Journal of Applied Analysis and Computation
Volume 13, Number 4, August 2023, 2121–2152
Website:http://www.jaac-online.com
DOI:10.11948/20220389

THE WITHIN-HOST VIRAL KINETICS OF SARS-COV-2

Meng Wang[1], Yafei Zhao[1], Chen Zhang[2] and Jie Lou[1,†]

Abstract Understanding the dynamics of SARS-COV-2 infection in vivo is crucial for exploring more effective treatments. This paper presents a series of dynamic models of viral infection in host. We use affine invariant set Monte Carlo algorithm to achieve parameter fitting and model selection, and study the structural identifiability of these models to determine if the clinical data could specify the model parameters. Then we analyze the actual identifiability and numerical simulation of the selected optimal model. Research shows that all models are structurally identifiable, and data noise has little effect on the actual identifiability of key parameters. Through numerical simulation we found the key factors that may cause cytokine storms. In addition, we also obtain some qualitative conclusions of the model, including the infection threshold, the stability of the equilibrium state and the periodic solution. Studies have found that viral load may exhibit complex periodic motions in some cases, which may provide new evidence to handle repeated reactivations among new corona virus infections.

Keywords Dynamics model, SARS-COV-2, identifiability analysis, model selection, periodic solution.

MSC(2010) 92D30, 92D40.

1. Introduction

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) [50]. SARS-COV-2 is an RNA virus whose S protein infects alveolar epithelial cells (mainly type II alveolar cells) by binding to angiotensin-converting enzyme 2 (ACE2) receptors [36, 45]. Destruction of alveolar epithelial cells and increased cellular permeability lead to viral release [32]. SARS-COV-2 infection causes T lymphocytes to release numerous extracellular molecular regulators, such as pro-inflammatory factors, chemokines, etc., to induce innate and adaptive immune responses [15, 31], and to promote the death of viruses and infected cells [33]. Indeed, in the early stages of SARS-COV-2 infection, T cell-mediated immunity is essential to control the spread of the virus [44]. However, in severe COVID-19 patients, T cells may over-induce immune responses, causing tissue damage [42].

†The corresponding author.
Email: m13563783270@163.com(M. Wang), yafei_025@163.com(Y. Zhao), chen_zhang@urmc.rochester.edu(C. Zhang), jie_lou@shu.edu.cn(J. Lou)
[1]Department of Mathematics, Shanghai University, 99 Shangda Road Shanghai, 200444, China
[2]School of Nursing, University of Rochester, Rochester, NY, USA

Many researchers have explored the dynamic model of the COVID-19 transmission among macroscopic individuals from various angles [7,18,21,37,46]. Still, there are few models studying the virus dynamics in the host of SARS-COV-2 infection [15]. Indeed, within-host models can not only explore how the infection process is affected by the model parameters, especially the parameters related to the host immune system and drug intervention. It also helps to guide the treatment steps in the viral life cycle and optimize the treatment strategy [23]. Ejima et al. [9]used a simple 2-ODE system to model the infection process of SARS-COV-2 in susceptible host cells, and used this model to estimate the infection time and distinguish the imported cases from the local secondary cases. Kim et al. [19] used the same model to simulate potential anti-SARS-COV-2 therapies. Miao et al. used viral load data from 11 patients from different countries [20, 30] and non-human primates [27] to explore how cell-mediated and humoral immunity affects the infection process and viral load dynamics [38]. Sahoo et al. developed a model to analyze the intra-host dynamics between virus-infected cells, identifying key parameters of various clinical phenotypes associated with COVID-19 [39]. However, none of these studies were fitted with real clinical data from SARS-COV-2 infected individuals and lacked medical support.

Even if reliable clinical data are collected, obtaining parameter estimation by establishing a model and fitting it is not rigorous. Researchers can describe the infection process of viruses in vivo by selecting different mathematical models. Choosing the model among several models that best fits the actual situation is a problem worth exploring. It is a new attempt to apply the model selection method in the infectious disease dynamics model. Different mathematical models will produce other dynamic behaviors of the virus. After the optimal model represents the observation data information to a certain extent, the model can be used for subsequent analysis to provide more valuable information.

The data help us to fit the proper parameters as closely as possible, provided that the parameters are identifiable. However, when fitting models with limited clinical data to estimate model parameters, this often results in the best-fit parameter values not being unique, i.e. multiple parameter values will result in the best fit to the data. Still, different parameter values may produce very different predictions. This phenomenon is known as an unidentifiable parameter. Parameter identifiability analysis includes structural identifiability analysis and actual identifiability analysis. The former is investigated under the assumption that the data are noise-free. See the literature [5,10,28] for details. The latter is an identifiability analysis when the data is affected by noise. Since the data collected is often noisy, the actual unidentifiable parameters are difficult to avoid. At this stage, the impact of unidentifiability on the reliability of parameter valuation is reduced by searching for more independent data sets or using better-fitting algorithms and models. The affine invariant set Monte Carlo algorithm (GWMCMC) is currently the most efficient method for fitting data [12, 34]. The algorithm combines affine invariant sets with the traditional MCMC algorithm, using multiple walkers to iterate simultaneously. The next position of a walker depends on the current positions of all other walkers. Its performance is independent of the spatial affine transformation, and it converges quickly to the target posterior distribution even in the face of unrecognizable problems.

In this paper, we model the biological mechanisms of SARS-COV-2 infection, replication, and interaction with the immune system in the host, perform model selection and parameter identifiability analysis, and investigate the dynamic changes

in viral load. In the second part, a four-variable infectious disease model (Full model) is developed based on the infection mechanism; in the third part, the Full model is analyzed for structural identifiability and four sub-models are developed based on different assumptions. In the fourth part, the equilibrium point, stability, and Hopf branch of the Full model are analyzed and validated by numerical simulations.

2. The model and sources of clinical data

2.1. The model

In this study, we hypothesized that type II alveolar cells are the main target cells of the SARS-COV-2 virus and developed a kinetic model for the infection process of this virus in vivo. This model depicts the interaction between healthy type II alveolar cells (C), infected type II alveolar cells (I), T lymphocytes (T), and free SARS-COV-2 virus particles (V). The S protein on the surface of SARS-COV-2 plays an important role in the infection of target cells such as alveolar type II cells. The S1 subunit of S protein contains a receptor binding domain (RBD). RBD binds to ACE2 on target cells, leading to cell infection [47]. Wang et al. [48] found that the antibody "47D11" also has an affinity for the RBD-binding domain of the S protein in their study of antiviral therapy . This dynamic response and 'cooperative binding' between organisms are often portrayed by the Hill function [29,40,48]. We assume that the mechanism of action between the virus and the immune system is as shown in Fig 1. This infection mechanism can also be described by the following nonlinear system (2.1).

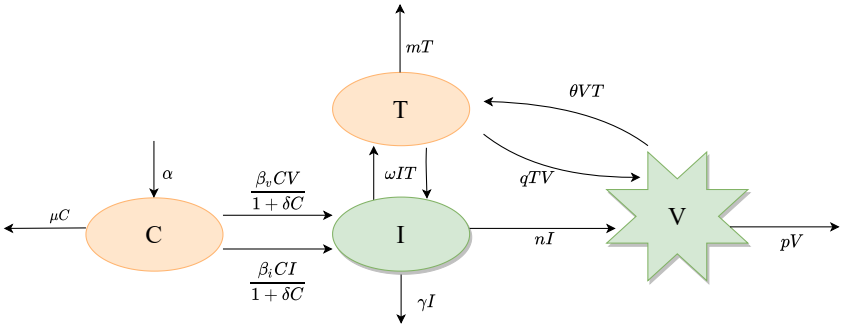

Figure 1. Diagram of the dynamics of SARS-COV-2 infection. Healthy type II alveolar cells are produced at a constant rate of α and can be infected by viruses as well as infected cells in the form of Hill functions. The entry of SARS-COV-2 into the body activates adaptive immunity, with ω and θ indicating the effect of the immune response induced by the infected cells and the virus, respectively. T lymphocytes clear the virus at a rate of q and infected cells at a rate of ω. μ, γ, m and p represent the natural apoptosis rates of healthy II type alveolar cells, infected II type alveolar cells, T lymphocytes and free SARS-COV-2 virus, respectively. See Table 1 for details of the meaning of the parameters.

$$\begin{cases} \dfrac{dC}{dt} = \alpha - \dfrac{\beta_v CV}{1+\delta C} - \dfrac{\beta_i CI}{1+\delta C} - \mu C, \\ \dfrac{dI}{dt} = \dfrac{\beta_v CV}{1+\delta C} + \dfrac{\beta_i CI}{1+\delta C} - \omega IT - \gamma I, \\ \dfrac{dT}{dt} = \omega IT + \theta VT - mT, \\ \dfrac{dV}{dt} = nI - qTV - pV. \end{cases} \tag{2.1}$$

Table 1. Description and unit of parameters.

Parameter	Description	Unit
α	Type II alveolar cells production rate	cells ml^{-1}day^{-1}
δ	Hill coefficient	ml cell^{-1}
β_v	Virus-to-type II alveolar cells infection rate	ml virion^{-1}day^{-1}
β_i	Infected cells-to-type II alveolar cells infection rate	ml cell^{-1}day^{-1}
θ	Virus-to-T lymphocytes infection rate	ml virion^{-1}day^{-1}
ω	Infected cells-to-T lymphocytes cells	ml cell^{-1}day^{-1}
n	Virion production rate per infected epithelial cell	virion cell^{-1} day^{-1}
q	The rate of lymphocyte attacking virus	ml virion^{-1}day^{-1}
μ	Death rate of healthy type II alveolar cells	day^{-1}
γ	Death rate of infected type II alveolar cells	day^{-1}
m	Death rate of T lymphocytes	day^{-1}
p	Virus clearance rate	day^{-1}

2.2. Sources of clinical data

The clinical data of a total of 198 adult patients ($\geq$18) that were recruited to the Yale IMPACT Biorepository between 18 March and 9 May 2020 were used in our study [25, 43]. These patients were positive for SARS-COV-2 by RT-PCR of nasal or pharyngeal swabs. All patients were admitted to the hospital with symptoms. The first time point sample of the study was collected for each patient at the time of positive diagnosis, and subsequent biological samples were collected for 48 of these patients at an average of every 3 to 7 days. Doctors scored the severity of COVID-19 disease (1-5) by reviewing electronic cases at each longitudinal time point. The literature shows a large gap between the biological indicators for mild patients (scores 1-3) and severe patients (scores 4-5), and clinical data for severe patients are severely missing in the time series. In addition, the variation in medication use from patient to patient may also affect the viral load data to some extent [25, 43].

This paper intends to use the clinical data of 'T lymphocyte count' and 'viral load' of some COVID-19 patients in Yale New Haven Hospital [25, 43] to fit the model parameters. To make use of the clinical data of the patients and to make the fitting results more effective, the data need to be homogenized. We screened data from patients with mild disease treated with both hydroxychloroquine (HCQ) and tocilizumab (Toci). At the same time, to reduce the effect of drug action time on biomass, we only selected patients who were treated within 7 days after symptoms and those who were treated for more than 3 days as the research objects. The T lymphocyte data and viral load data of patients every two days were averaged and applied to the system.

3. The identifiability and model selection

The predicting results of the model are highly dependent on accurate parameter estimation. The reasonableness of the parameter estimates requires a discussion of the identifiability of the model parameters. The identifiability analysis is generally divided into two parts: mathematical identifiability and statistical identifiability. Structural identifiability analysis investigates whether the parameters in an ODE model are identifiable under ideal measurement assumptions (no measurement error), while practical identifiability investigates whether experimental data with noise can identify (and reasonably estimate) the parameters in the model.

3.1. Structural identifiability analysis

Structural identifiability analysis examines whether unique estimates of the parameters can be obtained from the input-output equations, assuming ideal conditions of measurement data free of interference and error. Under ideal data conditions, a model is said to be structurally identifiable if parameter values can theoretically be uniquely determined from these data, or structurally unidentifiable if two or more parameter sets can lead to the same observed output.

Structural identifiability of the model is necessary for finding a solution to the parameter estimates from the actual data. Even if a model is structurally unidentifiable, it can reveal helpful information between parameters. This makes the structural identifiability of analytical models an essential consideration in epidemiological modeling [11]. To determine whether the model's parameters (2.1) are structurally identifiable under ideal conditions (noise-free data), a definition of structural identifiability analysis is introduced here.

Definition 3.1 ([28]). A parameter set $\boldsymbol{p}$ is said to be globally (or uniquely) structurally identifiable if for every $\boldsymbol{q}$ in the parameter space

$$\boldsymbol{g}(\boldsymbol{x}(t), \boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{x}(t), \boldsymbol{q}) \Rightarrow \boldsymbol{p} = \boldsymbol{q},$$

where $\boldsymbol{x}(t)$ is the state variable, $\boldsymbol{p}$ is the system parameter, and the observations (new diagnoses and new deaths) are defined by the model output $\boldsymbol{g}(\boldsymbol{x}(t), \boldsymbol{p})$.

