

Exploring the evolutionary dynamics of infectious diseases through SIS epidemic models

KING-YEUNG LAM, YUAN LOU, AND SHIZHAO MA

For the spread and evolution of infectious diseases, we incorporate phenotypic structure into Susceptible-Infected-Susceptible (SIS) epidemic models of the reaction-diffusion type. It is shown that the unique disease-free equilibrium is globally asymptotically stable when the basic reproduction number is less than one, and the infected population persists when the basic reproduction number is greater than one. The asymptotic profile of the endemic equilibrium is determined when the mutation rate of the infected population converges to zero. We integrate analytical results with numerical simulations to investigate how multiple phenotypic traits evolve. Our findings confirm that the susceptible population evolves to be primarily made up of individuals with low immunity, while the infected population is eventually comprised of highly infectious individuals with low mutation rate. These results indicate that as the disease infectivity continues to increase, the group immunity will decrease. In addition, if the virus mutation rate is initially small, it will first increase rapidly before eventually decreasing. Finally, those strains with low mutation rates are more advantageous, i.e. the virus might first employ the high mutation rate to increase infectivity and decrease immunity, and then use the low mutation rate to maintain high infectivity and low immunity.

1. Introduction

The coronavirus pandemic 2019 (COVID-19) is caused by the infection of SARS-CoV-2 disease [1]. First discovered in December of 2019, the disease spread rapidly across the globe. As of December 14, 2022, a total number of 650,661,649 confirmed cases are reported along with 6,656,920 deaths [2]. COVID-19 has become one of the largest epidemics in human history.

During the early phase of COVID-19, the virus evolved rapidly and maximized transmission between individuals [3, 4, 5]. As the vaccinated and infected population grows, host immune pressure increases the selection of SARS-CoV-2 variants that is more prone to immune escape [3, 4, 5].

Antigenic drift, which is caused by the high viral mutation rate, enables immune escape, and limits the duration of immunity conferred by infection or vaccination. This is especially relevant in the evolution of SARS-CoV-2 [3].

The World Health Organization has classified 5 major variants of concern (VOC) regarding SARS-CoV-2. They include the Alpha variant (December 2020), the Beta variant (December 2020), the Gamma variant (January 2021), the Delta variant (May 2021) and the Omicron variant (August 2021) [6].

Since the outbreak of COVID-19, the adaptive evolution of virus can be observed in terms of the changes of transmissibility, virulence and immune escape [6]. As the virus evolves, each VOC seemed to be more infectious than the strain it displaced [7, 6, 8]. Most of VOCs feature immune escape mutation [5, 9], especially Omicron, which spreads explosively between individuals with high immunity due to previous infection or vaccination [5, 10, 11].

While viral immune escape and transmissibility are also under strong evolutionary pressure, the evolution of virulence is typically a by-product of these effects and is hard to predict. Ultimately, the evolution of virulence depends on the complex interactions between factors in both the host and the pathogen [5].

Mathematical modeling [12, 13, 14, 15] has played an important role in describing the dynamics of infectious diseases and in policy making. Using systems of nonlinear differential equations, classical epidemic models aim to predict and provide a guide for policy-makers for the goals of disease prevention and control. The research of partial differential equation (PDE) epidemiological models mainly focuses on the influence of spatial (or phenotypic) heterogeneity on disease progression. See, for instance the spatial Susceptible-Infected-Susceptible (SIS) epidemic reaction-diffusion model [16] and a population evolution model with a continuously varying phenotypic trait [17].

In this paper, we introduce a set of partial differential equation models with phenotypic heterogeneity to better understand the spread and evolution of diseases. These PDE models take into account the heterogeneity in the susceptibility (determined by host immunological trait) of the healthy population as well as the heterogeneity in the infectiousness (determined by virus variants) of the population inflicted with the disease. In these models, we presume that gene expression affect the immunological trait, and the viral strain determines the infectivity of the infected population. In a long-lasting epidemic, it is commonly observed that the immunity of susceptible population decreases, while the infectivity of infected population increases.

Based upon the mutation of quantitative traits in infectious diseases, our models suggest some underlying mechanism for the evolution of infectivity, immunity and mutation rate.

The rest of the paper is organized as follows. In Sect. 2, we introduce several structured SIS PDE models for infectious diseases and perform the mathematical analysis of the continuous trait models, together with the numerical simulations of the dynamics of the models. We also consider the asymptotic profile of the endemic equilibrium as the mutation rate converges to zero. In Sect. 3, building upon the models in Sect. 2, we introduce another set of structured mathematical models to study the evolution of the mutation rate, and numerical simulations are presented to explore the evolutionary trend of the infectious diseases. Some discussions are given in Sect. 4.

2. Evolution of infectivity and immunity

We start by introducing the Kermack-McKendrick model in epidemiology [18, 19]. A population is divided into susceptible compartment and infected compartment according to the proportion of uninfected and infected populations over the time. To be more specific, let $S(t)$ and $I(t)$ be the number of susceptible individuals and infected individuals at time t , respectively. The susceptible individuals become infected with the transmission rate β and the infected individuals recover at the rate γ . It is assumed that each infected individual after recovery immediately becomes susceptible again. This process is illustrated in Fig. 1A. The corresponding Susceptible-Infected-Susceptible model is described by the system

$$(2.1) \quad \begin{cases} S'(t) = -\frac{\beta SI}{S+I} + \gamma I, \\ I'(t) = \frac{\beta SI}{S+I} - \gamma I, \end{cases} \quad t > 0.$$

This simple epidemic model (2.1) ignores individual variations in susceptibility and immunity. On the other hand, the heterogeneity of the population can be captured by introducing multiple compartments [20].

The properties which determine a host's infectivity, such as the viral load levels or the virus variants with higher infectivity, can differ among individuals over time [14]. To explore the effects of individual variations, we extend the single strain case (2.1) into the case of multiple infected groups I_j that is infected with different virus strains j ($j = 1, \dots, m$). Similarly, we take into account n immunological types (or other properties that determine the susceptibility, such as differential exposure to infection [12]) of susceptible

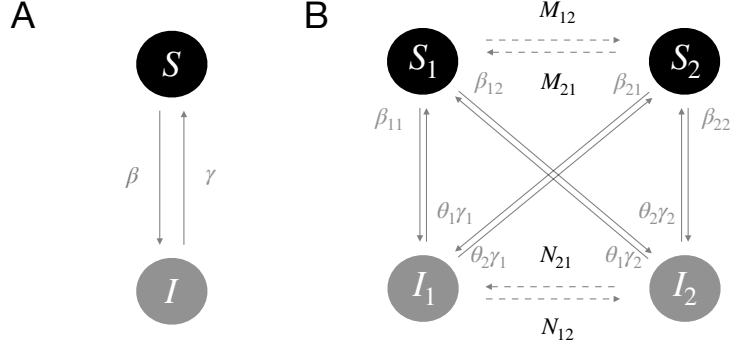


Figure 1: **Schematic illustrations of models.** (Panel A) S (or I) represents susceptible (or infected) individuals; (Panel B) S_i ($i = 1, \dots, n$) indicates the susceptible sub-populations with varied immunological traits, and I_j ($j = 1, \dots, m$) indicates the sub-populations infected with virus strain j . An illustration of $m = n = 2$ is given in Fig. 1B. The dashed lines represent the mutation process, while the solid lines represent the transmission and recovery.

individuals in order to model the individual variations in susceptibility to infection.

On the basis of the above assumption, we construct an epidemic model in which all individuals are divided into $n + m$ classes: The uninfected population is divided according to the type host immunological trait indexed by i ($S_i, i = 1, \dots, n$), and the infected population is divided according to the viral strain indexed by j ($I_j, j = 1, \dots, m$). Note that we do not consider infection of an individual by more than one strain. This model highlights population heterogeneity and link heterogeneity to disease transmission rate and recovery rate, that is, different individuals can be characterized by their transmission rate or recovery rate. We assume that $S_i(t)$ denotes the number of susceptible individuals of type i at time t ; $I_j(t)$ indicates the number of individuals infected with virus strain j at time t , respectively. Let $S = (S_1, S_2, \dots, S_n)$, $I = (I_1, I_2, \dots, I_m)$ and define the $n \times m$ matrix β , the $m \times 1$ matrix γ and the $n \times 1$ matrix θ as follows:

$$\beta = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nm} \end{pmatrix}, \quad \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix},$$

where β_{ij} represents the transmission rate at which susceptible individuals S_i become infected with the virus strain j ; γ_j is the recovery rate for the disease caused by virus strain j ; θ_i denotes the probability that, as an infected individual recovers, he/she enters the susceptible compartment S_i after recovery. Note that the vector θ satisfies $\sum_{i=1}^n \theta_i = 1$, i.e. every recovered person becomes susceptible again.

In order to incorporate the adaptation of individuals into the mathematical model, we use the mutation matrix to model the effect of the genetic or phenotypic variations occurring at rate d_S and d_I , respectively. We define the $n \times n$ mutation matrix M and $m \times m$ mutation matrix N as follows:

$$M = \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n1} & M_{n2} & \cdots & M_{nn} \end{pmatrix}, \quad N = \begin{pmatrix} N_{11} & N_{12} & \cdots & N_{1m} \\ N_{21} & N_{22} & \cdots & N_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ N_{m1} & N_{m2} & \cdots & N_{mm} \end{pmatrix},$$

where M and N satisfy

$$M_{ii} < 0, \quad M_{ij} \geq 0 \quad \text{for } i \neq j, \quad \text{and} \quad \sum_{i=1}^n M_{ij} = 0 \quad \text{for all } j,$$

and

$$N_{ii} < 0, \quad N_{ij} \geq 0 \quad \text{for } i \neq j, \quad \text{and} \quad \sum_{i=1}^m N_{ij} = 0 \quad \text{for all } j.$$

This process is shown in Fig. 1B for the case $m = n = 2$. The model is described by a system of $n + m$ ordinary differential equations as follows:

$$(2.2) \quad \left\{ \begin{array}{l} \frac{dS_i}{dt} = d_S \underbrace{\sum_{k=1}^n M_{ki} S_k}_{\text{Mutation}} - \underbrace{\frac{\sum_{j=1}^m \beta_{ij} I_j}{\sum_{i=1}^n S_i + \sum_{j=1}^m I_j}}_{\text{Infection}} S_i + \underbrace{\theta_i \sum_{j=1}^n \gamma_j I_j}_{\text{Recovery}}, \quad t > 0, \quad i = 1, \dots, n, \\ \frac{dI_j}{dt} = d_I \underbrace{\sum_{k=1}^m N_{kj} I_k}_{\text{Mutation}} + \underbrace{\frac{\sum_{i=1}^n \beta_{ij} S_i}{\sum_{i=1}^n S_i + \sum_{j=1}^m I_j}}_{\text{Infection}} I_j - \underbrace{\gamma_j I_j}_{\text{Recovery}}, \quad t > 0, \quad j = 1, \dots, m. \end{array} \right.$$

Next, we let $n, m \rightarrow \infty$ and extend the multi-compartment model (2.2) to a continuum (in trait) version. For this purpose, we introduce the continuous variables $x \in \Omega_1 = [0, L_1]$ and $y \in \Omega_2 = [0, L_2]$ for the phenotypic state of populations. As a result, $S(x, t)$ denotes the density of susceptible population with phenotypic state x at time t and $I(y, t)$ denotes the density of infected population with phenotypic state y at time t , respectively.

