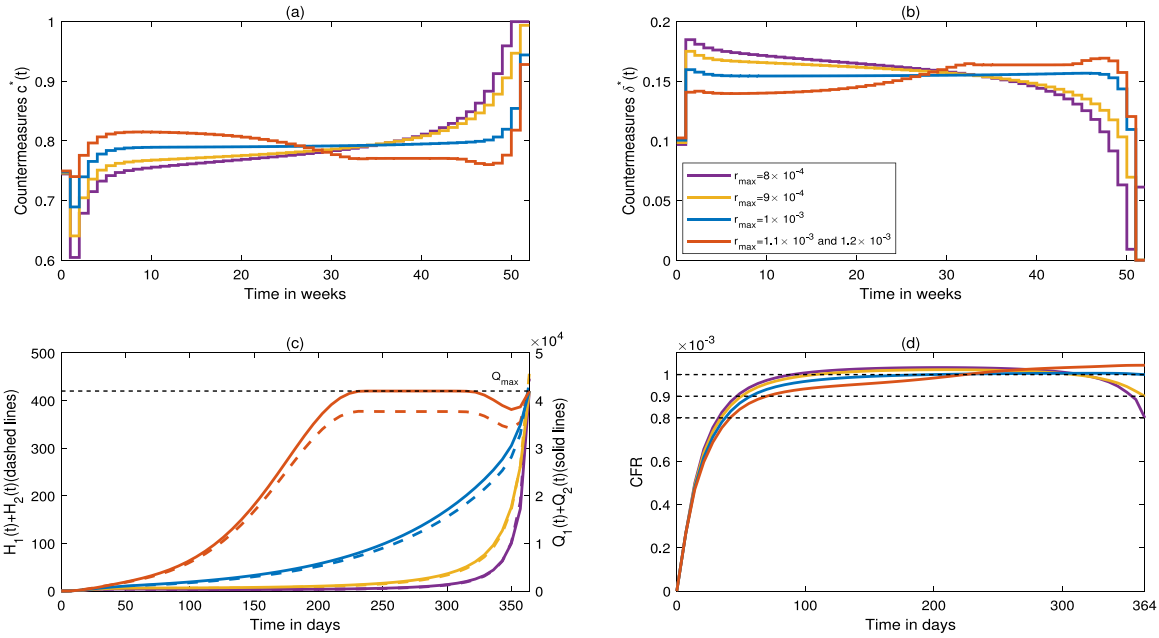

Fig. 4. Results for the OCP (2.5) under different cases. (a) and (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) under different cases; (c): The variation in hospital beds and isolated resources occupancy under different cases; (d): the case fatality rate within t days under different cases. Note that the curves of the case when $r_{max} = 1.2 \times 10^3$ and $r_{max} = 1.1 \times 10^3$ coincide. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In particular, for relatively low r_{max} ($r_{max} = 1 \times 10^{-3}$ or even less), the number of isolated individuals (Fig. 4(c)) and the yearly CFR (Fig. 4(d)) just reach their upper limits at the end of the control period. For relatively large r_{max} ($r_{max} = 1.1 \times 10^{-3}$ and $r_{max} = 1.2 \times 10^{-3}$), the optimal controls and the corresponding hospital beds and isolated resources occupancy are similar to those with the benchmark case (see details in Figs. B.3–B.4). This indicates that for relatively (relaxed) strict CFR restriction, i.e., r_{max} is relatively (large) small, the optimal controls are mainly determined by (the isolation resources) the limit of CFR. Note that when the yearly CFR restriction is strict, the instantly calculated CFR may exceed this limit, that is because our subjective conditions only limit the average CFR in a year. And hospital beds are never used up for various restrictions of CFR (Fig. 4(c)).

The corresponding quantities of interest under these scenarios are given in Table 3. We can get that for relatively large r_{max} ($r_{max} = 1.1 \times 10^{-3}$ or even greater), the values of each indicator are the same as those in the benchmark

Table 3
The cumulative number of deaths and confirmed cases within one year, the effective reproductive number at $T = 364$ and the total cost of taking countermeasures in a year of OCP (2.5) under different cases.

r_{max}	$D^1(T) + D^2(T)$	$C(T)$	$R_t(T)$	The total cost
1.2×10^{-3}	1381	1.32×10^6	1.8511	32.54
1.1×10^{-3}	1381	1.32×10^6	1.8511	32.54
1×10^{-3}	490	4.90×10^5	1.8500	32.80
9×10^{-4}	178	1.98×10^5	1.8467	33.54
8×10^{-4}	103	1.29×10^5	1.8522	34.62

The number of people in this table is rounded to an integer.
The total cost includes the normalized cost of taking social distancing and the relative cost of taking nucleic acid screening in a year.

case. And with the further decrease of r_{max} ($r_{max} = 1 \times 10^{-3}$ or even less), the cumulative number of deaths and confirmed cases decrease, while total costs of taking control measures increase. It can also be verified that the infection size and the cost under optimal countermeasures are mainly affected by isolation resources (the yearly CFR restriction) when the yearly CFR restriction is loose (strict).

Moreover, if the yearly CFR restriction is relaxed from 9×10^{-4} to 1×10^{-3}, the cumulative number of deaths will increase 1.8 times (Table 3). This motivates us to consider scenarios in which the constraint is relaxed somewhat, and consequently the yearly CFR restriction is replaced by the quarterly CFR restriction. Then a multi-stage problem (C.3) is established based on the OCP (2.3) and allows you to implement discrete events where a jump in the state trajectory happens [27], to seek the optimal prevention and control measures (See details in Appendix C).

In the following we initially examine the impact of seasonal restrictions on optimal controls (see Fig. 5) and corresponding deaths (see case 0 in Table C.2). It is worth noting that if the maximum allowable CFR is 9×10^{-4}, the optimal controls given quarterly CFR limit exhibit seasonal oscillations (blue curves in Fig. 5(a) and (b)) compared to yearly CFR limit (yellow curves in Fig. 4(a) and (b)). To examine whether it is possible to relax a certain quarter or several quarters of the constraint to reduce the scale of deaths, as opposed to relaxing the yearly constraint directly, we consider the possible relaxation scenarios listed in Table C.1 and obtain the cumulative number of deaths/confirmed cases listed in Table C.2. We find that the scale of death within a year of most cases in Table C.2 is smaller than the case with yearly CFR limit $r_{max} = 1 \times 10^{-3}$, especially cases 1, 2 and 5 in Table C.2 where the CFR restrictions for the first, second and first two quarters were relaxed. This implies that it is better to relax the restriction for a certain quarter or several quarters rather than for the yearly restriction in terms of the scale of deaths. Besides, it follows from Fig. 5 that the optimal controls for scenario 5 exhibit oscillations with averagely highest optimal contact rate and lowest detection rate. We can obtain that hospital beds are never fully occupied in any scenario, and scenario 5 requires minimal isolation resources/hospital beds.

3.3. Effect of booster vaccination on optimal controls

In order to explore the effect of booster vaccination on optimal controls, we initially analyze the impact of increasing vaccination rates. We re-solve the OCP (2.3) when the vaccination rate (v_1) increases from 0.1 (Table 1) to 0.12 and 0.15 while keeping the baseline restrictions for isolation and hospital capacity, respectively, as shown in Figs. 6 and 7. It follows from Fig. 6 that the higher the rate of booster vaccination, the weaker the prevention and control measures required; the later the restriction of quarantine resources been reached. And Fig. 7(a–b) shows that the higher the rate of booster vaccination, the more (less) the susceptible population with high (low) immunity.

In addition, the corresponding quantities of interest under various vaccination rates are given in Table 2. It shows that the higher the rate of booster vaccination, the lower the cumulative number of deaths, the lower the total cost of taking control strategies, and the lower the effective reproduction number at time T. And increasing vaccination rate by 20% and reducing the restriction of isolation resources by half (i.e., $v_1 = 0.12$ and $Q_{max} = 21,000$) will lead to a (significant decrease) decrease in (the death toll) the total cost. This suggests that effective and available vaccines can contribute to reducing control costs, relaxing control measures, alleviating medical resources and reducing the infection scales.

