RESEARCH ARTICLE

SQEAIR: an Improved Infectious Disease Dynamics Model

Chenxi Wang[1] · Yongchao Jin[1,2] · Lihui Zhou[1] · Wei Hou[1] · Dongmei Liu[1] ·
Jianjun Wang[3] · Xiyin Wang[1,2,3]

Abstract

The Covid-19 pandemic poses a significant threat to human health and life. Timely and accurate prediction of the epidemic's trajectory is crucial for devising effective prevention and control strategies. Traditional infectious disease models may not capture the complexity of modern epidemics, especially when governments implement diverse policies. Drawing from China's epidemic prevention strategies and Covid-19 transmission characteristics, this study introduces two distinct categories quarantined cases and asymptomatic cases to enhance the traditional SEIR model in depicting disease dynamics. To address the intricate nature of prevention and control efforts, the quarantined cases are further segmented into three subgroups: exposed quarantined, asymptomatic quarantined, and infected quarantined cases. Consequently, a novel SQEAIR model is proposed to model the dynamics of Covid-19. Evaluation metrics such as the Akaike information criterion (AIC) and Absolute Percentage Error (MAE) are employed to assess the efficacy and accuracy of both the newly proposed and traditional models. By fitting the models to the number of infected cases in Shanghai (March to May 2022) and Guangzhou (November 2022), it was observed that the SQEAIR model exhibited a lower AIC value compared to the SEIR model, indicating superior fitting accuracy for Covid-19 infections. Moreover, the high accuracy of the SQEAIR model enabled precise predictions of confirmed cases in Guangzhou. Leveraging the SQEAIR model, various parameters were tested to simulate the impact of different influencing factors, enabling the evaluation of defense strategies. These findings underscore the effectiveness of key epidemic control measures, such as quarantining exposed cases, in enhancing public health and promoting awareness of personal protection.

Keywords Infectious disease dynamics model · SEIR model · SQEAIR model · Epidemic prediction · Influencing factors

Yongchao Jin, Lihui Zhou, Wei Hou, Dongmei Liu, Jianjun Wang and Xiyin Wang these authors contributed equally to this work

✉ Xiyin Wang
   wangxiyin@vip.sina.com

Extended author information available on the last page of the article

 Springer

1 Introduction

The novel Coronavirus is a highly contagious viral strain found in late 2019, and the pneumonia caused by this virus was named novel Coronavirus disease 2019 (Covid-19) [1]. Typical symptoms of cases infected by Covid-19 include fever, fatigue, dry cough, etc. [2], and severe cases may result in respiratory distress syndrome or even death [3]. The main transmission routes of the Covid-19 are droplet transmission, contact transmission, and aerosol transmission [4]. The virus is characterized by rapid transmission, wide infection range, random mutation in its RNA, and other characteristics, and the prevention and control of its transmission is extremely difficult. The epidemic has been lingering around the world for more than three years, and fast-transmitting variants have been still affecting different regions. These showed that the Covid-19, no matter what kinds of new virus lines evolve or prevail, will continue to pose a critical threat to human life and health in near future.

Currently, conducting quantitative research and theoretical analysis on the spread of Covid-19 is essential for predicting its trends and implementing appropriate control measures. When it comes to epidemic forecasting, mathematical models play a crucial role in simulating the transmission dynamics of the outbreak. A well-designed mathematical model not only allows for the analysis of the transmission patterns of the epidemic and prediction of its trajectory but also offers valuable insights for epidemic prevention and control efforts. By leveraging mathematical models, informed recommendations can be made for future prevention and control strategies.

The traditional SEIR model considers the impact of factors such as contact rate, incidence rate, and latent cases on the epidemic. Hethcote discussed involving the basic reproduction number R_0, the contact number σ, and the replacement number R are reviewed for the classic SIR epidemic and endemic models [5]. Rock et al. focused on three critical aspects : heterogeneously structured populations, stochasticity, and spatial structure [6]. In the case of the Covid-19, different nations have implemented prevention and control measures, which to certain extent have contributed to hindering the Covid-19 transmission. However, these prevention and control measures have not been fully considered by traditional infectious disease models. It is not weird that using the traditional SEIR model could result in a large deviation between the prediction results and the number of infected cases.

Actually, many models were developed to describe the transmission of the Covid-19. Kuhl brought together modern concepts in mathematical epidemiology, computational modeling, physics-based simulation, data science, and machine learning to understand the outbreak dynamics and outbreak control of COVID-19 [7]. Frank [8] analyzed COVID-19 outbreaks in various countries around the globe, made use of mathematical models of epidemiology such as the SIR and SEIR models, and discussed the impacts of measures implemented to stop the spread of COVID-19 disease. Gatto [9] accounted for uncertainty in epidemiological reporting, and time dependence of human mobility matrices and awareness-dependent exposure probabilities and draw scenarios of different containment measures and their impact. Ngonghala et al. [10] developed a new mathematical model to assess the population-level impact of control and mitigation strategies, and emphasized the important role social-distancing plays

in curtailing the burden of COVID-19. Yang Z et al. [11] proposed a M-SEIR model based on the improved SEIR model. Based on the M-SEIR model, Wang J et al. [12] proposed a CSPE model with the community as the basic forecasting unit. Martcheva [13] proposed quarantine and isolation were typically modeled by introducing separate classes into the model. Wei Y et al. [14] established a SEIR+CAQ transmission dynamics model, considering the transmission mechanism, infection spectrum, and isolation measures. According to the actual situation in China, Zhao X et al. [15] added the transmission process of latent tracking admission and morbidity tracking admission on the basis of the traditional SEIR model to depict the development trend of the epidemic and created a CSEIR model. By assuming that infected but undetected individuals are sent for quarantine during the incubation period, Pal. D [16] proposed a SEQIR model to assess and manage the outbreak of the infectious disease Covid-19. Daniele P et al. [17] developed a SPQEIR model to link intervention categories against epidemic spread to epidemiological model compartments. Giulia G et al. [18] developed SIDARTHE model discriminates between infected individuals depending on whether they have been diagnosed and on the severity of their symptoms, this delineation also helps to explain misperceptions of the case fatality rate and of the epidemic spread. Luca G et al. [19] introduced a framework to quantify how the uncertainty in the data affects the determination of the parameters and the evolution of the unmeasured variables of a given model. Francoise K et al. [20] extended the SEIR model, and considered how the interplay between vaccinations and social measures could shape infections and hospitalizations.

While mathematical models serve as valuable tools for understanding epidemics and guiding response strategies, they are not without their limitations. Firstly, the spread of Covid-19 is influenced by a multitude of factors such as policy adjustments, levels of isolation, quality of medical care, and adherence to personal protective measures. These factors can introduce significant deviations in the fitting and prediction of confirmed cases. Secondly, the parameters within these models play a critical role in determining prediction outcomes. Setting default parameters based on existing knowledge and subjective judgment can lead to discrepancies in results, especially in scenarios with limited data.Furthermore, many models overlook the classification and analysis of different types of quarantined cases, which may not align with effective epidemic control policies and the unique transmission characteristics of Covid-19. This oversight can result in suboptimal prediction accuracy and other challenges.

To address these limitations, we developed a novel SQEAIR model that considers both quarantine policies and the presence of asymptomatic cases. This model categorizes quarantined individuals into exposed, asymptomatic, and infected groups. By incorporating these compartments and simulating contact tracing, isolation protocols, and the release of asymptomatic cases based on medical observation, the SQEAIR model offers a more comprehensive approach to epidemic prevention and control. The model also accounts for isolating exposed and asymptomatic cases, leading to improved predictive accuracy compared to the traditional SEIR model.Through parameter optimization and fitting the model to the number of infected cases, we demonstrated the superior performance of the SQEAIR model. Finally, by testing various parameterizations, we evaluated the effectiveness of epidemic control measures

implemented by governments, health authorities, and individuals in curbing the spread of Covid-19.

2 Model Introduction

The most widely used infectious disease dynamics modeling methods are grouped as compartment model, and a prominent one, namely the SEIR model. It was established by Kermack WO and McKendrick AG [21] in 1927 when studying the Black Death in London and the plague in Mumbai, which was developed into a specialized theory: infectious disease dynamics. These models describe the transmission process of infectious diseases, analyze the change rule of the number of infected individuals, and reveal the development pattern of infectious diseases through the quantitative relationship according to the general infectious disease transmission mechanism [22].