Generally speaking, x (or y) can refer to the expression levels of marker genes and can be measured by single cell sequencing techniques [21]. For simplicity, we refer to x as quantities that affect immunological trait and y as quantities that affect infectivity, that is, the coefficients β , γ and θ in model (2.2) are individual-specific and depending on the state x or y in the population. We suppose that $\beta(x, y)$ accounts for the rate of disease transmission with phenotypic state x and y ; Infected individuals with phenotypic state y recover at rate $\gamma(y)$. As in model (2.2), the probability of individuals becoming susceptible again, $\theta(x)$, satisfies $\int_0^{L_1} \theta(x) dx = 1$. The dynamics of the susceptible and infected populations is modeled by the following reaction-diffusion system, which generalizes (2.2):

$$(2.3) \quad \left\{ \begin{array}{l} S_t(x, t) = d_S S_{xx}(x, t) - \frac{\int_0^{L_2} \beta(x, y) I(y, t) dy}{\int_0^{L_1} S(x, t) dx + \int_0^{L_2} I(y, t) dy} S(x, t) \\ \quad + \theta(x) \int_0^{L_2} \gamma(y) I(y, t) dy, \quad 0 < x < L_1, \quad t > 0, \\ I_t(y, t) = d_I I_{yy}(y, t) + \frac{\int_0^{L_1} \beta(x, y) S(x, t) dx}{\int_0^{L_1} S(x, t) dx + \int_0^{L_2} I(y, t) dy} I(y, t) - \gamma(y) I(y, t), \\ \quad 0 < y < L_2, \quad t > 0, \\ S_x(x, t) = I_y(y, t) = 0, \quad x \in \{0, L_1\}, y \in \{0, L_2\}, t > 0, \\ S(x, 0) = S_0(x) \geq 0, I(y, 0) = I_0(y) \geq 0, \quad 0 < x < L_1, 0 < y < L_2. \end{array} \right.$$

The interpretations of all nonnegative parameters in model (2.3) are given in Table 1.

Table 1: The variables in model (2.3)

Notation	Description
x	Immunity status of susceptible individuals
y	Infectivity status of infected individuals
$S(x, t)$	The density of susceptible population with phenotypic state x at time t
$I(y, t)$	The density of infected population with phenotypic state y at time t
d_S	The mutation rate of immunity
d_I	The mutation rate of infectivity
$\beta(x, y)$	The rate of disease transmission
$\gamma(y)$	The rate of recovery
$\theta(x)$	The probability of individuals at state x becoming susceptible again
$[0, L_1]$	The mutation space of immunity x
$[0, L_2]$	The mutation space of infectivity y

We define the size of susceptible and infected populations at time t , respectively, as follows:

$$\bar{S}(t) = \int_0^{L_1} S(x, t) dx, \quad \bar{I}(t) = \int_0^{L_2} I(y, t) dy.$$

Moreover, we define the mean phenotypic state at time t as

$$\bar{x}(t) = \frac{\int_0^{L_1} xS(x, t) dx}{\int_0^{L_1} S(x, t) dx}, \quad \bar{y}(t) = \frac{\int_0^{L_2} yI(y, t) dy}{\int_0^{L_2} I(y, t) dy}.$$

Throughout this paper, we assume that the initial conditions satisfy

$$\int_0^{L_1} S(x, 0) dx + \int_0^{L_2} I(y, 0) dy = N,$$

where N is a positive constant which is fixed throughout this paper.

Integrating the equation of S over x and integrating the equation of I over y , summing two equations, we deduce that

$$\frac{\partial}{\partial t} \left(\int_0^{L_1} S(x, t) dx + \int_0^{L_2} I(y, t) dy \right) = 0, \quad t > 0.$$

Therefore, it follows that

$$(2.4) \quad \int_0^{L_1} S(x, t) dx + \int_0^{L_2} I(y, t) dy \equiv N \quad \text{for all } t \geq 0.$$

We are interested in the non-negative equilibrium solutions of model (2.3), that is, the non-negative solutions of the following system:

$$(2.5) \quad \begin{cases} d_S S_{xx}(x) - \frac{\int_0^{L_2} \beta(x,y) I(y) dy}{\int_0^{L_1} S(x) dx + \int_0^{L_2} I(y) dy} S(x) + \theta(x) \int_0^{L_2} \gamma(y) I(y) dy = 0, & 0 < x < L_1, \\ d_I I_{yy}(y) + \frac{\int_0^{L_1} \beta(x,y) S(x) dx}{\int_0^{L_1} S(x) dx + \int_0^{L_2} I(y) dy} I(y) - \gamma(y) I(y) = 0, & 0 < y < L_2, \\ S_x(x) = I_y(y) = 0, & x \in \{0, L_1\}, y \in \{0, L_2\}. \end{cases}$$

Here, $S(x)$ and $I(y)$ denote the density of susceptible and infected individuals at equilibrium, respectively. Recall (2.4), we have

$$(2.6) \quad \int_0^{L_1} S(x) dx + \int_0^{L_2} I(y) dy \equiv N.$$

In the population model (2.5), only solution $(S(x), I(y))$ satisfying $S(x) \geq 0$ on $[0, L_1]$ and $I(y) \geq 0$ on $[0, L_2]$ are of interest. A disease-free equilibrium (DFE) is a solution of (2.5)-(2.6) so that $I(y) = 0$ for every $y \in (0, L_2)$; An endemic equilibrium (EE) of (2.5)-(2.6) is a solution in which $I(y) > 0$ for some $y \in (0, L_2)$. We denote a DFE by $(\hat{S}, 0)$ and an EE by $(\tilde{S}, \tilde{I})$. By direct computations and condition (2.6), we get $\hat{S}(x) = N/L_1$. Thus (2.5) have a unique disease-free equilibrium, which is phenotypically homogeneous.

Following [16], we define the basic reproduction number for model (2.3) as follows:

$$(2.7) \quad \mathcal{R}_0 = \sup_{\substack{\varphi \in H^1([0, L_2]) \\ \varphi \neq 0}} \left\{ \frac{\frac{1}{L_1} \int_0^{L_1} \int_0^{L_2} \beta(x, y) \varphi^2(y) dx dy}{d_I \int_0^{L_2} (\varphi'(y))^2 dy + \int_0^{L_2} \gamma(y) \varphi^2(y) dy} \right\}.$$

Remark 2.1. It follows from (2.7) that $\mathcal{R}_0 > 1$ provided that

$$\frac{1}{L_1} \int_0^{L_1} \int_0^{L_2} \beta(x, y) dx dy > \int_0^{L_2} \gamma(y) dy.$$

2.1. Mathematical analysis of model (2.3)

In this subsection, we study the dynamics of model (2.3) when the basic reproduction number is less than one and greater than one, respectively. Subsect. 2.1.1 provides some preliminary estimates on the solutions of model (2.3). Subsect. 2.1.2 is devoted to the stability analysis of the disease-free equilibrium. In Subsect. 2.1.3 the existence of endemic equilibrium is established.

2.1.1. Preliminary estimates Let

$$X = \left\{ (S_0, I_0) \in C([0, L_1]; \mathbb{R}_+) \times C([0, L_2]; \mathbb{R}_+) : \int_0^{L_1} S_0 dx + \int_0^{L_2} I_0 dy = N \right\}.$$

We prove several *a priori* estimates of solutions to (2.3) in the next three lemmas.

Lemma 2.2. *Let $\int_0^{L_1} S_0 dx + \int_0^{L_2} I_0 dy = N$ for some constant $N > 0$. Then*

$$(2.8) \quad \liminf_{t \rightarrow \infty} \int_0^{L_1} S(x, t) dx \geq N \min \left\{ 1, \frac{\min_{[0, L_2]} \gamma}{\max_{[0, L_1] \times [0, L_2]} \beta} \right\}.$$

Proof. Denote

$$\bar{S}(t) = \int_0^{L_1} S(x, t) dx \quad \text{and} \quad \bar{I}(t) = \int_0^{L_2} I(y, t) dy$$

and

$$\beta^* = \sup_{[0, L_1] \times [0, L_2]} \beta \quad \text{and} \quad \gamma_* = \inf_{[0, L_2]} \gamma.$$

If $I_0 \equiv 0$, then it is easy to see that $S(x, t) \rightarrow N/L_1$ as $t \rightarrow \infty$, and we are done. We will henceforth assume that $I_0 \not\equiv 0$, i.e. $\bar{I}(0) > 0$ and $\bar{S}(0) < N$. Integrate the first equation of (2.3) over $x \in [0, L_1]$, then

$$\frac{d\bar{S}}{dt} \geq \bar{I} \left(\gamma_* - \frac{\beta^*}{N} \bar{S} \right) = (N - \bar{S}) \left(\gamma_* - \frac{\beta^*}{N} \bar{S} \right).$$

This, together with $\bar{S}(0) < N$, implies that (2.8) holds. $\square$

Lemma 2.3. *Let $t_1 > 1$. For any $p > 0$, there exists $C_0 > 0$ independent of t_1 such that*

$$(2.9) \quad \sup_{\substack{0 < x < L_1 \\ t_1 < t < t_1 + 1}} S \leq C_0 \left(\|S\|_{L^p([0, L_1] \times [t_1 - 1, t_1 + 1])} + \sup_{t \geq t_1 - 1} \|I(\cdot, t)\|_{L^1([0, L_2])} \right).$$

In particular,

$$(2.10) \quad \sup_{[0, L_1] \times [1, \infty)} S \leq 2C_0 N.$$

Proof. The estimate (2.9) is a direct consequence of the local maximum principle [22, Theorem 7.36]. Next, take $p = 1$ in (2.9), and use the fact that $\int_0^{L_1} S(x, t) dx + \int_0^{L_2} I(y, t) dy = N$, we deduce (2.10). $\square$

Lemma 2.4. *Let $t_1 \geq 2$. There exists $C_1 > 0$ such that for any solution (S, I) of (2.3) with initial data in X , we have*

$$(2.11) \quad \sup_{t \geq 2} [\|S(\cdot, t)\|_{C^1([0, L_1])} + \|I(\cdot, t)\|_{C^1([0, L_2])}] \leq C_1.$$

Proof. By the Harnack inequality [23, 24], there exists $C_2 > 1$ independent of initial data such that

$$(2.12) \quad \sup_{0 < y < L_2} I(y, t) \leq C_2 \inf_{0 < y < L_2} I(y, t) \quad \text{for } t \geq 1.$$

Note that we have used the fact that I satisfies a linear parabolic equation, with L^∞ bounded coefficient $\frac{1}{N} \int_0^{L_1} \beta(x, y) S(x, t) dx - \gamma(y)$, in the above. It follows that

$$\sup_{t > 1} \|I(\cdot, t)\|_{L^\infty([0, L_2])} \leq C_2 \sup_{t > 1} \int_0^{L_2} I(y, t) dy \leq C_2 N.$$

Combining with (2.10) we have

$$\sup_{t \geq 1} [\|S(\cdot, t)\|_{L^\infty([0, L_1])} + \|I(\cdot, t)\|_{L^\infty([0, L_2])}] \leq C_3.$$

By L^p estimates and Sobolev embedding, we obtain (2.11). $\square$

Definition 2.5. (i) We define Φ to be the semiflow generated by (2.3); i.e. for initial data $P_0 = (S_0, I_0) \in X$ and each $t \geq 0$ for which the solution remains in X , define $\Phi_t(P_0) = (S(\cdot, t), I(\cdot, t))$, where (S, I) is the solution of (2.3) with initial data (S_0, I_0) .
(ii) We say that Φ is point-dissipative if there exists $C > 0$ independent of initial condition such that

$$\limsup_{t \rightarrow \infty} (\|S(\cdot, t)\| + \|I(\cdot, t)\|) \leq C.$$

- (iii) We say that Φ is eventually bounded on compact subsets K of X if $\cup_{t \geq t_0} \Phi_t(K)$ is bounded for some $t_0 \geq 0$.
- (iv) For each $t > 0$, we say that $\Phi_t : X \rightarrow X$ is compact if $\Phi_t(B)$ is precompact for every bounded subset B of X .