For linear models, the problem can be solved globally via a transfer function approach and other linear algebra methods [6]. However, for nonlinear ODE models, identifiability analysis remains a challenge for applications beyond relatively simple models. One productive identifiability approach is the differential algebra method [4, 24]. We can employ the model not only to determine the overall identifiability but also reveal identifiable parameter combinations and the re-parameterization. Therefore, the structural identifiability of the model is discussed next using the differential-algebraic method.

Theorem 3.1. *With two sets of observations 'T-lymphocyte' and 'viral load', the model* (2.1) *with parameters* $\boldsymbol{p}$=$[\alpha, \delta, \beta_v, \beta_i, \theta, \omega, n, q, \mu, \gamma, m, p]$ *is structurally identifiable.*

Proof. Since the dataset contains both T-lymphocyte and viral load data, the corresponding outputs of the model can be given by y_1, y_2. Let $y_1 = T$, $y_2 = V$. For simplicity, the latter two equations of the model are expressed in terms of y_1,

y_2, y_1' and y_2'

$$y_1' = \omega I y_1 + \theta y_2 y_1 - m y_1, \tag{3.1}$$
$$y_2' = nI - p y_2 - q y_1 y_2. \tag{3.2}$$

From (3.1), we have

$$I = \frac{y_1' + m y_1 - \theta y_1 y_2}{\omega y_1}. \tag{3.3}$$

Derivation of (3.3)

$$I' = \frac{y_1 y_1' - \theta y_1^2 y_2'}{\omega y_1^2}. \tag{3.4}$$

Substitute (3.4) into (3.2)

$$y_2' = \frac{n(y_1' + m y_1 - \theta y_1 y_2)}{\omega y_1} - p y_2 - q y_1 y_2. \tag{3.5}$$

Simplifying (3.5) yields the first input-output equation

$$-n y_1' + \omega y_1 y_2' + q\omega y_1^2 y_2 + (p\omega + \theta n) y_1 y_2 - mn y_1 = 0. \tag{3.6}$$

To test the identifiability of the parameters (m, n, ω, q) in the input-output equation (3.6), suppose a set of parameters (b_1, b_2, b_3, b_4) yields the same output, then a new set of equations can be obtained

$$\begin{cases} -n = -b_2, \\ \omega = b_3, \\ q\omega = b_4 b_3, \\ -mn = -b_1 b_2. \end{cases}$$

Using (b_1, b_2, b_3, b_4) to solve for (m, n, ω, q), we get $m = b_1, n = b_2, \omega = b_3, q = b_4$, so the parameters m, n, ω, q are identifiable.

The second input-output equation is not given here because it is too lengthy. However, in order to determine the identifiability of the remaining unknown parameters $(\alpha, \beta_v, \beta_i, \delta, \mu, \gamma, \theta, p)$, we also assume that a set of alternative parameters $(b_5, b_6, b_7, b_8, b_9, b_{10}, b_{11}, b_{12})$ can produce the same output, then the following set of equations can be obtained from some of the terms in the second input-output equation:

$$\begin{cases} \beta_i \delta = b_7 b_8, & (3.7) \\ \omega\beta_v\delta - \theta\beta_i\delta = \omega b_6 b_8 - b_{11} b_7 b_8, & (3.8) \\ -\omega^2\beta_v\delta + 2\omega\theta\beta_i\delta + q\theta\beta_i\delta = -\omega^2 b_6 b_8 + 2\omega b_{11} b_7 b_8 + q b_{11} b_7 b_8, & (3.9) \\ \delta(-\omega\beta_v + \theta\delta) = b_8(-\omega b_6 + b_{11} b_8), & (3.10) \\ -3\gamma\omega + 3\omega m + 2\omega\beta_i + 3\theta n = -3 b_{10}\omega + 3\omega m + 2\omega b_7 + 3 b_{11} n, & (3.11) \\ \omega D + 6\theta mn = \omega D_1 + 6\theta mn, & (3.12) \\ \omega H - 6\theta^2 n\beta_i = \omega H_1 - 6n b_{11}^2 b_7, & (3.13) \\ \omega Q - 6\theta^2 mn\beta_i = \omega Q_1 - 6 b_{11}^2 mn b_7. & (3.14) \end{cases}$$

Of which,

$$
\begin{aligned}
D &= \omega\alpha - 6\gamma m + \omega\beta_i\delta + \omega\delta\mu + 4m\beta_i + 6\theta mn,\\
D_1 &= \omega b_5 - 6b_{10}m + \omega b_7 b_8 + \omega b_8 b_9 + 4mb_9,\\
H &= -3p\theta\beta_i - 3\gamma\omega\beta_v + 6\gamma\theta\beta_i + 4\omega\beta_v\beta_i + 3\theta n\beta_v - 4\theta\beta_i^2,\\
H_1 &= -3b_{12}b_{11}b_7 - 3\omega b_{10}b_6 + 6b_{10}b_{11}b_7 + 4\omega b_6 b_7 + 3nb_{11}b_6 - 4b_{11}b_7^2,\\
Q &= -3pm\theta\beta_i - 3m\gamma\omega\beta_v + 6m\gamma\beta_i + \omega^2\alpha\beta_v + \omega^2\beta_v\beta_i\delta + \omega^2\beta_v\delta\mu - \omega\theta\alpha\beta_i\\
&\quad - \omega\theta\beta_i^2\delta - \omega\theta\beta_i\delta\mu + 4\omega m\beta_i\beta_v + 3\theta mn\beta_v - 4\theta m\beta_i^2,\\
Q_1 &= -3mb_{12}b_{11}b_7 - 3m\omega b_{10}b_6 + 6mb_{10}b_7 + \omega^2 b_5 b_6 + \omega^2 b_6 b_7 b_8 + \omega^2 b_6 b_8 b_9\\
&\quad - \omega b_{11}b_5 b_7 + 4\omega m b_6 b_7 + 3mnb_{11}b_6 - 4mb_{11}b_7^2.
\end{aligned}
$$

From (3.7)–(3.9):

$$\theta = \frac{(\omega b_6 b_8 - b_{11}b_7 b_8)\omega - \omega^2 b_6 b_8 + 2\omega b_{11}b_7 a_8 + qb_{11}b_7 b_8}{b_7 b_8(\omega + q)} = b_{11}, \tag{3.15}$$

$$\beta_v\delta = \frac{(\omega b_6 b_8 - b_{11}b_7 b_8)(2\omega + q) - \omega^2 b_6 b_8 + 2\omega b_{11}b_7 b_8 + qb_{11}b_7 b_8}{\omega(\omega + q)} = b_6 b_8. \tag{3.16}$$

Substituting (3.15), (3.16) into (3.10), we get $\delta = b_8$. Then use $(b_5, b_6, b_7, b_9, b_{10}, b_{12})$ to solve for $(\alpha, \beta_v, \beta_i, \mu, \gamma, p)$ to get $\alpha = b_5$, $\beta_v = b_6$, $\beta_i = b_7$, $\mu = b_9$, $\gamma = b_{10}$, $p = b_{12}$. So substituting (3.15), (3.16) into (3.10) gives $\delta = b_8$, and then use $(b_5, b_6, b_7, b_9, b_{10}, b_{12})$ to solve for $(\alpha, \beta_v, \beta_i, \mu, \gamma, p)$ to get $\alpha = b_5$, $\beta_v = b_6$, $\beta_i = b_7$, $\mu = b_9$, $\gamma = b_{10}$, $p = b_{12}$. So, model (2.1) is structurally identifiable. □

3.2. Model selection

We often want to use simple models to describe the process of virus infection, but worry that 'simplicity' ignores some important factors, which requires modelers to select multiple models with different complexity. Choosing a model that can best reflect the actual infection process of the virus from the limited clinical data is of great significance. Based on the mechanism of SARS-COV-2 infection in host cells, we regard the system (2.1) as Full model. Under different biological assumptions, the model nested four different sub-models (see Table 2, and these according dynamic models can be found in the Appendix). We fit all five models using two sets of clinical statistics, 'lymphocyte count' and 'viral load', for COVID-19 patients at Yale New Haven Hospital, and then used the fit results to evaluate the different models to determine which model was more appropriate for the clinical data set.

Since some parameters of the model have been reasonably estimated in other literature, these parameters will be fixed as constants (see Table 3). For other unknown parameters in the model, we use the affine invariant set Monte Carlo algorithm(GWMCMC) [12] for parameter estimation. This algorithm is superior to the Metropolis-Hastings (M-H) algorithm and random walk M-H algorithm [8], especially when the parameters cannot be identified. Because the original Markov Chain Monte Carlo (MCMC) algorithms are used to approximate the posterior distribution of parameters by randomly sampling the parameter space [13]. The advantage of the GWMCMC algorithm in this paper is using multiple walkers. The position of walkers is updated based on the current situation of all other walkers

Table 2. Models and corresponding assumptions.

Model	Assumptions	Brief summary
Model 1	Full Model	Considering the infectivity of all warehouses
Model 2	$\beta_i = 0$	Not considering the attack of infected cells on target cells
Model 3	$\omega = 0$	Not considering the attack of infected cells on T lymphocytes
Model 4	$\beta_i = 0$, $\omega = 0$	Not considering the attack of infected cells on target cells and the interaction between infected cells and T lymphocytes
Model 5	$\beta_v = 0$, $\omega = 0$	Not considering the attack of SARS-COV-2 on target cells and the interaction between infected cells and lymphocytes

Table 3. Constant value of parameters and initial value of variables.

Parameters	Value	Source	Initial Condition	Value	Source
μ	0.001	[22]	C^0	6e4	[38]
γ	0.088	[17]	I^0	0.01	assumption
m	0.325	[17]	T^0	6e5	[25]
p	7.5	[38]	V^0	0.01	assumption

so that the fitting value is more accurate. See the literature for details on the algorithm [12, 34, 35].

Assuming that the priors of all parameters obey uniform distribution, the posterior distribution of all estimated parameters is generated by the GWMCMC algorithm. From these distributions, we obtain the best fitting value of the parameters to be estimated in the Full Model and its 95% confidence interval (95%CI), respectively. The results are shown in Table 4.

Table 4. Model parameters of the Full Model and its estimations.

Parameter	α	β_v	β_i	δ
Best-fit value	1.69e3	0.11	0.09	1.03
95%CI	(1.58e3, 1.70e3)	(0.10, 0.15)	(0.08, 0.14)	(0.91, 1.20)
Parameter	ω	θ	n	q
Best-fit value	5.47e-6	4.08e-6	757.21	2.16e-5
95%CI	(5.36e-6, 6.98e-6)	(4.02e-5, 5.11e-5)	(6.08e2, 7.78e2)	(1.36e-5, 2.42e-5)

The parameter estimations and 95% confidence interval of the remaining four sub-models are shown in Table 5.

Table 4 and Table 5 give the best parameter fitting value and 95% confidence interval of the five models, respectively. Based on the above parameter estimates, we can compare the five models to select the best model. Commonly used model selection criteria include the akaike information criterion (AIC) [2] and the Bayesian Information Criterion (BIC) [41], as well as the variant (AICc) of AIC, generally, the minimum value of AIC, BIC or AICc is the best model. In the likelihood function

Table 5. Model parameters and their estimations of the four submodels.

Par	Model 2		Model 3		Model 4		Model 5	
	Best-fit value	95%CI	Best-fit value	95%CI	Best-fit value	95%CI	Best-fit value	95%CI
α	2.64e3	(2.5, 2.69)e3	920.84	(0.89, 9.29)e2	926.39	(9.00, 1.09)e2	867.80	(0.85,1.04)e3
β_v	0.11	(0.08, 0.19)	0.08	(0.02, 0.09)	1.14	(1.01, 1.79)	NA	NA
β_i	NA	NA	0.29	(0.14, 1.30)	NA	NA	1.47	(1.03,2.00)
δ	0.83	(0.62, 1.50)	0.43	(0.18, 1.17)	2.52	(2.23, 3.93)	1.05	(0.74,1.42)
ω	6.44e-6	(5.72, 8.04)e-5	NA	NA	NA	NA	NA	NA
θ	4.42e-6	(3.95, 5.30)e-5	5.19e-6	(4.77, 5.67)e-6	5.32e-6	(5.00, 5.81)e-6	4.76e-6	(4.66,5.13)e-6
n	727.54	(6.02, 7.97)e2	91.36	(0.90, 1.19)e2	57.52	(54.86, 79.91)	69.63	(61.40,69.99)
q	3.90e-5	(2.64, 4.52)e-5	1.75e-4	(1.67, 2.59)e-4	9.89e-5	(0.87, 1.65)e-4	1.08e-4	(0.89,1.22)e-4

framework, these criteria can be written as follows:

$$\begin{aligned}\mathrm{AICc} &= \mathrm{AIC} + \frac{2K(K+1)}{n-K-1} = -2\ln L(\theta_{MLE}) + 2K;\\ \mathrm{BIC} &= -2\ln L(\theta_{MLE}) + K\ln(n);\end{aligned}$$

where L is the maximum likelihood function obtained when the best fit is θ_{MLE}, K denotes the number of parameters to be estimated in the model, and n denotes the sample size, i.e. the number of observed data at the corresponding time point. AIC is more suitable when $K < \frac{n}{40}$, i.e.when the number of time points corresponds to a larger number of parameters; AICc is more suitable when $K > \frac{n}{40}$. Also, as the sample size increases, AICc converges to AIC, and it is often recommended to use AICc [3].

