



A data-driven spatially-specific vaccine allocation framework for COVID-19

Zhaofu Hong¹ · Yingjie Li^{2,5} · Yeming Gong³  · Wanying Chen⁴

Accepted: 18 October 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Although coronavirus disease 2019 (COVID-19) vaccines have been introduced, their allocation is a challenging problem. We propose a data-driven, spatially-specific vaccine allocation framework that aims to minimize the number of COVID-19-related deaths or infections. The framework combines a regional risk-level classification model solved by a self-organizing map neural network, a spatially-specific disease progression model, and a vaccine allocation model that considers vaccine production capacity. We use data obtained from Wuhan and 35 other cities in China from January 26 to February 11, 2020, to avoid the effects of intervention. Our results suggest that, in region-wise distribution of vaccines, they should be allocated first to the source region of the outbreak and then to the other regions in order of decreasing risk whether the outcome measure is the number of deaths or infections. This spatially-specific vaccine allocation policy significantly outperforms some current allocation policies.

Keywords Data-driven decision making · COVID-19 · Spatially-specific SEIR model · Deep learning · Vaccine allocation

1 Introduction

Coronavirus disease 2019 (COVID-19) is an acute respiratory infectious disease and a growing public health problem. The severity of COVID-19 varies from asymptomatic infections to mild upper respiratory tract illnesses, severe interstitial pneumonia with respiratory failure, and even death (Rothan & Byrareddy, 2020). Safe and effective COVID-19 vaccines

✉ Yeming Gong
GONG@em-lyon.com

¹ School of Management, Northwestern Polytechnical University, Xi'an, People's Republic of China

² School of Civil Engineering, Central South University, Changsha, People's Republic of China

³ EMLYON Business School, Écully, France

⁴ School of Management and E-Business, Zhejiang Gongshang University, Hangzhou, People's Republic of China

⁵ School of Management, Lanzhou University, Lanzhou, People's Republic of China

are considered promising solutions to control the epidemic, although transmission mitigation strategies such as physical distance have protected countless people from being infected (Jeyanathan et al., 2020). With the help of vaccines, herd immunity can be achieved when 70–80% of people in the group gain immunity, thereby preventing large-scale outbreaks. People can then have normal contact and lives. Because of the severity of the COVID-19 outbreak and the scarcity of vaccines, the demand for vaccines tends to outstrip supply, necessitating optimal allocation to ensure that they can be used most effectively.

Several allocation policies have been proposed. Proposals have been made to prioritize medical workers, elderly people (age > 65), and those with certain serious medical conditions, as well as to consider differences in ethnicity and socioeconomic status (Subbaraman, 2020). The World Health Organization (WHO) suggests allocating vaccines according to the population of each country. Although population allocation is in line with the principle of equality, providing medical materials to severely affected regions according to international aid practices may be more reasonable (Emanuel et al., 2020). The spread of COVID-19 has typical characteristics of divergently spreading around the outbreak areas, and the severity of the epidemic varies in different cities (Huang et al., 2020). It is important to distinguish the epidemic risk level of each city and allocate vaccines according to the epidemic risk level. Zhou et al. (2021) defined five epidemic risk levels based on herd immunity rate. We further propose a regional risk classification model that considers the density of the newly infected population and the migration scale of the population from the source areas of the epidemic to each city. The regional risk classification model could identify areas with serious epidemics, and a vaccine allocation model based on the regional risk classification could control the epidemic in a targeted manner, reducing resource waste.

Considering the scarcity of vaccines in the context of the COVID-19 outbreak, we must develop a rigorous decision support system for allocating vaccines by considering the transmission characteristics of COVID-19. However, most of the available literature focuses on modeling to determine how to allocate vaccine supplies based on age (Sah et al., 2018). The risk level of each city and spatial optimization of vaccine allocation have not been fully explored in detail. Previous methods of vaccine allocation are worthy of improvement and expansion. This study fills such gaps with a data-driven and spatially-specific vaccine allocation method for the current practice. We address the following research questions: (i) *How should the COVID-19 risk level be determined for each city?* (ii) *How should vaccines be allocated to each risk-level region?*

To answer these questions, we propose a data-driven, spatially-specific vaccine allocation framework with the overall objective of minimizing infections or deaths. The framework incorporates a regional risk-level classification model that is established by a self-organizing map (SOM) neural network (a deep learning algorithm), a spatially-specific disease progression model that is trained by solving an inverse problem, and a vaccine allocation model that considers vaccine production capacity. Specifically, we first introduce a clustering algorithm to solve the regional risk-level classification model and partition the set of cities into regions according to the risk that the epidemic will spread in each city. We then train the spatially-specific susceptible-exposed-infectious-recovered (SEIR) model to predict deaths and infections. Finally, we apply a genetic algorithm to solve the vaccine allocation problem with vaccine capacity constraints. We investigate the vaccine allocation policy for 36 cities in China using data from January 26 to February 11, 2020. The analysis results show that if vaccines are distributed by region, the epidemic sources and high-risk regions should be prioritized for vaccine allocation.

The main contributions of this study are as follows. (i) We propose an integrated framework for designing a COVID-19 vaccine allocation plan combining a regional risk classification

model, a disease progression model, and an optimization method. (ii) We develop a hybrid algorithm to solve these models. (iii) We conduct a simulation case study and find that limited vaccine doses should be allocated first to the source region of the outbreak and then to other regions in the order of decreasing risk. We further find that the region-based vaccine allocation policy is better than the age-based policy.

The remainder of this paper is organized as follows. Section 2 reviews the literature and highlights this study's main contributions. In Sect. 3, we propose a framework for allocating vaccines to minimize the number of deaths or infections. In Sect. 4, we analyze the results of a simulation case study and verify the framework's effectiveness. Section 5 concludes the paper and discusses future research.

2 Literature review

This study is closely related to two research streams—COVID-19 epidemic prediction based on compartmental models, and the usage of mathematical models to solve the vaccine allocation problem. Thus, we review studies relevant to each stream and highlight the differences between this study and the existing studies.

2.1 COVID-19 epidemic prediction by compartmental models

Compartmental models have been used to predict the transmission of COVID-19 since the outbreak of the disease, and they have proven superior to other models, such as the regression model (Pandey et al., 2020). Compartmental models divide individuals into different compartments according to their epidemiological states, such as susceptible, infected, and recovered states, and the shifting relationships between different epidemiological states are represented by differential equations (Allen, 2017). Kozio et al. (2020) proposed a fractional-order SIR (susceptible-infected-recovered) epidemic model to predict the number of infected individuals and estimated the optimal parameters of the fractional-order SIR epidemic model using the Monte Carlo method. Okuonghae and Omame (2020) formulated an improved SEIR model with an asymptomatic infectious class, a symptomatic infectious class, and a detected infectious class to examine the effects of various non-pharmaceutical interventions (such as common social distancing, the use of face masks, and case detection) on the dynamic transmission of COVID-19 in Lagos. He et al. (2020) extended the SEIR model by considering quarantine and treatment and used a particle swarm optimization (PSO) algorithm to estimate the parameters of the model. However, few studies have considered the infectivity of subclinical and pre-symptomatic patients, as well as a subclinical class that carries the virus but remains asymptomatic from infection to recovery. These are also important features of COVID-19.

In addition to extending the SIR model from the perspective of the states of the model, some scholars also considered the differences in the population by dividing the population into subgroups according to age or location, and further improved the infectious disease prediction models. Mwalili et al. (2020) proposed an age-structured SEIR model to show different epidemic peaks of COVID-19 for various age structures, and examined the influence of age and social contact structures on mitigation strategies. Aràndiga et al. (2020) formulated a spatial-temporal suspected asymptomatic infected retired model to analyze the effect of mobility on the evolution of COVID-19 after an isolation period by modeling the flow of people between different regions. These improved SEIR models consider the structure of age

or spatial location. We further consider the specificity of the epidemic risk level, which can accurately reflect the characteristics of epidemic transmission and is worth considering in the prediction model.