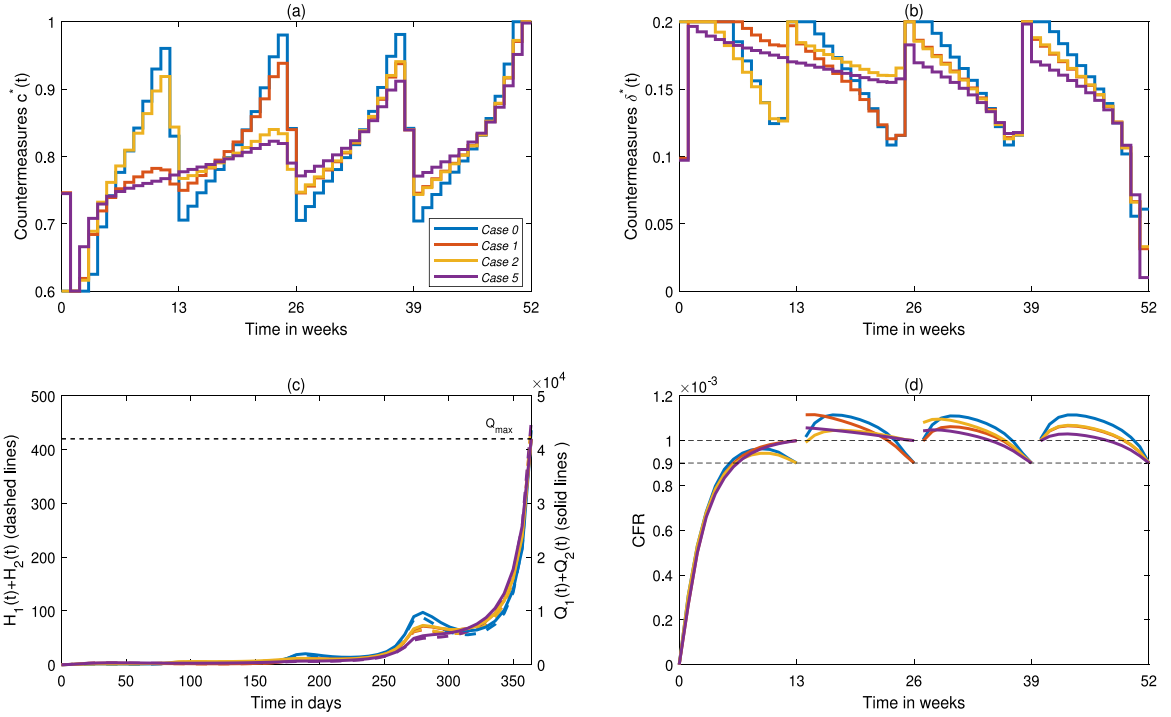

Fig. 5. Results for the multi-stage problem under cases 0 or 1,2,5 where strict limits are maintained every quarter or the CFR restrictions for the first, second and first two quarters were relaxed. (a) and (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) under those cases; (c): The variation in hospital beds and isolated resources occupancy under those cases; (d): the case fatality rate within t days of each quarter under different cases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

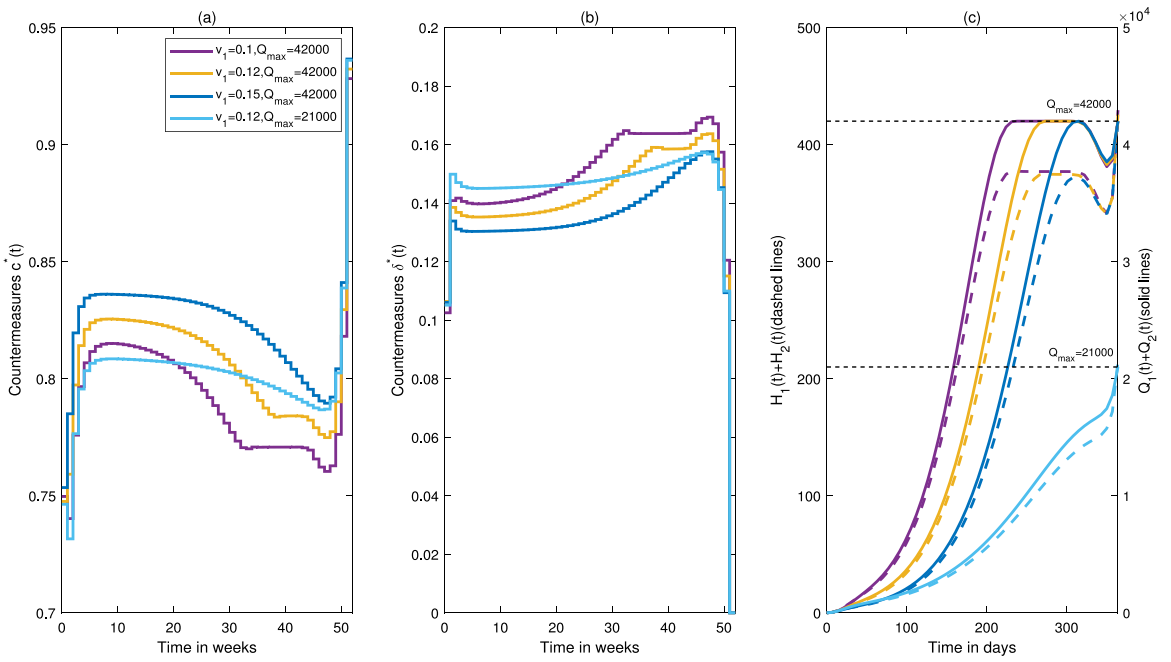

Fig. 6. (a) and (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) with different v_1 and Q_{max}; (c): Hospital beds and isolated resources occupancy under different v_1 and Q_{max}.

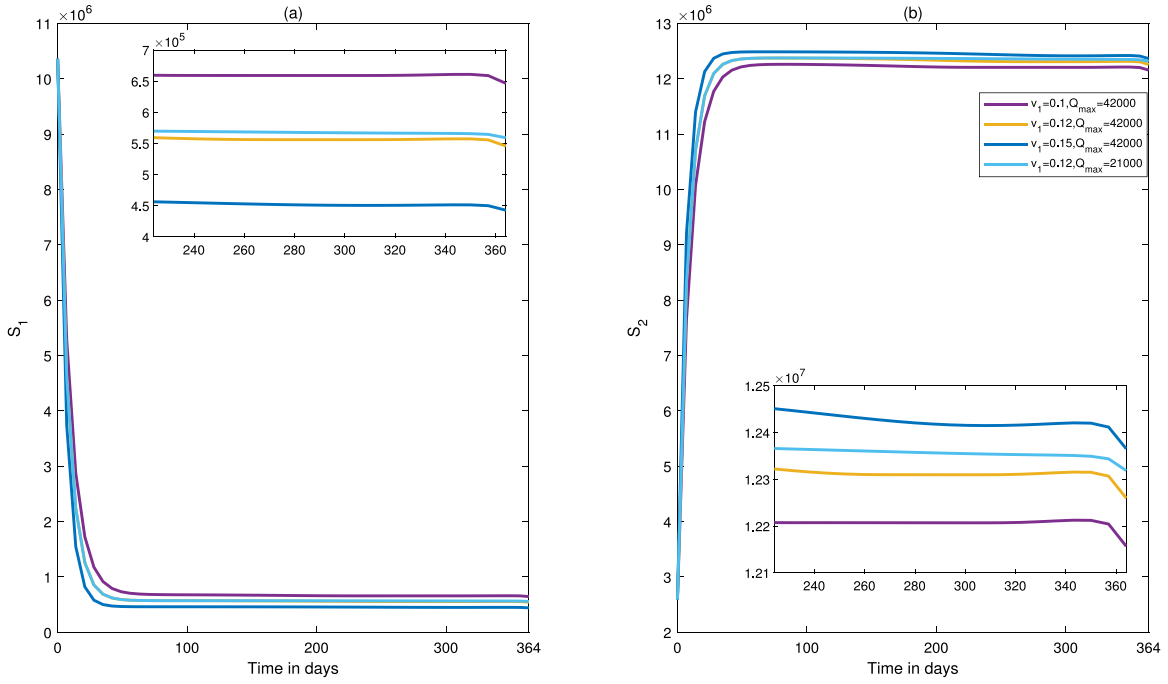

Fig. 7. The evolution of S_1 (a) and S_2 (b) under corresponding optimal controls with different v_1 and Q_{max}.

As a particular case we consider the case with no booster, i.e., $v_1 = 0$ (Fig. 8). Subplot (b) of Fig. 8 shows that the optimal contact rate $c^*(t)$ is obvious piece-wise functions with switching between 0 and certain levels, which corresponds to lockdown measure, opening, re-lockdown and re-opening and so on. In particular, Fig. 8(b) shows that lockdown needs to be implemented at week 7, 9, 12, 15, ..., and Fig. 8(c) shows that the total number of quarantined individuals reaches the restriction at the end of week 7, 9, 12, 15, ..., and hospital beds never run out. This means that given no booster, all the isolation resources will occupy when the lockdown measure ends. Note that the effective reproduction numbers R_t under optimal countermeasures show no tendency of decrease, shown in Fig. 8(d), indicating that the infection may not be controlled. A case with the quite low booster rate $v_1 = 0.001$ is also shown in Appendix D, and similar conclusions can be drawn. This indicates that given no vaccination or very low booster rate, intermittent lockdowns can ensure the isolated/hospitalized individuals no greater than the capacities in exchange for at least four times the cost of the benchmark case (Table 2), suggesting that vaccination and effective booster shots play a vital role in helping realize the smooth transition from a pandemic to an endemic phase and avoiding lockdown measures. To further analyze what intensity vaccination rates are effective in terms of avoiding lockdowns, we focus on the optimal combinations of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) for the OCP (2.3) under *the benchmark case* with different v_1, shown in Fig. 9. It shows that when the rate of vaccination v_1 reaches 0.01, no lockdown is necessary, and to keep people as half mobility as they were before the pandemic, the vaccination rate of 0.02 is needed.

4. Discussion

Designing our own path to smoothly transit from the COVID-19 pandemic to an endemic phase is challenging, given high susceptibility, persist vaccination waning, and relatively weak strength of health systems. In this study, we proposed a novel mathematical model with waning of immunity and examined the optimal contact rate and detection rate such that the cost is minimized, and as well as the number of hospitalized/isolated individuals no greater than the hospital/isolation capacity. Under the solved, optimal and dynamic intensifies of social distancing and detecting, the health system keeps normal operated, and hence these measures can act as our normalized control strategies which becomes a feasible transition path from a pandemic to an endemic phase has been suggested.