Proposition 2.6. *The system (2.3) generates a semiflow Φ in X . Moreover, Φ is (i) point-dissipative, (ii) eventually bounded on X , and (iii) $\Phi_t : X \rightarrow X$ is compact for each $t > 1$.*

Proof. This is a consequence of (2.11). $\square$

Corollary 2.7 (Existence of compact global attractor). *The semiflow Φ has a compact attractor A of X , i.e. $\text{dist}_X(\Phi_t(X), A) \rightarrow 0$ as $t \rightarrow \infty$.*

Proof. Since Φ is point-dissipative, eventually bounded in X and that $\Phi_t : X \rightarrow X$ is compact for some $t > 0$, it follows from [25, P.41, Theorem 2.30 and Remark 2.26(b)]. $\square$

Lemma 2.8. *Let λ_1 be the principal eigenvalue of the problem*

$$(2.13) \quad \begin{cases} d_I \varphi_{yy} + \varphi [\frac{1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y)] = \lambda \varphi & \text{for } 0 < y < L_2, \\ \varphi_y = 0 & \text{for } y = 0, L_2, \end{cases}$$

then

$$\text{sgn}(\mathcal{R}_0 - 1) = \text{sgn } \lambda_1.$$

Proof. By the definition of $\mathcal{R}_0$, there exists some positive function $\phi \in C^2([0, L_2])$ such that

$$(2.14) \quad \begin{cases} -d_I \phi_{yy} + \gamma(y) \phi = \frac{1}{\mathcal{R}_0 L_1} \left(\int_0^{L_1} \beta(x, y) dx \right) \phi & \text{for } 0 < y < L_2, \\ \phi_y = 0 & \text{for } y = 0, L_2. \end{cases}$$

The principal eigenvalue λ_1 satisfies

$$(2.15) \quad \begin{cases} d_I (\varphi_1)_{yy} + \varphi_1 [\frac{1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y)] = \lambda_1 \varphi_1 & \text{for } 0 < y < L_2, \\ (\varphi_1)_y = 0 & \text{for } y = 0, L_2. \end{cases}$$

Multiplying (2.14) by φ_1 and integrating by parts in $(0, L_2)$, we have

$$(2.16) \quad d_I \int_0^{L_2} \phi_x (\varphi_1)_y dy + \int_0^{L_2} \gamma(y) \phi \varphi_1 dy = \frac{1}{\mathcal{R}_0 L_1} \int_0^{L_1} \int_0^{L_2} \beta(x, y) \phi \varphi_1 dx dy.$$

Multiplying (2.15) by ϕ and integrating by parts in $(0, L_2)$, we have

$$(2.17) \quad -d_I \int_0^{L_2} \phi_x(\varphi_1)_y dy + \int_0^{L_2} \left[\int_0^{L_1} \frac{\beta(x, y)}{L_1} dx - \gamma(y) \right] \phi \varphi_1 dy = \lambda_1 \int_0^{L_2} \phi \varphi_1 dy.$$

Adding (2.16) and (2.17), we can obtain

$$(2.18) \quad \left(1 - \frac{1}{\mathcal{R}_0}\right) \frac{1}{L_1} \int_0^{L_1} \int_0^{L_2} \beta(x, y) \phi \varphi_1 dx dy = \lambda_1 \int_0^{L_2} \phi \varphi_1 dy.$$

Since $\frac{1}{L_1} \int_0^{L_1} \int_0^{L_2} \beta(x, y) \phi \varphi_1 dx dy > 0$ and $\int_0^{L_2} \phi \varphi_1 dy > 0$, we can obtain that $\text{sgn}(\mathcal{R}_0 - 1) = \text{sgn } \lambda_1$. This proves Lemma 2.8. $\square$

Definition 2.9. (i) Define a function $\rho : X \rightarrow [0, \infty)$ by

$$\rho((S_0, I_0)) = \inf_{0 < y < L_2} I_0(y).$$

(ii) We say that the semiflow Φ is uniformly weakly ρ -persistent (resp. uniformly ρ -persistent) if there exists $\eta_0 > 0$ independent of initial data $(S_0, I_0) \in X$ such that any solution to (2.3) satisfies

$$\limsup_{t \rightarrow \infty} \rho(S(\cdot, t), I(\cdot, t)) \geq \eta_0 \quad (\text{resp.} \quad \liminf_{t \rightarrow \infty} \rho(S(\cdot, t), I(\cdot, t)) \geq \eta_0).$$

Lemma 2.10. *If $\mathcal{R}_0 > 1$, then Φ is uniformly weakly ρ -persistent.*

Proof. Suppose $\mathcal{R}_0 > 1$, by Lemma 2.8, the principal eigenvalue λ_1 of (2.13) is positive. By continuous dependence on parameter, there exists $0 < \delta_1 < 1$ such that the principal eigenvalue $\hat{\lambda}_1$ of

$$(2.19) \quad \begin{cases} d_I \varphi_{yy} + \varphi \left[\frac{1-\delta_1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y) \right] = \lambda \varphi & \text{for } 0 < y < L_2, \\ \varphi_y = 0 & \text{for } y = 0, L_2 \end{cases}$$

is positive. We denote by $\hat{\phi}_1$ a positive eigenfunction corresponding to the principal eigenvalue $\hat{\lambda}_1$ of (2.19). Let $0 < \delta_2 < 1$ be a positive number to be specified later. Suppose to the contrary that for some $t_1 \geq 2$,

$$(2.20) \quad \inf_{0 < y < L_2} I(y, t) < \frac{\delta_2 N}{C_2 L_2} \quad \text{for all } t \geq t_1,$$

where C_2 is as in (2.12). By this choice of C_2 , we obtain

$$\sup_{0 < y < L_2} I(y, t) < \frac{\delta_2 N}{L_2} \quad \text{for all } t \geq t_1.$$

Decompose $S(x, t) = \bar{S}(t) + \tilde{S}(x, t)$, where $\bar{S}(t) = \int_0^{L_1} S(x, t) dx$. Then

$$(2.21) \quad \bar{S}(t) = N - \int_0^{L_2} I(y, t) dy \geq N(1 - \delta_2) \quad \text{for } t \geq t_1 - 2,$$

and

$$\begin{cases} \tilde{S}_t - d_S \tilde{S}_{xx} = F(x, t) - \int_0^{L_1} F(x, t) dx & \text{for } 0 < x < L_1, \ t > 1, \\ \tilde{S}_x = 0 & \text{for } x = 0, L_1, \ t > 1, \end{cases}$$

where

$$F(x, t) = -\frac{S(x, t)}{N} \int_0^{L_2} \beta(x, y) I(y, t) dy + \theta(x) \int_0^{L_2} \gamma(y) I(y, t) dy.$$

Using (2.10) we have

$$\|F(x, t)\|_{L^\infty([0, L_1] \times [t_1 - 2, \infty))} \leq (2C_0 \|\beta\|_\infty + \|\theta\|_\infty \|\gamma\|_\infty) \bar{I}(t) \leq C_4 \delta_2 N.$$

Let $L = -\partial_{xx}$ and $X_2 = \{\psi \in L^2([0, L_1]) : \int_0^{L_1} \psi dx = 0\}$, then

$$\tilde{S}(\cdot, t) = e^{-tL} \tilde{S}(\cdot, 1) + \int_{t_1}^t e^{-(t-s)L} \left[F(\cdot, s) - \int_0^{L_1} F(x, s) dx \right] ds.$$

By the fact that $e^{-tL} : X_2 \rightarrow X_2$ satisfies (see, e.g. [24, Theorem 4.22])

$$\|e^{-tL}\| \leq e^{-\sigma t} \quad \text{for some } \sigma > 0,$$

we deduce that there is $t_2 \in (t_1, \infty)$ and $C_5 > 0$ such that

$$\|\tilde{S}(\cdot, t)\|_{L^2([0, L_1])} \leq C_5 \delta_2, \quad \text{for } t \geq t_2,$$

where $t_2 = t_2(\delta_2)$ but C_5 can be chosen to be independent of δ_2 . Combining with (2.21), we have

$$\int_0^{L_1} \beta(x, y) S(x, t) dx = \int_0^{L_1} \beta(x, y) dx \bar{S}(t) + \int_0^{L_1} \beta(x, y) \tilde{S}(x, t) dx$$

$$\begin{aligned}
&\geq N(1 - \delta_2) \int_0^{L_1} \beta(x, y) dx - C_6 \sqrt{\delta_2} \\
&\geq N(1 - \delta_1) \int_0^{L_1} \beta(x, y) dx \quad \text{for } t \geq t_2,
\end{aligned}$$

where δ_1 is given in (2.19). Note that this is possible by choosing δ_2 small enough, as $\inf \beta > 0$ is a fixed positive constant. We deduce that $I(y, t)$ is a supersolution of

$$\begin{cases} w_t = d_I w_{yy} + \left[\frac{1-\delta_1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y) \right] w & \text{for } 0 < y < L_2, \ t > t_2, \\ w_y = 0 & \text{for } y = 0, L_2, \ t > t_2. \end{cases}$$

By taking $\eta > 0$ small enough, we observe that $\underline{I}(y, t) = \eta e^{\hat{\lambda}_1(t-t_2)} \hat{\phi}_1(y)$ is a subsolution of the above problem with $\hat{\lambda}_1 > 0$. It follows that

$$I(y, t) \geq \eta e^{\hat{\lambda}_1(t-t_2)} \hat{\phi}_1(y) \quad \text{for } 0 < y < L_2, \ t \geq t_2.$$

This is in contradiction with (2.20). $\square$

2.1.2. Stability of the Disease-Free Equilibrium (DFE) To study the stability of the DFE, we consider an eigenvalue problem associated with (2.3). We linearize (2.3) around DFE to obtain

$$(2.22) \quad \begin{cases} \eta_t(x, t) = d_S \eta_{xx}(x, t) - \frac{1}{L_1} \int_0^{L_2} \beta(x, y) \xi(y, t) dy \\ \quad + \theta(x) \int_0^{L_2} \gamma(y) \xi(y, t) dy, & 0 < x < L_1, \ t > 0, \\ \xi_t(y, t) = d_I \xi_{yy}(y, t) + \left(\frac{1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y) \right) \xi(y, t), & 0 < y < L_2, \ t > 0. \end{cases}$$

Suppose that $(\eta, \xi) = (e^{-\lambda t} \phi, e^{-\lambda t} \psi)$ is a solution of the linear system where $\lambda \in \mathbb{R}$, $\phi = \phi(x)$, and $\psi = \psi(y)$. We substitute this solution into the linearized equations and divide by $e^{-\lambda t}$ to get the following linear eigenvalue problem

$$(2.23) \quad \begin{cases} d_S \phi_{xx}(x) - \frac{1}{L_1} \int_0^{L_2} \beta(x, y) \psi(y) dy + \theta(x) \int_0^{L_2} \gamma(y) \psi(y) dy = \lambda \phi(x), & 0 < x < L_1, \\ d_I \psi_{yy}(y) + \left(\frac{1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y) \right) \psi(y) = \lambda \psi(y), & 0 < y < L_2, \end{cases}$$

with boundary conditions

$$(2.24) \quad \phi_x(x) = 0 \quad \text{for } x \in \{0, L_1\}, \quad \text{and} \quad \psi_y(y) = 0 \quad \text{for } y \in \{0, L_2\}.$$

By (2.6), we additionally impose that

$$(2.25) \quad \int_0^{L_1} \phi(x) dx + \int_0^{L_2} \psi(y) dy = 0.$$

Lemma 2.11. *If $\mathcal{R}_0 < 1$ then the DFE is stable, but if $\mathcal{R}_0 > 1$ then it is unstable.*

Proof. **1.** Suppose first that $\mathcal{R}_0 < 1$. We will show that the DFE is linearly stable. Suppose the conclusion is false, then we can find (λ, ϕ, ψ) which is a solution of (2.23) with the condition (2.24), with at least one of ϕ and ψ not identical zero, and that $\text{Re } \lambda \geq 0$. We consider the case (i) $\psi \not\equiv 0$ and (ii) $\psi \equiv 0$.