2.1 SEIR Model

In the SEIR model, the following assumptions are made:

1. The total population of the region remained unchanged during the disease observation period;
2. Each individual is equally likely to be infected, and only person-to-person transmission is considered;
3. After treatment and recovery, cases will not be transformed into susceptible cases in a short time.

SEIR model divides the total population into the following categories:

Susceptible (S): those who are not infected but have a chance of contracting the disease;

Exposed (E): refers to individuals who are in the incubation period and will become infected after the incubation period;

Infected (I): refers to individuals who are infectious and are showing symptoms;

Removed (R): refers to individuals who have either recovered or died among the infected individuals;

The model flow chart is as follows:

β is the effective contact rate of cases per unit time, ω is the probability of converting a exposed person into an infected person, and γ is the probability of recovery or death of cases (Fig. 1).
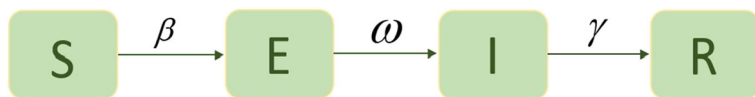

Fig. 1 Flow chart of SEIR infectious disease dynamics model

Based on the SEIR infectious disease dynamics model and the above assumptions, differential equations can be constructed:

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta SI}{N} \\ \dfrac{dE}{dt} = \dfrac{\beta SI}{N} - \omega E \\ \dfrac{dI}{dt} = \omega E - \gamma I \\ \dfrac{dR}{dt} = \gamma I \end{cases} \tag{1}$$

2.2 Improved SQEAIR Model

On the basis of the traditional SEIR model, we propose the SQEAIR infectious disease dynamics model by considering the characteristics of the transmission of the novel coronavirus pneumonia and China's epidemic prevention policies.

In SQEAIR model, the following assumptions are added:

1. The quarantined cases include some of the close contacts of exposed cases, asymptomatic cases and infected cases, and the quarantined cases will not infect others;
2. Latent cases are tested to have been affected (confirmed) or asymptomatic cases after the incubation period;
3. Latent, asymptomatic, and confirmed cases are all infectious, and the infection rate is the same;
4. Confirmed cases and asymptomatic cases were treated and eventually cured or died.

SQEAIR model is proposed on the basis of the traditional SEIR model, and the following four compartments are added for research:

Asymptomatic cases (A): cases who are infected but not showing symptoms;

Quarantined exposed cases (Q_E): cases who are exposed but not capable of infecting susceptible cases;

Quarantined asymptomatic cases (Q_A): cases who are asymptomatic but not capable of infecting susceptible cases;

Quarantined infected cases (Q_I): cases who are infected and not capable of infecting susceptible cases;

The model flow chart is as follows:

The effective contact rate per unit of time of a patient is β. When a susceptible individual is infected, they initially enter the incubation period or a latent state, with the rate of conversion from a latent individual to an infected individual denoted as ω. Infected individuals encompass both asymptomatic cases and diagnosed patients. The proportion of infected individuals who exhibit symptoms and progress to confirmed patients is p, while the proportion of asymptomatic infected individuals who remain symptom-free is $1 - p$. The rate at which an asymptomatic infected individual transitions to a confirmed patient is δ.

Individuals have a likelihood of entering the isolation category during the latent, asymptomatic infected, and confirmed patient phases, with the respective proportions denoted as q, λ, μ.

Individuals admitted to the isolation category remain there until they recover or pass away, becoming emigrants. γ_1 represents the rate that the confirmed but not quarantined patients recover or pass away, γ_2 represents the rate that asymptomatic infected but not quarantined ones recover or pass away, γ_3 represents the rate that the quarantined and treated ones recover or pass away, and γ_4 represents the rate that the quarantined asymptomatic infected ones recover or pass away (Fig. 2).

Based on the SQEAIR model and the above assumptions, differential equations can be constructed:

$$
\begin{cases}
\dfrac{dS}{dt} = -\dfrac{\beta S(E + A + I)}{N} \\[2mm]
\dfrac{dE}{dt} = \dfrac{\beta S(E + A + I)}{N} - [(1 - q)\omega + q]E \\[2mm]
\dfrac{dQ_E}{dt} = qE - \omega Q_E \\[2mm]
\dfrac{dA}{dt} = (1 - q)(1 - p)\omega E - (\lambda + \delta + \gamma_2)A \\[2mm]
\dfrac{dQ_A}{dt} = (1 - p)\omega Q_E + \lambda A - (\delta + \gamma_4)Q_A \\[2mm]
\dfrac{dI}{dt} = (1 - q)p\omega E + \delta A - (\mu + \gamma_1)I \\[2mm]
\dfrac{dQ_I}{dt} = p\omega Q_E + \delta Q_A + \mu I - \gamma_3 Q_I \\[2mm]
\dfrac{dR}{dt} = \gamma_1 I + \gamma_2 A + \gamma_3 Q_I + \gamma_4 Q_A
\end{cases}
\tag{2}
$$

Since the effective contact rate of patients per unit time β will continue to decrease with the increase and decrease of prevention and control measures and monitoring
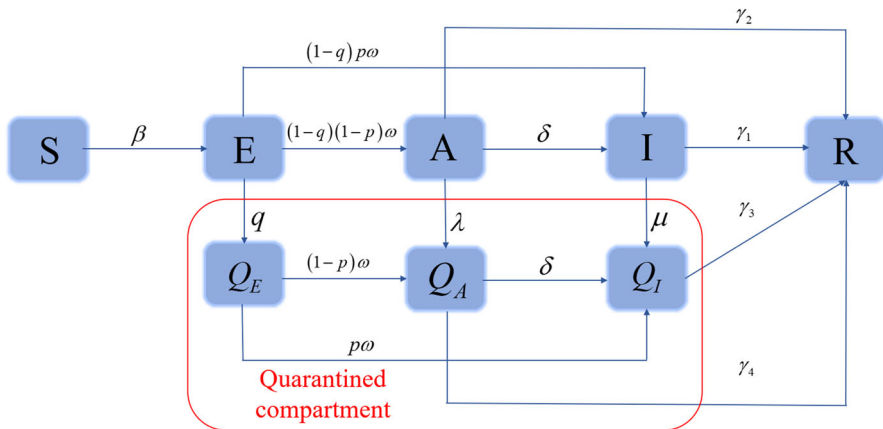

Fig. 2 Flow chart of dynamic model of SQEIAR infectious disease

efforts, this paper describes the change of effective contact rate of patients per unit time β over time by introducing inhibitory factors:

$$\beta(t) = \beta_0 e^{-\alpha} \tag{3}$$

Where α denotes the exponential decreasing rate of effective contact rate β, which is used to indicate the inhibiting effect of epidemic prevention work on the spread of the epidemic, and β_0 denotes the initial value of the effective contact rate, and the effective contact rate on the first day of the experimental data is used as the initial value of the effective contact rate, which can be introduced according to the first differential equation above:

$$\beta_0 = \frac{(S_0 - S_1)N}{S_0(E_0 + A_0 + I_0)} \tag{4}$$

Therefore, both β in the system of Eq. 2 are: $\frac{(S_0-S_1)N}{S_0(E_0+A_0+I_0)}$

2.3 Model Solving

Runge-Kutta Method is an important iterative method for solving nonlinear ordinary differential equations, it can be used to solve the SEIR model and SQEAIR model. This method is mainly used when the derivative and initial value information of the equation are known and computer simulation is used, which can save the complicated process of solving the differential equation. In the following, the SEIR model is taken as an example to introduce the solution process of Runge-Kutta Method.

First, the time interval $[0, t]$ is discretized into n step, each of which is $h = \frac{t}{n}$.

Second, initialize initial conditions: S_0, E_0, I_0, R_0 (the meaning of the arguments are shown in Table 1).