Table 6. List of AICc and BIC for different models.

Model	Number of fitted pars	AICc	BIC
Model 1	8	353.35	407.81
Model 2	7	342.15	401.66
Model 3	7	337.43	396.94
Model 4	6	323.18	394.75
Model 5	6	325.77	396.34

According to the estimated parameter values of all models in Table 4 and Table 5, Table 6 gives the AICc value and BIC value of all models. We know that the model with the smallest value of AICc (BIC) is the best. As seen in Table 6, the AICc value for Model 4 is the smallest of all the models, and its BIC value also confirms this result. Therefore, Model 4 was chosen as the best model. This suggests that the factors 'attack of infected cells on target cells' and 'interaction between infected cells and lymphocytes' are not essential. Although they can provide a more detailed picture of the COVID-19 infection process.

Fig 2 shows the fitting results of Model 4 to the lymphocyte count (left panel) and the viral load (right panel). The initial moment of the fitting (time 0) is when

the virus invades the individual. As seen in Fig 2 (a), there is a decline in T lymphocytes in the early stages of infection establishment. As the infection builds up, the T-lymphocyte levels gradually increase, but returning to the initial levels is difficult. During the incubation period, the viral load was low until the 10th day of infection, when it increased rapidly and peaked on the 13th day. Subsequently, viral load levels began to decrease in response to the drug.

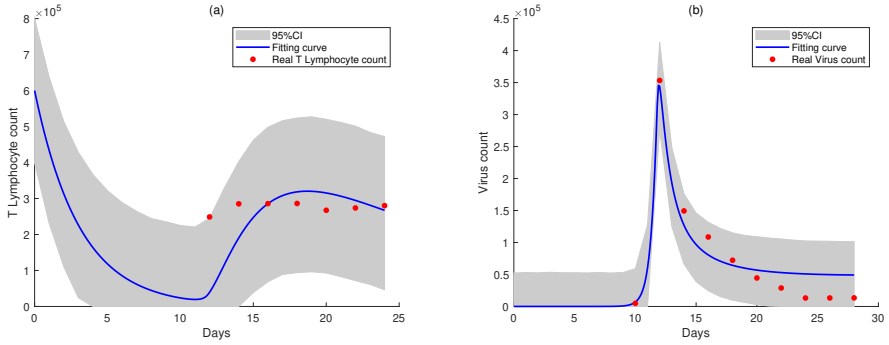

Figure 2. Fitting results. (a) Fitting curve of the T lymphocyte and its 95% confidence interval. (b) Fitting curve of the COVID-19 virus load and its 95% confidence interval.

3.3. Practical identifiability analysis

The analysis in Section 3.1 clearly shows that Model 4 is structurally identifiable. This is a necessary condition to be able to obtain reliable parameter estimates from the data. However, structurally identifiable parameters are not necessarily recognizable, which may be related to the fact that the data are not ideal.

The actual data collected is often noisy. A key problem when fitting a model to data is the effects of noise on the parameter estimates and model identifiability. To investigate the effect of different noise distributions on parameter estimates in real systems, we tested the following common error models through simulations: Poisson, Gaussian (standard deviation equal to 10% of the variance), Negative binomial (because of the overdispersion, a negative binomial distribution with a variance equal to 5 times the mean and a variance equal to 50 times the mean will be considered respectively). We took the best parameter values obtained in Model 4 as the true parameter values, added each of the four noises to the model output to generate 100 data, and re-estimated the model parameters; the results of the parameter estimation are shown in Table 7.

The second column in Table 7 shows the values of the best-fit parameters obtained from the Model 4 fit, and the next four columns represent the average of the best parameter values when the noise is Poisson, Gaussian (standard deviation equal to 10% of the variance), Negative binomial with variance 5 times the mean and Negative binomial with variance 50 times the mean, obtained by 100 replications of the simulation with maximum likelihood. We found that the estimation results were similar for all noise distributions, and all parameter estimation were close to the true parameter values, suggesting that the different noise distributions had little effect on the parameter fitting results. Further consideration was given to the practical identifiability of the Model 4 parameters under Poisson noise: the

Table 7. Parameter estimates.

Parameters	True value	Poisson	Gaussian (10%CV)	Negative binomial (5 times)	Negative binomial (50 times)
α	926.3887	926.4645	923.0535	925.9913	932.2961
β_v	1.1427	1.1420	1.1151	1.1496	1.1409
δ	2.5199	2.5192	2.4825	2.5356	2.5101
θ	5.3242e-6	5.3237e-6	5.3421e-6	5.3229e-6	5.3262e-6
n	57.5238	57.5640	59.8367	57.5092	57.3546
q	9.8904e-5	9.8989e-5	1.0409e-4	9.8802e-5	9.8722e-5

simulation was repeated 100 times using GWMCMC to obtain the best estimate of the parameter values and generate a scatter plot of the parameter estimates as shown in Fig 3.

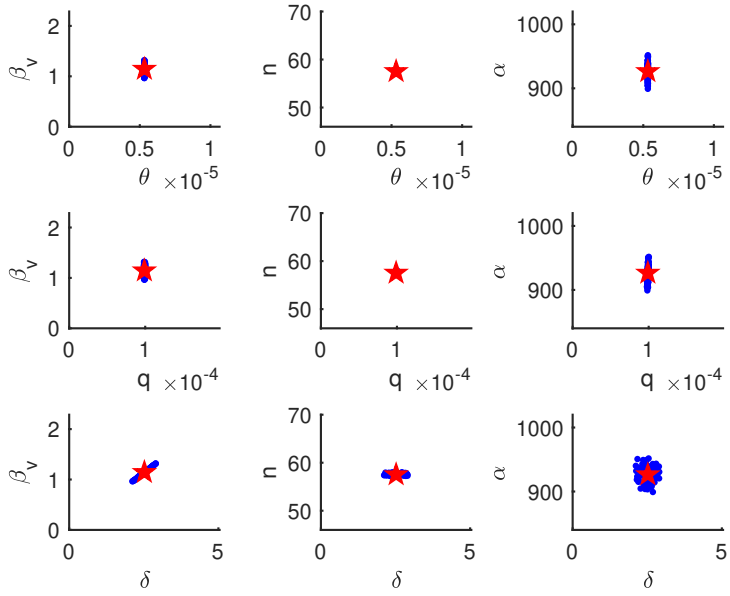

Figure 3. Scatterplots showing parameter estimates for 100 simulated data sets use Affine Invariant Ensemble Markov chain Monte Carlo algorithm for Poisson noise. Beat-fit parameters (indicated by red stars) are as given in Table 5. Note that the parameters θ,q and n can be identified. The parameter δ, β_v, α cannot be identified.

Fig. 3 with red pentagrams is the best fit for Model 4 in Table 5, and the blue points are the parameter estimation obtained from 100 iterations of the simulation. The first row of the figure plots the relationship between the parameter θ and the parameters β_v, n, and α, respectively, showing that the estimates of the parameter θ vary very little, all within a small range around the best-fit value, and we claim that the parameter θ is practically identifiable. Similarly, the parameter q is also practically identifiable. The third-row plots the relationship between the parameter δ and the parameters β_v, n and α, respectively, and shows that the estimated value of the parameter δ varies considerably, so the parameter δ is said to be practically

unidentifiable. In terms of columns, the first column shows the relationship between the parameter β_v and the parameters θ, q, and δ, respectively. The estimation of the parameter β_v varies considerably, so the parameter β_v is practically unidentifiable. The second column shows the relationship between the parameter n and the parameters θ, q, and δ, respectively. The estimation of the parameter n varies relatively little, so the parameter n is practically identifiable. The estimation of α is highly variable. So the parameter α is practically unidentifiable.

To verify our conclusions related to the practical identifiability of the parameters from the scatter plot, we applied Monte Carlo simulations further to analyze the practical identifiability of the parameters in Model 4. Monte Carlo simulations have also been used extensively for the practical identifiability of ODE models. We generated 100 synthetic data sets using the true parameter set $\boldsymbol{p}_0$ and added noise. The results of fitting the true parameter set of the model are shown in Table 8. In general, the Monte Carlo simulation procedure [28] can be summarised as follows.

(1) Solve the Model 4 numerically with the true parameters $\boldsymbol{p}_0$ and obtain the output vector $g(\boldsymbol{x}(t_i), \boldsymbol{p}_0)$ at the discrete data time points $\{t_k\}_{k=1}^n$.

(2) Generate $M = 1000$ simulated data from the statistical model with a given measurement error structure. Data sets are drawn from a normal distribution whose mean is the output vector obtained in step 1 and the standard deviation is the $\sigma_0\%$ of the mean.

(3) Fit the Model 4 to each of the M simulated data sets by the GWMCMC algorithm. Get the Parameter Set of Maximum Likelihood Function $\hat{\boldsymbol{p}}_j, j = 1, 2, \cdots, M$.

(4) Calculate the average relative estimation error (ARE) [26] for each parameter in the set $\boldsymbol{p}$ by

$$ARE(p^{(k)}) = 100\% \frac{1}{M} \sum_{j=1}^{M} \frac{|p_0^k - p_j^k|}{p_0^k},$$

where $p^{(k)}$ is the kth parameter in the set $\boldsymbol{p}$, $p_0^{(k)}$ is the kth parameter in the true parameter set $\boldsymbol{p}_0$ and $\hat{\boldsymbol{p}}$, and $p_j^{(k)}$ is the kth parameter in the estimated parameter set $\boldsymbol{p}_j$.

(5) Repeat steps 1 through 4 with increasing level of noise, that is take $\sigma_0\% = 0, 5, 10, 20\%$.

The calculated ARE values allow a discussion of the identifiability of the parameters to be fitted in Model 4. Since Model 4 is globally identifiable in structure, when $\sigma_0 = 0$, the largest ARE value of each parameter is 0.4456% (corresponding to the parameter β_v, see Table 8). All six parameters are well identified when the data are free of noise. This indicates that the practical identifiability obtained by Monte Carlo simulation is consistent with the structural identifiability analysis performed by differential algebraic methods. From the above table we can see that as the noise in the data increases, the value of ARE also increases. When the error level increases to 5%, the relative errors of θ, n and q remain below the error level of 5%, but the relative errors of α, β_v and δ are above the error level. The results are the same when the error increases to 10% and 20%. It is noteworthy that even at the error level of 20%, the relative error of θ is still below 5%, indicating that θ is very insensitive to the noise in the data.

We assume that a parameter is practically identifiable if its ARE value is less than the measurement error level. Thus, based on the Monte Carlo simulation results shown in the table above, we obtain that the parameters θ, n and q are

Table 8. Parameters identifiability analysis of the model 4. The output vector corresponding to the real parameters is taken as the mean value, and the σ_0 % of the mean value is taken as the standard deviation to generate 100 simulation data sets subject to normal distribution, which are brought into Model 4 to obtain new estimation parameters through GWMCMC simulation. The ARE calculation is based on 100 simulation runs with measurement error levels of $\sigma_0 = 0, 5, 10, 20\%$.

Parameters	**ARE** $\sigma_0 = 0\%$	**ARE** $\sigma_0 = 5\%$	**ARE** $\sigma_0 = 10\%$	**ARE** $\sigma_0 = 20\%$
α	0.2183 %	5.5726 %	14.7645%	29.1059%
β_v	0.4456%	6.7799%	11.2064%	21.6982%
δ	0.4437%	7.3191%	11.1876%	20.7077%
θ	0.0378%	1.5388%	2.8112%	5.6269%
n	0.0621%	2.3115%	6.1430%	17.0549%
q	0.0740%	1.6609%	6.1152%	13.0296%

practically identifiable, while the parameters α, β_v, and δ are practically unidentifiable. This is also in general agreement with the practical identifiability results of the parameters in the scatter plot corresponding to the data under Poisson noise.

Although it is not straightforward to conclude that the parameters α, β_v, δ are practically identifiable, due to the superiority of the GWMCMC algorithm in the presence of unidentifiable parameters and the degree of aggregation of the parameters, it can be assumed that the estimation of the parameters α, β_v, δ are to some extent reliable.