For the first case, $0 \leq \text{Re } \lambda \leq \lambda_1$, where λ_1 is the principal eigenvalue of (2.13). But this is in contradiction with Lemma 2.8. Therefore we must have $\psi \equiv 0$ on $[0, L_2]$ and $\phi \not\equiv 0$ on $[0, L_1]$. Hence,

$$(2.26) \quad \begin{cases} d_S \phi_{xx}(x) = \lambda \phi, & 0 < x < L_1, \\ \phi_x(x) = 0, & x \in \{0, L_1\}. \end{cases}$$

It is easy to see that λ is real and nonpositive (e.g. by multiplying (2.26) with the complex conjugate of $\phi(x)$ and integrating by parts). Since also $\text{Re } \lambda \geq 0$, we deduce that $\lambda = 0$ and ϕ is a constant. But then (2.25) implies that $\phi \equiv 0$. This is again a contradiction. Therefore, we must have $\text{Re } \lambda < 0$, i.e. the DFE is linearly stable.

2. Suppose that $\mathcal{R}_0 > 1$. We will show that DFE is linearly unstable. We will establish that there exists a solution (λ, ϕ, ψ) of (2.23) with $\lambda > 0$. Lemma 2.8 implies that the principal eigenvalue $\lambda_1 > 0$ and is associated with a positive eigenfunction $\psi_1(x) > 0$. Consider the first equation of (2.23) with $(\lambda, \psi) = (\lambda_1, \psi_1)$, i.e.,

$$(2.27) \quad \begin{cases} -d_S \phi_{xx}(x) + \lambda_1 \phi = -\frac{1}{L_1} \int_0^{L_2} \beta(x, y) \psi_1(y) dy + \theta(x) \int_0^{L_2} \gamma(y) \psi_1(y) dy, & 0 < x < L_1, \\ \phi_x(x) = 0, & x \in \{0, L_1\}. \end{cases}$$

By the invertibility of the operator $-d_S \partial_{xx} + \lambda_1 I$ with zero Neumann boundary condition, (2.27) has a unique solution ϕ_1 . Hence, the triple $(\phi_1, \psi_1, \lambda_1)$ satisfies (2.23). Hence, the DFE is linearly unstable. $\square$

Next, we show that if $\mathcal{R}_0 < 1$ then the DFE is globally asymptotically stable.

Lemma 2.12. *If $\mathcal{R}_0 < 1$ then $(\tilde{S}, \tilde{I}) \rightarrow (\hat{S}, 0)$ in $C([0, L_1]) \times C([0, L_2])$ as $t \rightarrow \infty$.*

Proof. Suppose that $\mathcal{R}_0 < 1$. By the equation of (2.3), we have

$$\tilde{I}_t(y, t) \leq d_I \tilde{I}_{yy}(y, t) + \left(\frac{1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y) \right) \tilde{I}(y, t), \quad 0 < y < L_2, t > 0.$$

Set $u(y, t) = M e^{-\lambda_1 t} \psi_1$ where $\lambda_1 > 0$ by Lemma 2.8, $\psi_1 > 0$ on $[0, L_2]$, and M is chosen so large that $\tilde{I}(y, 0) < u(y, 0)$ for every $y \in [0, L_2]$. Here, $u(y, t)$ satisfies

$$(2.28) \quad \begin{cases} u_t(y, t) = d_I u_{yy}(y, t) + \left(\frac{1}{L_1} \int_0^{L_1} \beta(x, y) dx - \gamma(y) \right) u(y, t), & 0 < y < L_2, t > 0, \\ u_y(y) = 0, & y \in \{0, L_2\}. \end{cases}$$

By the comparison principle, $\tilde{I}(y, t) \leq u(y, t)$ for every $y \in [0, L_2]$ and $t > 0$. Since $u(y, t) \rightarrow 0$ as $t \rightarrow \infty$ for every $y \in [0, L_2]$, we also have that $\tilde{I}(y, t) \rightarrow 0$ as $t \rightarrow \infty$ for every $y \in [0, L_2]$.

Finally we show that $\tilde{S}$ tends to $\hat{S}$ as $t \rightarrow \infty$. Observe the first equation of (2.3)

$$\begin{aligned} \tilde{S}_t(x, t) - d_S \tilde{S}_{xx}(x, t) = & - \frac{\int_0^{L_2} \beta(x, y) \tilde{I}(y, t) dy}{\int_0^{L_1} \tilde{S}(x, t) dx + \int_0^{L_2} \tilde{I}(y, t) dy} \tilde{S}(x, t) \\ & + \theta(x) \int_0^{L_2} \gamma(y) \tilde{I}(y, t) dy, \quad 0 < x < L_1, t > 0, \end{aligned}$$

By the continuity of $\beta(x, y)$ and $\gamma(y)$ on $[0, L_2]$, together with the above argument about $\tilde{I}(y, t)$ and $u(y, t)$, we have

$$|\tilde{S}_t(x, t) - d_S \tilde{S}_{xx}(x, t)| \leq C_8 e^{-\lambda_1 t}, \quad x \in [0, L_1], t > 0,$$

for some positive constant C_8 . Since the right-hand side tends to 0 exponentially, it follows that $\tilde{S}(x, t)$ tends to some positive function $\tilde{S}^*(x)$ as $t \rightarrow \infty$,

where $\tilde{S}^*$ satisfies

$$(\tilde{S}^*)_{xx} = 0 \text{ in } [0, L_1], \quad (\tilde{S}^*)_x = 0 \text{ for } x = 0, L_1, \quad \text{and} \quad \int_0^{L_1} \tilde{S}^* dx = N.$$

Thus, $\tilde{S}^*(x) = N/L_1 = \hat{S}(x)$. $\square$

The global asymptotic stability of the DFE when $\mathcal{R}_0 < 1$ implies that there can be no EE in this case. We consider the situation when $\mathcal{R}_0 > 1$ in the next subsection.

2.1.3. Existence of Endemic Equilibrium (EE) The main result of this subsection is given by the following theorem.

Theorem 2.13. *If $\mathcal{R}_0 > 1$, then the following statements hold.*

- (a) *The infected population is uniformly ρ -persistent, i.e. there exists $\eta_0 > 0$ (independent of initial data) such that for any solution (S, I) of (2.3) such that $I_0 \not\equiv 0$, we have*

$$\liminf_{t \rightarrow \infty} \left[\inf_{0 < y < L_2} I(y, t) \right] \geq \eta_0.$$

- (b) *There exists at least one EE.*
- (c) *If, in addition, $\beta(x, y) = \beta(y)$, then EE is unique, and is globally asymptotically stable among solutions of (2.3) with initial data in X satisfying $I_0 \not\equiv 0$.*

Proof. Suppose $\mathcal{R}_0 > 1$. First, we prove (a), i.e. Φ is uniformly ρ -persistent. Now, let

$$B = \{(S_0, I_0) \in X : \|S_0\|_{C^1([0, L_1])} + \|I_0\|_{C^1([0, L_2])} \leq C_1\},$$

where C_1 is given by Lemma 2.4. Observe that B satisfies the following:

- (i) For every $P_0 \in X$ such that $\rho(P_0) > 0$, we have $\text{dist}(\Phi_t(P_0), B) \rightarrow 0$;
- (ii) B is compact;
- (iii) If $P_0 \in X$ and $\rho(P_0) > 0$, then $\rho(\Phi_t(P_0)) > 0$ for all $t > 0$;
- (iv) Φ is uniformly weakly ρ -persistent.

Here (i) and (ii) follow from Lemma 2.4, (iii) follows from the strong maximum principle applied to the second equation of (2.3), and (iv) follows from Lemma 2.10. We can then apply [25, Theorem 4.13] to conclude assertion (a), i.e. Φ is uniformly ρ -persistent.

Next, we prove (b). We have shown that (v) the semiflow Φ is uniformly ρ -persistent, (vi) $\Phi_t : X \rightarrow X$ is compact for each $t > 1$, and (vii) Φ has a compact attractor of X . Observe, in addition, that (viii) X is a closed convex subset of the Banach space $C([0, L_1]; \mathbb{R}) \times C([0, L_2]; \mathbb{R})$, and (ix) $\rho : X \rightarrow [0, \infty)$ is continuous and concave. Here concave means

$$\rho(\tau(S_1, I_1) + (1 - \tau)(S_2, I_2)) \geq \tau\rho(S_1, I_1) + (1 - \tau)\rho(S_2, I_2)$$

for $\tau \in [0, 1]$, $(S_i, I_i) \in X$. Having verified (v) - (ix), the existence of an endemic equilibrium then follows from [25, P. 158, Theorem 6.2]. This proves (b).

Suppose $\beta(x, y) = \beta(y)$. Then the equation of I can be written as

$$(2.29) \quad I_t(y, t) = d_I I_{yy}(y, t) + \beta(y) \frac{\int_0^{L_1} S dx}{N} I(y, t) - \gamma(y) I(y, t)$$

for $0 < y < L_2$, $t > 0$. Using the conservation (2.4), we observe that $I(y, t)$ in fact satisfies a single PDE with nonlocal dependence:

$$(2.30) \quad \begin{cases} I_t(y, t) = d_I I_{yy}(y, t) + \left[\beta(y) - \gamma(y) - \frac{1}{N} \int_0^{L_2} I(y', t) dy' \right] I(y, t) & \text{for } 0 < y < L_2, \ t > 0, \\ I_y(y, t) = 0 & \text{for } y = 0, L_2, \ t > 0, \\ I(y, 0) = I_0(y) & \text{for } 0 < y < L_2. \end{cases}$$

It then follows from [24, Theorem 10.1.1] that when

$$(2.31) \quad \beta(y) > \gamma(y),$$

the nonlocal parabolic equation (2.30) has a unique positive equilibrium $I_e(y)$ and moreover that $I(\cdot, t) \rightarrow I_e$ in $C([0, L_2])$ as $t \rightarrow \infty$. This proves the assertion (c) when (2.31) holds. For the general case, observe that the hypothesis (2.31) is needed to conclude that the principal eigenvalue $\hat{\lambda}_1$ of

$$(2.32) \quad d_I \phi_{yy} + (\beta - \gamma)\phi + \hat{\lambda}_1 \phi = 0 \quad \text{in } [0, L_2], \quad \phi_y(0) = \phi_y(L_2) = 0,$$

is negative. Since (2.32) is a special case of (2.13), this is equivalent to $\mathcal{R}_0 > 1$ thanks to (2.8). Once we have that $\hat{\lambda}_1 < 0$, we can then repeat the proof of [24, Theorem 10.1.1] to show the existence, uniqueness and global attractivity of equilibrium I_e .