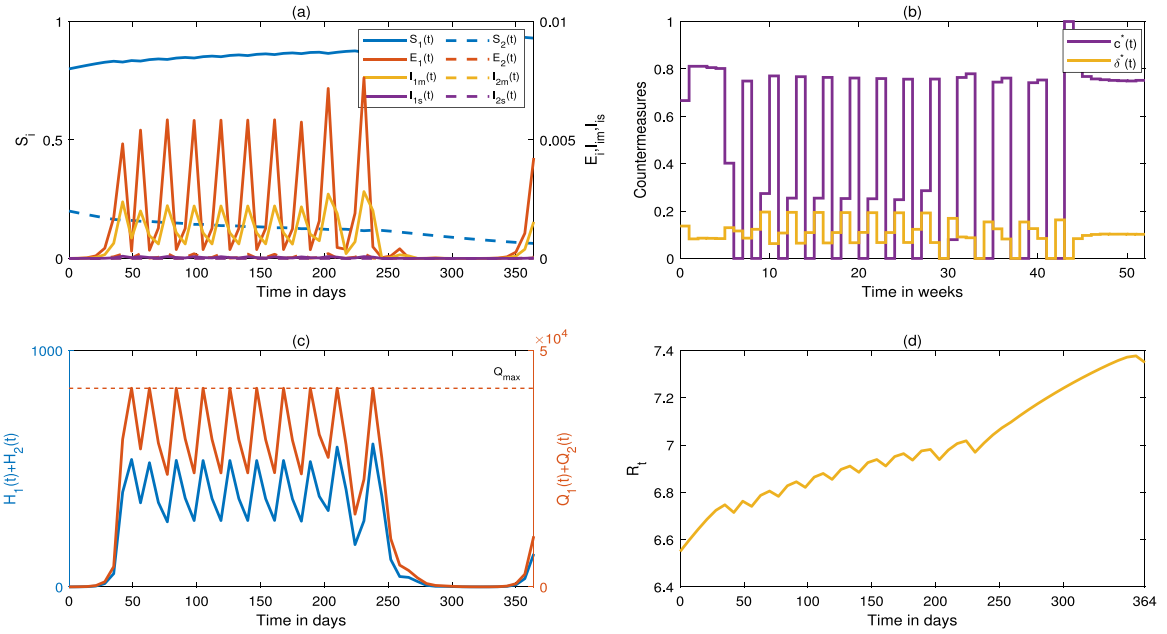

Fig. 8. Results for the OCP (2.3) when $v_1 = 0$, $H_{max} = 13,700$ and $Q_{max} = 42,000$. (a): The evolution of the epidemic under the optimal controls, where the left vertical axis corresponds to evolution of susceptible individuals (S_i), and the right vertical axis corresponds to evolution of latent (E_i), mild (I_{im}), and severe (I_{is}) patients (i = 1,2). (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$). (c): The variation in hospital beds and isolated resources occupancy. (d): The effective reproduction number (R_t) at time t.

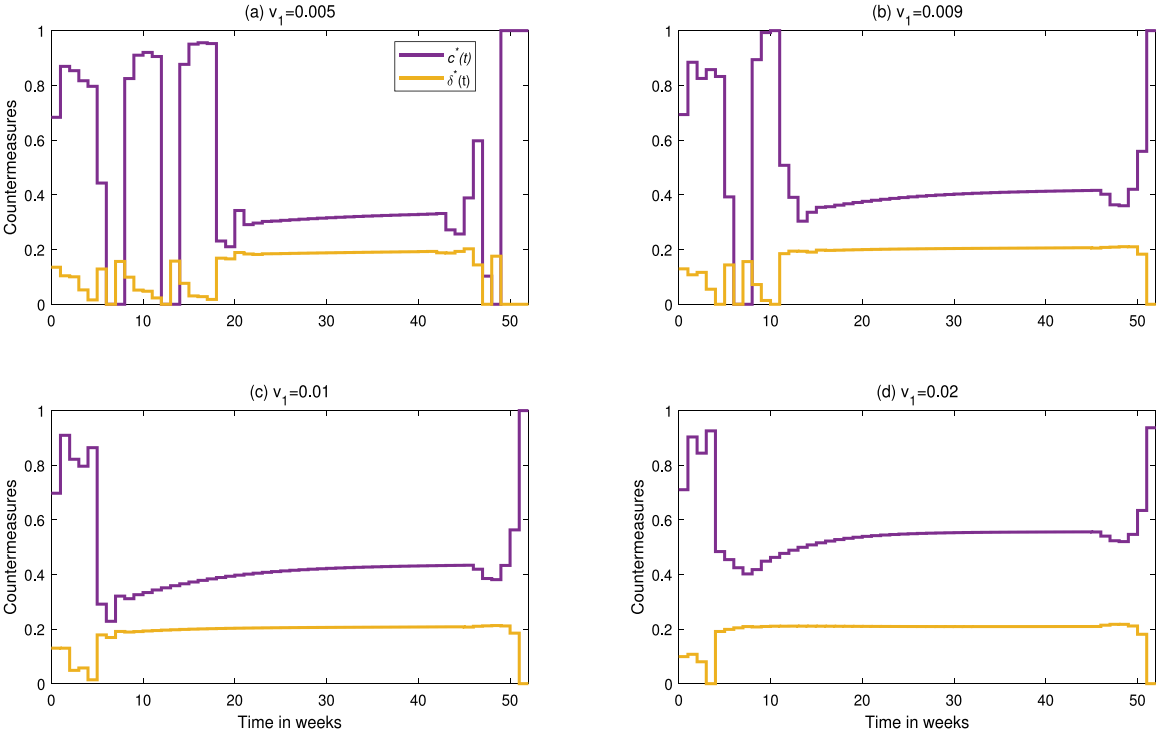

Fig. 9. Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) for the OCP (2.3) under *The benchmark case* with different v_1.

There are evidences showing that the quarantining of infected individuals or people who need specially monitoring [36,43] and providing more adequate medical resources are critical to mitigating the outbreak [38,40]. On the basis of reasonable parameter values tailored for Xi'an, the capital city of Shaanxi province, our study made it clear that it is the restriction of isolation resource rather than the capacity of hospital beds that significantly affects the optimal intensifies of contact reduction and detection. The greater the restriction of isolation resource, the looser the control measures.

When looking for the path of transition from a pandemic to an endemic, ones hope that the develop trend and/or death rate of infection caused by a novel variant are similar to those for influenza. We then further considered an additional restriction determined by the CFR of influenza and found that for relatively small (large) yearly CFR, the optimal controls are mainly determined by CFR (the isolation capacity). And the optimal controls under quarterly CFR restriction may show oscillations, and relaxing quarterly CFR restriction may cause a relatively low death rate.

In particular, we analyzed the impacts of booster vaccination, we obtained that if there is no booster or vaccination rate is extremely low, the optimal contact rates are obvious piece-wise functions with switching between 0 and certain levels. That is to say the optimal strategy is a control mode switching between lock down and opening with certain social distancing, demonstrating that vaccination and effective booster shots play a vital role in helping realize the smooth transition from a pandemic to an endemic phase and avoiding lockdown measures. This also reflects the importance and necessity of China's static management on a certain area during Covid-19 outbreak.

What the problem has been solved in this study is to find the optimal countermeasures within a fixed planning horizon. This could lead to the phenomenon that disease infection takes off at the end of control. We may solve this problem by adding a terminal term about the number of infected people at the end of control to the objective function or some terminal state constraints about the fraction of infected people. We leave this for future work.

Funding source

This research was funded by the National Natural Science Foundation of China (grant number:12220101001, 12031010).

CRediT authorship contribution statement

Tangjuan Li: Methodology, Software, Writing – original draft. **Yanni Xiao:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

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. Necessary conditions for a general problem and fast direct multiple shooting algorithms for optimal control problems

We consider the following optimal control problem:

$$\begin{array}{lll} \underset{\mu}{min} & \int_0^T F(x,\mu,t)\,dt, & (a) \\ \text{subject to} & \dot{x(t)} = f(x,\mu,t), x(0) = x_0 & (b) \\ & g(x,\mu,t) \le 0, & (c) \\ & h(x,t) \le 0, & (d) \\ & a(x(T),T) \le 0, & (e) \end{array} \tag{A.1}$$

where state $x(t) \in R^n$, control $\mu(t) \in R^m$, functions F from $E^n \times E^m \times E$ into E, f from $E^n \times E^m \times E$ into E^n, g from $E^n \times E^m \times E$ into E^q, h from $E^n \times E$ into E^p and a from $E^n \times E^n \times E$ into E^n. The special cases of the mixed constraint $g(x,\mu,t) \le 0$ are $\mu_i \in [N_i, M_i]$ with $N_i < M_i$, where M_i and N_i are constants. And this situation is the one covered in this article.

We first give some relevant notations. An interval $(\theta_1, \theta_2) \subset [0,T]$ with $\theta_1 < \theta_2$ is called an interior interval if $h_i(x(t),t) < 0$, $\forall t \in (\theta_1, \theta_2)$. If the optimal trajectory satisfies $h_i(x(t),t) = 0$ for $\tau_1 \le t \le \tau_2$ for some i, then $[\tau_1, \tau_2]$ is called a boundary interval. And the instant τ_1 is called an entry time if there is an interior interval ending

at $t = \tau_1$ and a boundary interval starting at τ_1. τ_2 is called an exit time if a boundary interval ends and an interior interval starts at τ_2. The time τ is a contact time for x if it is an isolated point such that $h(x(\tau), \tau) = 0$. The entry, exit and contact times are called junction times.

In order to formulate the minimum principle for the problem (A.1), we define the Hamiltonian function $H^d : E^n \times E^m \times E^1 \to E^1$ as

$$H^d = F(x, \mu, t) + \lambda^d f(x, \mu, t); \tag{A.2}$$

and the Lagrangian function $L^d : E^n \times E^m \times E^n \times E^q \times E^q \times E^1 \to E^1$ as

$$L^d(x, \mu, \lambda^d, \psi, \eta^d, t) = H^d(x, \mu, \lambda^d, t) + \psi g(x, \mu, t) + \eta^d h(x, t). \tag{A.3}$$

According to Theorem 9.3.1 in [29] or [34], we have the following minimum theorem underlying the direct method.