According to the transmission characteristics of COVID-19, the improved SEIR model proposed in this study considers epidemic risk-level subgroups, examines the transmissibility of asymptomatic patients, and is trained using real data. Specifically, the model distinguishes the epidemic situation at each risk level and considers the epidemic situation in each region. In addition, this model examines not only the transmissibility of infected people, but also the transmissibility of asymptomatic individuals who are pre-symptomatic and subclinical. In this study, we improved the prediction accuracy by training a region-specific SEIR model using real data. The improved SEIR model in this study is conducive to developing better disease mitigation plans based on the characteristics of COVID-19 and evaluating the effectiveness of disease intervention strategies.

2.2 Vaccine allocation based on mathematical models

Vaccination is an important strategy for disease intervention. Some studies have focused on modeling to determine how to allocate vaccine doses based on age, and few studies have considered vaccine allocation according to region. Mylius et al. (2008) investigated the effect of estimated infectious contact patterns within and among age groups, and the timing of vaccination on the effectiveness of vaccine allocation programs for various risk groups. Medlock and Galvani (2009) set up a compartmental model that included age structure and vaccination status. They developed a mathematical program to allocate influenza vaccine based on age to minimize the total number of infections, total number of deaths, years of life lost, contingent valuation, and total cost. The optimization problem was solved numerically using a nonlinear simplex method. Ndeffo Mbah et al. (2014) developed an age-specific compartmental model to consider human demography, four dengue stereotypes, and various profiles of dengue hemorrhagic fever (DHF) in different countries. They also compared the effect of an age-targeted policy with that of an infection-history-targeted policy on reducing the incidence over five years.

Some researchers have examined the problem of vaccine allocation by city (Araz et al., 2012). However, their geospatial models may not be suitable for targeted vaccine allocation among a large number of areas, because the specific policies of each city are considered, which is not feasible, and the calculations are too complicated. Other scholars have partitioned cities by geographical location and then performed vaccine allocation (Teytelman & Larson, 2013). This partitioning method is not optimal because it does not consider the city specificity of the epidemic. In particular, the severity of outbreaks may vary between cities within the same geographical area while being similar in cities in different areas.

With the development of COVID-19 vaccines, studies on their allocation have emerged. According to the literature, most COVID-19 vaccine allocation policies are based on age (Bubar et al., 2021; Matrajt et al., 2020). Foy et al. (2021) investigated the interaction of COVID-19 vaccine allocation policies in different age groups with non-pharmacological interventions, namely, social distancing, mandatory mask wearing, and blockade. Spatial characteristics, which are an important factor in the spread of COVID-19 and should be considered in vaccine allocation, are rarely considered in the vaccine allocation problem.

This study fills these research gaps by developing a data-driven and region-specific COVID-19 vaccine-allocation framework. In the first step, the proposed framework partitions cities into several regions according to the spread of the epidemic in each city, which

could identify the areas with serious epidemics and help control the epidemic in a targeted manner without wasting resources. In addition, the framework combines a machine learning technique, spatially-specific SEIR, and an optimization model that complements vaccine allocation methods, enriches the application of big data in vaccine allocation, and provides an effective vaccine allocation plan. The proposed vaccine allocation framework can solve existing problems in the spatial allocation of vaccines.

3 A spatially-specific vaccine allocation framework

In this section, we propose a data-driven, spatially-specific, vaccine allocation framework to derive spatial allocation policies using a regional risk level classification model, a spatially-specific disease progression model, and a vaccine allocation model (Fig. 1). In Sect. 3.1, we propose a regional risk classification model driven by regional epidemic data and migration data using a self-organizing map (SOM) neural network to solve the model. In Sect. 3.2, we train a spatially-specific disease progression model by solving an inverse problem. In Sect. 3.3, we formulate a vaccine allocation model that considers vaccine production capacity, and a genetic solution algorithm is presented. In Sect. 3.4, we summarize the algorithms used in this framework.

3.1 Regional risk classification model

A regional risk-level classification model, solved by a clustering method, is used to classify the risk levels of the cities. Clustering methods can be combined with frontier estimation techniques to identify objects with similar characteristics (Dai & Kuosmanen, 2014). We categorize cities into K risk levels. The first risk level (cluster) includes all cities where the epidemic first appeared. For example, at the start of the sample period for the 36 Chinese cities, the epidemic was only in Wuhan, so the first cluster only includes Wuhan. Using a deep learning algorithm, that is, the self-organizing map algorithm, other cities are grouped into $(K-1)$ clusters, according to the density of their newly infected population (D_i) and the migration scale of the population from the source areas of the epidemic to each city (PM_i). These two features reflect the speed of disease progression (Xiang and Wang, 2020) and are formulated as follows:

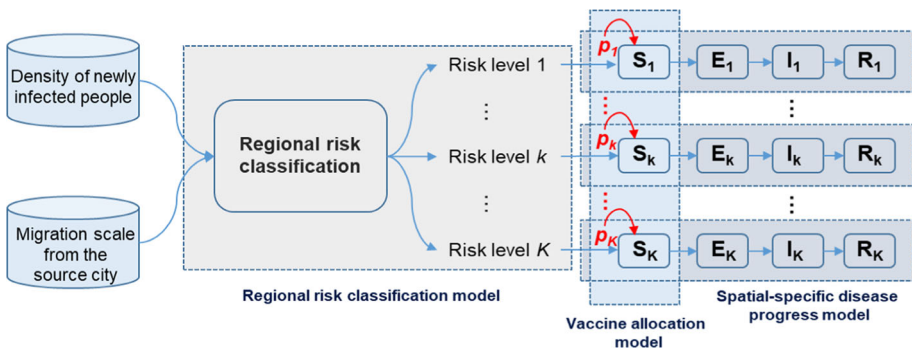


Fig. 1 Data-driven spatially-specific vaccine allocation framework

$$D_i = \frac{I'_i(N) - I'_i(0)}{I'_i(0)A_i}, \text{ and } PM_i = \sum_{t=0}^N 1000P_i(t)F_i(t).$$

where $I'_i(0)$ and $I'_i(N)$ denote the total number of historical infections in city i at the initial and final moments, respectively, covered by the available data. A_i is the population size of city i , $P_i(t)$ is the migration index, and $F_i(t)$ denotes the proportion of the population that migrated from the source cities of the epidemic to city i to the total migrant population from the outbreak cities.

The self-organizing map (SOM) neural network is essentially a neural network with only an input layer and a hidden layer; the latter is used to cluster data. A node in the hidden layer represents a class that must be clustered. The method of “competitive learning” is adopted during training, and each input example finds a node that best matches it in the hidden layer; that node is called the winning neuron. Then, the stochastic gradient descent method is used to update the parameters of the winning neuron. Meanwhile, the points close to the winning neuron also update their parameters appropriately, according to their distance from the winning neuron (Vesanto et al., 2000). Clustering using a self-organizing map neural network has advantageous features. First, the SOM is an unsupervised network that does not distinguish between training and test sets. Then, the function of “near excitation and far inhibition” in the nervous system of the brain is realized through competition between neurons, and it has the ability to map high-dimensional input to low-dimensional output. Compared with clustering methods, such as k -means, the SOM neural network is less affected by the initial value, can achieve more stable clustering results, and is less susceptible to noise.

3.2 Spatially-specific COVID-19 SEIR models

3.2.1 COVID-19 SEIR model

We use a disease progression model, together with the transmission probability and death rate prediction model, to estimate the fraction of people in different health stages in each city. The transmission probabilities of symptomatic patients (β_E), asymptomatic patients (β_I), and death rate (ν) are used to reflect the contagiousness of the disease (Hou et al., 2020; Yang et al., 2020). An inverse problem is appropriate for predicting these parameters (Libotte et al., 2020).