Finally, we prove convergence of S as $t \rightarrow \infty$. By compactness, we can pass to a subsequence $t_n \rightarrow \infty$ such that $S(x, t + t_n) \rightarrow \tilde{S}(x, t)$ in $C_{loc}([0, L_1] \times \mathbb{R})$, where $\tilde{S}$ is a bounded entire solution of

$$(2.33) \quad \begin{cases} \tilde{S}_t = d_S \tilde{S}_{xx} - p(x) \tilde{S} + q(x) & \text{for } 0 < x < L_1, t \in (-\infty, \infty), \\ \tilde{S}_x(0, t) = \tilde{S}(L_1, t) = 0 & \text{for } t \in (-\infty, \infty), \end{cases}$$

where $p(x) = \frac{1}{N} \int_0^{L_2} \beta(y) I_e(y) dy$ and $q(x) = \theta(x) \int_0^{L_2} \gamma(y) I_e(y) dy > 0$. Since $\sup_{t \in \mathbb{R}} \|\tilde{S}(\cdot, t)\|_\infty$ is bounded, it follows that $\tilde{S}(x, t) = S_e(x)$, where $S_e = [-d_S \partial_{xx} + p(x)I]^{-1}[q]$. This completes the proof of (c). $\square$

2.2. Numerical simulation

In this subsection, we present some numerical results of model (2.3). Set $\tilde{t} = \delta t, d_I = \epsilon^2$, where $\delta = \tau \epsilon$ indicates the time scale parameter for a fixed τ . For the sake of simplicity in notations, we drop the tilde, with the focus on the situation when ϵ is small.

$$(2.34) \quad \begin{cases} \delta S_t(x, t) = d_S S_{xx}(x, t) - \frac{\int_0^{L_2} \beta(x, y) I(y, t) dy}{\int_0^{L_1} S(x, t) dx + \int_0^{L_2} I(y, t) dy} S(x, t) \\ \quad + \theta(x) \int_0^{L_2} \gamma(y) I(y, t) dy, & 0 < x < L_1, t > 0, \\ \delta I_t(y, t) = \epsilon^2 I_{yy}(y, t) + \frac{\int_0^{L_1} \beta(x, y) S(x, t) dx}{\int_0^{L_1} S(x, t) dx + \int_0^{L_2} I(y, t) dy} I(y, t) - \gamma(y) I(y, t), \\ & 0 < y < L_2, t > 0, \\ S_x(x, t) = I_y(y, t) = 0, & x \in \{0, L_1\}, y \in \{0, L_2\}, t > 0, \\ S(x, 0) = S_0(x) \geq 0, I(y, 0) = I_0(y) \geq 0, & 0 < x < L_1, 0 < y < L_2. \end{cases}$$

We assume that initially, the immunity level of susceptible individuals, indicated by the variable x , is uniformly distributed. In addition, vaccines can prevent infection and render breakthrough cases less infectious to avert transmission [11], so we assume that the initial distribution of infected population remains in low infectivity (indicated by y) level. Hence, we consider the initial population density functions as shown below in order to appropriately describe the above assumptions:

$$(2.35) \quad S(x, 0) = N - \int_0^1 I(y, 0) dy, \quad I(y, 0) = C_\epsilon \exp\left(-\frac{(y - y_0)^2}{\epsilon}\right) \quad \text{for some } y_0 > 0.$$

In addition, we assume the followings.

- The transmission rate, β , decreases as immunity x increases or as infectivity y decreases. For simulation purpose, we choose $\beta(x, y) = \frac{1+y}{1+ax}$, where a is a positive number.
- All infected individuals, regardless of their infectivity, will recover at the same rate. That is, the recovery rate is given by $\gamma(y) \equiv 1$.
- The immunity level of the individual changes after infection, and for simulation purpose, assumed to be distributed with probability density $\theta(x) = 2(1 - x)$. We observe that the function $\theta(x)$ has a strong influence on the long term immunity of the susceptible population. As such, it is an important quantity to monitor in practice.

The results established by Theorem 2.13 are discussed in Subsect. 2.2.1 and illustrated by numerical simulations. In Subsect. 2.2.2, we are interested in how the dynamics and equilibrium solutions exhibit the dominant phenotype under the processes of selection and mutation.

Table 2: The parameter values in model (2.34) simulation

Symbol	Relation	Value
d_S	Mutation rate of immunity x	1
ϵ^2	Mutation rate of infectivity y	0.001 ²
y_0	Mean infectivity state at $t = 0$	0.1
a	Shape parameter for transmission rate	0
N	Total population size	1
τ	Time scale parameter	100
$[0, L_1]$	The mutation space of immunity x	$[0, 1]$
$[0, L_2]$	The mutation space of infectivity y	$[0, 1]$

2.2.1. Population dynamics Point (c) of Theorem 2.13 reveals that, if $\mathcal{R}_0 > 1$, i.e. $\int_0^1 \int_0^1 \beta(x, y) dx dy > \int_0^1 \gamma(y) dy$ is satisfied, the number of susceptible and infected populations ($\bar{S}(t), \bar{I}(t)$) converge to the stable positive value. The result is illustrated by Fig. 2A which presents the total number of infected individuals $\bar{I}(t) = \int_0^1 I(y, t) dy$ and susceptible individuals $\bar{S}(t) = \int_0^1 S(x, t) dx$ against time. We can observe that the total number of infected individuals $\bar{I}(t)$ increases while the number of susceptible individuals $\bar{S}(t)$ decreases against time, eventually attaining endemic equilibrium.

2.2.2. Evolution of phenotypic heterogeneity The distribution of susceptible population density $S(x, t)$ and infected population density $I(y, t)$

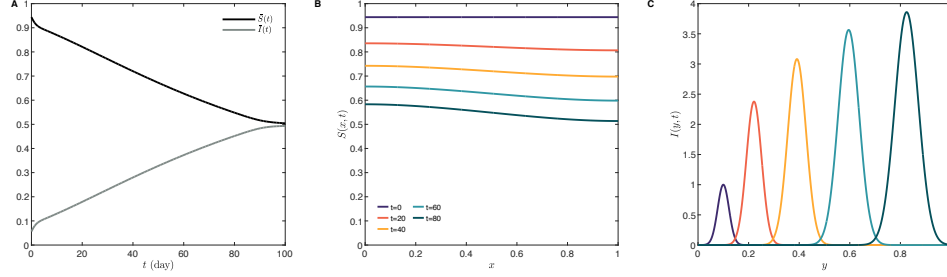


Figure 2: (A) The number of susceptible population $\bar{S}(t) = \int_0^1 S(x, t) dx$ and the number of infected population $\bar{I}(t) = \int_0^1 I(y, t) dy$ are indicated by black and grey curves, respectively. (B) Susceptible population density function $S(x, t)$ at $t = 0, 20, 40, 60, 80$ day. (C) Infected population density function $I(y, t)$ at $t = 0, 20, 40, 60, 80$ day. The value of parameters in simulation is the same with Table 2.

at various time points are shown in Figs. 2B and 2C, respectively. The immunity distribution $S(x, t)$ at a fixed time indicates that the density of susceptible individuals $S(x, t)$ increases as immunity (x) decrease. It is observed that the infectivity distribution $I(y, t)$ at a fixed time is unimodal with the mean phenotypic state $y = \bar{y}(t)$ being the maximum point of the distribution. Fig. 2C suggests that $I(y, t) \approx \bar{I}(t)\delta(y - \bar{y}(t))$, i.e., $I(y, t)$ behaves like a moving Dirac mass, supported at $y = \bar{y}(t)$.

For susceptible population, we can observe that as $\bar{S}(t)$ decreases over time, the proportion of individuals with lower immunity increases, while the proportion of individuals with higher immunity decreases as well. This finding may help explain the phenomenon wherein the immunity of susceptible people decreases in the process of increasing the infectivity of diseases. The above results for continuous time are shown in Figs. 3A and 3B.

Moreover, the mean immunity $\bar{x}(t)$ of susceptible individuals, given by

$$\bar{x}(t) = \frac{\int_0^1 xS(x, t) dx}{\int_0^1 S(x, t) dx},$$

declines against time (Fig. 3C). The mean infectivity $\bar{y}(t) = \frac{\int_0^1 yI(y, t) dy}{\int_0^1 I(y, t) dy}$ gradually rises over time, the result is depicted in Fig. 3D. Finally, the infected population is dominated by the individuals with the phenotypic state $y = 1$ (large phenotypic state correlating to the higher level of infectivity), as illustrated in Fig. 3B. In other words, the infected population evolves to be

mainly composed of highly infected individuals, in agreement with Fig. 2C. While the decline in mean immunity is possibly attributed to the choice of a decreasing $\theta(x)$, the increase in virulence seems to be a robust phenomenon independent of our choice of parameters [26]. This is possibly due to the choice of $\theta(x)$.

In addition, Fig. 3E shows the change in the heterogeneity of infectivity over time, which also indicates that $I(y, t)$ behaves like a moving Dirac mass, supported at $y = \bar{y}(t)$. Here $\bar{y}(t)$ denotes the peak location of $I(y, t)$, being an increasing function against time. This result is also in agreement with Fig. 2C and Fig. 3B and it will be further discussed in the next subsection.

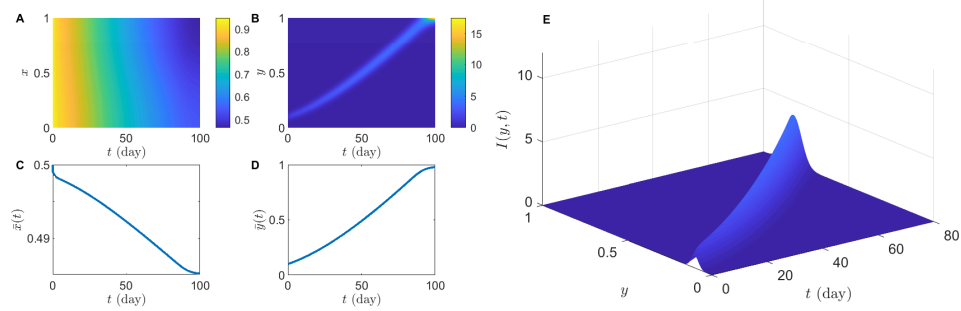


Figure 3: Yellow denotes high density and blue denotes low density, respectively. (A) Susceptible population density function $S(x, t)$ in $t \in [0, 100]$. (B) Infected population density function $I(y, t)$ in $t \in [0, 100]$. (C) Mean immunity state $\bar{x}(t)$ against time. (D) Mean infectivity state $\bar{y}(t)$ against time. (E) Infected population density function $I(y, t)$ in $t \in [0, 80]$ and $y \in [0, 1]$. The value of parameters in simulation is the same with Table 2.