Theorem 1. *Let (μ^*, x^*) be an optimal pair of functions. Then there exist an adjoint function λ^d, which may be discontinuous at a time in a boundary interval or a contact time, multiplier functions ψ, α, γ^d, η^d, and a jump parameter $\zeta^d(t)$ at each time τ where λ^d is discontinuous, such that the following conditions hold:*

(a) $\dot{x^*} = f(x^*, \mu^*, t)$, $x^*(0) = x_0$, satisfying constraints $g(x^*, \mu^*, t) \le 0$, $h(x^*, t) \le 0$, and the terminal constraints $a(x^*(T), T) \le 0$;

(b) the adjoint equation

$$\dot{\lambda^d} = -L_x[x^*, \mu^*, \lambda^d, \psi, \eta^d, t]$$

with the transversality conditions $\lambda^d(T^-) = \alpha a_x(x^*(T), T) + \gamma^d h_x(x^*(T), T)$, and $\alpha \ge 0$, $\alpha a(x^*(T), T) = 0$, $\gamma^d \ge 0$, $\gamma^d h(x^*(T), T) = 0$;

(c) the Hamiltonian minimizing condition

$$H^d[x^*(t), \mu^*(t), \lambda^d(t), t] \le H^d[x^*(t), \mu, \lambda^d(t), t]$$

at each $t \in [0, T]$ for all μ satisfying $g[x^*, \mu, t] \le 0$;

(d) the jump conditions at any time τ, where λ^d is discontinuous, are $\lambda^d(\tau^-) = \lambda^d(\tau^+) + \zeta^d(\tau) h_x(x^*(\tau), \tau)$ and

$$H^d[x^*(\tau), \mu^*(\tau^-), \lambda^d(\tau^-), \tau] = H^d[x^*(\tau), \mu^*(\tau^+), \lambda^d(\tau^+), \tau] - \zeta^d h_t(x^*(\tau), \tau);$$

(e) the Lagrange multipliers $\mu(t)$ are such that

$$\frac{\partial L^d}{\partial \mu}|_{\mu=\mu^*} = 0, \frac{dH^d}{dt} = \frac{dL^d}{dt} = \frac{\partial L^d}{\partial t};$$

(f) the complementary slackness conditions $\psi(t) \ge 0$, $\psi(t) g(x^*, \mu^*, t) = 0$, $\eta(t) \ge 0$, $\eta(t) h(x^*(t), t) = 0$, and $\zeta^d(\tau) \ge 0$, $\zeta^d(\tau) h(x^*(\tau), \tau) = 0$ hold.

Let us use OCP (2.5) as an example. Let $x = (x_1, x_2, \ldots, x_{15}) = (S_1, E_1, I_{1m}, I_{1s}, Q_1, H_1, D_1, S_2, E_2, I_{2m}, I_{2s}, Q_2, H_2, D_2, C)$ denote the state variables and $\mu(t) = (\delta(t), c(t))$ denote control variables. The ordinary differential equations describing the development of the disease at each stage satisfy Eqs. (2.1) and (2.4), which are denoted as

$$\dot{x}(t) = f(x(t), \mu(t), t). \tag{A.4}$$

Then the objective function is $F(x, \mu, t) = (1 - c(t))^2 + \kappa \delta^2(t) \sum_{j=1}^{2} (S_j + E_j + I_{jm} + I_{js})$. The corresponding constraints are

$$g(x, \mu, t) = \begin{pmatrix} -c(t) \\ c(t) - 1 \\ -\delta(t) \\ \delta(t) - 1 \end{pmatrix}, \quad h(x, t) = \begin{pmatrix} N \cdot (H_1 + H_2) - H_{max} \\ N \cdot (Q_1 + Q_2) - Q_{max} \end{pmatrix}.$$

The terminal constraint is

$$a(x(T), T) = D_1(T) + D_2(T) - r_{max} C(T).$$

Direct methods are basic approaches to address optimal control problems, which transform the original infinite optimal control problem into a finite dimensional nonlinear programming problem (NLP). Among them, the direct multiple shooting method is the method of choice for nonlinear optimal control problems [18]. In our paper, we assumed our controls are piecewise constant (see (2.3)(g–h)), i.e.,

$$\mu(t) = (\delta_{qn}, c_{qn}) \doteq q_n \quad \text{for} \quad t \in [n\Delta t, (n+1)\Delta t].$$

Then, we solve the ODE on each interval $[n\Delta t, (n+1)\Delta t]$ independently, starting with an artificial initial value s_n:

$$\dot{x}_n(t) = f(x_n(t), q_n), t \in [n\Delta t, (n+1)\Delta t] \quad \text{with} \quad x_n(n\Delta t) = s_n.$$

The corresponding trajectory pieces denoted as $x_n(t; s_n, q_n)$ and the integrals follow

$$l_n(s_n, q_n) = \int_{n\Delta t}^{(n+1)\Delta t} F(x_n(t; s_n, q_n), q_n, t)\, dt.$$

Thus, we arrive at the following NLP formulation:

$$\begin{array}{lllll} \underset{\mu}{min} & \sum_{n=0}^{N-1} l_n(s_n, q_n) & & & \\ \text{subject to} & s_0 - x_0 = 0, & & & \text{(initial value)} \\ & s_{n+1} - x_n\left((n+1)\Delta t; s_n, q_n\right) = 0, & n = 0, \ldots, N-1, & & \text{(continuity)} \\ & -g(x_n, q_n, t) \geq 0, & n = 0, \ldots, N, & & \text{(discretized path constraints)} \\ & -h(x_n, t) \geq 0, & n = 0, \ldots, N, & & \text{(discretized path constraints)} \\ & -a(s_N, T) \geq 0. & & & \text{(terminal constraints)} \end{array} \tag{A.5}$$

If we denoted all variables as $w = (s_0, q_0, s_1, q_1, \ldots, s_N, q_N)$, the following finite dimensional NLP can be obtained [18].

$$\underset{w}{min} \quad a(w) \quad \text{subject to} \quad b(w) = 0, c(w) \geq 0.$$

Then, we can use an iterative Sequential Quadratic Programming to solve this form of NLP [41].

Appendix B. Supplementary drawing

For relatively strict measures ($c^*(t) \equiv 0.77$ and $\delta^*(t) \equiv 0.16$), we obtain that neither the isolation resources nor the number of hospital beds would be used up (shown in Fig. B.1(b)), while the total cost of interventions is relatively high; for relatively loose measures ($c^*(t) \equiv 0.8$ and $\delta^*(t) \equiv 0.14$), we find that although the number of hospital beds would not be used up, but at least 2.7×10^5 isolated beds are needed for mild cases, more than six

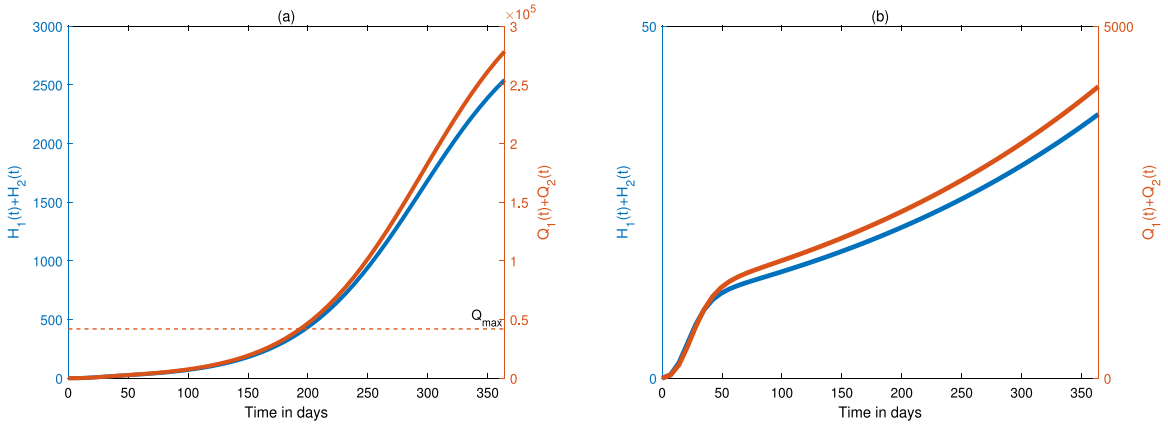

Fig. B.1. The variation in hospital beds and isolated resources occupancy when countermeasures are more lax ((a): $c^*(t) \equiv 0.8$ and $\delta^*(t) \equiv 0.14$) or more restrictive, ((b): $c^*(t) \equiv 0.77$ and $\delta^*(t) \equiv 0.16$) than the optimal combination of social distancing strategy and detection rate in Fig. 2(a). Note that neither the isolation resources nor the number of hospital beds would be used up in (a) and the number of hospital beds would not be used up in (b).