Considering the unique characteristics of COVID-19, such as the long incubation period, presence of subclinical patients, and infectiousness in asymptomatic patients, we establish a SEIR model using a system of differential equations that contains five epidemiological states: susceptible, asymptomatic (subdivided into subclinical and pre-symptomatic), symptomatic, recovered, and dead states. Figure 2 illustrates the shifting relationships between the different epidemiological states. Let $S[t]$, $E_1[t]$, $E_2[t]$, $I[t]$, $R[t]$, and $D[t]$ be the numbers of susceptible, subclinical, pre-symptomatic, symptomatic, recovered, and dead individuals, respectively. The force of infection (λ) is given by the contact rate (φ), transmission probability given contact by an asymptomatic patient (β_E), transmission probability given contact with a symptomatic patient (β_I), and the total number of people in each state. The progress rate from the pre-symptomatic state to the symptomatic state is τ , whereas that from the subclinical state to recovery is ζ , the recovery rate is γ , and the COVID-19-induced death rate is ν . Our COVID-19 SEIR model is given by:

$$\frac{dS}{dt} = -\lambda S, \quad (1)$$

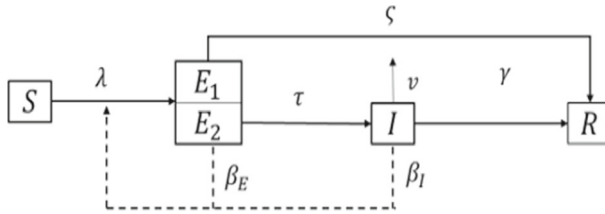


Fig. 2 Model diagram (solid lines indicate individual transitions to different states, and dashed lines represent the transmission routes)

$$\frac{dE}{dt} = \lambda S - \tau E_2 - \zeta E_1, \tag{2}$$

$$\frac{dI}{dt} = \tau E_2 - v I - \gamma I, \tag{3}$$

$$\frac{dR}{dt} = \gamma I + \zeta E_1, \tag{4}$$

$$\frac{dD}{dt} = v I, \tag{5}$$

where $\lambda = \frac{\varphi\beta_E E + \varphi\beta_I I}{S + E + I + R}$.

In the above, Eqs. (1)–(5) represent the changes in the number of susceptible people, latent people, infected people, recovered people, and dead people per unit of time, respectively. Specifically, Eq. (2) indicates that the change in the latent population per time unit is the number of susceptible people converting to a latent population minus the number of latent people converting to recovered or infected people. Equation (3) shows that the change in the infected population per time unit is the number of latent people converted to the infected population minus the number of infected people converted to recovered or dead people.

Subsequently, model parameters describing the characteristics of a particular infectious disease must be determined. Key parameters such as $\varphi\beta_E$, $\varphi\beta_I$, and v are obtained by solving the inverse problem formulated in Eq. (6), which is similar to the equation proposed by Libotte et al. (2020). The inverse problem arises from the requirement to determine the parameters of a theoretical model such that it can be employed to simulate the behavior of the system under different operating conditions (Libotte et al., 2020). The probability of transmission given contact by a symptomatic patient may be higher than that of an asymptomatic patient (Hou et al., 2020) (constraint 7). The actual numbers of infections and deaths are $I^a(t)$ and $D^a(t)$, while $I^e(t)$ and $D^e(t)$ denote the estimated numbers of infections and deaths, respectively. Parameter N is the total number of periods (e.g., days in our numerical example) in the dataset.

$$\min_{\varphi\beta_E, \varphi\beta_I, v} \frac{\sum_{t=0}^N [I^a(t) - I^e(t)]^2}{\sum_{t=0}^N \frac{[I^a(t)]^2}{N}} + \frac{\sum_{t=0}^N [D^a(t) - D^e(t)]^2}{\sum_{t=0}^N \frac{[D^a(t)]^2}{N}} \tag{6}$$

$$\text{Subject to } \beta_E < \beta_I \tag{7}$$

We use the linear approximation (COBYLA) algorithm to solve the problem above and used R-squared to test the fitting effect under the optimal parameters (Ogut & Bishop, 2010).

3.2.2 Spatially-specific SEIR model

We develop a spatially-specific SEIR model by extending the basic model. For notational convenience, we use subscript k to refer to a region, $k = 1, 2, \dots, K$. The mortality rate (v_k) and force of infection (λ_k) are key parameters for vaccine allocation decisions. The force of infection (λ_k) is given by the contact rate per person in region k with individuals in region κ ($\varphi_{k\kappa}$), contact rate per person in region κ with individuals in region k ($\varphi_{\kappa k}$), transmission probability given contact by an asymptomatic patient (β_E), transmission probability given contact by a symptomatic patient (β_I), and total number of people in each region at various stages (Medlock & Galvani, 2009). Parameters such as v_k , β_E , and β_I are obtained by solving an inverse problem. The spatially-specific SEIR model is given by:

$$\begin{aligned} \frac{dS_k}{dt} &= -\lambda_k S_k, \\ \frac{dE_k}{dt} &= \lambda_k S_k - \tau E_{2k} - \varsigma E_{1k}, \\ \frac{dI_k}{dt} &= \tau E_{2k} - v_k I_k - \gamma I_k, \\ \frac{dR_k}{dt} &= \gamma I_k + \varsigma E_{1k}, \\ \frac{dD_k}{dt} &= v_k I_k, \\ \lambda_k &= \sum_{\kappa=1}^K \frac{\beta_I I_\kappa + \beta_E E_\kappa}{2} \left(\frac{\varphi_{k\kappa}}{S_k + E_\kappa + I_\kappa + R_\kappa} + \frac{\varphi_{\kappa k}}{S_k + E_k + I_k + R_k} \right), \\ S(t) &= \sum_{k=1}^K S_k(t), \\ E(t) &= \sum_{k=1}^K E_{2k}(t) + \sum_{k=1}^K E_{1k}(t), \\ I(t) &= \sum_{k=1}^K I_k(t), \text{ and} \\ R(t) &= \sum_{k=1}^K R_k(t). \end{aligned}$$

3.3 Vaccine allocation model

The vaccine intervention begins at $t = 0$. Vaccines can directly reduce the force of infection or indirectly affect it by reducing the number of infected individuals. The force of infection of unvaccinated people is λ'_k , and the force of infection of vaccinated people decreases to $\lambda'_k(1 - \varepsilon)$ in region k , considering the effectiveness of the vaccine against infection (ε). As the vaccination dose is the bottleneck of the mass vaccination policy because of the limited vaccine production capacity, the vaccine allocation part of our technique is only subject to the constraint of the vaccine amount (Medlock & Galvani, 2009; Ndeffo Mbah et al., 2014). A spatially specific ring, p_k , denotes the proportion of people in each vaccinated cluster. The number of people vaccinated must meet the vaccine dose constraints, $\left(\sum_{k=1}^K p_k S_k(0) \leq W \right)$.

In addition, after an outbreak of the disease, people may self-isolate; therefore, we suppose that the vaccine may work immediately with full effectiveness. The postponement of vaccine effectiveness affects the people in each city equally; therefore, the vaccine allocation decision will not change. With the objective $\min_{p_k} b * (S_{all}(0) - S_{all}(T)) + S_{all}(0) + E_{all}(0) + I_{all}(0) + R_{all}(0) - S_{all}(T) - E_{all}(T) - I_{all}(T) - R_{all}(T)$, we minimized the number of infected individuals or deaths (Medlock & Galvani, 2009). The spatially-specific COVID-19 vaccine allocation model is as follows:

$$\min_{p_k} b * [S_{all}(0) - S_{all}(T)] + S_{all}(0) + E_{all}(0) + I_{all}(0) + R_{all}(0) - S_{all}(T) - E_{all}(T) - I_{all}(T) - R_{all}(T),$$

subject to $0 \leq p_k \leq 1$,

$$\sum_{k=1}^K p_k S_k(0) \leq W,$$

$$S_{all}(t) = \sum_{k=1}^K S_k^U(t) + \sum_{k=1}^K S_k^V(t),$$

$$E_{all}(t) = \sum_{k=1}^K E_k^U(t) + \sum_{k=1}^K E_k^V(t),$$

$$I_{all}(t) = \sum_{k=1}^K I_k^U(t) + \sum_{k=1}^K I_k^V(t),$$

$$R_{all}(t) = \sum_{k=1}^K R_k^U(t) + \sum_{k=1}^K R_k^V(t),$$

$$\frac{dS_k^U}{dt} = -\lambda'_k S_k^U,$$

$$\frac{dE_k^U}{dt} = \lambda'_k S_k^U - \tau E_{2k}^U - \varsigma E_{1k}^U,$$

$$\frac{dI_k^U}{dt} = \tau E_{2k}^U - v_k I_k^U - \gamma I_k^U,$$

$$\frac{dR_k^U}{dt} = \gamma I_k^U + \varsigma E_{1k}^U,$$

$$\frac{dD_k^U}{dt} = v_k I_k^U,$$

$$\frac{dS_k^V}{dt} = -\lambda'_k (1 - \varepsilon) S_k^V,$$

$$\frac{dE_k^V}{dt} = \lambda'_k (1 - \varepsilon) S_k^V - \tau E_{2k}^V - \varsigma E_{1k}^V,$$