2.3. Dirac asymptotics of infected populations: formal analysis

In this subsection, we formally determine the asymptotic profile of $I^\epsilon(y, t)$, where $(S^\epsilon(x, t), I^\epsilon(y, t))$ denotes solutions of model (2.34) with initial condition (2.35). For model (2.34), as in Subsect. 2.2, we choose the following formulas and parameters:

$$\beta(x, y) = \frac{1+y}{1+ax}, \quad \gamma(y) \equiv 1, \quad \theta(x) = 2(1-x), \quad L_1 = L_2 = 1, \quad \delta = \tau\epsilon.$$

Based upon those assumptions, we will formally establish the asymptotic behavior of I^ϵ as $\epsilon \rightarrow 0$. To this end, consider the WKB-ansatz

$$u^\epsilon(y, t) = -\epsilon \log I^\epsilon(y, t) \quad \text{or} \quad I^\epsilon(y, t) = \exp\left(-\frac{u^\epsilon(y, t)}{\epsilon}\right).$$

Since $\int_0^1 I^\epsilon(y, t) dy$ is uniformly bounded away from zero and infinity, we deduce that

$$\inf_{0 < y < 1} u^\epsilon(y, t) = o(1).$$

For $0 < \epsilon \ll 1$, we will formally derive the law governing the moving Dirac solution, i.e.

$$(2.36) \quad I^\epsilon(y, t) \approx \bar{I}^\epsilon(t) \delta(y - \bar{y}^\epsilon(t)).$$

Imposing (2.36) into the equation of S^ϵ , we have

$$o(1) = S_{xx}^\epsilon(x, t) + \left[-\frac{1}{N} S^\epsilon(x, t) \beta(x, \bar{y}^\epsilon(t)) + \gamma(\bar{y}^\epsilon(t)) \theta(x) \right] \bar{I}^\epsilon(t).$$

Integrating in x , and using the Neumann boundary condition, we obtain

$$(2.37) \quad \frac{1}{N} \int_0^1 S^\epsilon(x, t) \beta(x, \bar{y}^\epsilon(t)) dx \approx \gamma(\bar{y}^\epsilon(t)) \int_0^1 \theta(x) dx,$$

and

$$(2.38) \quad S^\epsilon(x, t) = \gamma(\bar{y}^\epsilon(t)) \bar{I}^\epsilon(t) \left[-\partial_{xx} + \frac{1}{N} \beta(x, \bar{y}^\epsilon(t)) \bar{I}^\epsilon(t) \right]^{-1} [\theta(\cdot)].$$

Integrating in x , and using the constraint $\bar{S}^\epsilon(t) + \bar{I}^\epsilon(t) = N$, we can determine $\bar{I}^\epsilon(t)$ by

$$(2.39) \quad N - \bar{I}^\epsilon(t) = \gamma(\bar{y}^\epsilon(t)) \bar{I}^\epsilon(t) \int_0^1 \left[-\partial_{xx} + \frac{1}{N} \beta(x, \bar{y}^\epsilon(t)) \bar{I}^\epsilon(t) \right]^{-1} [\theta(\cdot)] dx.$$

Using the form $\gamma(y) = 1$ and substituting (2.37) into the equation of I^ϵ , we obtain

$$\delta I_t^\epsilon(y, t) = \epsilon^2 I_{yy}^\epsilon(y, t) + \frac{1}{N} I^\epsilon(y, t) \left[\int_0^1 (\beta(x, y) - \beta(x, \bar{y}^\epsilon(t)) S^\epsilon(x, t) dx \right].$$

Now, we may derive the equation for the rate function $u^\epsilon(y, t) = -\epsilon \log I^\epsilon(y, t)$:

$$\begin{cases} \tau u_t^\epsilon - \epsilon u_{yy}^\epsilon + |u_y^\epsilon|^2 + \frac{1}{N} \int_0^1 (\beta(x, y) - \beta(x, \bar{y}^\epsilon(t))) S^\epsilon(x, t) dx \approx 0 \\ \quad \text{for } (y, t) \in [0, 1] \times \mathbb{R}^+, \\ \inf u^\epsilon(\cdot, t) \approx 0 \quad \text{for } t \in \mathbb{R}^+. \end{cases}$$

Suppose that $u^\epsilon(y, t) \rightarrow u(y, t)$ locally uniformly, and that $\bar{y}^\epsilon(t) \rightarrow \bar{y}(t)$, we deduce that u is a solution, in viscosity sense, to the following equation

$$(2.40) \quad \begin{cases} \tau u_t + |u_y|^2 + \frac{1}{N} \int_0^1 (\beta(x, y) - \beta(x, \bar{y}(t))) S(x, t) dx = 0 \\ \quad \text{for } (y, t) \in [0, 1] \times \mathbb{R}^+, \\ \inf u(\cdot, t) = u(\bar{y}(t), t) = 0 \quad \text{for } t \in \mathbb{R}^+, \\ u(y, 0) = (y - y_0)^2 \quad \text{for } y \in [0, 1], \end{cases}$$

with Neumann boundary condition. Here $\bar{I}(t)$ and $S(x, t)$ can be uniquely determined (for given $\bar{y}(t)$) by

$$N - \bar{I}(t) = \gamma(\bar{y}(t)) \bar{I}(t) \int_0^1 [-\partial_{xx} + \frac{1}{N} \beta(x, \bar{y}(t)) \bar{I}(t)]^{-1} [\theta(\cdot)] dx,$$

$$S(\cdot, t) = \gamma(\bar{y}(t)) \bar{I}(t) [-\partial_{xx} + \frac{1}{N} \beta(x, \bar{y}(t)) \bar{I}(t)]^{-1} [\theta(\cdot)].$$

If $u(y, t) = \sigma(t)(y - \bar{y}(t))^2 + O(|y - \bar{y}(t)|^3)$, for some $\sigma(t) > 0$, then we can differentiate the following identity

$$u_y(\bar{y}(t), t) = 0 \quad \text{for } t \geq 0,$$

to obtain

$$u_{yy}(\bar{y}(t), t) \frac{d}{dt} \bar{y}(t) = -u_{yt}(\bar{y}(t), t) = \frac{1}{\tau N} \int_0^1 \beta_y(x, \bar{y}(t)) S(x, t) dx.$$

Hence, we obtain the following equation governing the dynamics of $\bar{y}(t)$:

$$\frac{d}{dt} \bar{y}(t) = \frac{1}{\tau N u_{yy}(\bar{y}(t), t)} \int_0^1 \beta_y(x, \bar{y}(t)) S(x, t) dx.$$

Using the form $\beta(x, y) = \frac{1+y}{1+ax}$, we have

$$(2.41) \quad \frac{d}{dt} \bar{y}(t) = \frac{1}{\tau N u_{yy}(\bar{y}(t), t)} \int_0^1 \frac{1}{1+ax} S(x, t) dx > 0.$$

(2.41) indicates that $\bar{y}(t)$ is an increasing function. To support the result, we numerically solve the Hamilton-Jacobi equation (2.40). The value of parameters in simulation is the same with Table 2. Numerically we found that $\bar{y}(t)$, the peak location of $I(y, t)$, increases against time in Fig. 4A (indicated by blue curve), which is consistent with the analytical description in (2.41).

In Subsect. 2.2.2, we noted that when $\epsilon = 0.001$, the mean infectivity $\bar{y}(t)$ in model (2.34) increases over time, as seen in Fig. 3D. Now we choose $\epsilon = 0.01, 0.02, 0.03, 0.04, 0.05$ to simulate individually and the other parameters in (2.34) are the same as those in Table 2.

Let $error(\epsilon)$ be the sum of squared errors between the mean infectivity $\bar{y}^\epsilon(t)$ of model (2.34) for different ϵ and $\bar{y}(t)$ in Hamilton-Jacobi equation (2.40), that is,

$$error(\epsilon) = \|\bar{y}^\epsilon(t) - \bar{y}(t)\| = \sqrt{\sum_{i=1}^T (\bar{y}^\epsilon(t_i) - \bar{y}(t_i))^2},$$

where $t \in [0, 60]$ and the time interval is discretised with the uniform step $\Delta t = 0.01$. T indicates the number of time points.

In Fig. 4, we compare the results of $\bar{y}^\epsilon(t)$ in model (2.34) for different ϵ values with the result of Hamilton-Jacobi equation (2.40), which supports that the results of $\bar{y}^\epsilon(t)$ in model (2.34) for different ϵ values converge to the result of Hamilton-Jacobi equation (2.40) as ϵ decreases.

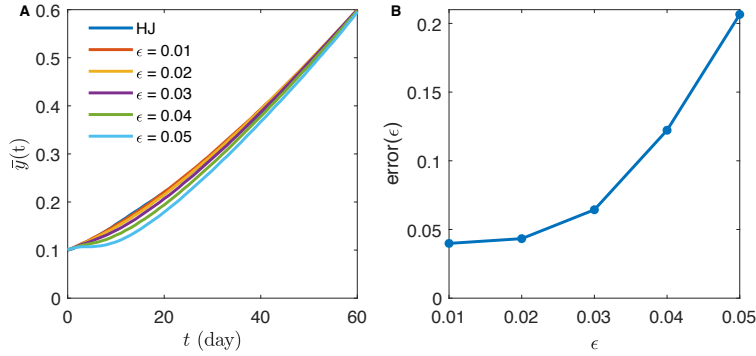


Figure 4: **The results comparison of Hamilton-Jacobi equation and model (2.34).** (A) shows $\bar{y}^\epsilon(t)$ at different ϵ values and (B) shows $error(\epsilon)$ at different ϵ values.

3. Evolution of mutation rate

In this section, we first consider k phenotypes of infected population with continuous trait y (i.e. infectivity) based upon the model (2.3) introduced in Sect. 2 and the mutation-selection model considered by Dockery et al. [27]. We assume that the phenotypes differ only in mutation rates, i.e. phenotype i has mutation rate ϵ_i^2 . $\sigma^2 M_{ij}$ represents the mutation process from phenotype i to j . When $\sigma = 0$, there is no mutation. The model has the following form:

$$(3.1) \quad \begin{cases} \frac{\partial}{\partial t} S(x, t) = d_S \frac{\partial^2}{\partial x^2} S(x, t) - \frac{\sum_{i=1}^k \int_0^{L_2} \beta(x, y) I_i(y, t) dy}{\int_0^{L_1} S(x, t) dy + \sum_{i=1}^k \int_0^{L_2} I_i(y, t) dy} S(x, t) \\ \quad + \theta(x) \sum_{i=1}^k \int_0^{L_2} \gamma(y) I_i(y, t) dy, & 0 < x < L_1, t > 0, i = 1, \dots, k, \\ \frac{\partial}{\partial t} I_i(y, t) = \epsilon_i^2 \frac{\partial^2}{\partial y^2} I_i(y, t) + \frac{\int_0^{L_1} \beta(x, y) S(x, t) dx}{\int_0^{L_1} S(x, t) dy + \sum_{i=1}^k \int_0^{L_2} I_i(y, t) dy} I_i(y, t) \\ \quad - \gamma(y) I_i(y, t) + \sigma^2 \sum_{j=1}^k M_{ji} I_j(y, t), & 0 < y < L_2, t > 0, i = 1, \dots, k, \\ \frac{\partial}{\partial x} S(x, t) = \frac{\partial}{\partial y} I_i(y, t) = 0, & x \in \{0, L_1\}, y \in \{0, L_2\}, t > 0, i = 1, \dots, k, \\ S(x, 0) = S_0 \geq 0, I_i(y, 0) = I_{i,0} \geq 0, & 0 < x < L_1, 0 < y < L_2, i = 1, \dots, k, \end{cases}$$

where $0 < \epsilon_1 < \epsilon_2 < \dots < \epsilon_k$ are constants. M is a constant $k \times k$ matrix which satisfies (i) $M_{ii} < 0$ for all i and $M_{ij} \geq 0$ for $i \neq j$; (ii) $M_{ii} = -\sum_{j \neq i} M_{ij}$. $\sigma^2 \geq 0$ is the mutation rate. Other parameters are the same with model (2.3).