Third, for each time step, calculate the k_1 of the current time step:

$$\begin{cases} k_{1S} = -\dfrac{\beta S_n I_N}{N} \\ k_{1E} = \dfrac{\beta S_n I_N}{N} - \omega E_n \\ k_{1I} = \omega E_n - \gamma I_n \\ k_{1R} = \gamma I_n \end{cases} \tag{5}$$

Table 1 The situation in Shanghai, Mar. 28, 2022

Arguments	Meaning	Initial value	Source
N	Total population of Shanghai	24894000	Official data
E_0	Exposed at the beginning	28937	Estimated data
A_0	Asymptomatic at the beginning	18531	Official data
I_0	Infected at the beginning	363	Official data
R_0	Removed at the beginning	535	Official data

calculate the k_2 of the current time step:

$$
\begin{cases}
k_{2S} = -\dfrac{\beta(S_n + \frac{h}{2}k_{1S})(I_N + \frac{h}{2}k_{1I})}{N} \\[2mm]
k_{2E} = \dfrac{\beta(S_n + \frac{h}{2}k_{1S})(I_N + \frac{h}{2}k_{1I})}{N} - \omega(E_n + \dfrac{h}{2}k_{1E}) \\[2mm]
k_{2I} = \omega(E_n + \dfrac{h}{2}k_{1E}) - \gamma(I_n + \dfrac{h}{2}k_{1I}) \\[2mm]
k_{2R} = \gamma(I_n + \dfrac{h}{2}k_{1I})
\end{cases}
\tag{6}
$$

calculate the k_3 of the current time step:

$$
\begin{cases}
k_{3S} = -\dfrac{\beta(S_n + \frac{h}{2}k_{2S})(I_n + \frac{h}{2}k_{2I})}{N} \\[2mm]
k_{3E} = \dfrac{\beta(S_n + \frac{h}{2}k_{2S})(I_n + \frac{h}{2}k_{2I})}{N} - \omega(E_n + \dfrac{h}{2}k_{2E}) \\[2mm]
k_{3I} = \omega(E_n + \dfrac{h}{2}k_{2E}) - \gamma(I_n + \dfrac{h}{2}k_{2I}) \\[2mm]
k_{3R} = \gamma(I_n + \dfrac{h}{2}k_{2I})
\end{cases}
\tag{7}
$$

calculate the k_4 of the current time step:

$$
\begin{cases}
k_{4S} = -\dfrac{\beta(S_n + hk_{3S})(I_n + hk_{3I})}{N} \\[2mm]
k_{4E} = \dfrac{\beta(S_n + hk_{3S})(I_n + hk_{3I})}{N} - \omega(E_n + hk_{3E}) \\[2mm]
k_{4I} = \omega(E_n + hk_{3E}) - \gamma(I_n + hk_{3I}) \\[2mm]
k_{4R} = \gamma(I_n + hk_{3I})
\end{cases}
\tag{8}
$$

Thus, the next value is determined by the product of the present values, the time interval (h), and a slope estimated by the weighted average of the slopes of k_1, k_2, k_3, and k_4, where k_1 is the slope at the beginning of the time period; k_2 is the slope of the midpoint of the time period, and the slope k_1 is used to determine the value of y. k_3 is also the slope of the midpoint, and the slope k_2 is used to determine y. k_4 is the slope of the end point of the time period, whose y is determined by k_3. When the four

slopes are averaged, the slope of the midpoint has a greater weight. The Runge-Kutta method for SEIR model is given by the following equation:

$$
\begin{cases}
S_{n+1} = S_n + \dfrac{h}{6}(k_{1S} + 2k_{2S} + 2k_{3S} + k_{4S}) \\[2mm]
E_{n+1} = E_n + \dfrac{h}{6}(k_{1E} + 2k_{2E} + 2k_{3E} + k_{4E}) \\[2mm]
I_{n+1} = I_n + \dfrac{h}{6}(k_{1I} + 2k_{2I} + 2k_{3I} + k_{4I}) \\[2mm]
R_{n+1} = R_n + \dfrac{h}{6}(k_{1R} + 2k_{2R} + 2k_{3R} + k_{4R})
\end{cases}
\tag{9}
$$

Through several iterations, the approximate values of each time step S, E, I, R, can be obtained to solve the SEIR model. Using the same method, arguments in the SQEAIR model can also be solved.

In this paper, the 4-order Runge-Kutta Method is used to solve the differential equation. The 5-order method controls the error, and the overall truncation error is $\circ(h^5)$.

2.4 Basic Reproduction Number

Basic reproduction number, usually denoted by R_0, it is an important indicator in epidemiology of the ability of an infectious disease to spread in a given population. The basic regeneration number indicates the average number of susceptible individuals in a population who develop the disease per infected individual during the infectious period. When $R_0 > 1$, it indicates that an infected person transmits the virus to more than one newly infected individual during their illness. Consequently, the number of people afflicted with the disease increases, and the disease persists. Conversely, when $R_0 < 1$, the infected person transmits the virus to fewer than one newly infected individual throughout the course of the disease. As a result, the number of people suffering from the disease decreases, and the disease is gradually eradicated. When $R_0 = 1$, the number of people suffering from the disease will be constant.

Van den Driessche P [23] gave an exact procedure for the computation of the fundamental regeneration number R_0 based on the propagation of a system of ordinary differential equations for hamster models, which has been widely used, as follows:

Let $x = (x_1, x_2, \cdots, x_n)^T$, where $x_i \geq 0 (i = 1, 2, \cdots, n)$, denotes the number of individuals in the i, compartment. Let $F_i(x)$ be the probability of a new infection in the i compartment, $V_i^+(x)$ be the rate of movement of individuals into the i compartment, and $V_i^-(x)$ be the rate of movement of individuals out of the i compartment. Assuming that each function is continuously differentiable at least twice in each variable, the disease transmission model consists of non-negative initial conditions and the following equations:

$$
\frac{dS}{dt} = F_i(x) - V_i(x), i = 1, 2, \cdots, n
\tag{10}
$$

where $V_i(x) = V_i^+(x) - V_i^-(x)$? $V_i^-(x)$ denotes the rate of individuals moving out of the i compartment, $V_i^+(x)$ denotes the rate of individuals moving into the i compartment, and $F_i(x)$ denotes the rate of new infections.

Calculate the full derivatives of $F_i(x)$ and $V_i(x)$ with respect to $x = (x_1, x_2, \cdots, x_n)^T$ respectively, the full derivative of $F_i(x)$ is $F = [\frac{\partial F_i}{\partial x_i}(x_0)]$ and the full derivative of $V_i(x)$ is $V = [\frac{\partial V_i}{\partial x_i}(x_0)]$, where x_0 denotes the disease-free equilibrium point.

The formula for the basic regeneration number R_0 is obtained:

$$R_0 = \rho(FV^{-1}) \tag{11}$$

Here, V^{-1} denotes the inverse matrix of matrix V, and $\rho(FV^{-1})$ denotes the spectral radius of matrix FV^{-1}, i.e., the upper definitive bound on the absolute value of the eigenvalues of the matrix.