$$\frac{dI_k^V}{dt} = \tau E_{2k}^V - v_k I_k^V - \gamma I_k^V,$$

$$\frac{dR_k^V}{dt} = \gamma I_k^V + \varsigma E_{1k}^V,$$

$$\frac{dD_k^V}{dt} = v_k I_k^V,$$

$$\lambda'_k = \sum_{\kappa=1}^K \frac{\beta_I I_\kappa^V + \beta_I I_\kappa^U + \beta_E E_\kappa^U + \beta_E E_\kappa^V}{2} \left(\frac{\varphi_{k\kappa}}{S_\kappa + E_\kappa + I_\kappa + R_\kappa} + \frac{\varphi_{\kappa k}}{S_k + E_k + I_k + R_k} \right).$$

where $S_k^U(t)$, $E_k^U(t)$, $I_k^U(t)$, $R_k^U(t)$, and $D_k^U(t)$ are the numbers of unvaccinated susceptible, asymptomatic, symptomatic, recovered, and dead individuals, respectively, whereas $S_k^V(t)$, $E_k^V(t)$, $I_k^V(t)$, $R_k^V(t)$, and $D_k^V(t)$ are defined similarly but for vaccinated people. The parameters $S_{all}(t)$, $E_{all}(t)$, $I_{all}(t)$, and $R_{all}(t)$ are the total number of people in these states. The number of available vaccine doses for allocation is W . When $b = 0$, the outcome measure is the total number of deaths after T ; when $b = 1$, the outcome measure is the total number of infections after time T . The spatially-specific COVID-19 vaccine allocation model (combining the spatially-specific SEIR and optimization models) distinguishes the epidemic situation of each risk level by fully considering the epidemic situation in each region and predicts the minimum number of deaths or infections to determine an effective vaccine allocation.

A complex prediction model is embedded in this allocation model with k decision variables, which makes it difficult to solve. Therefore, we use a genetic algorithm to obtain the value of the spatially-specific ring in each zone, which is verified to be efficient in solving vaccine distribution problems (Patel et al., 2005). Let p_1, p_2, \dots, p_{k-1} , and p_k be genes that together form a chromosome. After the emergence of the initial population, a better approximate solution can be produced according to the principle of the survival of the fittest. In each generation, chromosomes are selected according to their fitness, and combined crossover and mutation are performed with the help of natural genetic operators to generate a population representing a new solution set. The optimal chromosome in the last generation population can be used as an approximately optimal solution to the problem after decoding (Min et al., 2006). The computational results show that the algorithm converges efficiently and effectively to the best solutions because the number of decision variables is normally not too large.

3.4 Summary of the algorithms for our framework

In this framework, the regional risk classification problem is solved by SOM, the training required in the spatially-specific disease model is tackled by a linear approximation algorithm, and the vaccine allocation problem combined with our spatially-specific disease model is solved by our genetic algorithm. We summarize the hybrid algorithm and present its procedure as follows.

Input: Migration scale from the source region, regional epidemic data, and regional population data.

Output: City classification and vaccine allocation decisions.

Begin

Step 1 Calculate the optimal number of clusters $K-1$ and predict the classification for each city by SOM.

Step 2 Calculate the parameters of the spatially-specific disease model by linear approximation algorithm.

Step 3 Calculate the optimal proportion of each cluster vaccinated against COVID-19 p_k , and predict the minimal number of COVID-19-related deaths or infections after time T by GA.

- i.* Initialize p_k randomly within the specified range
- ii.* Predict the number of COVID-19-related deaths or infections after time T
- iii.* Update and optimize p_k through GA until the predicted number of deaths or infections after time T is the least.

End

4 Case study

In this section, we explore the use of the proposed framework to allocate a limited number of vaccine doses to different regions at the beginning of an outbreak to minimize the total number of deaths and infections. We also compare our results with those obtained using an age-specific vaccine allocation method to determine the effectiveness of the proposed framework.

4.1 Data sources

In December 2019, an outbreak of COVID-19 was reported in Wuhan, and the epidemic spread from Wuhan to neighboring cities in China. We selected 35 cities with a large number of migrants from Wuhan. We collected data on deaths, infections, rehabilitation, and populations of these 36 cities, as well as the migration index of Wuhan and the proportion of the population moving out of Wuhan to the other 35 cities, from January 26 to February 11, 2020. We took the numbers of infected, recovered, dead, and latent people in 36 cities on January 26 as the initial values. We limited our data extraction to the period from January 26 to February 11, 2020, as this was a period of disease progression without intervention. The number of latent individuals was the total number of new infections over the next five days. The epidemic and migration data were obtained from the websites of the National Health Commission of the People's Republic of China and the Health Commission of the related provinces of China, and population data were obtained from the China National Bureau of Statistics in 2020.

4.2 Results of spatially-specific vaccine allocation

4.2.1 Classification of cities according to epidemic risk levels

In the simulation, Wuhan was the epicenter of an epidemic outbreak in China. Considering that disease progression in Wuhan may be much faster than in other cities, Wuhan is placed in a separate category. For the other 35 cities, population density of newly infected people and scale of population migration to each city are calculated in a standardized way. As shown in Fig. 3, 35 cities are divided into four clusters using the SOM algorithm. This classification result is consistent with the result obtained by the k -means algorithm given in Appendix 3. Therefore, we classify 36 cities into five risk levels, with risk increasing from level 1 to 5. We classify Wuhan as the only city at risk level 5 (Fig. 4). Based on Fig. 4, we find that the risk level is spatially dependent; being closer to the epicenter of the epidemic outbreak corresponds to a higher risk level. This finding is consistent with the circular diffusion trend reported for COVID-19 (Huang et al., 2020).

4.2.2 Spatially-specific SEIR training

We predict the number of infected persons and deaths in each city based on the parameter values listed in Table 1. In the simulation, we cannot obtain the value of the force of infection (λ'_k) because data related to the contact rates are lacking. Therefore, we used the ratio of infections in a region to the total infections in all regions per day (μ_k), which includes population contact information, to replace the force of infection (λ'_k) as a characteristic that distinguishes each region. This replacement is reasonable and does not affect the insights and conclusions of our model if time T is relatively short. The value of μ_k also remains the

Table 1 Epidemiological parameter values and computation results in region-specific SEIR model

Parameter	Value	Sources
Progress rate from subclinical state to recovery ς	0.25	Hu et al. (2020)
Progress rate from pre-symptomatic state to symptomatic state τ	0.196	Lauer et al. (2020)
Recovery rate γ	0.1	Roda et al. (2020)
Effectiveness of the vaccine against infection ϵ	0.9	Callaway (2020)
$\varphi\beta E$	0.21	$R^2 = 0.91$
$\varphi\beta I$	0.23	$R^2 = 0.91$
Ratio of subclinical patients to pre-symptomatic patients $\frac{E_1}{E_2}$	0.41	Hu et al. (2020)
Ratio of number of infections in each cluster to total number of infections per day $\mu_k, k = 1, 2, 3, 4, 5$	0.110, 0.202, 0.139, 0.023, 0.526,	-
Death rate $\nu_k, k = 1, 2, 3, 4, 5$	0.0003 0.0019 0.0030 0.0066 0.0075	$R^2 = 0.90$ $R^2 = 0.87$ $R^2 = 0.93$ $R^2 = 0.67$ $R^2 = 0.93$

same when T is small. To simplify the model, we do not consider this indirect effect. The spatially-specific COVID-19 vaccine allocation model for the data is provided in Appendix 1.