The continuum (in trait) version of the mutation-selection model considered by Dockery et al. [27] is studied in [28, 29, 30, 17, 31]. Similarly, we further generalize that the rate of mutation ϵ is itself a continuous trait varying in the range $E = [J_*, J^*] \subseteq (0, \infty)$. We then obtain a continuum version of the model (3.1) as follows:

$$(3.2) \quad \left\{ \begin{array}{l} S_t(x, t) = d_S S_{xx}(x, t) - \frac{\int_{J_*}^{J^*} \int_0^{L_2} \beta(x, y) I(\epsilon, y, t) dy d\epsilon}{\int_0^{L_1} S(x, t) dx + \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, t) dy d\epsilon} S(x, t) \\ \quad + \theta(x) \int_{J_*}^{J^*} \int_0^{L_2} \gamma(y) I(\epsilon, y, t) dy d\epsilon, \quad 0 < x < L_1, J_* < \epsilon < J^*, t > 0, \\ I_t(\epsilon, y, t) = \epsilon^2 I_{yy}(\epsilon, y, t) + \frac{\int_0^{L_1} \beta(x, y) S(x, t) dx}{\int_0^{L_1} S(x, t) dx + \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, t) dy d\epsilon} I(\epsilon, y, t) \\ \quad - \gamma(y) I(\epsilon, y, t) + \sigma^2 I_{\epsilon\epsilon}(\epsilon, y, t), \quad 0 < y < L_2, J_* < \epsilon < J^*, t > 0, \\ S_x(x, t) = 0, \quad x \in \{0, L_1\}, t > 0, \\ I_y(\epsilon, y, t) = 0, \quad y \in \{0, L_2\}, J_* < \epsilon < J^*, t > 0, \\ I_\epsilon(\epsilon, y, t) = 0, \quad 0 < y < L_2, \epsilon \in \{J_*, J^*\}, t > 0, \\ S(x, 0) = S_0(x) \geq 0, \quad 0 < x < L_1, \\ I(\epsilon, y, 0) = I_0(\epsilon, y) \geq 0, \quad 0 < y < L_2, J_* < \epsilon < J^*, \end{array} \right.$$

where $S(x, t)$ denotes the density of susceptible population with immunity x at time t ; $I(\epsilon, y, t)$ indicates the density of infected population with mutation rate ϵ^2 and infectivity y at time t , respectively. We again impose the initial value when $t = 0$ and the no-flux boundary condition in all other variables.

We now rescale time as $\tilde{t} = \delta t$, where $\delta = \tau\sigma$ indicates the time scale parameter for a fixed τ . To be consistent with model (2.34), we drop the tilde and obtain the following model:

$$(3.3) \quad \left\{ \begin{array}{l} \delta S_t(x, t) = d_S S_{xx}(x, t) - \frac{\int_{J_*}^{J^*} \int_0^{L_2} \beta(x, y) I(\epsilon, y, t) dy d\epsilon}{\int_0^{L_1} S(x, t) dx + \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, t) dy d\epsilon} S(x, t) \\ \quad + \theta(x) \int_{J_*}^{J^*} \int_0^{L_2} \gamma(y) I(\epsilon, y, t) dy d\epsilon, \quad 0 < x < L_1, J_* < \epsilon < J^*, t > 0, \\ \delta I_t(\epsilon, y, t) = \epsilon^2 I_{yy}(\epsilon, y, t) + \frac{\int_0^{L_1} \beta(x, y) S(x, t) dx}{\int_0^{L_1} S(x, t) dx + \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, t) dy d\epsilon} I(\epsilon, y, t) \\ \quad - \gamma(y) I(\epsilon, y, t) + \sigma^2 I_{\epsilon\epsilon}(\epsilon, y, t), \quad 0 < y < L_2, J_* < \epsilon < J^*, t > 0, \end{array} \right.$$

where the boundary conditions and initial conditions are the same as in (3.2).

Similar to the definitions in the model (2.34), we define the size of the susceptible and infected populations at time t , respectively, as follows:

$$\bar{S}(t) = \int_0^{L_1} S(x, t) dx, \quad \bar{I}(t) = \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, t) dy d\epsilon.$$

We also define the mean phenotype state at time t as

$$\bar{x}(t) = \frac{\int_0^{L_1} x S(x, t) dx}{\int_0^{L_1} S(x, t) dx},$$

$$\bar{y}(t) = \frac{\int_0^{L_2} \int_{J_*}^{J^*} y I(\epsilon, y, t) d\epsilon dy}{\int_0^{L_2} \int_{J_*}^{J^*} I(\epsilon, y, t) d\epsilon dy}, \quad \bar{\epsilon}(t) = \frac{\int_{J_*}^{J^*} \int_0^{L_2} \epsilon I(\epsilon, y, t) dy d\epsilon}{\int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, t) dy d\epsilon}.$$

We assume that the initial infected population composed of low infectivity (y) and low mutation rate (ϵ) individuals, while susceptible population composed of individuals with uniformly distributed immunity. Let the initial population density functions be given by

$$S(x, 0) = N - \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, 0) dy d\epsilon,$$

$$I(\epsilon, y, 0) = \exp\left(-\frac{(y - y_0)^2}{\epsilon} - \frac{(\epsilon - \epsilon_0)^2}{\sigma}\right)$$

and fix an $N > 0$ such that

$$\int_0^{L_1} S(x, 0) dx + \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, 0) dy d\epsilon = N,$$

where y_0, ϵ_0 are positive numbers.

Recall (2.4), for $t > 0$, we can obtain the total population size

$$\int_0^{L_1} S(x, t) dx + \int_{J_*}^{J^*} \int_0^{L_2} I(\epsilon, y, t) dy d\epsilon \equiv N.$$

Then we present the numerical simulation results of the model (3.3). The choices of disease transmission rate $\beta(x, y)$, recovery rate $\gamma(y)$, and probability distribution $\theta(x)$ are identical to those in Subsect. 2.2. The value of parameters in simulation are also identical with Table 3. The numerical results can be summarized as follows.

- The susceptible population evolves to be primarily made up of individuals with low immunity, which indicates that as diseases continue to evolve, group immunity decreases. It could be important to boost group immunity through techniques like ongoing immunization.
- The majority of the infected population is eventually comprised of highly infectious individuals with a low mutation rate. This indicates that the disease transmissibility will continue to increase.

- When the ceiling of virulence is reached in the model, disease strains with low mutation are more advantageous, assuming that the mutation rates have no impact on disease transmission and recovery rates. In other words, the virus will use the low mutation rate to maintain its advantageous evolutionary position. We do note that death caused by the disease is not included in the model. In fact, highly virulent strains may not be selected if death rate is included.

These findings are discussed and illustrated by numerical simulation in the following subsections. Because virus mutations can persist in human population for years or decades [9], the time length of the simulation is set at 120 months.

Table 3: The parameter values in model (2.34) simulation

Symbol	Relation	Value
y_0	Mean infectivity state at $t = 0$	0.3
ϵ_0	Mean mutation rate at $t = 0$	0.1
d_S	Mutation rate of immunity x	1
σ^2	Mutation rate of trait ϵ	0.01^2
N	Total population size	1
τ	Time scale parameter	1500
$[0, L_1]$	The mutation space of immunity x	$[0, 1]$
$[0, L_2]$	The mutation space of infectivity y	$[0, 10]$
$[J_*, J^*]$	The mutation space of mutation rate ϵ	$[0.01, 0.5]$

3.1. Convergence to endemic equilibrium

The number of infected individuals $\bar{I}(t)$ and susceptible individuals $\bar{S}(t)$ throughout time is depicted in Fig. 5. We can observe that the number of infected individuals $\bar{I}(t)$ increases while the number of susceptible individuals $\bar{S}(t)$ decreases against time, until reaching the endemic equilibrium (EE).

3.2. Evolution of phenotypic heterogeneity

The distribution of susceptible population density $S(x, t)$ and mean immunity state $\bar{x}(t)$ in $t \in [0, 120]$ are displayed in Fig. 6A. The mean immunity of susceptible individuals declines $\bar{x}(t)$ over time, and then stabilizes at a low level. It appears that the distribution of immunity level in susceptible population reaches an equilibrium distribution resembling the function $\theta(x) = 2(1 - x)$.

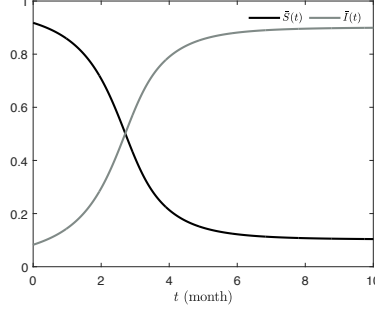


Figure 5: **The populations size.** The number of susceptible population $\bar{S}(t)$ and the number of infected population $\bar{I}(t)$ against time are indicated by black and grey curves, respectively. All parameter values are the same as in Table 3.

Similar to the results in model (2.34), when $\bar{S}(t)$ declines over time, susceptible individuals with lower immunity tend to become susceptible again, which reduces the average immunity of susceptible population. This finding highlights the importance of tracking the change of individual immunity post infection. When current infection does not provide better immunity against future infection (e.g. when $\theta(x)$ is decreasing in x), vaccination is a way to improve the overall immunity level.

The mean infectivity level and the mean mutation rate in infected population I both increase rapidly in the first phase $0 \leq t \leq 10$. These results (indicated by yellow curves) are shown in Figs. 6B and 6C, where the heat map of $\hat{I}_1(y, t) = \int_{J_*}^{J^*} I(\epsilon, y, t) d\epsilon$ and $\hat{I}_2(\epsilon, t) = \int_0^{L_2} I(\epsilon, y, t) dy$ are displayed. That is, the number of individuals with high infectivity and mutation rate increase rapidly at the initial time. This phenomenon can also be illustrated in Figs. 7A-D, which depict the density distribution of infected population $I(\epsilon, y, t)$ at $t = 0, 2.5, 5, 10$ month.

For $t > 10$, we observe in Figs. 6B and 6C that the infected individuals maintain high infectivity state, while the mutation rate of infected individuals starts to decrease over time. These results are consistent with Figs. 7E-H, which are the density distribution of infected population at $t = 15, 25, 40, 50$ month.

Fig. 6C indicates that the increase in mean mutation rate $\bar{\epsilon}(t)$ is only an initial transient: when $t > 10$, infected individuals with low mutation rate will gradually dominate the population. Fig. 8 supports the conclusion that lower mutation rates provide a competitive advantage. That is, the infected

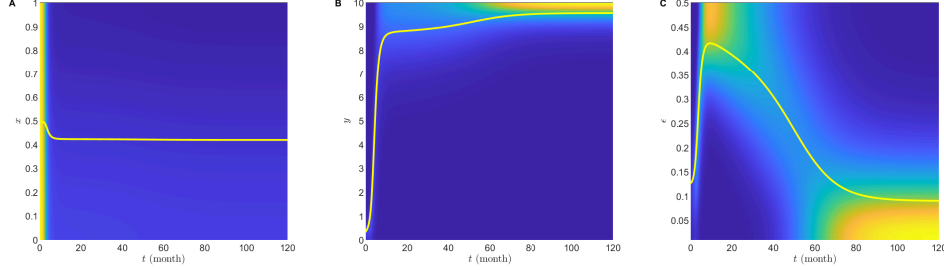


Figure 6: (A) Susceptible population density $S(x, t)$ in $t \in [0, 120]$. (B) The infectivity distribution of infected individuals $\hat{I}_1(y, t) = \int_{J_*}^{J_*^*} I(\epsilon, y, t) d\epsilon$ in $t \in [0, 120]$. (C) The mutation rate distribution of infected individuals $\hat{I}_2(\epsilon, t) = \int_0^{L_2} I(\epsilon, y, t) dy$ in $t \in [0, 120]$. Yellow denotes high density and blue denotes low density, respectively. The yellow line highlights the mean immunity $\bar{x}(t)$ of susceptible population, mean infectivity $\bar{y}(t)$ and mean mutation rate $\bar{\epsilon}(t)$ of infected population, respectively. All parameter values are the same as in Table 3.

population gradually evolves to be mainly composed of individuals with low mutation rate.

4. Discussion

In this work, we propose to study a Susceptible-Infected-Susceptible (SIS) model (2.3) for populations structured by phenotypical traits. Our analytical results demonstrate that the existence of endemic equilibrium when the basic reproduction number $\mathcal{R}_0 > 1$. Based upon asymptotic analysis of evolutionary dynamics, the simulation results generated from model (2.34) indicate that the mean immunity of susceptible individuals gradually decreases over time and eventually remains at a low level, while high infectivity individuals have competitive advantage. There is good agreement between our numerical simulations and analytical results. The results of our analysis and numerical simulations are consistent with the trend that more virulent strains are selected over the less virulent wild type.