Since in the SQEAIR model, only three bins, E, A, and I, are infected, only the following new system consisting of the second, fourth, and sixth equations needs to be utilized to calculate the basic regeneration number:

$$\begin{cases} \dfrac{dE}{dt} = \dfrac{\beta S(E + A + I)}{N} - [(1-q)\omega + q]E \\ \dfrac{dA}{dt} = (1-q)(1-p)\omega E - (\lambda + \delta + \gamma_2)A \\ \dfrac{dI}{dt} = (1-q)p\omega E + \delta A - (\mu + \gamma_1)I \end{cases} \tag{12}$$

Let $x = (E, A, I)^T$, then the above system of ordinary differential equations can be written as $\frac{dx}{dt} = F(x) - V(x)$, where:

$$F(x) = \begin{bmatrix} \frac{\beta S(E+A+I)}{N} \\ 0 \\ 0 \end{bmatrix}.$$

$$V(x) = \begin{bmatrix} [(1-q)\omega + q]E \\ -(1-q)(1-p)\omega E + (\lambda + \delta + \gamma_2) \\ -(1-q)p\omega E - \delta A + (\mu + \gamma_1)I \end{bmatrix}.$$

Calculating the full derivatives of $F(x)$ and $V(x)$ with respect to $x = (E, A, I)^T$, respectively, yields:

$$F = \begin{bmatrix} \frac{\beta S}{N} & \frac{\beta S}{N} & \frac{\beta S}{N} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

$$V = \begin{bmatrix} [(1-q)\omega + q] & 0 & 0 \\ -(1-q)(1-p)\omega & \lambda + \delta + \gamma_2 & 0 \\ -(1-q)p\omega & \delta & \mu + \gamma_1 \end{bmatrix}.$$

from Eq. 11:

$$R_0 = \frac{\beta}{(1-q)\omega+q}[1+\frac{(1-q)(1-p)\omega}{\lambda+\delta+\gamma_2}+\frac{(1-q)p\omega}{\mu+\gamma_1}+\frac{(1-q)(1-p)\omega\delta}{(\lambda+\delta+\gamma_2)(\mu+\gamma_1)}] \tag{13}$$

Since $\beta = \beta_0 e^{-\alpha t}$, the expression for the fundamental regeneration number R_0 is

$$R_0 = \frac{\beta_0 e^{-\alpha t}}{(1-q)\omega+q}[1+\frac{(1-q)(1-p)\omega}{\lambda+\delta+\gamma_2}+\frac{(1-q)p\omega}{\mu+\gamma_1}+\frac{(1-q)(1-p)\omega\delta}{(\lambda+\delta+\gamma_2)(\mu+\gamma_1)}] \tag{14}$$

2.5 Evaluating Indicators

For the prediction of the number of existing infected cases and model selection, Akaike information criterion (AIC) is adopted to test the accuracy of the two models. AIC was proposed by Akaike H [24] in 1974. It measures the quality of relative statistical models, based on reducing information dissipation. In addition, it can help determine the best model, thereby reducing model overfitting and improving model accuracy. Its prominet advantage is that it can obtain more accurate results through the gain of parameters or the reduction of information dissipation in the selection of models.

The specific formula are as follows:

$$AIC = 2p + n \ln \frac{RSS}{n} \tag{15}$$

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{16}$$

where n is the observed number, RSS is residual sum of squares, p is the number of parameters, y_i is the actual number of infected cases, and $\hat{y}_i$ is the observed number of infected cases.

To further assess the prediction effectiveness of the number of existing infected cases, Mean Absolute Percentage Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are also used to test the accuracy of the two models. The specific formula are as follows:

$$MAE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{17}$$

$$MSE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{18}$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{19}$$

$$MAPE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}| \tag{20}$$

3 Numerical Simulation and Prediction

3.1 Numerical Simulation and Model Comparison

Since the outbreak of the epidemic, the National Health Commission of the People's Republic of China (NHC) has reported data on a daily basis, including the number of new confirmed cases, the number of new asymptomatic cases, the number of asymptomatic cases converted to confirmed cases, the number of newly cured asymptomatic cases removed from medical observation, the number of existing confirmed cases, the number of asymptomatic cases, and the number of cases quarantined. We retrieved epidemic-related data from Shanghai, China for 65 days from March 28 to May 31, 2022, as well as data related to the model such as population of shanghai. The preliminary values of cases groups are shown in Table 1, with official data from NHC, exposed at the beginning is on the basis of the rate of exposed cases transforming into infected cases to estimate, the actual data of infected cases and asymptomatic cases are shown in Fig. 3.
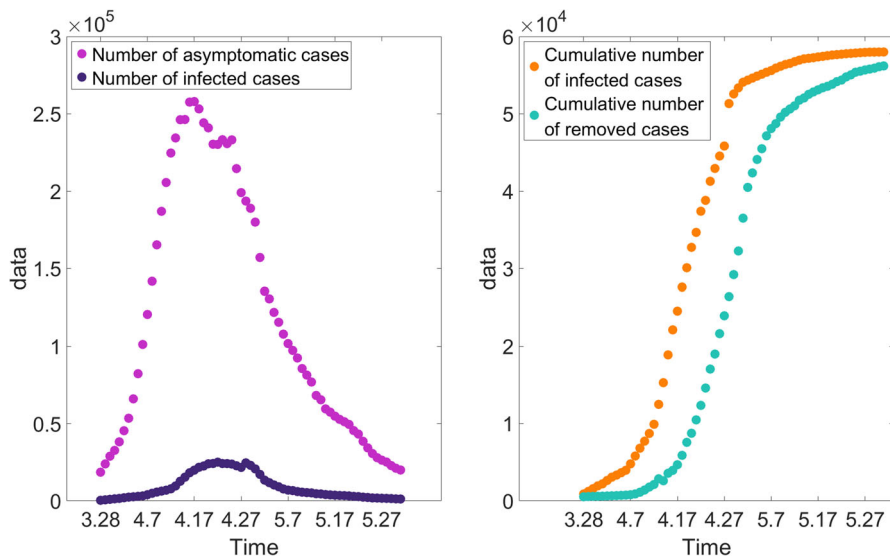
Fig. 3 Infected and asymptomatic cases in Shanghai

After obtaining the data, we first conducted stationarity test on the original data. Augmented Dickey-Fuller (ADF) test is also called the unit root test. The ADF test is to determine whether the sequence has a unit root: if the sequence is stationary, there is no unit root; Otherwise, there will be a unit root. Therefore, the null hypothesis of ADF test is the existence of unit root. If the obtained significance test statistic is less than 0.05, there is no unit root, and the sequence is stable. The ADF test values of both the existing confirmed cases and the existing asymptomatic cases were less than 0.05, negating the null hypothesis and believing that the original sequence was stable. Secondly, white noise test was carried out on the original data. The p value of the existing confirmed cases was 9.992×10^{-6} and the value of the existing asymptomatic cases was 1.221×10^{-15}, indicating that neither of the two groups of original data was white noise and could reflect the real trend and rule of the time series.

As to the data collected above, MATLAB software was used to conduct numerical simulation for the number of existing infections (number of existing infections are equal to the cumulative number of confirmed minus cumulative number of removed), and the values of parameters in the model (Table 2), the values of evaluating indicators (Table 3), and the fitting curve (Figs. 4 and 5) were obtained.

As to the fitting results (Figs. 4 and 5) and the AIC evaluation indexes (Table 3), although the improved SQEAIR model has four more chambers and added parameters than the original SEIR model, the fitting value is closer to the real data, the error is relatively low, and the fitting accuracy has improved, with better fitting effect. Notably, the fitting effect of SEIR model is accurate in the early stage, the estimated value in the middle stage is slightly lower than the real value, while the estimated value in the

Table 2 Model parameters for Shanghai

Arguments	Meaning	Initial value	Source
β_0	Effective patient contact rate at the beginning	1.1788	Calculate
α	Exponential decreasing rate of the effective contact rate	0.053	Fitting of parameters
q	The quarantined rate of exposed cases	0.5996	Fitting of parameters
λ	The quarantined rate of asymptomatic cases	0.6638	Fitting of parameters
μ	The quarantined rate of infected cases	0.7594	Fitting of parameters
ω	The rate of exposed becoming Infected cases	$\frac{1}{5.2}$	Reference [25]
p	The percentage of infected cases showing symptoms	0.64	Fitting of parameters
δ	The probability of asymptomatic cases transforming into infected cases	0.1077	Fitting of parameters
γ_1	The recovery rate of infected cases	$\frac{1}{8}$	Reference [26]
γ_2	The recovery rate of asymptomatic cases	$\frac{1}{12}$	Reference [26]
γ_3	The recovery or death rate of quarantined infected cases	0.4791	Fitting of parameters
γ_4	The recovery or death rate of quarantined asymptomatic cases	0.4504	Fitting of parameters

Table 3 Evaluating indicators of the models

Model	RSS	AIC	MAE	MSE	RMSE	MAPE
SEIR	2.03×10^8	978.18	1.34×10^3	3.13×10^6	1.77×10^3	28.11%
SQEAIR	1.19×10^8	953.35	1.01×10^3	1.83×10^6	1.35×10^3	20.18%

later stage is higher than the real value. Thus, the SQEAIR model is more suitable for the prediction of the number of cases infected by the Covid-19 epidemic in Shanghai.