To further illustrate the effect of non-identifiability on model prediction, based on the approximately linear relationship between the parameters δ and β_v in Fig. 4 and the circular dispersion between the parameters δ and α, we take four different combinations: $\min(\delta, \beta_v) = (2.1178,\ 0.9643)$, $\max(\delta, \beta_v) = (2.9020,\ 1.3192)$, $\min(\delta, \alpha) = (2.1178,\ 899.2132)$ and $\max(\delta, \alpha) = (2.9020,\ 951.6838)$. The maximum and minimum values are taken from the 100 sets of estimated parameter values for δ, β_v and α obtained by 100 iterations of the simulation. Under these four different combinations, we plotted Fig.4.

The blue curve in Fig. 4 shows the trend of virus load over time for the optimal parameters. The green and red curves in the left panel are the best curves when δ and β_v have both maximum ($\max(\delta, \beta_v)$) and minimum ($\min(\delta, \beta_v)$) values respectively. It can be seen that the combination of these two parameters has minimal effect on the viral load. The green and red curves on the right are the optimal curves when δ and α have a maximum ($\max(\delta, \alpha)$) and a minimum ($\min(\delta, \alpha)$), respectively. We found that although the time to peak viral load is delayed and advanced by one day in both cases, the time to stabilization and the value of stabilization are almost identical, suggesting to some extent that the parameter values we estimated earlier are robust.

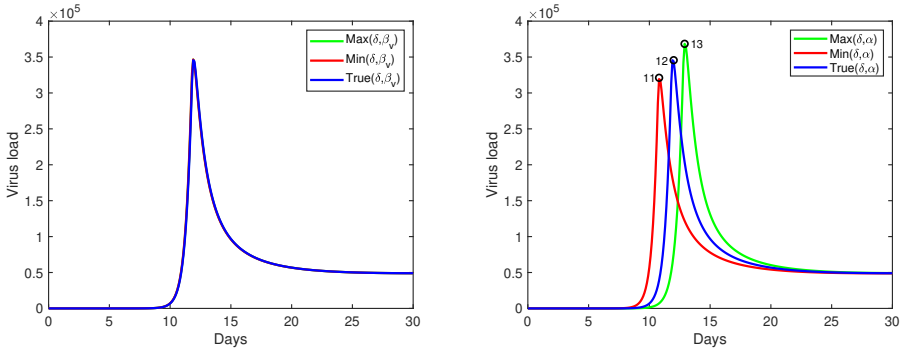

Figure 4. The curve of viral load under different combinations of parameters. The black dots correspond to the peak of curves.

3.4. Simulations

The main drug used in the treatment of COVID-19 is hydroxychloroquine (HCQ), which works by inhibiting the binding of SARS-COV-2 to cellular receptors and the replication of nucleic acids, affecting the release of the virus. To investigate the effect of HCQ drug effects on patients' treatment progress, we made different assumptions about two parameters of HCQ drug effects, n and β_v, assuming that other parameters remained unchanged as column 6 in Table 5, and the simulation results are shown in Fig 5 (left). Assuming that without drug intervention or weakened drug effect, the ability of virus binding cell receptor and replication ability reach 1.25 times (blue curve) or 1.5 times (purple curve) of the original, then the peak number of newly infected target cells caused by virus attack will greatly increase, and the time to reach the peak will also be advanced with the decrease of drug control. If the drug is assumed to be more potent than it is, for example, if a factor of 0.5 (red suppresses the ability of the virus to bind cellular receptors and replicate curve) and 0.75 (green curve), respectively, the peak number of newly infected target cells is significantly reduced and the time to peak is significantly delayed. As a result, an appropriate increase in the effectiveness of HCQ drugs can effectively control the disease in infected patients, allowing more time for treatment and reducing mortality.

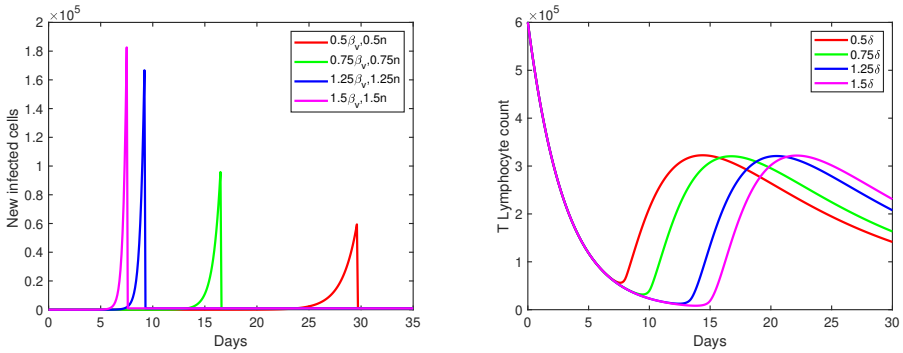

Figure 5. The figure on the left shows the number of cells infected by the virus under different combinations of parameters. The figure on the right shows the number of T lymphocytes under different Hill coefficients.

During SARS-COV-2 infection, the virus rapidly replicates and activates immune cells to release many inflammatory cytokines, which is easy to induce a 'cytokine storm'. At this time, the number of T lymphocytes decreases [49]. This is the main reason for the death of patients with COVID-19 [16]. At the same time, some studies have found that the RBD domain of S protein and '47D11' antibody has affinity, which may lead to increased infection ability citehuman. The decrease of Hill coefficient δ will lead to the enhancement of infection ability, that is, the number of newly infected cells increases, so how does the number of lymphocytes that play an important role in the infection change to explore this problem, we make several different assumptions about δ. The results are shown in Fig. 5 (right). It can be seen that as δ increases, the number of T lymphocytes declines more, and when δ is large, T lymphocytes will maintain a low level for a long period, which shows that the increase of δ is more likely to cause cytokine storm. Therefore, the treatment of immunosuppression during the hospitalization of patients is also worthy of attention.

4. Stability of equilibrium and Hopf bifurcations of the Full model

Next, we study the qualitative results of the system. We conclude that the equilibrium point of the system (2.1) and its stability are as follows. Although this conclusion is for the system (2.1), it also holds for all the other four subsystems.

Theorem 4.1. *Let*

$$\mathcal{R}_0 = \frac{\alpha(p\beta_i + n\beta_v)}{pr(\mu + \delta\alpha)}, \qquad \mathcal{R}_1 = 1 + \frac{\gamma p\big[\mu + (\mathcal{R}_0 - 1)(\mu + \delta\alpha)\big]}{\alpha}.$$

Then

1. *The system* (2.1) *always has a disease-free equilibrium* $E_0\left(\frac{\alpha}{\mu}, 0, 0, 0\right)$. *When* $\mathcal{R}_0 < 1$, E_0 *is locally asymptotically stable.*
2. *When* $\mathcal{R}_0 > 1$, E_0 *becomes unstable. At this point the system* (2.1) *has a boundary equilibrium point*

$$E_b\left(\frac{\alpha}{(\mu + \delta\alpha)(\mathcal{R}_0 - 1) + \mu}, \frac{\alpha(\mu + \delta\alpha)(\mathcal{R}_0 - 1)}{\gamma[(\mu + \delta\alpha)(\mathcal{R}_0 - 1) + \mu]}, 0, \frac{\alpha n(\mu + \delta\alpha)(\mathcal{R}_0 - 1)}{\gamma p[(\mu + \delta\alpha)(R_0 - 1) + \mu]}\right).$$

   When $1 < \mathcal{R}_0 < \mathcal{R}_1$, *the boundary equilibrium* E_b *is locally asymptotically stable;*

3. *When* $\mathcal{R}_0 > \mathcal{R}_1 > 1$, E_b *becomes unstable and the system* (2.1) *has a positive equilibrium point* $E^*(C^*, I^*, T^*, V^*)$.

Proof. We first prove the existence of a positive equilibrium point. Let the right-hand side of the model (2.1) be equal to zero and rewrite the system of equations

as:

$$\begin{cases} \alpha = \dfrac{(\beta_v V + \beta_i I)C}{1+\delta C} + \mu C, & (4.1) \\ \dfrac{(\beta_v V + \beta_i I)C}{1+\delta C} = (\omega T + \gamma) I, & (4.2) \\ \omega I + \theta V = m, & (4.3) \\ nI = (qT + p)V. & (4.4) \end{cases}$$

Substituting the relationship between I and V in (4.4) into (4.2) gives

$$\frac{[\beta_i(qT+p) + n\beta_v]C}{1+\delta C} = (\omega T + \gamma)(qT + p). \tag{4.5}$$

Solve the first expression of C by (4.5):

$$C = f_1(T) = \frac{(\omega T + \gamma)(qT + p)}{\beta_i(qT+p) + n\beta_v - \delta(\omega T + \gamma)(qT + p)}. \tag{4.6}$$

Solving from (4.3) and (4.4) gives

$$I = \frac{m(qT + p)}{n\theta + \omega(qT + p)}. \tag{4.7}$$

From (4.1) and (4.2) we have

$$\alpha = \mu C + (\omega T + \gamma) I. \tag{4.8}$$

The second expression for C is obtained by combining (4.7) and (4.8):

$$C = f_2(T) = \frac{1}{\mu}\left[\alpha - \frac{m(\omega T + \gamma)(qT + p)}{n\theta + \omega(qT + p)}\right]. \tag{4.9}$$

Next, determine the intersection of the functions (4.6) and (4.9) by their monotonicity. From (4.6) we have

$$\begin{aligned} f_1(T)' &= -\frac{1}{\mu}\frac{m[\omega(qT+p) + (\omega T + \gamma)q][n\theta + \omega(qT+p)] - m\omega q(\omega T + \gamma)(qT + p)}{[n\theta + \omega(qT+p)]^2} \\ &= -\frac{1}{\mu}\frac{m[\omega n\theta(qT+p) + \omega^2(qT+p)^2 + n\theta q(\omega T + \gamma)]}{[n\theta + \omega(qT+p)]^2}. \end{aligned}$$

Clearly, $f_1(T)' < 0$, so $f_1(T)$ is monotonically decreasing at $(0,+\infty)$ and

$$f_1(0) = \frac{1}{\mu}\left(\alpha - \frac{m\gamma p}{n\theta + \omega p}\right).$$

From (4.9), we have

$$f_2(T)' = \frac{\omega\beta_i(qT+p)^2 + n\beta_v\omega(qT+p) + n\beta q(\omega T + \gamma)}{[\beta_i(qT+p) + n\beta_v - \delta(\omega T + \gamma)(qT + p)]^2}.$$

Clearly, $f_2(T)' > 0$, then $f_2(T)$ is monotonically increasing at $(0,+\infty)$ and

$$f_2(0) = \frac{\gamma p}{p\beta_i + n\beta_v - \delta\gamma p}.$$

At this point, $f_2(0) > 0$. If $f_1(0) > f_2(0)$, then $f_1(T)$ and $f_2(T)$ must have an intersection (T^*, C^*) at $(0,+\infty)$. Next, explore under what circumstances, $f_1(0) > f_2(0)$. To make $f_1(0) > f_2(0)$, i.e.

$$\alpha(p\beta_i + n\beta_v)(n\theta + \omega p) > \gamma p(\mu + \delta\alpha)(n\theta + \omega p) + m\gamma p(p\beta_i + n\beta_v - \delta\gamma p).$$

Namely

$$\alpha(p\beta_i + n\beta_v) > \gamma p(\mu + \delta\alpha) + \frac{m\gamma p(p\beta_i + n\beta_v - \delta\gamma p)}{(n\theta + \omega p)}.$$

That is:

$$\mathcal{R}_0 = \frac{\alpha(p\beta_i + n\beta_v)}{\gamma p(\mu + v\delta\alpha)} > 1 + \frac{m(p\beta_i + n\beta_v - \delta\gamma p)}{(\mu + \delta\alpha)(n\theta + \omega p)} = 1 + \frac{\gamma p\big[\mu + (\mathcal{R}_0 - 1)(\mu + \delta\alpha)\big]}{\alpha} = \mathcal{R}_1.$$

Therefore $\mathcal{R}_0 > \mathcal{R}_1$ is equivalent to $f_1(0) > f_2(0)$, at which point $f_1(T)$ and $f_2(T)$ have an intersection (T^*, C^*) on $(0,+\infty)$. According to (4.4) and (4.7), if $T^* > 0$, $C^* > 0$, then there must be $I^* > 0$, $V^* > 0$. In summary, at $\mathcal{R}_0 > \mathcal{R}_1$, a positive equilibrium point exists. Secondly, the local stability of the disease-free equilibrium point E_0 is proved. The corresponding characteristic equation of the system (2.1) at E_0 is easily known as:

$$(\lambda + m)(\lambda + \mu)[(-\alpha\delta - \mu)\lambda^2 + (\alpha\beta_i - (p + \gamma)(\alpha\delta + \mu))\lambda + \big(p\gamma(\alpha\delta + \mu) - \alpha(\beta_i p + \beta_v n)\big)(\alpha\delta + \mu)] = 0.$$

The eigenvalues of E_0 are easily obtained as:

$$\lambda_1 = -m, \lambda_2 = -\mu, \lambda_3 = \frac{\Psi_1 + \sqrt{\Psi_1^2 - 4\Psi_2}}{2(\alpha\delta + \mu)}, \lambda_4 = \frac{\Psi_1 - \sqrt{\Psi_1^2 - 4\Psi_2}}{2(\alpha\delta + \mu)}.$$