The coefficient of determination (R^2) was used to test the fitting effect of the predicted and true values (Ogut & Bishop, 2010; Teles, 2020; Ivanov & Leonenko, 2018). As shown in Fig. 5, only the coefficient of determination of v_4 was less than 0.9 ($R^2 = 0.67$). The intuition behind this is that there is only one city (i.e., Ezhou) in the 4th-level risk area and the number of deaths is small; thus, the result is easily affected by the noise of the data. The average

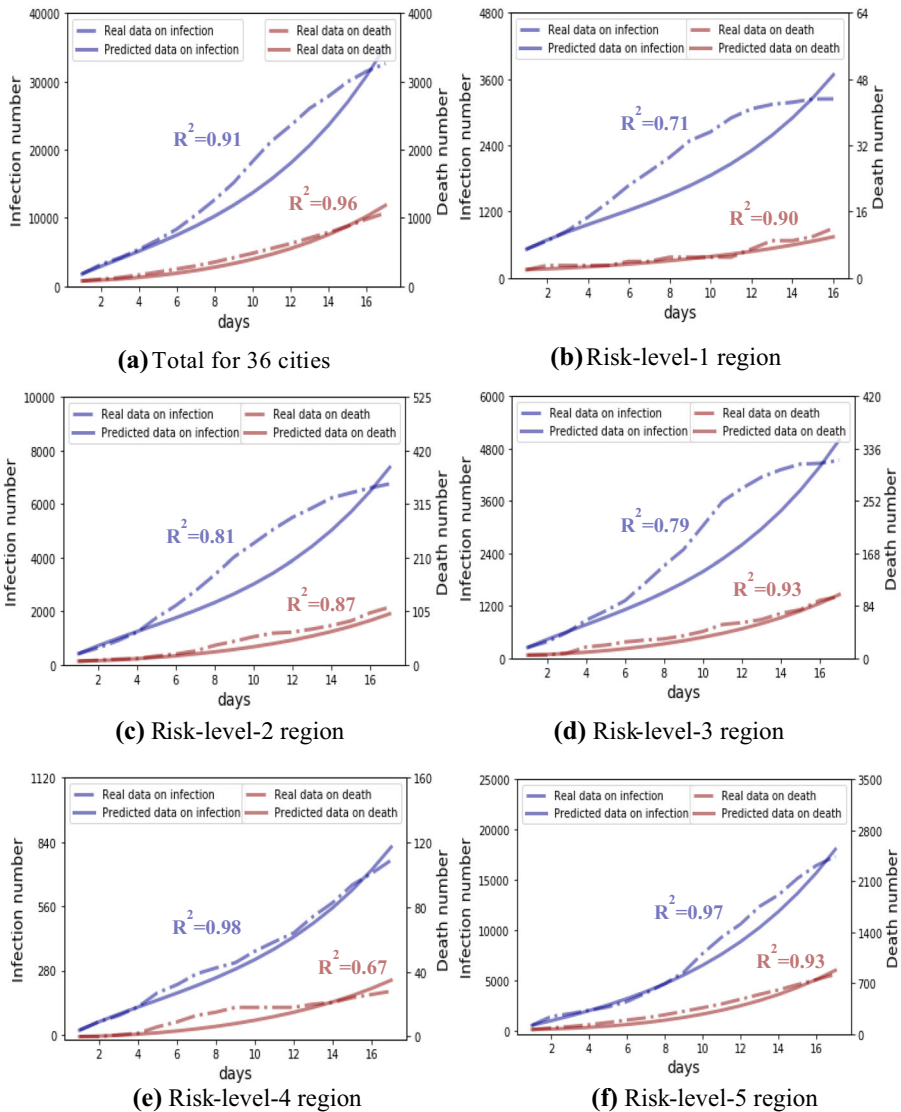


Fig. 5 Coefficient of determination (R^2) in estimating the transmission probability and death rate

coefficient of determination of the five clusters between the predicted and real infections was above 0.85, which implies that it is appropriate and effective to replace λ'_k with μ_k in the model.

4.2.3 Spatially-specific vaccine allocation

The vaccine allocation results are given in Fig. 6, which shows that the vaccines should be allocated first to the source region of the epidemic and then to the others in order of decreasing risk when the vaccine doses are limited. Our results can also help decision-makers determine the optimal quantity of vaccines to be used in each region. For example, with 40 million vaccine doses available, decision makers may first allocate p_1 , p_2 , and p_3 at 100% coverage and then p_4 at 50% coverage. Figure 7 shows the specific quantity of vaccines allocated to each region when the available vaccine doses are 20 million, 40 million, 60 million, and 80 million. We find that the allocation plans with the targets of death and infection minimization are the same.

Fig. 6 Proportion of each age group vaccinated to minimize the number of deaths or infections (The results are the same whether the objective is minimizing deaths or infections)

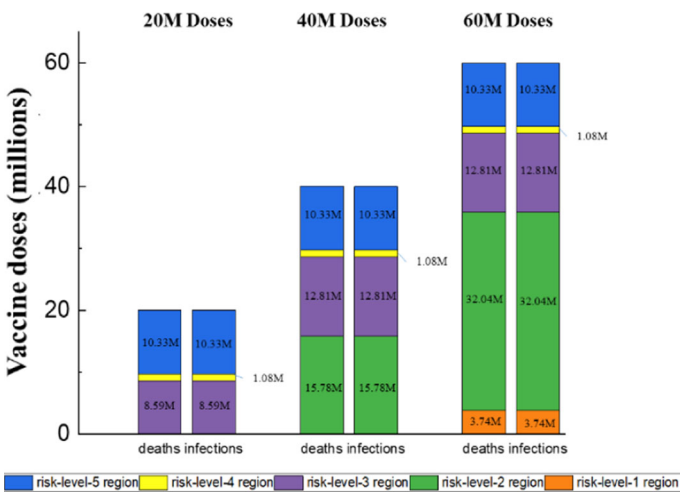
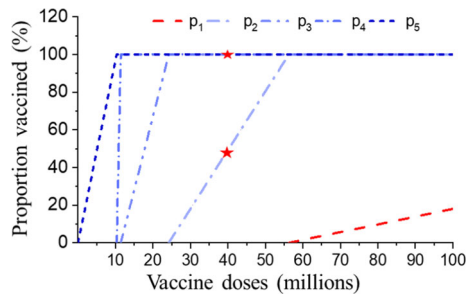


Fig. 7 Number of vaccine doses optimally allocated to each region to minimize the number of deaths or infections

4.3 Results of age-specific vaccine allocation

In this section, we further analyze an age-dependent model by substituting different age groups for the risk zones considered in our model. We divide the individuals into six age groups: 0–29, 30–39, 40–49, 50–59, 60–69, and 70+ years (Epidemiology Working Group for NCIP epidemic response, 2020). We use μ_a (the ratio of infections in each age group to the total infections per day) and d_a (the ratio of deaths in each age group to the total deaths per day) to calculate the number of infections and deaths per day. Appendix 2 presents the age-specific COVID-19 vaccine allocation model. The values of the parameters are obtained according to the computation results and references, which are presented in Table 2.

We allocate the vaccines first to those in the 60–69 age group, followed by those in the 70+, 50–59, 40–49, 30–39, and 0–29 age groups, with the objective of minimizing the total number of infections (Fig. 8). The elderly should receive the vaccination first, followed by the 60–69, 50–59, 40–49, 30–39, and 0–29 age groups. This is an intuitive way to minimize the total number of deaths because elderly people have weaker immune systems and die faster without vaccine intervention than other groups.

When the number of available vaccine doses are 20 million, 40 million, 60 million, and 80 million, we can obtain specific allocation policies for each age group (Fig. 9). When there are 20 million available vaccine doses, the government could use different allocation plans for two outcome measures. In contrast, when the vaccine doses are 40, 60, and 80 million, the allocation plans for the two results are the same.