In order to test the accuracy of the model, we also collected 32 days of epidemic data in Guangzhou, China, from November 1 to December 2, 2022 on the official website of Guangzhou Health Commission (Table 4). The actual data of the existing asymptomatic number, the existing confirmed number, the cumulative confirmed number, and the cumulative cured number are shown in Fig. 6.

The accuracy of SQEAIR model is verified again by comparing the fitting effect diagram and the error evaluation index (Figs. 7 and 8).

In addition, SQEAIR model can also fit and predict the number of asymptomatic cases because of the addition of asymptomatic cases compartment. Figure 9 shows the fitting results of SQEAIR model on the number of asymptomatic cases in the epidemic in Shanghai. It can be seen that the overall fitting effect is good, and the specific error indicators are shown in Tables 3, 5 and 6.

3.2 Prediction

Based on the SQEAIR model with high accuracy and the estimated values of the parameters obtained above, the future development trend of the epidemic in Guangzhou was predicted, and the prediction results were obtained on December 3 and in the following five days (Table 7) and the prediction effect diagram (Fig. 10). As shown in Table 7
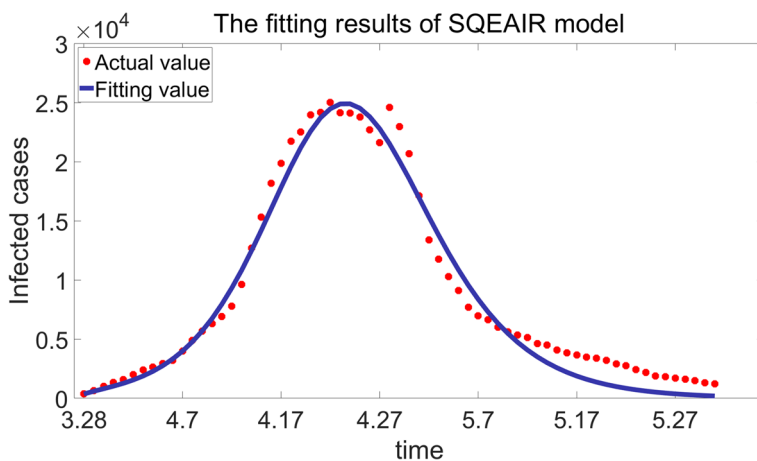
Fig. 4 The fitting results of SQEAIR model
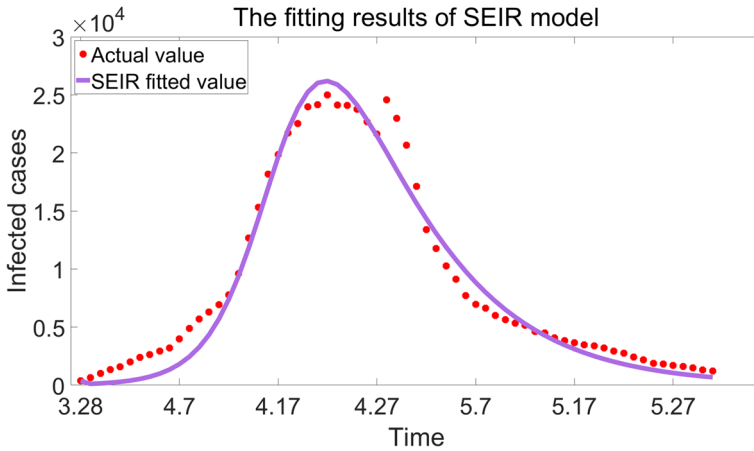
Fig. 5 The fitting results of SEIR model

and Fig. 10, the predicted number of cases is not much different from the real number of cases, average error rate is 2.93%, the overall prediction effect is relatively good, indicating that the parameter estimation is relatively accurate, and the SQEAIR model can simulate the development law of the epidemic situation well. Besides the number of confirmed cases in Guangzhou will gradually rise in the next five days, but has been actually kept within a manageable range.

The basic regeneration number R_0 is a key indicator for portraying whether there is an outbreak of an infectious disease, so it is necessary to utilize the calculations of the epidemic data from Guangzhou City to determine whether there is a possibility of a large-scale outbreak.

The SQEAIR model was utilized to fit the number of existing infections in Guangzhou City, and the values of each parameter were obtained as shown in the Table 8.

According to the formula and the values of the parameters in the above table, the transformed graph of the basic regeneration number can be obtained (Fig. 11).

Since the beginning of the Guangzhou epidemic on November 1, the value of the basic regeneration number is 2.06, and then it gradually decreases. In fact, the basic regeneration number $R_0 = 1$ on the 17th day, and then the values become smaller

Table 4 The situation in Guangzhou, Nov. 1, 2022

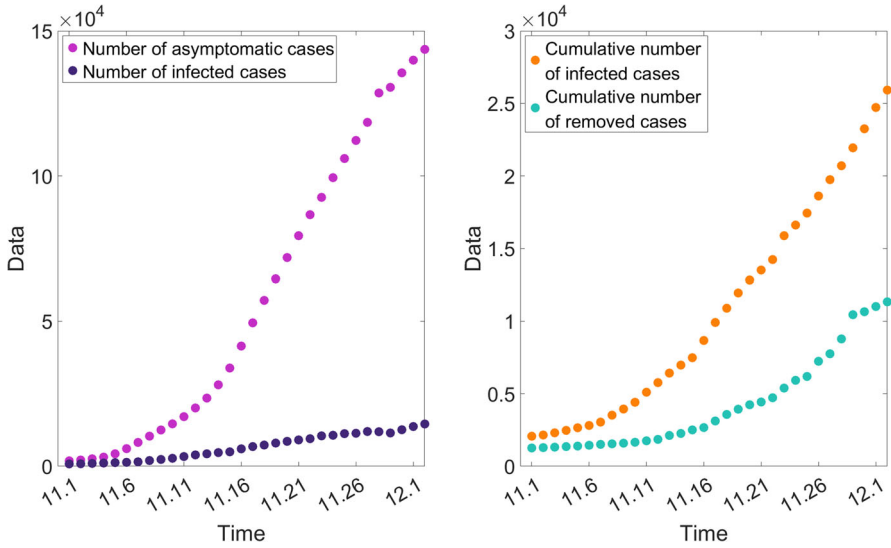| Arguments	Meaning	Initial value	Source
N	Total population of Guangzhou	18811000	Official data
E_0	Exposed at the beginning	5277	Estimated data
A_0	Asymptomatic at the beginning	1492	Official data
I_0	Infected at the beginning	811	Official data
R_0	Removed at the beginning	1637	Official data

Fig. 6 Infected and asymptomatic cases in Guangzhou

than 1, which means that from the 17th day onwards, on average, each infected person causes fewer than 1 person to be infected of the disease during, meaning that the disease's infectious capacity has begun to decline. Due to the incubation period and the time required for asymptomatic infected persons to turn into confirmed patients, the epidemic will eventually be brought under control, although the number of infected persons is still gradually increasing.
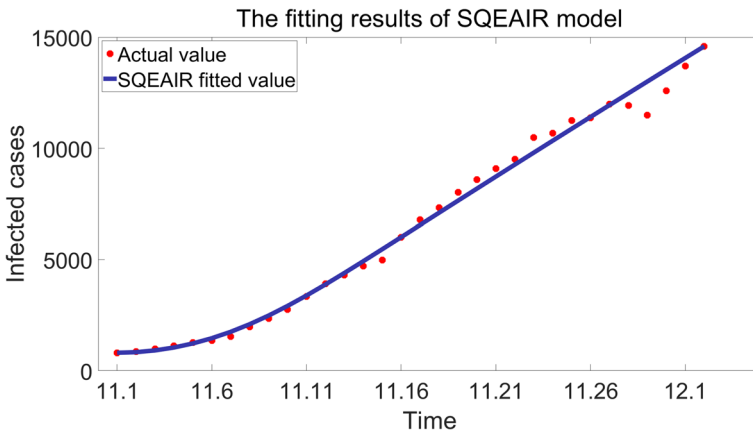
Fig. 7 The fitting results of SQEAIR model
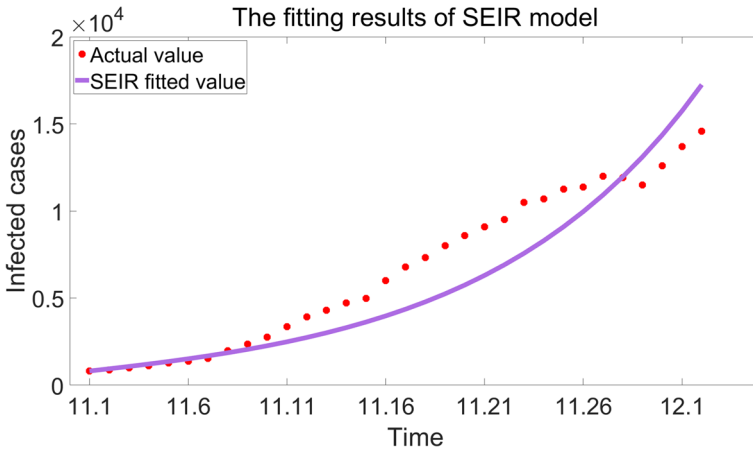
Fig. 8 The fitting results of SEIR model