Of which,

$$\Psi_1 = -\left[\frac{(1 - R_0) + \alpha n\beta_v}{p} + p(\alpha\delta + \mu)\right], \qquad \Psi_2 = p\gamma(1 - R_0)(\alpha\delta + \mu)^2.$$

When $\mathcal{R}_0 < 1$, $\Psi_1 < 0$, $\Psi_2 > 0$, the disease-free equilibrium point E_0 of the system (2.1) is thus locally asymptotically stable. When $\mathcal{R}_0 > 1$, $\Psi_2 < 0$ and $\lambda_{3,4}$ has a positive value, E_0 is unstable. Next, we prove the local stability of the boundary equilibrium point E_b. The characteristic equation of the system (2.1) at the boundary equilibrium point E_b can be calculated as

$$(\lambda^3 + K_2\lambda^2 + K_1\lambda + K_0)\left(\lambda + \frac{\alpha(\mathcal{R}_1 - \mathcal{R}_0)(\alpha\delta + \mu)}{p\gamma\big[(\mathcal{R}_0 - 1)(\alpha\delta + \mu) + \mu\big]}\right) = 0.$$

Of which,

$$
\begin{aligned}
K_0 &= \frac{p^2\gamma^2\big[(\mathcal{R}_0-1)^2(\mu+\delta\alpha)^2+\mu(\mathcal{R}_0-1)(\mu+\delta\alpha)\big]}{p\beta_i+n\beta_v},\\
K_1 &= \frac{p\mu(p\beta_i+n\beta_v)+\gamma\mu n\beta_v}{p\beta_i+n\beta_v},\\
K_2 &= \frac{\alpha B^2+p\gamma\big[p^2\beta_i+n\beta_v\gamma+p\gamma\big(n\beta_v+\gamma\delta\mu)\big]}{p\gamma(p\beta_i+n\beta_v)},\\
B &= n\beta_v+p(\beta_i-\gamma\delta).
\end{aligned}
$$

It is easy to see that when $1<\mathcal{R}_0<\mathcal{R}_1$, we have $\lambda_1=-\dfrac{\alpha(\mathcal{R}_1-\mathcal{R}_0)(\alpha\delta+\mu)}{p\gamma\big[(\mathcal{R}_0-1)(\alpha\delta+\mu)+\mu\big]}<0$. At this point, $K_0>0, K_1>0$, $K_2>0$. Also because

$$
\begin{aligned}
K_1K_2-K_0 =& \frac{p^3\gamma(p\beta_i+n\beta_v)\big[\alpha B^2+p\gamma^2\mu(p\delta-\beta_i+\gamma\delta)+B\mu\gamma^2\big]}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
&+\frac{p\gamma^2(n\beta_v+p\delta\mu)\big[\alpha pB^2+\alpha\gamma B^2+p^2\gamma^2\mu(p\delta-\beta_i+\gamma\delta)\big]}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
&+\frac{\alpha B^2\big[\alpha pB^2+\alpha\gamma B^2+p^2\gamma^2\mu(p\delta-\beta_i+\gamma\delta)\big]}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
>&\frac{p^3\gamma(p\beta_i+n\beta_v)(\alpha B^2+p^2\gamma^2\mu\delta+n\beta_v\mu\gamma^2)}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
&+\frac{p\gamma^2(n\beta_v+p\delta\mu)\big[\alpha pB\frac{p\gamma\mu}{\alpha}+\alpha\gamma B\frac{p\gamma\mu}{\alpha}+p^2\gamma^2\mu(p\delta-\beta_i+\gamma\delta)\big]}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
&+\frac{\alpha B^2\big[\alpha pB\frac{p\gamma\mu}{\alpha}+\alpha\gamma B\frac{p\gamma\mu}{\alpha}+p^2\gamma^2\mu(p\delta-\beta_i+\gamma\delta)\big]}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
=&\frac{\big[p^2\gamma^3\mu(p\beta_i+p\delta\mu)+\alpha B^2p\gamma\mu\big]\big[p(p\beta_i+n\beta_v)+\gamma n\beta_v\big]}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
&+\frac{p^3\gamma(p\beta_i+n\beta_v)\big[\alpha B^2+\mu\gamma^2(p^2\delta+n\beta_v)\big]}{p^2\gamma^2(p\beta_i+n\beta_v)^2}\\
>&0,
\end{aligned}
$$

by the Routh-Hurwitz criterion [1], the boundary equilibrium point E_b of the system (2.1) is locally asymptotically stable. When $\mathcal{R}_0>\mathcal{R}_1$, we have $\lambda_1>0$, so the boundary equilibrium point E_b is unstable. □

Next, we turn to consider possible bifurcations which may occur from the disease equilibrium E^*, including static bifurcation and Hopf bifurcation.

For ease of operation, first make a time transformation of the system (2.1)

$d\tau = \dfrac{dt}{1+\delta C}$, we get

$$\begin{cases} \dfrac{dC}{d\tau} = (\alpha - \mu C)(1+\delta C) - \beta_v CV - \beta_i CI, \\ \dfrac{dI}{d\tau} = \beta_v CV + \beta_i CI - (\omega IT + \gamma I)(1+\delta C), \\ \dfrac{dT}{d\tau} = (\omega IT + \theta VT - mT)(1+\delta C), \\ \dfrac{dV}{d\tau} = (nI - qTV - pV)(1+\delta C). \end{cases} \tag{4.10}$$

According to Theorem 4.1, when $\mathcal{R}_0 > \mathcal{R}_1$, there is a positive equilibrium in the system(4.10) :

$$E^*\left(\frac{(\omega T^*+\gamma)(qT^*+p)}{\beta_i(qT^*+p)+n\beta_v-\delta(\omega T^*+\gamma)(qT^*+p)}, \frac{m(qT^*+p)}{n\theta+\omega(qT^*+p)}, T^*, \frac{nm}{n\theta+\omega(qT^*+p)}\right),$$

where T^* is given in the form of the system parameter δ. The relationship between the two is determined by the following implicit functions on T^* and δ:

$$\begin{aligned} F(T^*,\delta) =& \big(1.00000000\times10^{-26}T^{*5}\delta - (1.00000000\times10^{-21} - 7.53017560\times10^{-17}\delta)T^{*4} \\ & - (0.05655160 - 0.00664649\delta)T^{*2} - (568729.51338458 + 10033.92167360\delta)T^* \\ & - (1.50300877\times10^{-11} - 1.51607276\times10^{-9}\delta)T^{*3} + 7.96595012\times10^{11} \\ & - 9.33426912\times10^{8}\big)(3751.00000000 + 1.00000000\times10^{-7}T^*)^{-1} \\ & \times(-5.00000000\times10^{-8}T^* - 375.50000000 + 5.00000000\times10^{-13}\delta T^{*2} \\ & + 5.04400000\times10^{-6}\delta T^* + 0.44000000\delta)^{-2} \\ =& 0. \end{aligned} \tag{4.11}$$

The rational numbers given in this equation are obtained by symbolic calculation, except for the branching parameter δ, the other parameters in the system are given in Table 9. Next, consider the stability at the positive equilibrium point E^*. The Jacobian matrix of the system at E^* is:

$$\mathrm{J(E^*)} = \begin{pmatrix} \delta(\alpha - 2\mu C^*) - \mu - \Sigma & -\beta_i C^* & 0 & -\beta_v C^* \\ \Sigma - \delta(\omega I^* T^* + \gamma I^*) & \beta_i C^* - (\gamma + \omega T^*)\Phi & -\omega I^*\Phi & \beta_v C^* \\ \Omega T^*\delta & \omega T^*\Phi & \Omega\Phi & \theta T^*\Phi \\ (nI^* - qT^*V^* - pV^*)\delta & n\Phi & -qV^*\Phi & -(qT^*+p)\Phi \end{pmatrix},$$

of which $\Phi = (1+\delta C^*), \Sigma = (\beta_v + \beta_i), \Omega = (\omega I^* + \theta V^* - m)$. The fourth order characteristic equation at E^* can then be obtained as:

$$P(\lambda, T^*, \delta) = \lambda^4 + a_1(T^*,\delta)\lambda^3 + a_2(T^*,\delta)\lambda^2 + a_3(T^*,\delta)\lambda + a_4(T^*,\delta) = 0, \tag{4.12}$$

where the coefficients $a_1(T^*,\delta), a_2(T^*,\delta), a_3(T^*,\delta)$ and $a_4(T^*,\delta)$ of the characteristic polynomial are shown below:

$$a_1(T^*,\delta) = \big((-8.00000000\times10^{-17}\delta^2 + 2.00000000\times10^{-14}\delta - 5.00000000\times10^{-22})T^{*3}$$

$$
\begin{aligned}
&+ (0.00015040\delta - 3.00160704 \times 10^{-6}\delta^2 - 2.06001500 \times 10^{-9})T^{*2}\\
&+(1515.595341 08\delta - 30.27214080\delta^2 - 31.58770368)T^{*} - 2.64070400 \times 10^{6}\delta^2\\
&- 1.12801734 \times 10^{11} + 2.38577680 \times 10^{9}\delta\big)(3751.00000000\\
&+ 1.00000000 \times 10^{-7}T^{*})^{-1}(-5.00000000 \times 10^{-8}T^{*} - 375.50000000\\
&+ 5.00000000 \times 10^{-13}T^{*2}\delta + 0.44000000\delta + 5.04400000 \times 10^{-6}T^{*}\delta)^{-1},\\
a_2(T^{*},\delta) =&\big((3.02640400 \times 10^{-25}\delta^2 - 2.10000000 \times 10^{-27}\delta)T^{*5}\\
&+(3.12744836 \times 10^{-18}\delta^2 - 2.40643608 \times 10^{-17}\delta + 1.10001000 \times 10^{-22})T^{*4}\\
&+(1.24944608 \times 10^{-9}\delta^2 - 6.23747673 \times 10^{-8}\delta + 2.47842758 \times 10^{-12})T^{*3}\\
&+(13455.95835162\delta^2 - 1.60375888 \times 10^{6}\delta + 4.66107442 \times 10^{7})T^{*}\\
&+(0.013719 16\delta^2 - 0.68834071\delta + 0.01861460)T^{*2}1.00000000 \times 10^{-32}T^{*6}\delta^2\\
&- 9.74601454 \times 10^{11}\delta + 1.07873440 \times 10^{9}\delta^2 + 4.60838455 \times 10^{13}\big)\\
&\times (0.44000000\delta - 5.00000000 \times 10^{-8}T^{*} - 375.50000000\\
&+ 5.00000000 \times 10^{-13}T^{*2}\delta + 5.04400000 \times 10^{-6}T^{*}\delta)^{-1}\\
&\times (3751.00000000 + 1.00000000 \times 10^{-7}T^{*})^{-1},\\
a_3(T^{*},\delta) =&\big(-5.00000000 \times 10^{-54}T^{*9}\delta^2 + (1.00000000 \times 10^{-48}\delta\\
&- 2.62801320 \times 10^{-43}\delta^2)T^{*8} + (6.00801760 \times 10^{-38}\delta\\
&- 3.10695188 \times 10^{-33}\delta^2 - 5.00005000 \times 10^{-44})T^{*7}\\
&- (3.38005693 \times 10^{-33} - 1.01603807 \times 10^{-27}\delta + 1.06717078 \times 10^{-23}\delta^2)T^{*6}\\
&- (7.33636936 \times 10^{-23} + 3.21082378 \times 10^{-16}\delta^2 - 6.79582232 \times 10^{-18}\delta)T^{*5}\\
&- (7.21057166 \times 10^{-13} + 3.26107687 \times 10^{-9}\delta^2 - 1.60458621 \times 10^{-8}\delta)T^{*4}\\
&- (0.01206377 \times \delta^2 - 0.35332145\delta + 0.00335005)T^{*3}\\
&- (9286.50234027\delta^2 - 1.9666331610^{6}\delta + 6.0192718910^{6})T^{*2}\\
&- (8.07908258 \times 10^{8}\delta^2 - 7.88031514 \times 10^{11}\delta + 8.41109631 \times 10^{13})T^{*}\\
&- 7.20850093 \times 10^{12}(\delta - 853.40909091)^{2}\big)(3751.00000000\\
&+ 1.00000000 \times 10^{-7}T^{*})^{-2}(5.00000000 \times 10^{-13}T^{*2}\delta - 5.00000000 \times 10^{-8}T^{*}\\
&- 375.50000000 + 0.44000000\delta + 5.04400000 \times 10^{-6}T^{*}\delta)^{-3},\\
a_4(T^{*},\delta) =&\big(0.00002560T^{*}(2.00669128 \times 10^{-18}T^{*5}\delta^2 + 282138.57065529T^{*2}\delta^2\\
&+ 0.01879761\delta^2T^{*3} - 0.94082215\delta T^{*3} + 476044.61019270T^{*2}\\
&+ 0.00013042T^{*3} + 8.26640926 \times 10^{16}\delta^2 - 7.46839365 \times 10^{19}\delta\\
&+ 1.76370298 \times 10^{-35}T^{*7}\delta^2 + 1.17187502 \times 10^{-45}T^{*8}\delta^2\\
&+7.02413389 \times 10^{14}T^{*} + 3.90625001 \times 10^{-46}T^{*7} + 1.32085172 \times 10^{-24}T^{*5}\\
&+ 1.81879733 \times 10^{-14}T^{*4} + 4.39726564 \times 10^{-35}T^{*6}\\
&- 5.27617532 \times 10^{-40}T^{*7}\delta + 6.66269834 \times 10^{-26}T^{*6}\delta^2\\
&- 1.25813445 \times 10^{-19}T^{*5}\delta - 5.60527757 \times 10^{-10}T^{*4}\delta\\
&+ 3.53110680 \times 10^{21} + 9.64047333 \times 10^{11}T^{*}\delta^2 - 6.22741704 \times 10^{13}T^{*}\delta\\
&- 1.41437616 \times 10^{7}T^{*2}\delta - 3.90624999 \times 10^{-51}T^{*8}\delta
\end{aligned}
$$

$$
\begin{aligned}
&- 1.27732225 \times 10^{-29} T^{*6}\delta) + 2.75779648 \times 10^{-11} T^{*4}\delta^2\big) \\
&\times (-5.00000000 \times 10^{-8} T^* - 375.50000000 \\
&+ 5.00000000 \times 10^{-13} T^{*2}\delta + 5.04400000 \times 10^{-6} T^*\delta \\
&+ 0.44000000\delta)^{-4} (3751.00000000 + 1.00000000 \times 10^{-7} T^*)^{-2}.
\end{aligned}
$$