Table 2 Epidemiological parameter values and computation results in age-specific SEIR model

Parameter	Value	Sources
Progress rate from subclinical state to recovery ζ	0.25	Hu et al. (2020)
Progress rate from pre-symptomatic state to symptomatic state τ	0.196	Lauer et al. (2020)
Recovery rate γ	0.1	Roda et al. (2020)
Effectiveness of the vaccine against infection ε	0.9	Callaway (2020)
$\varphi\beta_E$	0.21	$R^2 = 0.91$
$\varphi\beta_I$	0.23	$R^2 = 0.91$
Ratio of subclinical patients to pre-symptomatic patients $\frac{E_1}{E_2}$	0.41	Hu et al. (2020)
Ratio of the number of infections in each age group to the total number of infections per day μ_a , $a = 1, 2, 3, 4, 5, 6$	0.102, 0.170, 0.192, 0.224, 0.192, 0.120	–
Ratio of number of deaths in each age group to total number of infections per day d_a , $a = 1, 2, 3, 4, 5, 6$	0.0078 0.0176 0.0371 0.1271 0.3021 0.5083	–

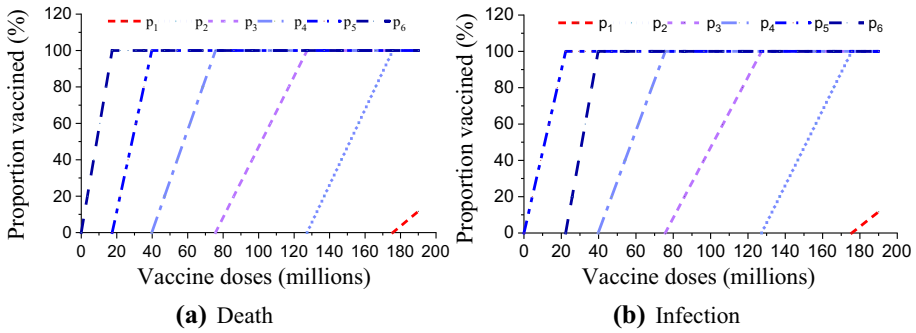


Fig. 8 Proportion of each age group vaccinated for minimization of the number of deaths or infections

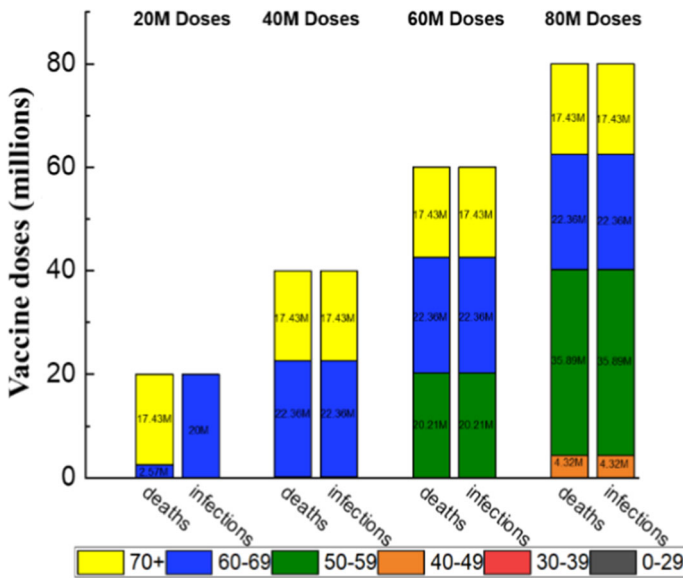


Fig. 9 Number of vaccines allocated to each age group to minimize the number of deaths or infections

4.4 Comparison and analysis

In this section, the proposed framework’s results are compared with those of an age-specific vaccine allocation model. Furthermore, the effects of both types of allocation policies (age-based and region-based) are analyzed over time, under different vaccine dose assumptions.

Figure 10 shows how both vaccine allocation policies reduce deaths and infections, comparing with the situation without intervention. If T is relatively short, the reduction of deaths is always more than the reduction of infections under these two policies (see Fig. 10). We also analyze the reduction in the number of deaths or infections for various levels of vaccine doses. We find that reduction in infections under the region-based policy for 20, 40, and 60 million vaccine doses increases evenly (see Fig. 11).

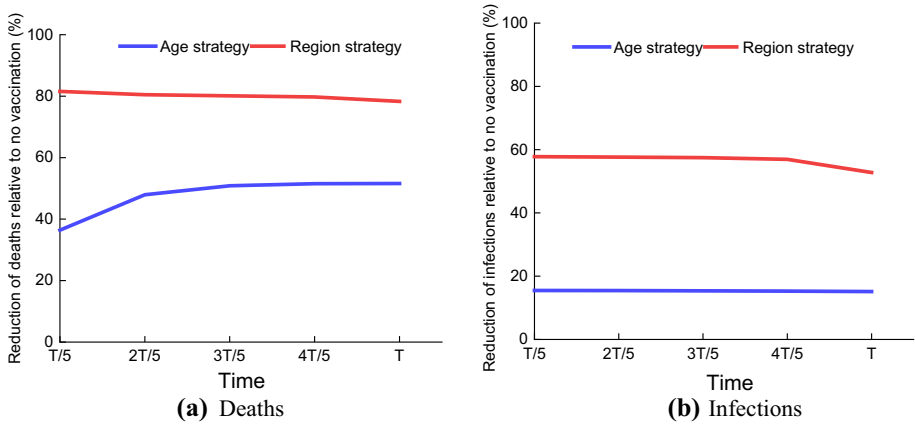


Fig. 10 Reduction in the numbers of deaths and infections over time compared with no vaccination

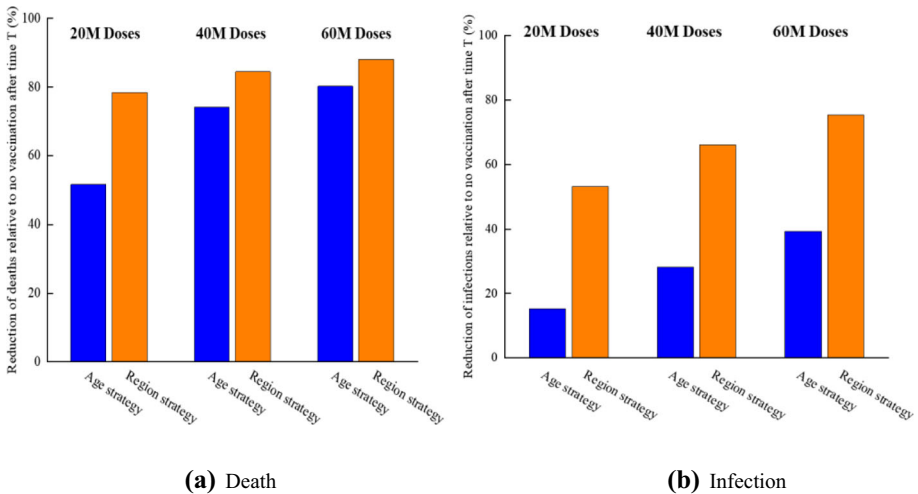


Fig. 11 Reduction in the numbers of deaths and infections under different vaccine doses compared with no vaccinations

Based on Figs. 10 and 11, a region-based policy reduces deaths and infections compared to age-based vaccine allocation. Specifically, when there are 20 million doses, our regional-priority allocation policy reduces the number of deaths and infections by more than 30 and 40%, respectively, compared to the age-priority policy. The former policy can prevent the spread of disease in low-risk cities and control the spread of disease in high-risk cities. Therefore, regional-priority vaccine allocation may be more appropriate when the objective is to minimize the number of deaths or infections.

5 Discussion and conclusion

Selecting vaccine allocation policies for infectious diseases is a critical administrative issue. Vaccine allocation policies influence the severity of the subsequent epidemic development, especially for malignant, infectious diseases such as COVID-19, with high infectivity, long incubation period, rapid spread, and a wide range of susceptible populations. We propose a data-driven and spatially-specific vaccine allocation framework that can allocate COVID-19 vaccine supplies to minimize the total number of infections or deaths. In particular, we provide a data-driven method for COVID-19 vaccine allocation by combining a regional risk classification model, spatially-specific disease progress model, and vaccine allocation model. Clustering cities based on a self-organizing map (SOM) neural network minimizes the computational burden of the allocation process and improves the efficiency of vaccine allocation. Training the SEIR model with real data improves the accuracy of the vaccine allocation policies.

We verified the applicability and effectiveness of the framework by simulating the vaccine allocation during the COVID-19 pandemic in China. The results show that vaccines should be allocated first to the source region of the outbreak and then to the other regions, in the order of decreasing risk. This method of priority allocation yields the best performance in limiting epidemic transmission and controlling the epidemic. In addition, our framework significantly outperforms the age-based vaccine allocation framework currently in use for COVID-19 vaccine allocation, and has been widely discussed in the literature. We developed region-specific allocation policies based on the spatial transmission trend of COVID-19 to minimize the ineffective use of vaccines. These findings support vaccine allocation for COVID-19 and other epidemics, and contribute to research on policies for spatial vaccine allocation.