4 Influencing Factors

Based on the SQEAIR model proposed above, respectively considering the impact of tracking isolation measures, medical treatment level and personal protection measures on the number of infected cases, and the effect of prevention and control was evaluated by changing parameters. The effectiveness of prevention and control strategies was evaluated mainly by the number of cases and the peak of the epidemic. The following prediction assumes a local population of 100,000.
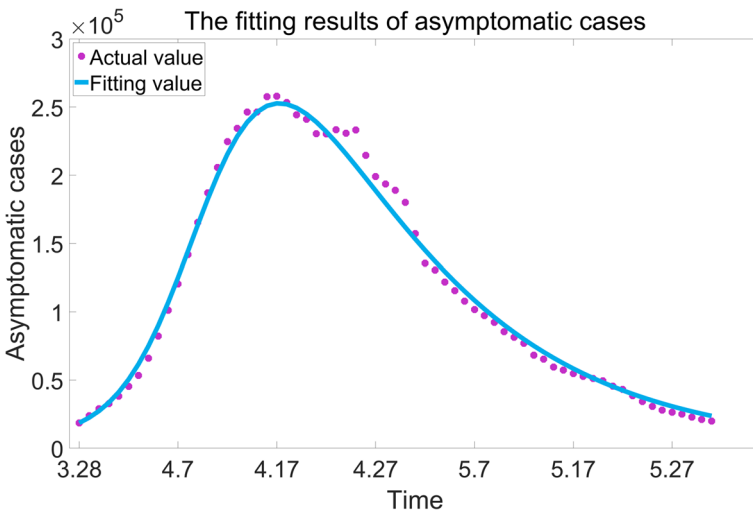
Fig. 9 SQEAIR model was used to fit the number of asymptomatic cases in the Shanghai

Table 5 Evaluating indicators of the models

Model	RSS	AIC	MAE	MSE	RMSE	MAPE
SEIR	5.73×10^7	466.74	1.07×10^3	1.79×10^6	1.34×10^3	15.68%
SQEAIR	3.43×10^6	394.58	189.8317	1.0702×10^5	327.1407	2.72%

4.1 Tracking Quarantine Measures

Local epidemic prevention authorities implemented control measures for close contacts and sub-close contacts, reflecting the effectiveness of the government?s tracking and quarantine measures. Scientific tracking quarantine measures can reduce the rate of infection effectively. Here, we depict the impact of different quarantined rate of exposed cases on the number of cases who got sick in an outbreak (Fig. 12). When the quarantined rate of exposed cases is 0, the number of asymptomatic cases and confirmed cases would greatly increase, and the peak of the epidemic would come in advance, which would have a great impact on local medical supplies and production and life. When the quarantined rate of exposed cases increases, the number of asymptomatic cases and confirmed cases would be significantly reduced, the peak of the epidemic would come later, and the impact on local medical institutions would be greatly reduced (Table 9).

4.2 Medical and Health Level

Local cure rates of infected cases reflecting the effectiveness of the Medical and health level. Figures 13 and 14 depicts the impact of the recovery rate of different infected cases on the number of sick cases in the epidemic. When medical and health level is inferior, the number of asymptomatic cases and confirmed cases will increase greatly, and the peak of the epidemic will come in advance, which will have a great impact on local medical supplies and production and life. When the medical level is high, the number of asymptomatic cases and confirmed cases will be significantly reduced, the peak of the epidemic will come later, and the impact on local medical institutions will be greatly reduced. And with the improvement of medical level, the peak number of infected cases decreased significantly (Tables 10 and 11).

Table 6 Evaluating indicators of the models

Indicators	MAE	MSE	RMSE	MAPE
	5.83×10^3	5.74×10^7	7.58×10^3	6.5%

Table 7 Evaluating indicators of the models

Data	Dec3	Dec4	Dec5	Dec6	Dec7	Dec8
Predicted number of cases	15116	15640	16163	16684	17204	17723
Real number of cases	15427	15456	16462	17059	17524	17865
Error rate	2.01%	1.13%	1.82%	2.20%	1.83%	0.79%
Average error rate	1.63%					

4.3 Personal Protection Measures

Individual epidemic prevention behaviors, such as wearing masks when going out, washing hands frequently at home, and using public spoons and chopsticks, reflect citizens' attention to the Covid-19 epidemic, and scientific protective measures can effectively reduce the rate of infection. The figure below depicts the effect of different effective exposure rates on the number of cases in an outbreak. As can be seen from the Fig. 15, the lower the effective contact rate of cases, the smaller the number of asymptomatic infected and confirmed cases, and the later the peak of the epidemic, the more conducive to the control of the epidemic. And with the reduction of the effective contact rate of cases, the peak number of infected cases has been significantly reduced (Table 12)

5 Conclusion

This study focuses on fitting, predicting, and scientifically managing the number of infected individuals using an enhanced infectious disease dynamics model tailored to the specific circumstances of the epidemic in China. Initially, an improved SQEAIR
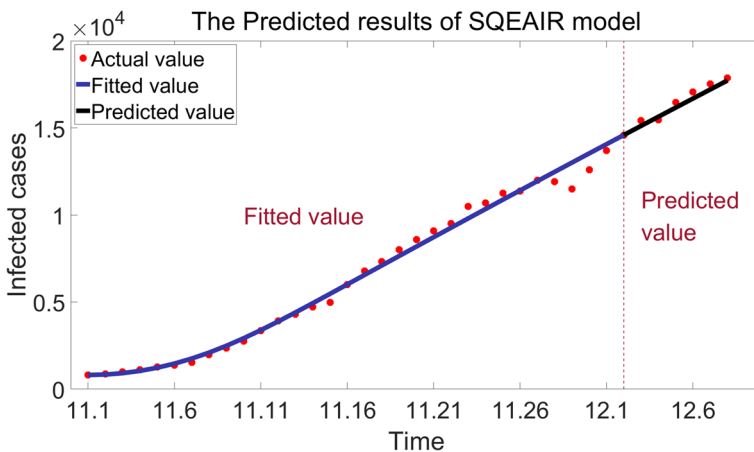
Fig. 10 Prediction results of SQEAIR model

Table 8 Model parameters for Guangzhou

Arguments	Meaning	Initial value	Source
β_0	Effective patient contact rate at the beginning	1.2088	Calculate
α	Exponential decreasing rate of the effective contact rate	0.042	Fitting of parameters
q	The quarantined rate of exposed cases	0.7237	Fitting of parameters
λ	The quarantined rate of asymptomatic cases	0.8799	Fitting of parameters
μ	The quarantined rate of infected cases	0.9629	Fitting of parameters
ω	The rate of exposed becoming Infected cases	$\frac{1}{5.2}$	Reference [25]
p	The percentage of infected cases showing symptoms	0.0126	Fitting of parameters
δ	The probability of asymptomatic cases transforming into infected cases	0.775	Fitting of parameters
γ_1	The recovery rate of infected cases	$\frac{1}{8}$	Reference [26]
γ_2	The recovery rate of asymptomatic cases	$\frac{1}{12}$	Reference [26]
γ_3	The recovery or death rate of quarantined infected cases	0.6124	Fitting of parameters
γ_4	The recovery or death rate of quarantined asymptomatic cases	0.4221	Fitting of parameters

infectious disease dynamics model was developed by augmenting the traditional SEIR model with three isolation compartments and one asymptomatic infected compartment to create a more realistic representation.