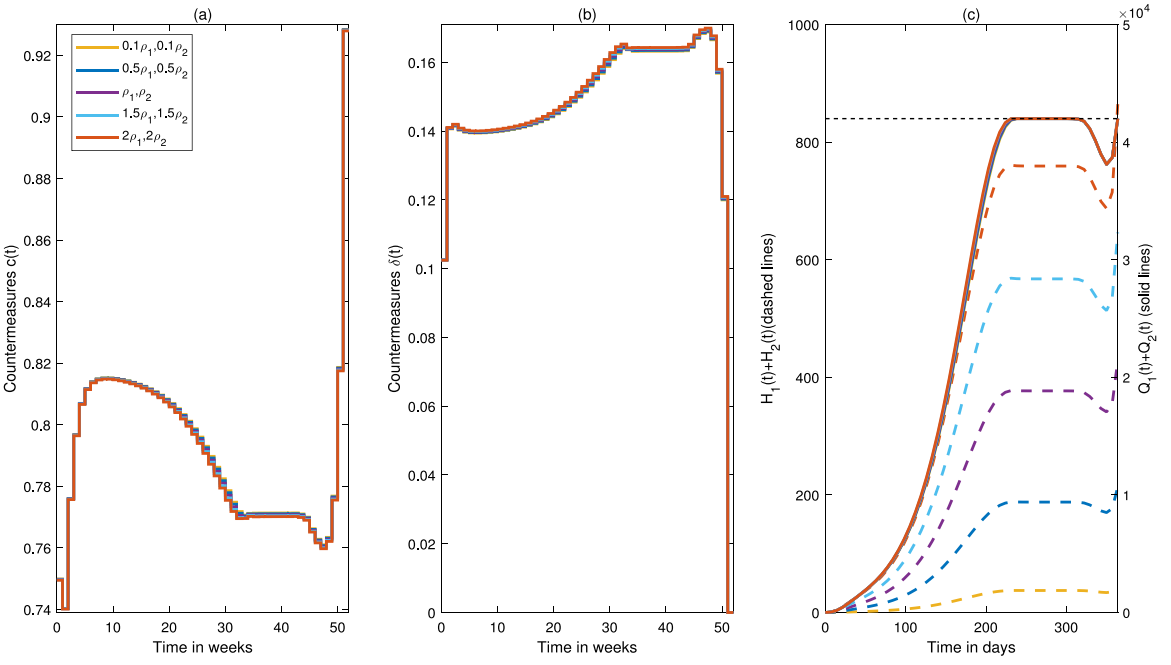

Fig. B.2. Results for the OCP (2.3) under *The benchmark case* with different probability of severe symptoms. (a) and (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$) under those cases; (c): The variation in hospital beds and isolated resources occupancy under those cases. Note that the curves of the benchmark case and case III coincide.

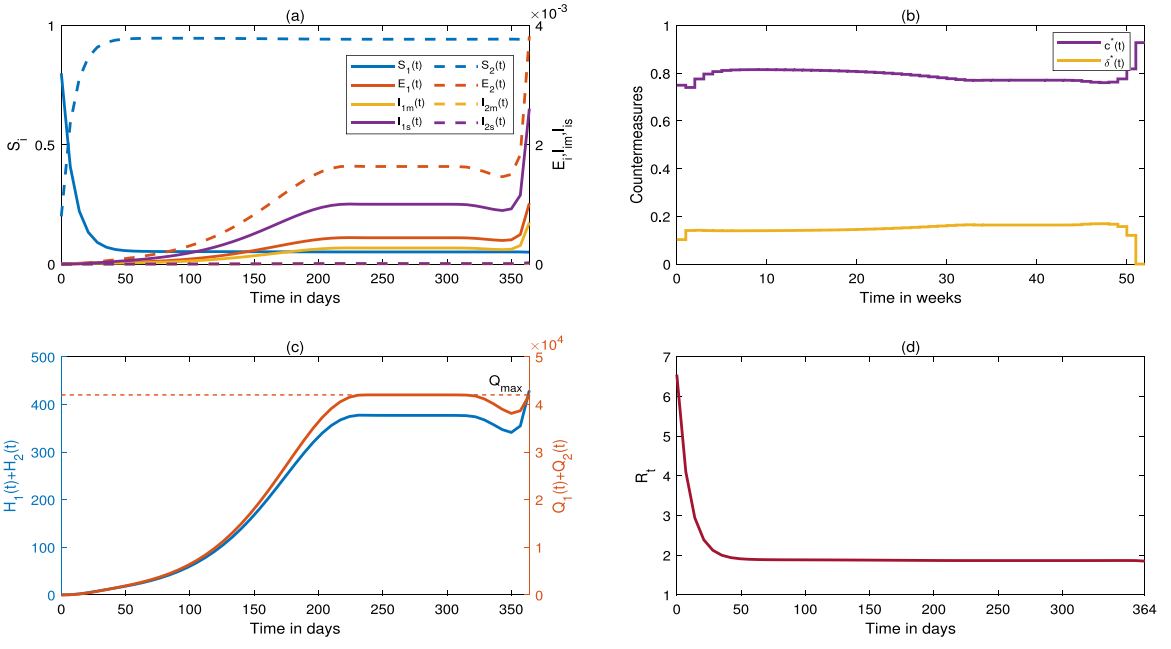

Fig. B.3. Results for the OCP (2.5) when $r_{max} = 1.2 \times 10^{-3}$. (a): The evolution of the epidemic under the optimal controls, where the left vertical axis corresponds to evolution of susceptible individuals (S_i), and the right vertical axis corresponds to evolution of latent (E_i), mild (I_{im}), and severe (I_{is}) patients (i = 1,2). (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$). (c): The variation in hospital beds and isolated resources occupancy. (d): The corresponding effective reproduction number (R_t) at time t.

Table B.1
The cumulative number of deaths ($D_1(T) + D_2(T)$), the cost of taking countermeasures in a year and the effective reproduction number at time T of OCP (2.3) under *The benchmark case* with different probability of severe symptoms.

	$D_1(T) + D_2(T)$	The total cost	$R_t(T)$
$0.1\rho_1,\ 0.1\rho_2$	137	32.35	1.8461
$0.5\rho_1,\ 0.5\rho_2$	686	30.43	1.8483
$\rho_1,\ \rho_2$	1381	32.54	1.8511
$1.5\rho_1,\ 1.5\rho_2$	2086	32.64	1.8538
$2\rho_1,\ 2\rho_2$	2798	32.75	1.8566

The number of people in this table is rounded to an integer.
The total cost includes the normalized cost of taking social distancing and the relative cost of taking nucleic acid screening in a year.

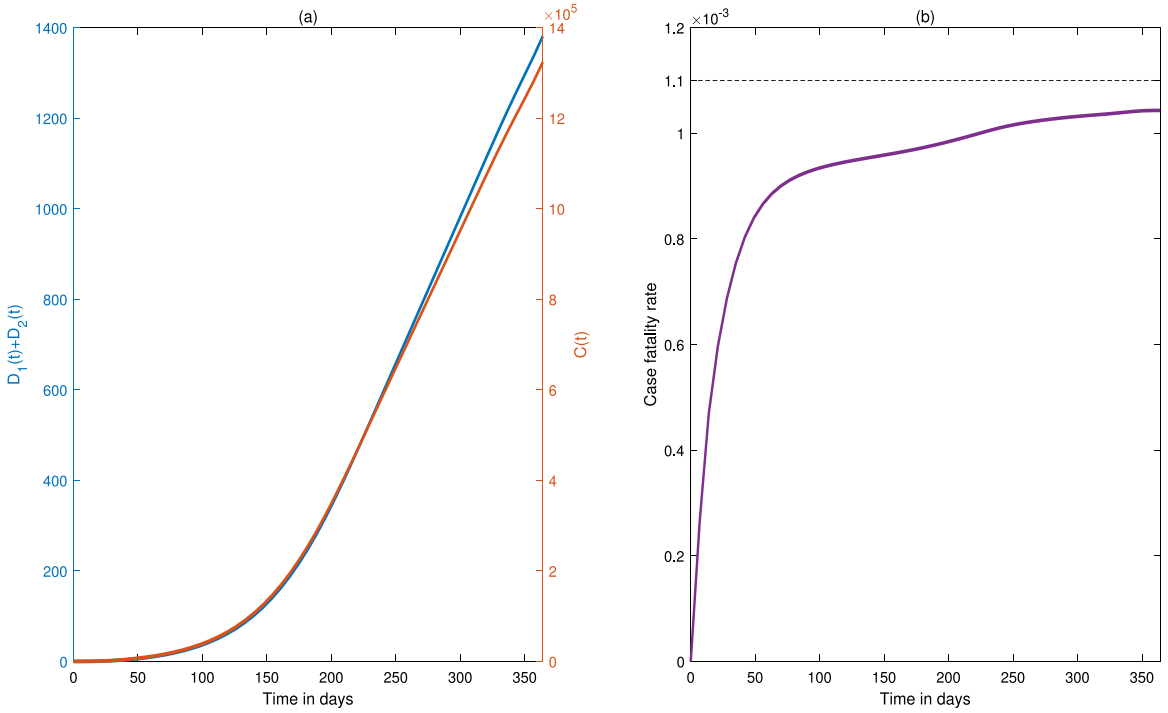

Fig. B.4. Results for the OCP (2.5) when $r_{max} = 1.2 \times 10^{-3}$. (a): The cumulative number of deaths and confirmed cases within one year. (b): The case fatality rate within t days.

times as the isolation capacity (Fig. B.1(a)). Those mean that any combination of constant controls are not indeed as well as our solved optimal and dynamic control in terms of contribution to our objectives.

We gave the results for the OCP (2.3) under the benchmark case with different probability of severe symptoms in Fig. B.2. The corresponding quantities of interest are listed in Table B.1. This shows that small changes in the probability of severe symptoms ($\rho_1,\ \rho_2$) have little impact on the required intensity of prevention and control measures, but have a significant impact on the number of hospitalizations and deaths.

Results for the OCP (2.5) when $r_{max} = 1.2\times10^{-3}$ are showing in Figs. B.3–B.4. It shows that the optimal controls and the corresponding hospital beds and isolated resources occupancy are similar to those with the benchmark case.