It will be interesting to consider related topics in the future. First, the single-level heterogeneity model in the proposed framework could be extended to a multilevel heterogeneity model. For instance, a two-level heterogeneity model of age and regional grouping could be integrated into our data-driven vaccine allocation framework. Second, infectious diseases usually last for a long time and vaccines may be distributed in multiple phases. In this case, the optimal allocation policies for each region in each period must be studied to minimize long-term social losses. Third, the proposed vaccine allocation framework can be used to study other epidemics.

Acknowledgements This research was supported in part by National Natural Science Foundation of China (Grant No. 72171104, 71771108, 72001189), Zhejiang Social Science Foundation (Grant No. 23NDJC164YB), and Zhejiang Business Research Institute (Grant No. 22ZSKT02YB). Yeming GONG is supported by AIM Institute and BIC Center at emlyon.

Appendix 1: The spatially-specific COVID-19 vaccine allocation model of data

The spatially-specific COVID-19 vaccine allocation model used currently in practice is slightly different from the spatially-specific COVID-19 vaccine allocation model specified in our framework owing to the transformation of a feature. Specifically, the ratio of infections in a region to the total infections in all regions per day (μ_k), which includes the population contact information, replaces the force of infection (λ_k) as a characteristic that distinguishes each region. The spatially-specific COVID-19 vaccine allocation model of data is shown

below.

$$\begin{aligned} \min_{p_k} b * [S_{all}(0) - S_{all}(T)] + S_{all}(0) + E_{all}(0) + I_{all}(0) \\ + R_{all}(0) - S_{all}(T) - E_{all}(T) - I_{all}(T) - R_{all}(T), \\ \text{subject to } 0 \leq p_k \leq 1, \end{aligned}$$

$$\sum_{k=1}^5 p_k S_k(0) \leq W,$$

$$S_{all}(t) = \sum_{k=1}^5 S_k^U(t) + \sum_{k=1}^5 S_k^V(t),$$

$$E_{all}(t) = \sum_{k=1}^5 E_k^U(t) + \sum_{k=1}^5 E_k^V(t),$$

$$I_{all}(t) = \sum_{k=1}^5 I_k^U(t) + \sum_{k=1}^5 I_k^V(t),$$

$$R_{all}(t) = \sum_{k=1}^5 R_k^U(t) + \sum_{k=1}^5 R_k^V(t),$$

$$\frac{dS_k^U}{dt} = -\frac{\lambda S \mu_k}{S_k} S_k^U,$$

$$\frac{dE_k^U}{dt} = \frac{\lambda S \mu_k}{S_k} S_k^U - \tau E_{2k}^U - \varsigma E_{1k}^U,$$

$$\frac{dI_k^U}{dt} = \tau E_{2k}^U - v_k I_k^U - \gamma I_k^U,$$

$$\frac{dR_k^U}{dt} = \gamma I_k^U + \varsigma E_{1k}^U,$$

$$\frac{dD_k^U}{dt} = v_k I_k^U$$

$$\frac{dS_k^V}{dt} = -\frac{\lambda S \mu_k}{S_k} (1 - \varepsilon) S_k^V,$$

$$\frac{dE_k^V}{dt} = \frac{\lambda S \mu_k}{S_k} (1 - \varepsilon) S_k^V - \tau E_{2k}^V - \varsigma E_{1k}^V,$$

$$\frac{dI_k^V}{dt} = \tau E_{2k}^V - v_k I_k^V - \gamma I_k^V,$$

$$\frac{dR_k^V}{dt} = \gamma I_k^V + \varsigma E_{1k}^V, \text{ and}$$

$$\frac{dD_k^V}{dt} = v_k I_k^V.$$

Appendix 2: The age-specific COVID-19 vaccine allocation model

Our age-specific COVID-19 vaccine allocation model resembles our spatially-specific COVID-19 vaccine allocation model. The ratio of infections in each age group to the total infections per day μ_a and the ratio of deaths in each age group to the total deaths per day d_a are used to calculate the numbers of infections and deaths per day, respectively. The age-specific COVID-19 vaccine allocation model is shown below.

$$\begin{aligned} \min_{p_a} b * [S'_{all}(0) - S'_{all}(T)] + S'_{all}(0) + E'_{all}(0) + I'_{all}(0) \\ + R'_{all}(0) - S'_{all}(T) - E'_{all}(T) - I'_{all}(T) - R'_{all}(T), \\ \text{subject to } 0 \leq p_a \leq 1, \end{aligned}$$

$$\begin{aligned} \sum_{a=1}^6 p_a S_a(0) &\leq W, \\ S'_{all}(t) &= \sum_{a=1}^6 S_a^U(t) + \sum_{a=1}^6 S_a^V(t), \\ E'_{all}(t) &= \sum_{a=1}^6 E_a^U(t) + \sum_{a=1}^6 E_a^V(t), \\ I'_{all}(t) &= \sum_{a=1}^6 I_a^U(t) + \sum_{a=1}^6 I_a^V(t), \\ R'_{all}(t) &= \sum_{a=1}^6 R_a^U(t) + \sum_{a=1}^6 R_a^V(t), \\ \frac{dS_a^U}{dt} &= -\frac{\lambda S \mu_a}{S_a} S_a^U, \\ \frac{dE_a^U}{dt} &= \frac{\lambda S \mu_a}{S_a} S_a^U - \tau E_{2a}^U - \varsigma E_{1a}^U, \\ \frac{dI_a^U}{dt} &= \tau E_{2a}^U - \frac{v I d_a}{I_a} I_a^U - \gamma I_a^U, \\ \frac{dR_a^U}{dt} &= \gamma I_a^U + \varsigma E_{1a}^U, \\ \frac{dD_a^U}{dt} &= \frac{v I d_a}{I_a} I_a^U, \\ \frac{dS_a^V}{dt} &= -\frac{\lambda S \mu_a}{S_a} (1 - \varepsilon) S_a^V, \\ \frac{dE_a^V}{dt} &= \frac{\lambda S \mu_a}{S_a} (1 - \varepsilon) S_a^V - \tau E_{2a}^V - \varsigma E_{1a}^V, \\ \frac{dI_a^V}{dt} &= \tau E_{2a}^V - \frac{v I d_a}{I_a} I_a^V - \gamma I_a^V, \\ \frac{dR_a^V}{dt} &= \gamma I_a^V + \varsigma E_{1a}^V, \text{ and} \\ \frac{dD_a^V}{dt} &= \frac{v I d_a}{I_a} I_a^V. \end{aligned}$$

Appendix 3: The classification results obtained by the k -means algorithm

The number of clusters is the key parameter of the k -means algorithm, and the sum of squared errors (SSE) and silhouette coefficient techniques are used to determine it. The optimal number of clusters obtained by two methods is the same in the case study, which is 4. So 35 cities (the sample excluding Wuhan) should be divided into four clusters (Fig. 12). Next, we can get the optimal clustering results (Fig. 13) by minimizing the sum of the squared error densities of the four clusters.

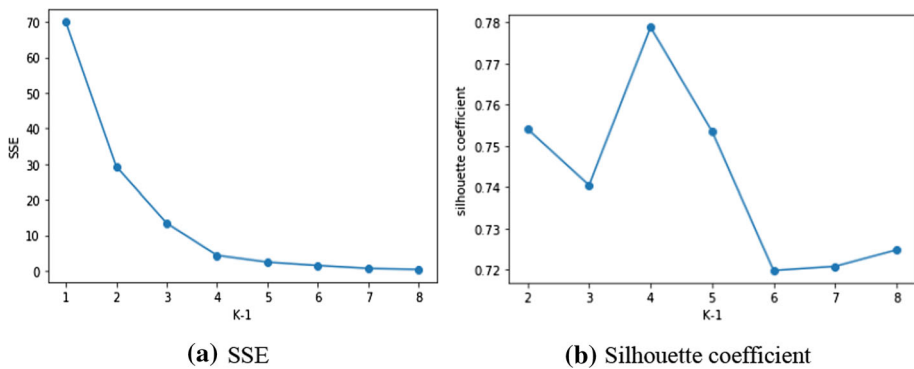


Fig. 12 $K-1$ with respect to SSE (the value of $K-1$ corresponding to the elbow is appropriate) and silhouette coefficient (the value of $K-1$ corresponding to a larger coefficient is appropriate)