Based on the characteristic polynomial (4.12), we consider possible bifurcations from E^*, including both static bifurcation and dynamic (Hopf) bifurcations. The static bifurcation occurs when $P(\lambda, T^*, \delta) = 0$ has zero roots (zero eigenvalues). In particular, when $a_4(T^*, \delta) = 0$, the characteristic polynomial (4.12) must have a single zero root, where T^* and δ should satisfy the (4.11). So, $(\delta_t, T_t) \approx (853.40909025, 0)$ is a static bifurcation point. The subscript 't' stands for transcritical bifurcation point. And, at that point,except for the Hurwitz criterion $\Delta_4 = 0$, all three are positive: $\Delta_1 = 1.399616481 \times 10^9$, $\Delta_2 = 2.51563118 \times 10^{24}$, $\Delta_3 = 2.27882202 \times 10^{37}$. This indicates that the boundary equilibrium point E_b meets the positive equilibrium point E^* at the static branch point (where $\mathcal{R}_0 = \mathcal{R}_1 = 1$) and the positive equilibrium point E^* appears. That is, the equilibrium point E^* exists only for $\delta \geq \delta_t$, and no further static branching of E^* is possible for $\delta > \delta_t$.

Table 9. Parameter values used in model.

Parameters	Value	Parameters	Value
α	1600	θ	5×10^{-6}
β_v	0.1	m	4
β_i	0.1	n	750
μ	0.001	p	1
γ	0.088	q	1×10^{-7}
ω	1×10^{-6}	δ	Bifurcation parameter

Next, consider the possible Hopf bifurcation at E^*. Obviously, a Hopf bifurcation occurs when $P(\lambda, T^*, \delta) = 0$ has a pair of purely imaginary eigenvalues. However, for high-dimensional systems, the eigenvalues are difficult to solve. The next lemma states the conditions for generating Hopf bifurcations and having stable periodic solutions without computing eigenvalues.

Lemma 4.1 (Lemma [14]). *Suppose that the system $\dot{x} = f_\mu(x)$, $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}$ has an equilibrium (x_0, μ_0) at which the following properties are satisfied:*
(H1) $D_x f_{\mu_0}(x_0)$ has a simple pair of pure imaginary eigenvalues and no other eigenvalues with zero real parts.

Then (H1) implies that there is a smooth curve of equilibria $(x(\mu), \mu)$ with $x(\mu_0) = x_0$. The eigenvalues $\lambda(\mu)$, $\bar{\lambda}(\mu)$ of $D_x f_{\mu_0}(x)$ which are imaginary at $\mu = \mu_0$ vary smoothly with μ. If, moreover,
(H2)

$$\frac{d}{d\delta} Re(\lambda(\mu))|_{\mu=\mu_0} = d \neq 0,$$

then there is a unique three-dimensional center manifold passing through (x_0, μ_0) in $\mathbb{R}^n \times \mathbb{R}$ and a smooth system of coordinates (preserving the planes μ= const.) for

which the Taylor expansion of degree 3 on the center manifold is given by

$$\dot{x} = (d\mu + a(x^2 + y^2))x - (\omega + c\mu + b(x^2 + y^2))y,$$
$$\dot{y} = (\omega + c\mu + b(x^2 + y^2))x + (d\mu + a(x^2 + y^2))y.$$

If $a \neq 0$, there is a surface of periodic solutions in the center manifold which has quadratic tangency with the eigenspace of $\lambda(\mu)$, $\bar{\lambda}(\mu)$ agreeing to second order with the paraboloid $\mu = -(a/d)(x^2 + y^2)$. If $a < 0$, then these periodic solutions are stable limit cycles, while if $a > 0$, the periodic solutions are repelling.

$$a = \frac{1}{16}[f_{xxx}+f_{xyy}+g_{xxy}+g_{yyy}]+\frac{1}{16\omega}[f_{xy}(f_{xx}+f_{yy})-g_{xy}(g_{xx}+g_{yy})-f_{xx}g_{xx}+f_{yy}g_{yy}].$$

By Lemma 4.1, in order to explore possible Hopf bifurcations arising from E^*, starting from (H1), consider the third order principal subformula $\Delta_3(T^*, \delta)$ of the fourth order characteristic polynomial $P(\lambda, T^*, \delta)$ and solve the system of equations

$$\begin{cases} \Delta_3(T^*, \delta) = a_1a_2a_3 - a_3^2 - a_1^2a_4 = 0, \\ F(T^*, \delta) = 0. \end{cases}$$

This gives the Hopf branch point $(\delta_H, T_H) \approx (853.40909023, 1.28938689 \times 10^{-6})$, where the subscript "H" represents the Hopf bifurcation. At the branch point,

$$a_1(\delta_H) \approx 1.39961648 \times 10^9, a_2(\delta_H) \approx 1.79737179 \times 10^{15}, a_3(\delta_H) \approx 1.10592583 \times 10^{16},$$
$$a_4(\delta_H) \approx 1.42021754 \times 10^{22}, \Delta_2(\delta_H) \approx 2.51563117 \times 10^{24}, \Delta_3(\delta_H)$$
$$\approx -3.56960583 \times 10^{-104}.$$

At this point, $P(\lambda, T^*, \delta)$ has a pair of purely imaginary characteristic roots, so the first condition (H1) in Lemma 4.1 is satisfied. By $a_1(\delta_H)a_2(\delta_H)a_3(\delta_H) - a_3^2(\delta_H) - a_1^2(\delta_H)a_4(\delta_H) = 0$, the characteristic polynomial(4.12) can be transformed into:

$$\left(\lambda^2 + \frac{a_3(\delta_H)}{a_1(\delta_H)}\right)\left(\lambda^2 + a_1(\delta_H)\lambda + \frac{a_1(\delta_H)a_2(\delta_H) - a_3(\delta_H)}{a_1(\delta_H)}\right) = 0.$$

Its four characteristic roots are

$$\lambda_{1,2} = \pm i\sqrt{\frac{a_3(\delta_H)}{a_1(\delta_H)}},\ \lambda_{3,4} = \frac{1}{2}\left[-a_1(\delta_H) \pm \sqrt{a_1^2(\delta_H) - \frac{4[a_1(\delta_H)a_2(\delta_H) - a_3(\delta_H)]}{a_1(\delta_H)}}\right].$$

So, the eigenvalues of the Jacobian matrix at E^* are a pair of pure imaginary roots $\pm 2810.98466597i$, and two negative real eigenroots: -1.28536923×10^6 and -1.39833111×10^9. Thus, E^* is stable within $\delta \in (\delta_H, \delta_t) \approx (853.40909023, 853.40909025)$. But when $\delta = \delta_H$, E^* loses its stability and the Hopf branch appears. We then investigate whether periodic solutions are generated at the Hopf bifurcation, and investigate the stability of the limit rings generated at the Hopf bifurcation using center manifold theory and normal form theory.

The first question is whether a periodic solution arises at the Hopf bifurcation. The first condition for a periodic solution is that the characteristic equation has purely imaginary roots, which have already been verified. We now confirm that the cross-sectional conditions required to produce a periodic solution are satisfied.

After a small perturbation of the parameter values at the branching point, the four characteristic roots of the characteristic polynomial (4.12) can be expressed in terms of the branching parameter δ as follows:

$$\begin{aligned}
\lambda_1(\delta) =& \nu_1(\delta) + i\nu_2(\delta),\\
\lambda_2(\delta) =& \nu_1(\delta) - i\nu_2(\delta),\\
\lambda_3(\delta) =& \frac{1}{2}\left[-a_1(\delta) + \sqrt{a_1(\delta)^2 - \frac{4(a_1(\delta)a_2(\delta) - a_3(\delta))}{a_1(\delta)}}\right],\\
\lambda_4(\delta) =& \frac{1}{2}\left[-a_1(\delta) - \sqrt{a_1(\delta)^2 - \frac{4(a_1(\delta)a_2(\delta) - a_3(\delta))}{a_1(\delta)}}\right].
\end{aligned}$$

The Hopf bifurcation must satisfy the transversal condition for it to occur,

$$\frac{d}{d\delta} Re(\lambda_j(\delta))|_{\delta=\delta_H} \neq 0, \qquad j = 1, 2.$$

Substituting $\lambda_1(\delta) = \nu_1(\delta) + i\nu_2(\delta)$ into the characteristic equation (4.12) and calculating the derivative, we get

$$\begin{aligned}
&\Gamma(\delta)\nu_1'(\delta) - \Theta(\delta)\nu_2'(\delta) + \Upsilon(\delta) = 0,\\
&\Theta(\delta)\nu_1'(\delta) + \Gamma(\delta)\nu_2'(\delta) + \Lambda(\delta) = 0.
\end{aligned}$$

Of which

$$\begin{aligned}
\Gamma(\delta) =& 4\nu_1^3(\delta) - 12\nu_1(\delta)\nu_2^2(\delta) + 3a_1(\delta)\nu_1^2(\delta) - 3a_1(\delta)\nu_2^2(\delta) + 2a_2(\delta)\nu_1(\delta) + a_3(\delta),\\
\Theta(\delta) =& 12\nu_1^2(\delta)\nu_2(\delta) + 6a_1(\delta)\nu_1(\delta)\nu_2(\delta) + 2a_2(\delta)\nu_2(\delta) - 4\nu_2^3(\delta),\\
\Upsilon(\delta) =& a_1'(\delta)\nu_1^3(\delta) - 3a_1'(\delta)\nu_1(\delta)\nu_2^2(\delta) + a_2'(\delta)\nu_1^2(\delta) - a_2'(\delta)\nu_2^2(\delta) + a_3'(\delta)\nu_1(\delta) + a_4'(\delta),\\
\Lambda(\delta) =& 3a_1'(\delta)\nu_1^2(\delta)\nu_2(\delta) - a_1'(\delta)\nu_2^3(\delta) + 2a_2'(\delta)\nu_1(\delta)\nu_2(\delta) + a_3'(\delta)\nu_2(\delta).
\end{aligned}$$

From $\Theta(\delta_H)\Lambda(\delta_H) + \Gamma(\delta_H)\Upsilon(\delta_H) \neq 0$,we get

$$\frac{d}{d\delta} Re(\lambda_j(\delta))|_{\delta=\delta_H} = -\frac{\Theta(\delta)\Lambda(\delta) + \Gamma(\delta)\Upsilon(\delta)}{\Gamma^2(\delta) + \Theta^2(\delta)}|_{\delta=\delta_H} \neq 0, \quad j = 1, 2.$$

and

$$\lambda_{3,4}(\delta_H) = \frac{1}{2}\left[-a_1(\delta) \pm \sqrt{a_1(\delta)^2 - \frac{4(a_1(\delta)a_2(\delta) - a_3(\delta))}{a_1(\delta)}}\right]|_{\delta=\delta_H} \neq 0.$$

So when $\delta = \delta_H$, the system satisfies the transversal condition generated by the periodic solution, and when $\delta < \delta_H$, the system (2.1) generates a Hopf bifurcation around E^*, resulting in a limit cycle.