Appendix C. A multi-stage problem

We now extend the OCP (2.3) to establish multi-stage problems (also called multi-phase problems), which allows you to implement discrete events where a jump in the state trajectory happens [27], to seek the optimal prevention

and control measures. For convenience, it is assumed that M stages are considered, the state variable of stage i is denoted as

$$\begin{aligned} X_i &= (x_1^i, x_2^i, \ldots, x_{15}^i) \\ &= (S_1^i, E_1^i, I_{1m}^i, I_{1s}^i, Q_1^i, H_1^i, D_1^i, S_2^i, E_2^i, I_{2m}^i, I_{2s}^i, Q_2^i, H_2^i, D_2^i, C^i), \end{aligned} \tag{C.1}$$

and the corresponding control variable of stage i is $\mu_i(t) = (\delta_i(t), c_i(t))$. The risk of infection in different stages is indicated by β_l^i $(l = 1, 2; i = 1, \ldots, M)$, where the subscript l represents the susceptible population with high or low level of immunity, and superscript i represents the ith stage. The ordinary differential equations describing the development of the disease at each stage satisfy Eqs. (2.1) and (2.4), which are denoted as

$$X_i'(t) = f(X_i(t), \mu_i(t), \beta_l^i), \tag{C.2}$$

where $X_i'(t) \doteq \frac{dX_i}{dt}$ and $i = 1, \ldots, M$. Then the corresponding multi-stage optimal control problem is expressed as:

$$\begin{aligned} &\underset{\substack{c_1,\ldots,c_M \\ \delta_1,\ldots,\delta_M}}{min} && \textstyle\sum_{i=1}^{M} \left(\int_{T_{i-1}}^{T_i} \left((1 - c_i(t))^2 + \kappa \delta_i^2(t) \sum_{j=1}^{2} (S_j^i + E_j^i + I_{jm}^i + I_{js}^i) \right) dt \right), && (a) \\ &\text{subject to} \\ &\quad for && i \in \{1, \ldots, M\} \\ &\qquad && X_i'(t) = f(X_i(t), \mu_i(t), \beta_l^i), && (b) \\ &\qquad && \tfrac{D_1^i(T_i) + D_2^i(T_i)}{C^i(T_i)} \le r_{max}^i, && (c) \\ &\qquad && 0 \le \delta_i(t) \le \delta_{max}, s_{max} \le c_i(t) < 1 && (d) \\ &\qquad && N \cdot (H_1^i(t) + H_2^i(t)) \le H_{max}, && (e) \\ &\qquad && N \cdot (Q_1^i(t) + Q_2^i(t)) \le Q_{max}, && (f) \\ &\qquad && c_i(t) = c_i(n\Delta t), t \in [n\Delta t, (n+1)\Delta t], n = 0, 1, \ldots, K, K\Delta t = T_i - T_{i-1} && (g) \\ &\qquad && \delta_i(t) = \delta_i(n\Delta t), t \in [n\Delta t, (n+1)\Delta t], n = 0, 1, \ldots, K, K\Delta t = T_i - T_{i-1} && (h) \\ &\quad for && i \in \{1, \ldots, M-1\} \\ &\qquad && X_j^{i+1}(T_i) = X_j^i(T_i) \quad for \quad j \in \{1, \ldots, 6, 8, \ldots, 13\}, && (i) \\ &\qquad && X_j^{i+1} = 0 \quad for \quad j \in \{7, 14, 15\}, && (g) \end{aligned} \tag{C.3}$$

where (c) indicates that the staged CFR does not exceed the corresponding upper limit r_{max}^i and (d) represents restrictions on control measures of each stage. Equations (i) and (g) represent the relationship between the initial value of the equation describing the next stage and the corresponding compartment value at the end of the previous stage. Specially, equations (i) means that the first state X_j^{i+1} of stage $i+1$ is equivalent to the last stage X_j^i of stage i at timepoint T_i.

In the following simulation we consider four quarters in a year to illustrate. We let $0 < \delta(t) < \delta_{max} = 0.2$, $s_{max} = 0.6 < c(t) < 1$ and $r_{max}^i < 1.1 \times 10^{-3}$ to further limit the intensity of countermeasures and the quarterly CFR restrictions $(r_{max}^i, i = 1, 2, 3, 4)$ respectively, which is determined by the results of OCP (2.5). We still fixed $\beta_1^i = 0.2$ $(i = 1, 2, 3, 4)$ and $\beta_2^i = 0.2\beta_1^i$.

We first examine the impact of considering seasonal restrictions on optimal controls and corresponding deaths. The corresponding quantities of interest under some cases with same seasonal limits (case 0 and 15 in Table C.1) are given in case 0 and 15 in Table C.2. It is worth noting that if the maximum allowable CFR is 9×10^{-4}, the optimal controls given quarterly CFR limit exhibit seasonal oscillations (blue curves in Fig. 5(a) and (b)) compared to yearly CFR limit (yellow curves in Fig. 4(a) and (b)), and the cumulative number of deaths and confirmed cases in a year are greater for quarterly CFR limit (case 0 in Table C.2) than those for yearly CFR limit ($r_{max} = 9 \times 10^{-4}$ in Table 2). While for relatively loose CFR limit (say, 1×10^{-3}), the numbers of deaths and total costs in a year do not differ much between the yearly ($r_{max} = 9 \times 10^{-3}$ in Table 3) or quarterly CFR limits (case 15 in Table C.2).

To investigate the effect of various quarterly CFR limits on optimal controls and cumulative number of deaths/confirmed cases, and examining whether it is possible to relax a certain quarter or several quarters of the constraint to save on the cost of countermeasures or reduce the scale of deaths, as opposed to relaxing the yearly constraint directly, we consider the possible relaxation scenarios listed in Table C.1 and obtain the cumulative number of deaths/confirmed cases listed in Table C.2. We find that the scale of death in a year of most cases in

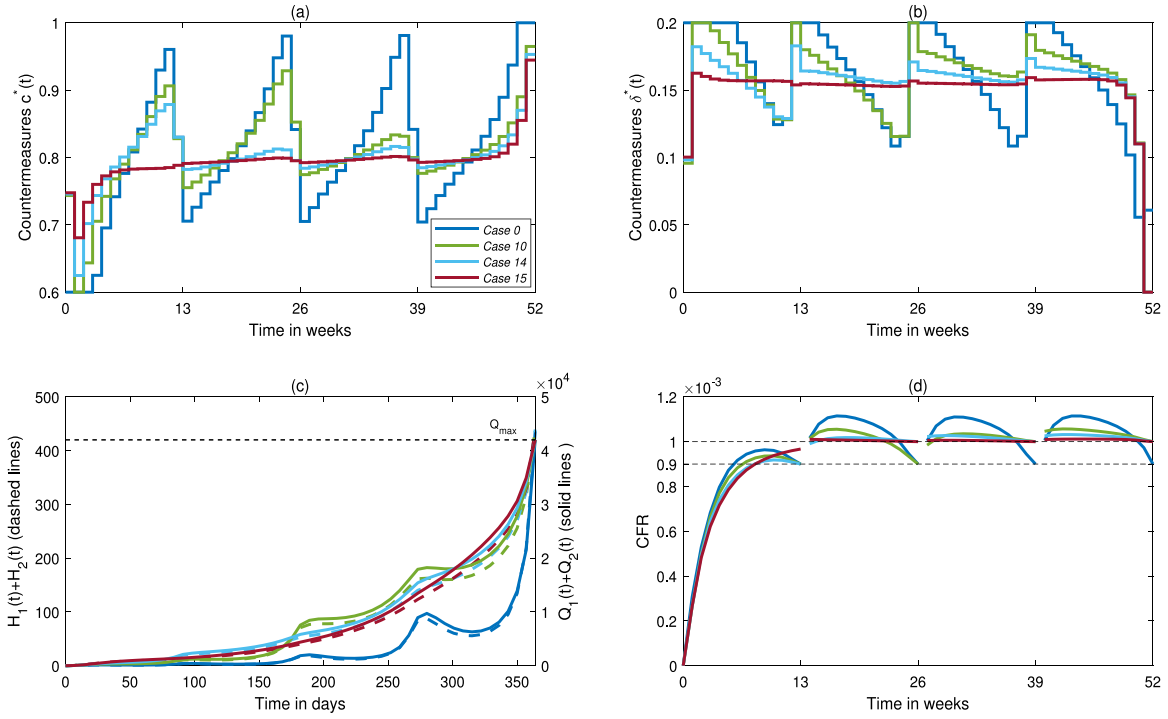

Fig. C.1. Results for the OCP (C.3) under cases 0,10,14,15 when $\beta_1^i = 0.2$ $(i = 1, 2, 3, 4)$. (a) and (b): Optimal combination of social distancing strategy $(c^*(t))$ and detection rate $(\delta^*(t))$ under those cases; (c): Hospital beds and isolated resources occupancy under those cases; (d): the case fatality rate within t days of each quarter under different cases.

Table C.1
All relaxation cases.

	r_{max}^1	r_{max}^2	r_{max}^3	r_{max}^4
0	0.9	0.9	0.9	0.9
1	1	0.9	0.9	0.9
2	0.9	1	0.9	0.9
3	0.9	0.9	1	0.9
4	0.9	0.9	0.9	1
5	1	1	0.9	0.9
6	1	0.9	1	0.9
7	1	0.9	0.9	1
8	0.9	1	1	0.9
9	0.9	1	0.9	1
10	0.9	0.9	1	1
11	1	1	1	0.9
12	1	1	0.9	1
13	1	0.9	1	1
14	0.9	1	1	1
15	1	1	1	1

Table C.2 was smaller than the case with yearly CFR limit $r_{max} = 1 \times 10^{-3}$, which means that it is better to relax the restriction for a certain quarter or several quarters rather than to relax the yearly restriction in terms of the scale of deaths.