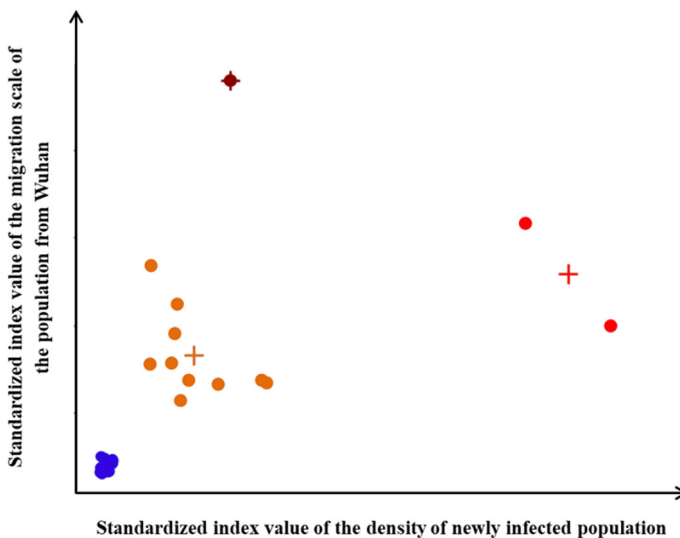


Fig. 13 Four clusters for 35 cities of the sample (excluding Wuhan)

References

- Allen, L. J. S. (2017). A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2, 128–142.
- Aràndiga, F., Baeza, A., Cordero-Carrión, I., Donat, R., Martí, M. C., Mulet, P., & Yáñez, D. F. (2020). A spatial-temporal model for the evolution of the COVID-19 pandemic in Spain including mobility. *Mathematics*, 8, 1677.
- Araz, O. M., Galvani, A., & Meyers, L. A. (2012). Geographic prioritization of distributing pandemic influenza vaccines. *Health Care Management Science*, 15, 175–187.
- Bubar, K. M., Reinholt, K., Kissler, S. M., Lipsitch, M., Cobey, S., Grad, Y. H., & Larremore, D. B. (2021). Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science*, 371(6532), 916–921.
- Callaway, E. (2020). What Pfizer's landmark COVID vaccine results mean for the pandemic. *Nature*. <https://doi.org/10.1038/d41586-020-03166-8>
- Dai, X., & Kuosmanen, T. (2014). Best-practice benchmarking using clustering methods: Application to energy regulation. *Omega*, 42(1), 179–188.
- Emanuel, E. J., Persad, G., Kern, A., Buchanan, A., Fabre, C., Halliday, D., Heath, J., Herzog, L., Leland, R. J., Lemango, E. T., & Luna, F. (2020). An ethical framework for global vaccine allocation. *Science*, 369(6509), 1309–1312.
- Epidemiology Working Group for NCIP Epidemic Response. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chinese Journal of Epidemiology*, 41, 145–151.
- Foy, B. H., Wahl, B., Mehta, K., Shet, A., Menon, G. I., & Britto, C. (2021). Comparing COVID-19 vaccine allocation strategies in India: A mathematical modelling study. *International Journal of Infectious Diseases*, 103, 431–438.
- He, S., Peng, Y., & Sun, K. (2020). SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics*, 101(3), 1667–1680.
- Hou, C., Chen, J., Zhou, Y., Hua, L., Yuan, J., He, S., Guo, Y., Zhang, S., Jia, Q., Zhao, C., Zhang, J., Xu, G., & Jia, E. (2020). The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis. *Journal of Medical Virology*, 92, 841–848.
- Hu, Z., Song, C., Xu, C., Jin, G., & Shen, H. (2020). Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China Life Sciences*, 63(5), 706–711.
- Huang, R., Liu, M., & Ding, Y. (2020). Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. *The Journal of Infection in Developing Countries*, 14, 246–253.
- Ivanov, S. V., & Leonenko, V. N. (2018). Prediction of influenza peaks in Russian cities: Comparing the accuracy of two SEIR models. *Mathematical Biosciences & Engineering*, 15(1), 209–232.
- Jeyanathan, M., Afkhami, S., Smaill, F., Miller, M. S., Lichty, B. D., & Xing, Z. (2020). Immunological considerations for COVID-19 vaccine strategies. *Nature Reviews. Immunology*, 20, 615–632.
- Kozio, K., Stanisawski, R., & Bialic, G. (2020). Fractional-order sir epidemic model for transmission prediction of covid-19 disease. *Applied Sciences*, 10(23), 8316.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172, 577–582.
- Libotte, G. B., Lobato, F. S., Platt, G. M., & Neto, A. J. S. (2020). Determination of an optimal control strategy for vaccine administration in COVID-19 pandemic treatment. *Computer Methods and Programs in Biomedicine*, 196, 105664.
- Matrajt, L., Eaton, J., Leung, T., & Brown, E. R. (2020). Vaccine optimization for COVID-19, who to vaccinate first?. *medRxiv*.
- Medlock, J., & Galvani, A. P. (2009). Optimizing influenza vaccine distribution. *Science*, 325, 1705–1708.
- Min, H., Ko, H. J., & Ko, C. S. (2006). A genetic algorithm approach to developing the multi-echelon reverse logistics network for product returns. *Omega*, 34(1), 56–69.
- Mwalili, S., Kimathi, M. E. M., Ojiambo, V. N., Gathungu, D. K., & Achia, T. N. O. (2020). Age-structured impact of mitigation strategies on COVID-19 severity and deaths in Kenya.
- Mylus, S. D., Hagenaars, T. J., Lugner, A. K., & Wallinga, J. (2008). Optimal allocation of pandemic influenza vaccine depends on age, risk and timing. *Vaccine*, 26, 3742–3749.
- Ndeffo Mbah, M. L., Durham, D. P., Medlock, J., & Galvani, A. P. (2014). Country- and age-specific optimal allocation of dengue vaccines. *Journal of Theoretical Biology*, 342, 15–22.

- Ogut, H., & Bishop, S. C. (2010). A stochastic modelling approach to describing the dynamics of an experimental furunculosis epidemic in Chinook salmon, *Oncorhynchus tshawytscha* (Walbaum). *Journal of Fish Diseases*, 30(2), 93–100.
- Okuonghae, D., & Omame, A. (2020). Analysis of a mathematical model for COVID-19 population dynamics in Lagos, Nigeria. *Chaos Solitons & Fractals*, 139, 110032.
- Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. arXiv preprint [arXiv:2004.00958](https://arxiv.org/abs/2004.00958).
- Patel, R., Longini, I. M., & Halloran, M. E. (2005). Finding optimal vaccination strategies for pandemic influenza using genetic algorithms. *Journal of Theoretical Biology*, 234(2), 201–212.
- Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, 5, 271–281.
- Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity*, 109, 102433.
- Sah, P., Medlock, J., Fitzpatrick, M. C., Singer, B. H., & Galvani, A. P. (2018). Optimizing the impact of low-efficacy influenza vaccines. *Proceedings of the National Academy of Sciences*, 115(20), 5151–5156.
- Subbaraman, N. (2020). Who gets a COVID vaccine first? Access plans are taking shape. *Nature*, 585(7826), 492–493.
- Teles, P. (2020). A time-dependent seir model to analyse the evolution of the sars-covid-2 epidemic outbreak in Portugal. arXiv preprint [arXiv:2004.04735](https://arxiv.org/abs/2004.04735).
- Teytelman, A., & Larson, R. C. (2013). Multiregional dynamic vaccine allocation during an influenza epidemic. *Service Science*, 5, 197–215.
- Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3), 586–586.
- Xiang, Y., & Wang, S. (2020). Spatial relationship between epidemic spread and population outflow of the Corona Virus Disease 2019 (COVID-19) that impacted Chinese urban public health classification. *Tropical Geography*, 40, 408–421.
- Yang, Z., Zeng, Z., Wang, K., Wong, S. S., & He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of covid-19 in china under public health interventions. *Journal of Thoracic Disease*, 12(3), 165–174.
- Zhou, S., Zhou, S., Zheng, Z., & Lu, J. (2021). Optimizing spatial allocation of covid-19 vaccine by agent-based spatiotemporal simulations. *GeoHealth*, 5(6), 1–16.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.