Furthermore, we consider an extension (3.2) of the model (2.3) to account for the variability in mutation rate. Numerical simulations are carried out to demonstrate that, if the virus mutation rate is low initially, it will increase rapidly and then decrease slowly over time, resulting in the individuals with the lowest mutation rate eventually dominating the infected population, assuming the virulence ceiling have been reached.

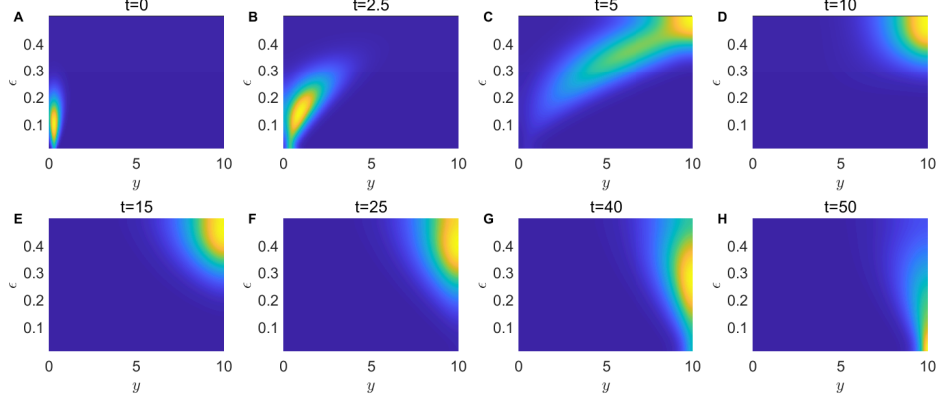


Figure 7: Infected population density function $I(\epsilon, y, t)$ at different time points. Consider Fig. 7B as an illustration, the mean infectivity $\bar{y}(t)$ and mean mutation rate $\bar{\epsilon}(t)$ increase compared to Fig. 7A ($t = 0$). The maximum point of infected population density $I(\epsilon, y, t)$ is attained in phenotypic state $(y, \epsilon) = (0.9, 0.14)$. Yellow denotes high density and blue denotes low density, respectively. All parameter values are the same as in Table 3.

In model (3.2), we impose that the reaction term be independent of mutation rate ϵ^2 , i.e., the correlation between mutation rate ϵ^2 and disease transmission rate $\beta(x, y)$ and recovery rate $\gamma(y)$ are not taken into account. The assumption may not hold true in actual pandemic scenarios. The advantage of a low mutation rate may be strengthened or diminished by the current lack of clarity regarding the relationship between mutation rate, disease transmission rate and recovery rate.

We also discuss several possible extensions of current work. From the perspective of mathematical modeling, it should be noted that our model does not examine the spatial distribution of population. Spatial movement of populations could play major role in speeding up the disease spreading. Moreover, spatial environmental heterogeneity of population could also impact disease spread. Our current model somewhat mimics the scenario when the populations are quickly mixed in space, which may introduce error in our estimate of the basic reproduction number. While current work is focused on the dynamics of susceptible and infected individuals, we leave the role of exposed individuals and asymptomatic individuals in disease dynamics for future work. In this work we did not include the death caused by the disease so that the total population size remains to be constant in time, which

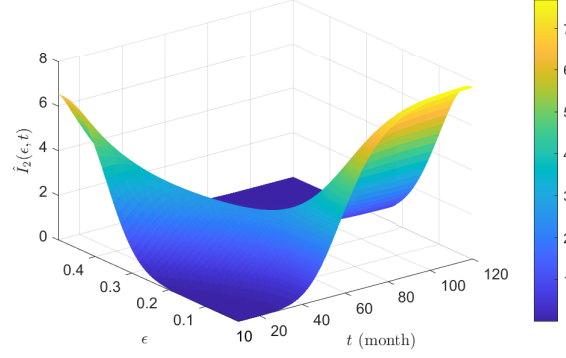


Figure 8: **The evolution of mutation rate ϵ .** The mutation rate distribution of infected individuals $\hat{I}_2(\epsilon, t) = \int_0^{L^2} I(\epsilon, y, t) dy$ in $t \in [10, 120]$. The infected sub-population with smallest mutation rate dominates the dynamics for large t . Yellow denotes high density and blue denotes low density, respectively. All parameter values are the same as in Table 3.

is accessible for mathematical analysis. Biologically, while COVID has significant death rate in the early stage of pandemics, the current death rate caused by the Omicron strain seems to be decreasing dramatically, so our model might also help provide some insight into the current development of COVID pandemics. From an epidemiological perspective, the model may further consider the recruitment rate and mortality rate of populations, and by integrating with the actual biological process and epidemiological data, to achieve more accurate modeling of the epidemic. Last but not least, we point out that the disease transmission rate $\beta(x, y)$ is assumed to be stationary in time in this work. In epidemiology, the disease transmission rate is usually time-dependent, and the time dependence corresponds to a variety of factors, including changes in non-pharmaceutical interventions (NPIs), behavioral changes and seasonal changes (influenced by weather or human population migration schedules) [11].

Acknowledgements

This work is partially supported by NSF grant DMS-1853561 and DMS-2325195 (KYL), NSFC grants No. 12250710674, 12261160366 (YL), NSFC grant No. 12071476, 12171478 (SZM).

References

- [1] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020) A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*.
- [2] Dong, E., Du, H., and Gardner, L. (2020) An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, **20**(5), 533–534.
- [3] Yewdell, J. W. (2021) Antigenic drift: understanding COVID-19. *Immunity*, **54**(12), 2681–2687.
- [4] Wang, R., Chen, J., Hozumi, Y., Yin, C., and Wei, G.-W. (2022) Emerging vaccine-breakthrough SARS-CoV-2 variants. *ACS Infectious Diseases*, **8**(3), 546–556.
- [5] Markov, P. V., Katzourakis, A., and Stilianakis, N. I. (2022) Antigenic evolution will lead to new SARS-CoV-2 variants with unpredictable severity. *Nature Reviews Microbiology*, **20**(5), 251–252.
- [6] Telenti, A., Hodcroft, E. B., and Robertson, D. L. (2022) The evolution and biology of SARS-CoV-2 variants. *Cold Spring Harbor Perspectives in Medicine*, **12**(5), a041390.
- [7] Callaway, E. Beyond Omicron: what’s next for COVID’s viral evolution. (2021).
- [8] Lippi, G., Mattiuzzi, C., and Henry, B. M. (2022) Updated picture of SARS-CoV-2 variants and mutations. *Diagnosis*, **9**(1), 11–17.
- [9] Jackson, C. B., Farzan, M., Chen, B., and Choe, H. (2022) Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology*, **23**(1), 3–20.
- [10] Liu, L., Iketani, S., Guo, Y., Chan, J. F.-W., Wang, M., Liu, L., Luo, Y., Chu, H., Huang, Y., Nair, M. S., et al. (2022) Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature*, **602**(7898), 676–681.
- [11] Koelle, K., Martin, M. A., Antia, R., Lopman, B., and Dean, N. E. (2022) The changing epidemiology of SARS-CoV-2. *Science*, **375**(6585), 1116–1121.
- [12] Gomes, M. G. M., Ferreira, M. U., Corder, R. M., King, J. G., Souto-Maior, C., Penha-Gonçalves, C., Gonçalves, G., Chikina, M., Pegden,

- W., and Aguas, R. (2022) Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *Journal of Theoretical Biology*, **540**, 111063.
- [13] Neipel, J., Bauermann, J., Bo, S., Harmon, T., and Jülicher, F. (2020) Power-law population heterogeneity governs epidemic waves. *PLoS One*, **15**(10), e0239678.
- [14] Großmann, G., Backenköhler, M., and Wolf, V. (2021) Heterogeneity matters: Contact structure and individual variation shape epidemic dynamics. *PLoS One*, **16**(7), e0250050.
- [15] Miranda, J. G. V., Silva, M. S., Bertolino, J. G., Vasconcelos, R. N., Cambui, E. C. B., Araújo, M. L. V., Saba, H., Costa, D. P., Duverger, S. G., de Oliveira, M. T., et al. (2021) Scaling effect in COVID-19 spreading: The role of heterogeneity in a hybrid ODE-network model with restrictions on the inter-cities flow. *Physica D: Nonlinear Phenomena*, **415**, 132792.
- [16] Allen, L. J., Bolker, B. M., Lou, Y., and Nevai, A. L. (2008) Asymptotic profiles of the steady states for an SIS epidemic reaction-diffusion model. *Discrete & Continuous Dynamical Systems*, **21**(1), 1–20.
- [17] Magal, P. and Webb, G. (2000) Mutation, selection, and recombination in a model of phenotype evolution. *Discrete & Continuous Dynamical Systems*, **6**(1), 221–236.
- [18] Kermack, W. O. and McKendrick, A. G. (1927) A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **115**(772), 700–721.
- [19] Kermack, W. O. and McKendrick, A. G. (1932) Contributions to the mathematical theory of epidemics. II.—The problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **138**(834), 55–83.
- [20] Hethcote, H. W. (1978) An immunization model for a heterogeneous population. *Theoretical Population Biology*, **14**(3), 338–349.
- [21] Lei, J. (2020) A general mathematical framework for understanding the behavior of heterogeneous stem cell regeneration. *Journal of Theoretical Biology*, **492**, 110196.
- [22] Lieberman, G. M. (1996) Second order parabolic differential equations, World Scientific Publishing Co., Inc., River Edge, NJ.

- [23] Húska, J. (2006) Harnack inequality and exponential separation for oblique derivative problems on Lipschitz domains. *J. Differential Equations*, **226**(2), 541–557.
- [24] Lam, K.-Y. and Lou, Y. (2022) Introduction to Reaction-Diffusion Equations: Theory and Applications to Spatial Ecology and Evolutionary Biology, Lecture Notes on Mathematical Modelling in the Life Sciences, Springer, Cham.
- [25] Smith, H. L. and Thieme, H. R. (2011) Dynamical systems and population persistence, Vol. 118 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI.
- [26] van Baalen, M. and Sabelis, M. W. (1995) The dynamics of multiple infection and the evolution of virulence. *The American Naturalist*, **146**(6), 881–910.
- [27] Dockery, J., Hutson, V., Mischaikow, K., and Pernarowski, M. (1998) The evolution of slow dispersal rates: a reaction diffusion model. *Journal of Mathematical Biology*, **37**(1), 61–83.
- [28] Lam, K.-Y. and Lou, Y. (2017) An integro-PDE model for evolution of random dispersal. *Journal of Functional Analysis*, **272**(5), 1755–1790.
- [29] Perthame, B. (2006) Transport equations in biology, Springer Science & Business Media.
- [30] Perthame, B. and Souganidis, P. E. (2016) Rare mutations limit of a steady state dispersal evolution model. *Mathematical Modelling of Natural Phenomena*, **11**(4), 154–166.
- [31] Lam, K.-Y., Lou, Y., and Perthame, B. (2023) A Hamilton-Jacobi approach to evolution of dispersal. *Communications in Partial Differential Equations*, **48**(1), 86–118.

KING-YEUNG LAM

DEPARTMENT OF MATHEMATICS, OHIO STATE UNIVERSITY
COLUMBUS, OH 43210, USA

E-mail address: Lam.184@osu.edu

YUAN LOU

SCHOOL OF MATHEMATICAL SCIENCES, CMA-SHANGHAI
SHANGHAI JIAO TONG UNIVERSITY
SHANGHAI 200240, CHINA

E-mail address: yuanlou@sjtu.edu.cn

SHIZHAO MA
INSTITUTE FOR MATHEMATICAL SCIENCES
RENMIN UNIVERSITY OF CHINA
BEIJING 100872, CHINA
E-mail address: shizhaoma@ruc.edu.cn

RECEIVED 26 JUNE 2023