It is worth noting that scenarios 1, 2, and 5 (bold black font in Table C.2) not only reduce the number of deaths (by 3.17%, 0.51% and 4.52%, respectively) and confirmed cases (by 3.31%, 0.98% and 4.97%, respectively) within a year, but also resulted in cost savings (by 4.40%, 4.73% and 7.26%, respectively) compared to scenario 0, and scenario 5 has the largest reduction in death toll (or confirmed cases) and cost. It shows from Fig. 5 that the optimal

Table C.2
The case $\beta_1^1 = \beta_1^2 = \beta_1^3 = \beta_1^4 = 0.2$.

Case	D_1	D_2	D_3	D_4	D_{total}	C_1	C_2	C_3	C_4	C_{total}	$Cost_1$	$Cost_2$	$Cost_3$	$Cost_4$	$Cost_{total}$
0	1.98	7.30	35.16	153.87	198.31	2194.46	8115.81	39062.30	170967.70	220340.26	12.46	8.32	8.33	7.70	36.80
1	2.86	4.10	26.63	158.42	**192.03**	2863.83	4560.67	29592.22	176025.49	**213042.21**	13.43	7.39	7.56	6.79	**35.18**
2	2.59	9.84	26.66	158.26	**197.35**	2880.18	9837.92	29617.65	175842.90	**218178.66**	11.66	9.09	7.47	6.84	**35.06**
3	2.64	13.27	62.79	155.57	234.27	2934.22	14742.27	62788.65	172856.69	253321.83	11.64	7.56	9.09	6.73	35.03
4	2.49	12.11	72.03	321.27	407.90	2766.75	13454.20	80030.23	321265.66	417516.83	11.79	7.67	7.65	7.92	35.03
5	3.69	5.25	20.10	160.29	**189.34**	3692.47	5254.81	22336.33	178095.69	**209379.30**	12.48	8.54	6.97	6.13	**34.13**
6	3.77	8.27	44.61	158.92	215.57	3770.58	9184.17	44613.56	176575.61	234143.93	12.46	6.89	8.66	6.09	34.11
7	3.66	7.79	63.27	325.60	400.31	3657.75	8658.72	70296.30	325593.37	408206.14	12.55	6.95	6.99	7.54	34.04
8	5.69	23.30	52.66	157.29	238.94	6317.81	23301.85	52657.73	174766.64	257044.03	10.10	8.81	8.78	6.43	34.11
9	5.48	22.28	67.10	322.73	417.58	6083.85	22276.10	74553.20	322724.47	425637.63	10.18	8.91	7.24	7.72	34.05
10	5.46	30.85	151.29	311.52	499.12	6065.72	34281.36	151290.18	311515.77	503153.02	10.21	7.34	8.83	7.63	34.01
11	6.12	12.17	34.22	160.38	212.90	6162.32	12173.30	34223.93	178197.36	230756.91	11.09	8.30	8.34	5.70	33.42
12	6.03	11.84	56.80	330.26	404.93	6068.55	11838.32	63115.27	330257.91	411280.06	11.13	8.30	6.53	7.37	33.34
13	6.10	20.83	125.79	322.19	474.91	6146.47	23147.79	125785.84	322186.45	477266.56	11.09	6.52	8.36	7.32	33.29
14	9.78	48.97	128.38	320.00	507.13	10871.37	48969.07	128379.99	319993.69	508214.12	8.90	8.45	8.38	7.39	33.12
15	11.25	36.62	113.62	326.45	487.94	11629.65	36620.94	113618.43	326446.89	488315.91	9.36	8.13	8.12	7.20	32.82

D_i, C_i and $Cost_i$ $(i = 1, 2, 3, 4)$ represent the cumulative number of deaths, the cumulative number of confirmed cases and the total cost of prevention and control measures in quarter i respectively.
D_{total}, C_{total} and $Cost_{total}$ represent the cumulative number of deaths, the cumulative number of confirmed cases and the total cost of prevention and control measures in the whole year.

controls for scenario 5 exhibit oscillations with averagely highest optimal contact rate and lowest detection rate. We can obtain that hospital beds are never fully occupied in any scenario, and scenario 5 requires minimal isolation resources/hospital beds in these four scenarios.

Moreover, by comparing scenario 15 with the most relaxed case per quarter, both scenario 10 and 14 not only have more deaths and infections in a year than scenario 15, but also their costs within a year are more than that for scenario 15, and hence are less feasible or undesirable. Fig. C.1 shows the variation of the optimal social distancing strategy (a) and detection rates (b), hospital beds and isolated resources occupancy (c) and the CFR within t days of each quarter for scenarios 10, 14, 15. We can find that the consumption of medical resources in these cases is far more than that in case 0.

Appendix D. Results for the OCP (2.3) with different $v_1 = 0.001$

Results for the OCP (2.3) when $v_1 = 0.001$ are shown in Fig. D.1. Subplots (b) of Fig. D.1 show that the optimal contact rate $c^*(t)$ is obvious piece-wise functions with switching between 0 and certain levels, which corresponds to lockdown measure, opening, re-lockdown and re-opening and so on. Note that the effective reproduction numbers R_t under optimal countermeasures show no tendency of decrease, shown in Fig. D.1(d), indicating that the infection may not be controlled, which is also consistent with oscillations of the corresponding optimal solutions (shown in Fig. D.1(a)). This indicates that given very low booster rate, intermittent lockdowns can ensure the isolated/hospitalized individuals no greater than the capacities in exchange for at least four times the cost of *the benchmark case* (shown in Table 2), suggesting that vaccination and effective booster shots play a vital role in helping realize the smooth transition from a pandemic to an endemic phase and avoiding lockdown measures.

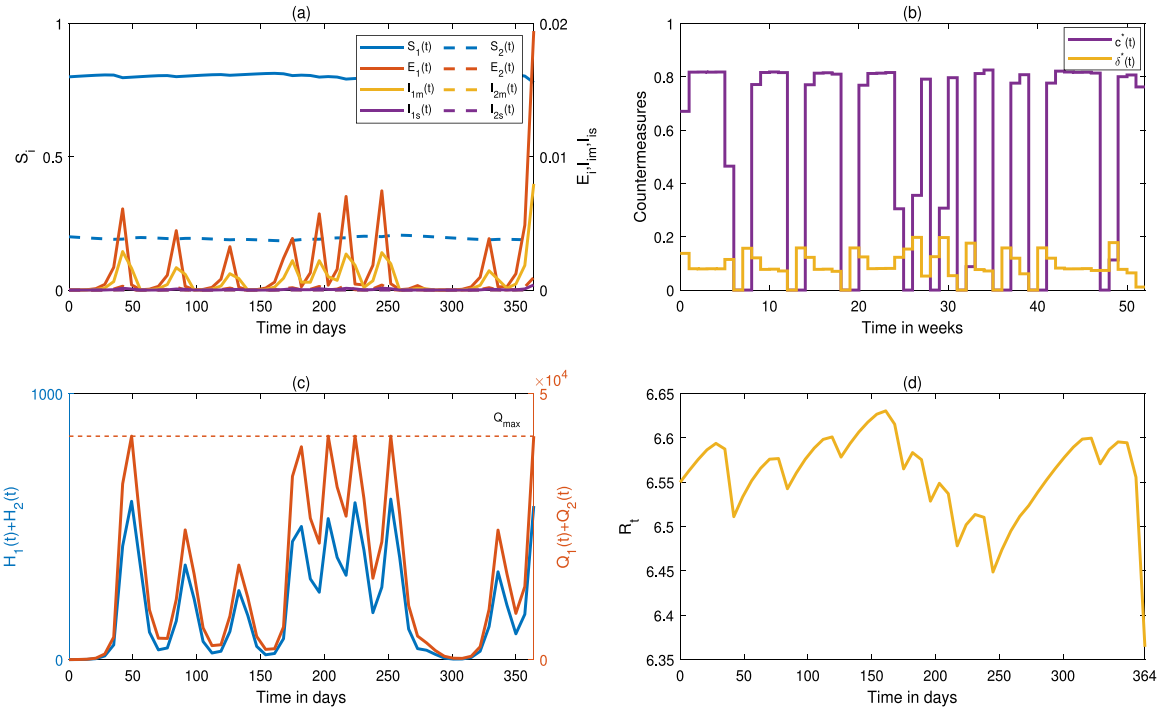

Fig. D.1. Results for the OCP (2.3) when $v_1 = 0.001$, $H_{max} = 13,700$ and $Q_{max} = 42,000$. (a): The evolution of the epidemic under the optimal controls, where the left vertical axis corresponds to evolution of susceptible individuals (S_i), and the right vertical axis corresponds to evolution of latent (E_i), mild (I_{im}), and severe (I_{is}) patients (i = 1,2). (b): Optimal combination of social distancing strategy ($c^*(t)$) and detection rate ($\delta^*(t)$). (c): The variation in hospital beds and isolated resources occupancy. (d): The effective reproduction number (R_t) at time t.