Next, we use the central pop theory and normal form theory to consider the stability of bifurcation limit cycles. First, we perform a translation transformation $Z = \Pi - E^*$, which translates the equilibrium point E^* to the origin.

$$\begin{pmatrix} z_1\\ z_2\\ z_3\\ z_4 \end{pmatrix} = \begin{pmatrix} C\\ I\\ T\\ V \end{pmatrix} - \begin{pmatrix} C^*\\ I^*\\ T^*\\ V^* \end{pmatrix}.$$

Let

$$\mathrm{P}=\begin{pmatrix} 0.11653 & -0.00025 & 0.00011 & -1.00000 \\ -0.00132 & -2.58370 & -0.00011 & 4.45531\times10^{-11} \\ -1.52333\times10^{-15} & 2.92867\times10^{-6} & -5.92599\times10^{-12} & -2.15696\times10^{-16} \\ -0.99319 & 0 & 1.00000 & 3.34483\times10^{-8} \end{pmatrix}.$$

Then let $Z=PX$ and consider the coordinate transformation to get

$$\dot{X}=G_i(x_1,x_2,x_3,x_4;\mu_H). \tag{4.13}$$

Among them,

$$\begin{aligned} G_1 \approx & 2810.98466597x_2 \\ & + 6.64444126\times10^{-9}x_1^2 - 4.82373735\times10^{-7}x_2^2 + 2.38741494\times10^{-8}x_3^2 \\ & + 2.87146129\times10^{-8}x_4^2 + 0.00817761x_1x_2 - 0.01180142x_2x_3 - 0.00186633x_2x_4 \\ & - 6.03998798\times10^{-8}x_1x_4 + 5.39916853\times10^{-8}x_3x_4 - 2.23866803\times10^{-8}x_1x_3, \\ G_2 \approx & -2810.98466597x_1 \\ & - 0.00021417x_1^2 - 3.32101036\times10^{-6}x_2^2 - 0.01300408x_3^2 - 1.13613328\times10^{-26}x_4^2 \\ & - 6384.75498892x_1x_2 + 0.01291567x_1x_3 + 0.00186680x_1x_4 + 0.00021502x_2x_4 \\ & + 6426.69547693x_2x_3 - 4.73761865\times10^{-7}x_3x_4, \\ G_3 \approx & -1.39833111\times10^{9}x_3 \\ & + 6.59897082\times10^{-9}x_1^2 - 6.69938531\times10^{-9}x_2^2 - 0.10008025x_3^2 \\ & + 2.85189481\times10^{-8}x_4^2 + 0.00827970x_1x_2 - 108.19053051x_1x_3 \\ & - 5.9988335610^{-8}x_1x_4 + 0.22450490x_2x_3 - 2.64421420\times10^{-10}x_2x_4 \\ & + 928.40909023x_3x_4, \\ G_4 \approx & -1.28536923\times10^{6}x_4 \\ & - 8.09187284\times10^{-11}x_1^2 + 5.95706860\times10^{-11}x_2^2 + 3.31683884\times10^{-6}x_3^2 \\ & + 0.85340909x_4^2 + 1.62806567x_1x_2 - 3.2941768910^{-6}x_1x_3 - 0.09945054x_1x_4 \\ & - 1.63920216x_2x_3 + 0.00021743x_2x_4 - 0.00009199x_3x_4. \end{aligned}$$

Thus the Jacobian matrix of the system (4.13) at equilibrium $x_i=0$ $(i=1,2,3,4)$ is the following Jordan standard type:

$$\mathrm{J}=\begin{pmatrix} 0 & 2810.98466597 & 0 & 0 \\ -2810.98466597 & 0 & 0 & 0 \\ 0 & 0 & -1.39833111\times10^{9} & 0 \\ 0 & 0 & 0 & -1.28536923\times10^{6} \end{pmatrix}.$$

From the above matrix, the system contains a two-dimensional central subspace as well as a two-dimensional stable subspace. The following is a downscaling of the

system. Let

$$\begin{cases} x_3 = h_3(x_1, x_2) = k_1 x_1^2 + k_2 x_1 x_2 + k_3 x_2^2 + o(r^2), \\ x_4 = h_4(x_1, x_2) = l_1 x_1^2 + l_2 x_1 x_2 + l_3 x_2^2 + o(r^2). \end{cases} \tag{4.14}$$

Then

$$\begin{cases} \dfrac{dx_3}{d\tau} = 2k_1 x_1 x_1' + k_2 x_1' x_2 + k_2 x_1 x_2' + 2k_3 x_2 x_2' + o(r^2), \\ \dfrac{dx_4}{d\tau} = 2l_1 x_1 x_1' + l_2 x_1' x_2 + l_2 x_1 x_2' + 2l_3 x_2 x_2' + o(r^2). \end{cases} \tag{4.15}$$

The solution of (4.15) is

$$\begin{aligned} &k_1 \approx -1.66220821 \times 10^{-17}, \quad k_2 \approx 5.92112929 \times 10^{-12}, \quad k_3 \approx -1.66938923 \times 10^{-17}, \\ &l_1 \approx 2.76991395 \times 10^{-9}, \quad l_2 \approx 1.26658899 \times 10^{-6}, \quad l_3 \approx -2.76991397 \times 10^{-9}. \end{aligned}$$

Substituting (4.14) into (4.13) for $G_1(x_1, x_2, x_3, x_4)$ and $G_2(x_1, x_2, x_3, x_4)$, we get

$$\begin{cases} x_1' = 2810.98466597 x_2 + f(x_1, x_2), \\ x_2' = -2810.98466597 x_1 + g(x_1, x_2). \end{cases}$$

Of which

$$\begin{aligned} f(x_1, x_2) \approx & 0.00817761 x_1 x_2 + 2.27799536 \times 10^{-25} x_1^4 - 3.49212618 \times 10^{-20} \\ & - 4.82373735 \times 10^{-7} x_2^2 + 5.16956847 \times 10^{-12} x_2^3 \\ & - 1.67622819 \times 10^{-16} x_1^3 - 1.21849458 \times 10^{-23} x_2^4 + 6.64444126 \times 10^{-9} x_1^2 \\ & + 1.72851847 \times 10^{-33} x_2^5 + 2.98595494 \times 10^{-36} x_1^5 \\ & + 3.91173910 \times 10^{-33} x_1^4 x_2 + 6.10637748 \times 10^{-10} x_1^2 x_2 \\ & + 5.90496694 \times 10^{-26} x_1^2 x 2^3 - 1.46389230 \times 10^{-17} x_1^3 x_2 \\ & + 1.30300583 \times 10^{-28} x_1^3 x_2^2 + 1.46446109 \times 10^{-17} x_1 x_2^3 \\ & - 2.65396392 \times 10^{-47} x_1^5 x_2 - 1.29926614 \times 10^{-28} x_1 x_2^4 \\ & - 2.36528722 \times 10^{-9} x_1 x_2^2 - 2.66919558 \times 10^{-47} x_1 x_2^5 \\ & - 6.21584175 \times 10^{-42} x_1^3 x_2^3 + 2.57251862 \times 10^{-44} x_1^2 x_2^4 \\ & - 6.69394284 \times 10^{-15} x_1^2 x_2^2 - 2.56864001 \times 10^{-44} x_1^4 x_2^2 \\ & - 2.05078906 \times 10^{-52} x_1^6 + 3.52683920 \times 10^{-52} x_2^6, \\ g(x_1, x_2) \approx & - 1.29264653 \times 10^{-27} x_2^5 - 2.39501322 \times 10^{-30} x_1^5 + 1.84894133 \times 10^{-38} \\ & - 0.00049399 x_1^2 x_2 + 1.83345860 \times 10^{-46} x_1^6 - 2.63749363 \times 10^{-46} x_2^6 \\ & - 0.00021417 x_1^2 - 5.98860318 \times 10^{-21} x_2^4 - 6384.75498892 x_1 x_2 \\ & - 3.23002752 \times 10^{-27} x_1^4 x_2 + 1.17419379 \times 10^{-11} x_1^3 x_2 \\ & + 5.36919961 \times 10^{-9} x_1^2 x_2^2 - 3.22295655 \times 10^{-20} x_1^2 x_2^3 \\ & + 7.10741955 \times 10^{-23} x_1 x_2^4 - 7.14593124 \times 10^{-23} x_1^3 x_2^2 \\ & + 1.11864297 \times 10^{-6} x_1 x_2^2 + 1.46379276 \times 10^{-41} x_1 x_2^5 \\ & - 1.40577277 \times 10^{-38} x_1^2 x_2^4 + 3.39263339 \times 10^{-36} x_1^3 x_2^3 \end{aligned}$$

$$
\begin{aligned}
&+ 1.40366014 \times 10^{-38} x_1^4 x_2^2 - 1.17419444 \times 10^{-11} x_1 x_2^3 \\
&+ 1.45549603 \times 10^{-41} x_1^5 x_2 + 4.12573940 \times 10^{-29} x_2 \\
&- 6.00677289 \times 10^{-21} x_1^4 - 3.32101036 \times 10^{-6} x_2^2 \\
&- 7.02322853 \times 10^{-13} x_2^3 + 5.42781920 \times 10^{-12} x_1^3.
\end{aligned}
$$

Therefore, the Lyapunov coefficient is

$$
\begin{aligned}
\rho &= \frac{1}{16}\Big\{\frac{1}{\delta_H}\big[f_{x_1x_2}(f_{x_1x_1} + f_{x_2x_2}) - g_{x_1x_2}(g_{x_1x_1} + g_{x_2x_2}) - f_{x_1x_1}g_{x_1x_1} + f_{x_2x_2}g_{x_2x_2}\big] \\
&\quad + f_{x_1x_1x_1} + f_{x_1x_2x_2} + g_{x_1x_1x_2} + g_{x_2x_2x_2}\Big\}\big|_{x_1=x_2=0} \\
&= -0.00026514 < 0.
\end{aligned}
$$

The Hopf bifurcation is therefore a supercritical Hopf bifurcation, and the bifurcation limit cycle is stable.

In summary, the disease-free equilibrium point E_0 is locally asymptotically stable when $\delta > 853.40909028$. As δ decreases, the boundary equilibrium point E_1 begins to appear and E_1 is locally asymptotically stable when $\delta_t = 853.40909025 < \delta < 853.40909028$. According to theorem (4.1), the positive equilibrium point E^* starts to appear when $\delta < \delta_t$ and E^* is stable when $\delta_H = 853.40909023 < \delta < \delta_t$. At $\delta < \delta_H$, E^* becomes unstable and produces a stable periodic solution. Fig 6 shows the periodic solution when the branching parameter $\delta = 850 < \delta_H$, other parameters are given in Table 9.

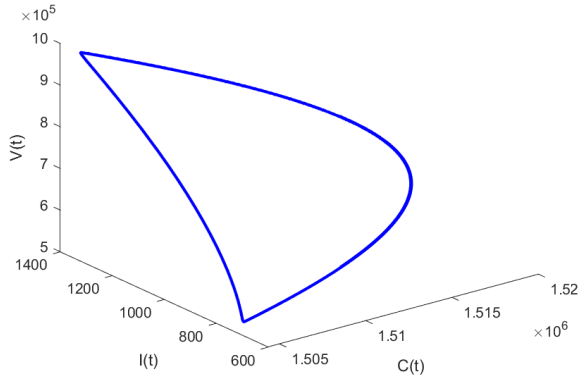

Figure 6. When $\delta = 850 < \delta_H$, an orbitally asymtotically stable periodic solution has appeared.

The results of this characterization suggest that, in a particular immune state, a cyclical oscillation of virus concentrations may occur in infected individuals. This may explain the recurrence of recurrence in some infected individuals after seemingly recovering.

5. Conclusions

Immunology is studied using a non-linear systems theory, which views the immune system as a complex system consisting of four broad types of components: 1. a

variety of immune cells; 2. a variety of immune molecules; 3. major histocompatibility complexes and autoantigens associated with genes; and 4. external invasive antigens. There are complex, non-linear interactions between them. To study this complex system's static and dynamic behavior, we need to observe several variables (e.g., the concentration or quantity of a particular bacterium, immune cell, immune molecule, or chemical component). They are selected to describe the system's state, and equations or other non-linear mathematical tools also characterize the relationships between these variables. The study of the Spatio-temporal evolution of variables, such as the system's stationary state and its stability and bifurcation behavior. A further step is to compare theoretical results with experimental observations and analyze them.

We propose a set of kinetic models to describe the process of SARS-COV-2 infection in individuals and the resulting immune response. The best model was selected by fitting five models to two sets of viral load and lymphocyte data from mildly ill patients at Yale New Haven Hospital using an information criterion. The model shows that the attack of infected cells on target cells is much less effective than that of viruses on target cells and can be ignored in the modeling.

To evaluate the reliability of parameter estimation, we first analyze the structural identifiability of the Full model. If there is no data noise, then the Full model is structurally identifiable for both viral load and Tlymphocyte data. Therefore all sub-models under the Full model are also structurally identifiable, i.e., all parameters can be uniquely determined. A practical identifiability analysis of the optimal model using the GWMCMC algorithm revealed that the parameters θ, n, q are identifiable in the case of two data sets, while the parameters α, β_v and δ are not. This suggests that even if the model is structurally identifiable, it may still be unidentifiable in practice due to data noise or data size.