Subsequently, data pertaining to the number of infected individuals in Shanghai, China, between March and May 2022 were gathered and fitted using both the SEIR and SQEAIR models. The estimates of crucial parameters in the models and the associated error values were then calculated. The results indicated that the SQEAIR
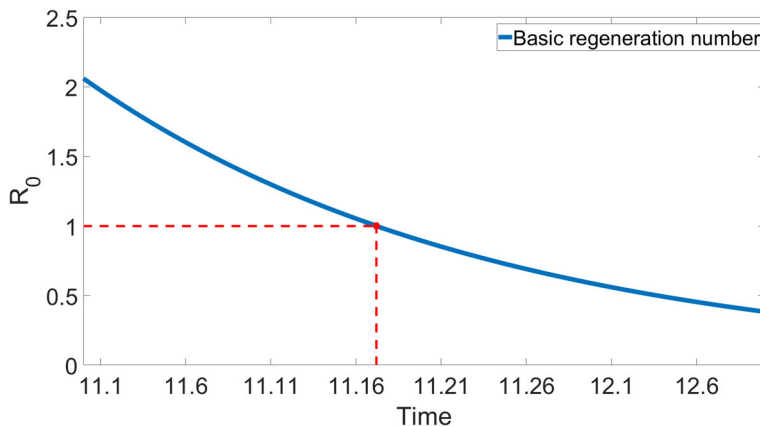

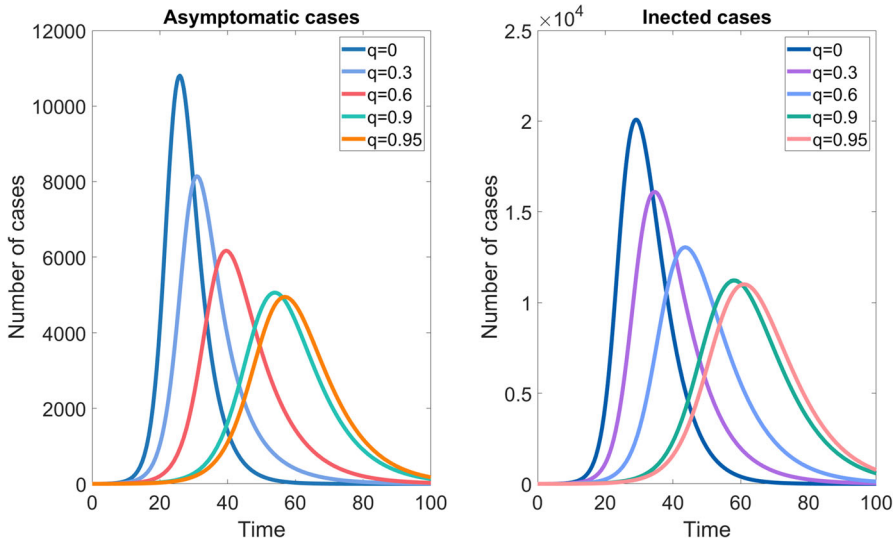
Fig. 11 Basic regeneration number change curve

Fig. 12 The impact of quarantine measures on the number of cases infected in the outbreak

model exhibited a lower fitting error and a reduced AIC value, suggesting its superior fit compared to the traditional infectious disease dynamics model in capturing the epidemic trend in Shanghai. Moreover, when the SQEAIR model was applied to fit data on asymptomatic infected individuals in Shanghai, the error rate was notably lower at 6.5%.

To further validate the model's accuracy, data on infected individuals in Guangzhou City in November 2022 were collected, with results consistently favoring the SQEAIR model over the SEIR model. Subsequent predictions using the improved model for confirmed cases in Guangzhou over the next five days yielded a low prediction error of only 2.53%. Calculations of the basic reproduction number of the epidemic in Guangzhou indicated that it had already reached 1 and was gradually decreasing, suggesting that while confirmed cases might rise, they would remain manageable.

Finally, leveraging the enhanced SQEAIR infectious disease dynamics model, the study explored the impact of varying latent isolation rates, infected person recovery rates, and effective patient contact rates on the number of infected individuals. Simulation results highlighted that higher isolation rates, lower recovery rates, and reduced

Table 9 Table of epidemic peaks corresponding to different quarantined rate of exposed cases

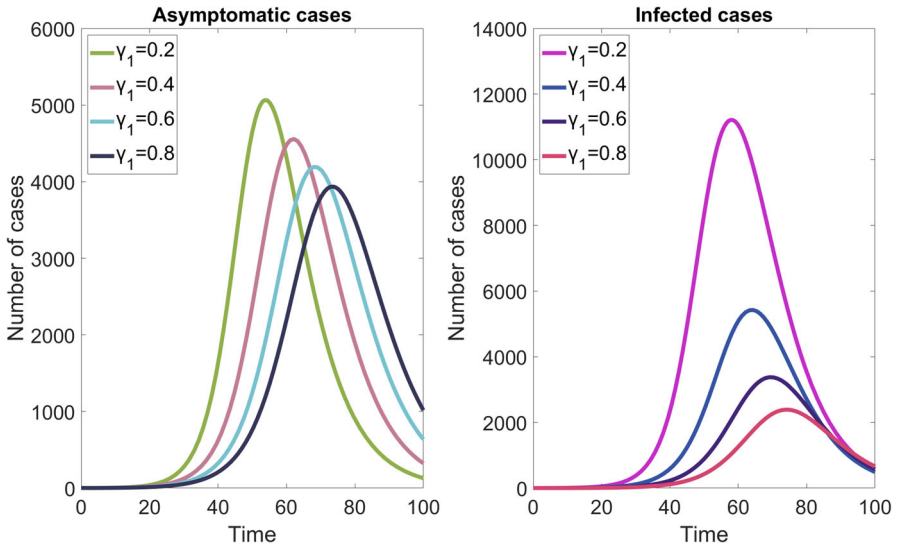| Quarantined rate of exposed cases	0	0.3	0.6	0.9	0.95
Peak days for asymptomatic cases	25	31	39	53	56
Maximum number of asymptomatic cases	10805	8141	6169	5066	4954
Peak days for infected cases	28	34	43	58	60
Maximum number of infected cases	20081	16098	13038	11215	11011

Fig. 13 Influence of cure rates of infected cases on the numbers of infected cases in the outbreak

patient contact rates led to fewer infections and delayed epidemic peaks, underscoring the importance of effective public management in outbreak prevention and control efforts.
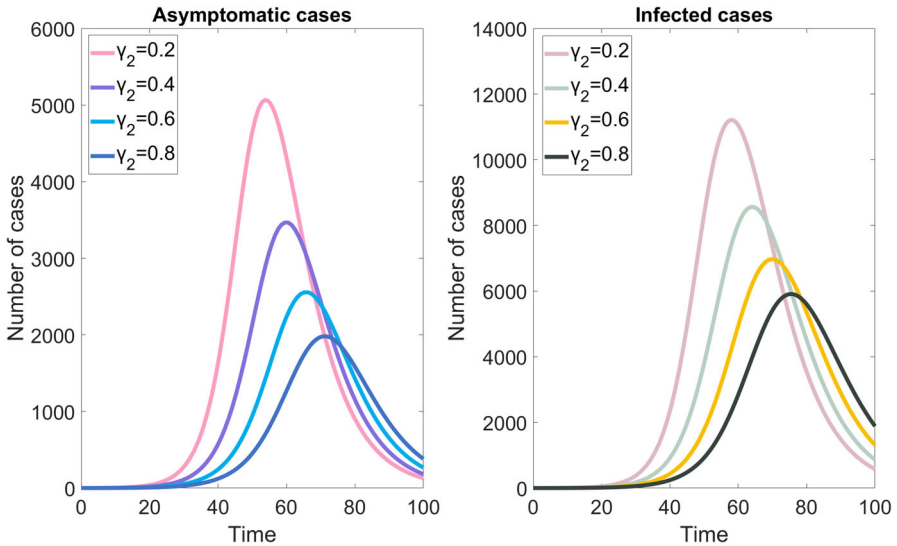