References

[1] Science brief: SARS-CoV-2 infection-induced and vaccine-induced immunity, https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/vaccine-induced-immunity.html(Retrieved 28 July 2022).
[2] National Bureau of statistics of China, http://www.stats.gov.cn/english/Statisticaldata/CensusData/.
[3] The Statistical yearbook of Xi 'an - 2021, http://tjj.shaanxi.gov.cn/.
[4] Xi 'an held the 47th press conference on COVID-19 prevention and control, http://www.shaanxi.gov.cn/xw/sxyw/202201/t20220106_2206716.html.
[5] Past seasons estimated influenza disease burden, https://www.cdc.gov/flu/about/burden/past-seasons.html.
[6] S. Abbott, K. Sherratt, M. Gerstung, S. Funk, Estimation of the test to test distribution as a proxy for generation interval distribution for the Omicron variant in England, MedRxiv (2022).
[7] M. Bin, P.Y. Cheung, E. Crisostomi, P. Ferraro, H. Lhachemi, R. Murray-Smith, C. Myant, T. Parisini, R. Shorten, S. Stein, et al., Post-lockdown abatement of COVID-19 by fast periodic switching, PLoS Comput. Biol. 17 (1) (2021) e1008604.
[8] A. Boccia, M.D. de Pinho, R.B. Vinter, Optimal control problems with mixed and pure state constraints, SIAM J. Control Optim. 54 (6) (2016) 3061–3083.
[9] L.T. Brandal, E. MacDonald, L. Veneti, T. Ravlo, H. Lange, U. Naseer, S. Feruglio, K. Bragstad, O. Hungnes, L.E. Ødeskaug, et al., Outbreak caused by the SARS-CoV-2 omicron variant in Norway, november to december 2021, Eurosurveillance 26 (50) (2021) 2101147.
[10] J. Cai, X. Deng, J. Yang, K. Sun, H. Liu, Z. Chen, C. Peng, X. Chen, Q. Wu, J. Zou, et al., Modeling transmission of SARS-CoV-2 omicron in China, Nature Med. (2022) 1.
[11] R. Carli, G. Cavone, N. Epicoco, P. Scarabaggio, M. Dotoli, Model predictive control to mitigate the COVID-19 outbreak in a multi-region scenario, Annu. Rev. Control 50 (2020) 373–393.
[12] S. Contreras, J. Dehning, M. Loidolt, J. Zierenberg, F.P. Spitzner, J.H. Urrea-Quintero, S.B. Mohr, M. Wilczek, M. Wibral, V. Priesemann, The challenges of containing SARS-CoV-2 via test-trace-and-isolate, Nature Commun. 12 (1) (2021) 1–13.
[13] S. Contreras, J. Dehning, S.B. Mohr, S. Bauer, F.P. Spitzner, V. Priesemann, Low case numbers enable long-term stable pandemic control without lockdowns, Sci. Adv. 7 (41) (2021) eabg2243.
[14] S. Contreras, V. Priesemann, Risking further COVID-19 waves despite vaccination, Lancet Infect. Dis. 21 (6) (2021) 745–746.
[15] P. Danza, T.H. Koo, M. Haddix, R. Fisher, E. Traub, K. OYong, S. Balter, SARS-CoV-2 infection and hospitalization among adults aged $\geq$ 18 years, by vaccination status, before and during SARS-CoV-2 b. 1.1. 529 (omicron) variant predominance—los angeles county, California, november 7, 2021-january 8, 2022, Morb. Mortal. Wkly. Rep. 71 (5) (2022) 177.
[16] D.K. Das, A. Khatua, T.K. Kar, S. Jana, The effectiveness of contact tracing in mitigating COVID-19 outbreak: A model-based analysis in the context of India, Appl. Math. Comput. 404 (2021) 126207.
[17] F. Di Lauro, I.Z. Kiss, J.C. Miller, Optimal timing of one-shot interventions for epidemic control, PLoS Comput. Biol. 17 (3) (2021) e1008763.
[18] M. Diehl, H.G. Bock, H. Diedam, P.-B. Wieber, Fast direct multiple shooting algorithms for optimal robot control, in: Fast Motions in Biomechanics and Robotics, Springer, 2006, pp. 65–93.
[19] C. Gaimon, Optimal control theory: Applications to management science and economics, 2002.
[20] Y. Goldberg, M. Mandel, Y.M. Bar-On, O. Bodenheimer, L. Freedman, N. Ash, S. Alroy-Preis, A. Huppert, R. Milo, Protection and waning of natural and hybrid COVID-19 immunity, MedRxiv (2021).
[21] V. Grimm, F. Mengel, M. Schmidt, Extensions of the SEIR model for the analysis of tailored social distancing and tracing approaches to cope with COVID-19, Sci. Rep. 11 (1) (2021) 1–16.
[22] S.M. Grundel, S. Heyder, T. Hotz, T.K. Ritschel, P. Sauerteig, K. Worthmann, How to coordinate vaccination and social distancing to mitigate SARS-CoV-2 outbreaks, SIAM J. Appl. Dyn. Syst. 20 (2) (2021) 1135–1157.
[23] A.D. Iuliano, J.M. Brunkard, T.K. Boehmer, E. Peterson, S. Adjei, A.M. Binder, S. Cobb, P. Graff, P. Hidalgo, M.J. Panaggio, et al., Trends in disease severity and health care utilization during the early omicron variant period compared with previous SARS-CoV-2 high transmission periods—United States, december 2020-january 2022, 2022.
[24] M. Kantner, T. Koprucki, Beyond just "flattening the curve": Optimal control of epidemics with purely non-pharmaceutical interventions, J. Math. Ind. 10 (1) (2020) 1–23.
[25] D. Kim, J. Jo, J.-S. Lim, S. Ryu, Serial interval and basic reproduction number of SARS-CoV-2 omicron variant in South Korea, MedRxiv (2021).
[26] A. Kleiderman, Covid: Travel and mask rules tightened over Omicron variant, https://www.bbc.com/news/uk-59445124.
[27] J. Koenemann, G. Licitra, M. Alp, M. Diehl, Openocl–open optimal control library, 2017.
[28] J. Köhler, L. Schwenkel, A. Koch, J. Berberich, P. Pauli, F. Allgöwer, Robust and optimal predictive control of the COVID-19 outbreak, Annu. Rev. Control 51 (2021) 525–539.
[29] F.L. Lewis, D. Vrabie, V.L. Syrmos, Optimal Control, John Wiley & Sons, 2012.
[30] X. Lü, H.-w. Hui, F.-f. Liu, Y.-l. Bai, Stability and optimal control strategies for a novel epidemic model of COVID-19, Nonlinear Dynam. 106 (2) (2021) 1491–1507.
[31] L. Matrajt, J. Eaton, T. Leung, E.R. Brown, Vaccine optimization for COVID-19: Who to vaccinate first? Sci. Adv. 7 (6) (2021) eabf1374.
[32] H. Maurer, Tutorial on control and state constrained optimal control problems, in: SADCO Summer School 2011-Optimal Control, 2011.
[33] T.A. Perkins, G. España, Optimal control of the COVID-19 pandemic with non-pharmaceutical interventions, Bull. Math. Biol. 82 (9) (2020) 1–24.

[34] S.P. Sethi, The maximum principle: Pure state and mixed inequality constraints, in: Optimal Control Theory, Springer, 2019, pp. 125–158.
[35] Z.-H. Shen, Y.-M. Chu, M.A. Khan, S. Muhammad, O.A. Al-Hartomy, M. Higazy, Mathematical modeling and optimal control of the COVID-19 dynamics, Results Phys. 31 (2021) 105028.
[36] B. Tang, X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao, J. Wu, Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions, J. Clin. Med. 9 (2) (2020) 462.
[37] B. Tang, F. Xia, S. Tang, N.L. Bragazzi, Q. Li, X. Sun, J. Liang, Y. Xiao, J. Wu, The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemics in the final phase of the current outbreak in China, Int. J. Infect. Dis. 95 (2020) 288–293.
[38] C. Tsay, F. Lejarza, M.A. Stadtherr, M. Baldea, Modeling, state estimation, and optimal control for the US COVID-19 outbreak, Sci. Rep. 10 (1) (2020) 1–12.
[39] R. Verschueren, G. Frison, D. Kouzoupis, N. van Duijkeren, A. Zanelli, R. Quirynen, M. Diehl, Towards a modular software package for embedded optimization, IFAC-PapersOnLine 51 (20) (2018) 374–380.
[40] X. Wang, Q. Li, X. Sun, S. He, F. Xia, P. Song, Y. Shao, J. Wu, R.A. Cheke, S. Tang, et al., Effects of medical resource capacities and intensities of public mitigation measures on outcomes of COVID-19 outbreaks, BMC Public Health 21 (1) (2021) 1–11.
[41] S. Wright, J. Nocedal, et al., Numerical optimization, Springer Sci. 35 (67–68) (1999) 7.
[42] Y. Xiao, B. Tang, J. Wu, R.A. Cheke, S. Tang, Linking key intervention timing to rapid decline of the COVID-19 effective reproductive number to quantify lessons from mainland China, Int. J. Infect. Dis. 97 (2020) 296–298.
[43] M.-Z. Yin, Q.-W. Zhu, X. Lü, Parameter estimation of the incubation period of COVID-19 based on the doubly interval-censored data model, Nonlinear Dynam. 106 (2) (2021) 1347–1358.
[44] J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, et al., Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China, Science 368 (6498) (2020) 1481–1486.
[45] J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen, et al., Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside hubei province, China: A descriptive and modelling study, Lancet Infect. Dis. 20 (7) (2020) 793–802.
[46] M. Zhang, J. Xiao, A. Deng, Y. Zhang, Y. Zhuang, T. Hu, J. Li, H. Tu, B. Li, Y. Zhou, et al., Transmission dynamics of an outbreak of the COVID-19 delta variant b. 1.617. 2—Guangdong province, China, may–june 2021, China CDC Wkly. 3 (27) (2021) 584.