To further analyze the impact of unrecognized parameters on the model prediction, we selected the maximum and minimum combinations of the actual unrecognized parameters α, β_v and δ for simulation, and found that different parameter combinations had little impact on viral load, indicating that the estimation of model parameters was robust. On this basis, we further explored the dynamic behavior of viral load under different medications.

It is worth noting that the best model selected is only relatively optimal among the five models when the data of New Haven Hospital are used to fit the model. In order to explore the dynamic behavior of the model in general sense, we continue to analyze the existence, stability and bifurcation of equilibrium points of Full model (so do other submodels). We find that the two threshold parameters $\mathcal{R}_0$ and $\mathcal{R}_1$ completely determine the dynamics of the model. Theorem 4.1 shows that the disease-free equilibrium point E_0 is locally asymptotically stable when $\mathcal{R}_0 < 1$, implying that the individual's immune system can eradicate SARS-COV-2 when the system parameters satisfy this relationship, otherwise, infection is established. Further analysis shows that under certain conditions, the system can have a more complex Hopf bifurcation phenomenon, and the system will produce a stable periodic solution. This means that in some cases, components such as viral load and T lymphocytes do not necessarily tend to be stable, and may show more complex periodic motions. This may explain why some infected people repeatedly return to positive without contacting new antigens after seemingly recovering. In some cases components such as viral load and T lymphocytes are not necessarily stable and may exhibit more complex periodic motions. This may explain why some infected

people repeatedly return to positive after seemingly recovering without contacting new antigens.

Our model is only a simple dynamic model. Considering that the novel coronavirus has constantly been changing, more experimental or clinical data can be further improved in the future to explore further infection mechanism of the novel coronavirus.

Appendix

Model 2:

$$\begin{cases} \dfrac{dC}{dt} = \alpha - \dfrac{\beta_v CV}{1+\delta C} - \mu C \\ \dfrac{dI}{dt} = \dfrac{\beta_v CV}{1+\delta C} - \omega IT - \gamma I \\ \dfrac{dT}{dt} = \omega IT + \theta VT - mT \\ \dfrac{dV}{dt} = nI - qTV - pV. \end{cases}$$

Model 3:

$$\begin{cases} \dfrac{dC}{dt} = \alpha - \dfrac{\beta_v CV}{1+\delta C} - \dfrac{\beta_i CI}{1+\delta C} - \mu C \\ \dfrac{dI}{dt} = \dfrac{\beta_v CV}{1+\delta C} + \dfrac{\beta_i CI}{1+\delta C} - \gamma I \\ \dfrac{dT}{dt} = \theta VT - mT \\ \dfrac{dV}{dt} = nI - qTV - pV. \end{cases}$$

Model 4:

$$\begin{cases} \dfrac{dC}{dt} = \alpha - \dfrac{\beta_v CV}{1+\delta C} - \mu C \\ \dfrac{dI}{dt} = \dfrac{\beta_v CV}{1+\delta C} - \gamma I \\ \dfrac{dT}{dt} = \theta VT - mT \\ \dfrac{dV}{dt} = nI - qTV - pV. \end{cases}$$

Model 5:

$$\begin{cases} \dfrac{dC}{dt} = \alpha - \dfrac{\beta_i CI}{1+\delta C} - \mu C \\ \dfrac{dI}{dt} = \dfrac{\beta_i CI}{1+\delta C} - \gamma I \\ \dfrac{dT}{dt} = \theta VT - mT \\ \dfrac{dV}{dt} = nI - qTV - pV. \end{cases}$$

References

[1] L. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, CRC Press, 2010.

[2] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, in *Selected papers of hirotugu akaike*, Springer, 1998. DOI: 10.1007/978-1-4612-1694-0_15.

[3] K. Burnham and D. Anderson, *Multimodel inference: Understad ing AIC and BIC in Model Selection*, 2004, 33(2), 261–304.

[4] G. Bellu, M. P. Saccomani, S. Audoly, et al., *DAISY: A new software tool to test global identifiability of biological and physiological systems*, Computer Methods and Programs in Biomedicine, 2007, 88(1), 52–61.

[5] O. T. Chis, J. R. Banga and E. Balsa-Canto, *Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods*, Plos One, 2011, 6(11), e27755–e27755.

[6] C. Cobelli and J. J. Distefano, *Parameter and Structural Identifiability Concepts and Ambiguities -a Critical-Review and Analysis*, The American journal of physiology, 1980, 239(1), R7–24.

[7] T. Chen, J. Rui, Q. Wang, et al., *A mathematical model for simulating the phase-based transmissibility of a novel coronavirus*, Infectious Diseases of Poverty, 2020, 9, 24–24.

[8] M. Chen, Q. Shao and J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, Springer, 2000. DOI: 10.1007/978-1-4612-1276-8.

[9] K. Ejima, K. S. Kim, Y. Ito, et al., *Inferring Timing of Infection Using Within-host SARS-CoV-2 Infection Dynamics Model: Are "Imported Cases" Truly Imported?* 2020. DOI: 2020,10.1101/2020.03.30.20040519.

[10] M. C. Eisenberg, S. L. Robertson and J. H. Tien, *Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease*, Journal of Theoretical Biology, 2013, 324(Complete), 84–102.

[11] N. D. Evans, L. J. White, M. J. Chapman, et al., *The structural identifiability of the susceptible infected recovered model with seasonal forcing*, Mathematical Biosciences, 2005, 194(2), 175–197.

[12] J. Goodman and J. Weare, *Ensemble samplers with affine invariance*, Communications in Applied Mathematics and Computational Science, 2010, 5(1), 65–80.

[13] J. K. Ghosh, *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists by Scott M. Lynch*, International Statal Review, 2010, 76(2), 311–312.

[14] M. Golubitsky and P. H. Rabinowitz, *Abzweigung einer periodischen Lösung von einer stationaeren Lösung eines Differentialsystems*, Akad. Wiss. (Leipzig), 1942, 94(1), 3–22.

[15] L. F. García, *Immune Response, Inflammation, and the Clinical Spectrum of COVID-19*. Frontiers in Immunology, 2020, 11, 1441–1441.

[16] C. Huang, Y. Wang, X. Li, et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*, The Lancet, 2020, 395(10223), 496–496.

[17] K. Hattaf and N. Yousfi, *Dynamics of SARS-CoV-2 infection model with two modes of transmission and immune response*, Mathematical Biosciences and Engineering, 2020, 17(5), 5326–5340.

[18] A. J. Kucharski, T. W. Russel, C. Diamond, et al., *Early dynamics of transmission and control of COVID-19: a mathematical modelling study*, The Lancet Infectious Diseases, 2020, 20(5), 553–558.

[19] K. S. Kim, K. Ejima, Y. Ito, et al., *Modelling SARS-CoV-2 Dynamics: Implications for Therapy*, Cold Spring Harbor Laboratory Press, 2020. DOI: 10.1101/2020.03.23.20040493.

[20] J. Y. Kim, J. H. Ko, Y. Kim, et al., *Viral load kinetics of SARS-CoV-2 infection in first two patients in Korea*, Journal of Korean Medical Science, 2020, 35(7), e86–e86.

[21] T. Liu, J. Hu, M. Kang, et al., *Transmission Dynamics of 2019 Novel Coronavirus (2019-nCoV)*, Social Science Electronic Publishing, 2020. DOI: 10.1101/2020.01.25.919787.

[22] C. Li, J. Xu, J. Liu, et al., *The within-host viral kinetics of SARS-CoV-2*, Mathematical Biosciences and Engineering, 2020, 17(4), 2853–2861.

[23] R. N. Leander, Y. Wu, W. Ding, et al., *A model of the innate immune response to SARS-CoV-2 in the alveolar epithelium*, Royal Society open science, 8(8), 210090–210090.

[24] L. Ljung and T. Glad, *Testing Global Identifiability for Arbitrary Model Parameterizations*, IFAC Proceedings Volumes, 1991, 24(3), 1085–1090.

[25] C. Lucas, P. Wong, J. Klein, et al., *Longitudinal analyses reveal immunological misfiring in severe COVID-19*, Nature, 2020, 584(7821), 463–469.

[26] H. Miao, X. Xia, A. S. Perelson, et al., *On Identifiability Of Nonlinear Ode Models And Applications In Viral Dynamics*, SIAM Review, 2011, 53(1), 3–39.

[27] V. J. Munster, F. Feldmann, B. N. Williamson, et al., *Respiratory disease in rhesus macaques inoculated with SARS-CoV-2*, Nature, 2020, 585(7824), 268–272.

[28] H. Miao, X. Xia, A. S. Perelson, et al., *On Identifiability Of Nonlinear Ode Models And Applications In Viral Dynamics*, SIAM Review, 2011, 53(1), 3–39.

[29] A. Mi, B, Si, A. As, et al., *AI- modelling of molecular identification and feminization of wolbachia infected Aedes aegypti*, Progress in Biophysics and Molecular Biology, 2020, 150, 104–111.

[30] Y. Pan, D. Zhang, P. Yang, et al., *Viral load of SARS-CoV-2 in clinical samples*, The Lancet Infectious Diseases, 2020, 20(4), 411–412.

[31] E. Prompetchara, C. Ketloy and T. Palaga, *Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic*, Asian Pac J Allergy Immunol, 2020, 38(1),1–9.

[32] R. F. Reis, A. B. Pigozzo, C. R. B. Bonin, et al., *A Validated Mathematical Model of the Cytokine Release Syndrome in Severe COVID-19*, Frontiers in Molecular Biosciences, 2021, 8, 639423–639423.

[33] Y. Ren, T. Shu, D. Wu, et al., *The ORF3a protein of SARS-CoV-2 induces apoptosis in cells*, Cellular & molecular immunology, 2020, 17(8), 1–3.

[34] W. C. Roda, *Bayesian inference for dynamical systems*, Infectious Disease Modelling, 2020, 5, 221–232.

[35] W. C. Roda, M. B. Varughese, D. Han, et al., *Why Is It Difficult to Accurately Predict the COVID-19 Epidemic?* Infectious Disease Modelling, 2020, 5, 271–281.

[36] J. Shang, Y. Wan, C. Luo, et al., *Cell entry mechanisms of SARS-CoV-2*, Proceedings of the National Academy of Sciences, 2020, 117(21), 11727–11734.

[37] M. Shen, Z. Peng, Y. Xiao, et al., *Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China*, 2020. DOI:10.1101/2020.01.23.916726.

[38] A. Sw, P. Yang, C. Qwb, et al., *Modeling the viral dynamics of SARS-CoV-2 infection - ScienceDirect*, Mathematical Biosciences, 2020, 328, 108438–108438.

[39] S. Sahoo, K. Hari, S. Jhunjhunwala, et al., *Mechanistic modeling of the SARS-CoV-2 and immune system interplay unravels design principles for diverse clinicopathological outcomes*, Public Health Intervention for the COVID-19, 2022. DOI: 10.1142/9789811249723_0002.

[40] L. Sherin, S. Farwa, A. Sohail, et al., *Cancer drug therapy and stochastic modeling of "nano-motors"*, International Journal of Nanomedicine, 2018, 13, 6429–6440.

[41] G. E. Schwarz, *Estimating the Dimension of a Model*, The Annals of Statistics, 1978, 6(2), 461–464.

[42] M. Z. Tay, C. M. Poh, L. Rénia, et al., *The trinity of COVID-19: immunity, inflammation and intervention*, Nature reviews. Immunology, 2020, 20(6), 1–12.

[43] T. Takahashi, M. K. Ellingson, P. Wong, et al., *Sex differences in immune responses that underlie COVID-19 disease outcomes*, Nature, 2020, 588(7837), 315–320.

[44] S. A. Vardhana and J. D. Wolchok, *The many faces of the anti-COVID immune response*, Journal of Experimental Medicine, 2020, 217(6), e20200678–e20200678.

[45] X. Wang, W. Xu, G. Hu, et al., *SARS-CoV-2 infects T lymphocytes through its spike protein-mediated membrane fusion*, Cellular & molecular immunology, 2020, 17(8), 894–894.

[46] J. Wu, K. Leung, M. Bushman, et al., *Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China*, Nature medicine, 2020, 26(4), 1149–1150.

[47] A. Wu, Y. Peng, B. Huang, et al., *Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China*, Cell Host & Microbe, 2020, 27(3), 325–328.

[48] C. Wang, W. Li, D. Drabek, et al., *A human monoclonal antibody blocking SARS-CoV-2 infection*, Nat. Commun., 2020, 11(1), 2251–2251.

[49] X. Zhang, Y. Tan, Y. Ling, et al., *Viral and host factors related to the clinical outcome of COVID-19*, Nature, 2020, 583, 437–440.

[50] P. Zhou, X. Yang, X. Wang, et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin*, Nature, 2020, 579(7798), 270–273.