Fig. 14 Influence of cure rates of asymptomatic cases on the numbers of infected cases in the outbreak

Table 10 Epidemic peaks corresponding to the cure rates of different infected cases

Cure rate of infected cases	0.2	0.4	0.6	0.8
Peak days for asymptomatic cases	53	62	68	72
Maximum number of asymptomatic cases	5045	4556	4193	3933
Peak days for infected cases	59	63	69	73
Maximum number of infected cases	11206	5415	3377	2377

Table 11 Epidemic peaks corresponding to the cure rates of different asymptomatic cases

Cure rate of asymptomatic cases	0.2	0.4	0.6	0.8
Peak days for asymptomatic cases	53	60	65	70
Maximum number of asymptomatic cases	5045	3466	2556	1977
Peak days for infected cases	58	64	70	76
Maximum number of infected cases	11215	8562	6973	5887
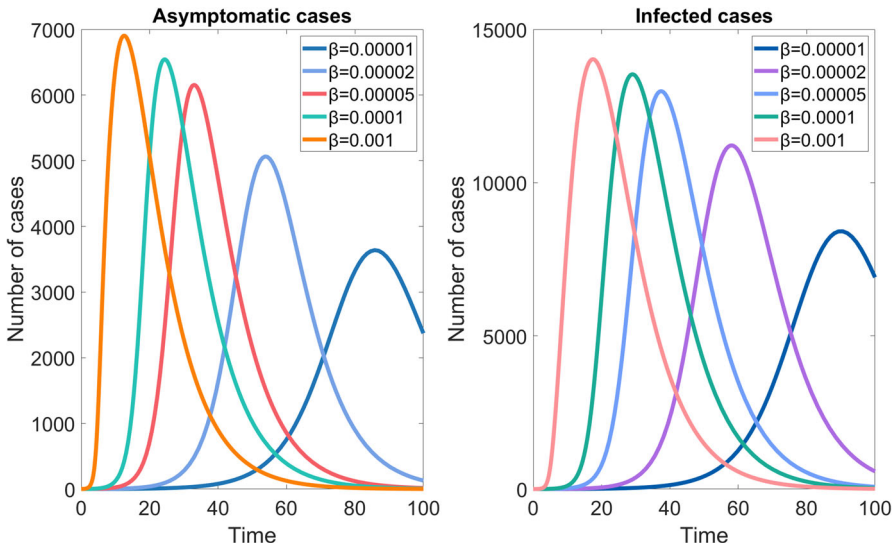
Fig. 15 Impact of personal protection measures on the numbers of cases infected in an outbreak

Table 12 Table of epidemic peaks corresponding to different effective contact rates

Effective contact rate	0.00001	0.00002	0.00005	0.0001	0.001
Peak days for asymptomatic cases	87	54	33	24	13
Maximum number of asymptomatic cases	3639	5066	6154	6543	6904
Peak days for infected cases	89	58	38	29	17
Maximum number of infected cases	8412	11215	12983	13535	14027

Declarations

Competing interests The authors declare no competing interests.

References

1. World, O. Health, et al.: 2019 novel coronavirus global research and innovation forum: Towards a research roadmap (2020)
2. Huang, C., Wang, Y., Li, X., et al.: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet **395**(10223), 497–506 (2020)
3. Read, J.M., Bridgen, J.R., Cummings, D.A., et al.: Novel coronavirus 2019-ncov (covid-19): early estimation of epidemiological parameters and epidemic size estimates. Philos. Trans. R. Soc. B **376**(1829), 20200265 (2011)
4. Shi, L., Shen, Y., Zhao, X., et al.: Infection mechanism of sars-cov-2 and its control of transmission route. Genomics Appl. Biol., 3874–3880 (2020)
5. Hethcote, H.W.: The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
6. Rock, K., Brand, S., Moir, J., Keeling, M.J.: Dynamics of infectious diseases. Rep. Prog. Phys. **77**(2), (2014)
7. Kuhl, E.: Computational Epidemiology. Springer, ??? (2021)
8. Frank, T.D.: COVID-19 Epidemiology and Virus Dynamics. Springer, ??? (2022)
9. Gatto, M., Bertuzzo, E., Mari, L., et al.: Spread and dynamics of the covid-19 epidemic in Italy: effects of emergency containment measures. Proc. Natl. Acad. Sci. **117**(19), 10484–10491 (2020)
10. Ngonghala, C.N., Iboi, E., Eikenberry, S., et al.: Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel coronavirus. Math. Biosci. **325**, 108364 (2020)
11. Yang, Z., Zeng, Z., Wang, K., et al.: Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. J. Thorac. Dis. **12**, 165 (2020)
12. Wang, J., Zhang, H., Jia, P., et al.: City-level structured prediction and simulation model of covid-19. J. Comput. Aided Des. Comput. Graph. **34**(08), 1302–1312 (2022)
13. Martcheva, M.: An Introduction to Mathematical Epidemiology. Springer, ??? (2015)

14. Wei, Y., Lu, Z., Du, Z., et al.: Fitting and forecasting the trend of covid-19 by seir (+ caq) dynamic model. Zhonghua Liu Xing Bing Xue Za Zhi= Zhonghua Liuxingbingxue Zazhi **41**(4), 470–475 (2020)

15. Sun, G., Zhao, Y.: Prediction of covid-19 outbreak and assessment of prevention and control measures based on improved seir model. J. Qingdao Univ.: Nat. Sci. Ed. **34**(2), 1–8 (2021)

16. Han, H., Zhang, Z.: Mathematical analysis of a covid-19 epidemic model by using data driven epidemiological parameters of diseases spread in India. Biophysics **67**(2), 231–244 (2022)

17. Proverbio, D., Kemp, F., Magni, S., et al.: Dynamical spqeir model assesses the effectiveness of non-pharmaceutical interventions against covid-19 epidemic outbreaks. PloS One **16**(5), 0252019 (2021)

18. Giordano, G., Blanchini, F., Bruno, R., et al.: Modelling the covid-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. **26**(6), 855–860 (2020)

19. Gallo, L., Frasca, M., Latora, V., et al.: Lack of practical identifiability may hamper reliable predictions in covid-19 epidemic models. Sci. Adv. **8**(3), 5234 (2022)

20. Kemp, F., Proverbio, D., Aalto, A., et al.: Modelling covid-19 dynamics and potential for herd immunity by vaccination in Austria, Luxembourg and Sweden. J. Theor. Biol. **530**, 110874 (2021)

21. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proc. R. soc. Lond. Ser. A-Contain. Pap. Math. Phys. Character **115**(772), 700–721 (1927)

22. Siettos, C.I., Russo, L.: Mathematical modeling of infectious disease dynamics. Virulence **4**(4) (2013)

23. Driessche P, V., J, W.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**(1-2) (2002)

24. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Control **19**(6) (1974)

25. Li, Q., Guan, X., Wu, P., et al.: Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. **382**(13), 1199–1207 (2020)

26. Tang, B., Wang, X., Li, Q., et al.: Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. J. Clin. Med. **9**(2), 462 (2020)

Authors and Affiliations

Chenxi Wang[1] · Yongchao Jin[1,2] · Lihui Zhou[1] · Wei Hou[1] · Dongmei Liu[1] · Jianjun Wang[3] · Xiyin Wang[1,2,3]

Chenxi Wang
chenxiwang123@outlook.com

Yongchao Jin
jinyongchao@ncst.edu.cn

Lihui Zhou
zhoulh324@163.com

Wei Hou
yjphhw@qq.com

Dongmei Liu
dmliu@ncst.edu.cn

Jianjun Wang
wangjianjun0916@163.com

[1] College of Sciences, North China University of Science and Technology, Tangshan 063210, Hebei, China

[2] Hebei Key Laboratory of Data Science and Application, Tangshan 063210, China

[3] School of Public Health, North China University of Science and Technology, Tangshan 063210, Hebei, China