CSIAM Trans. Life Sci.
doi: 10.4208/csiam-ls.SO-2024-0001

Vol. **1**, No. 1, pp. 1-21
March 2025

Modelling the Effect of Human Heterogeneity on Infectious Disease Transmission Dynamics

Pengfei Song[1], Jianhong Wu[2] and Yanni Xiao[1,*]

[1] *School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, ShaanXi 710049, P.R. China.*
[2] *Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Ontario, Toronto, Canada.*

Received 13 December 2023; Accepted 22 April 2024

Abstract. Human heterogeneity is a critical issue in infectious disease transmission dynamics modelling, and it has recently received much attention in COVID-19 studies. In this article, a general human heterogeneous disease model with mutation is proposed to comprehensively study the effects of human heterogeneity on basic reproduction number, final epidemic size and herd immunity. We show that human heterogeneity may increase or decrease herd immunity level, strongly depending on some convexity of the heterogeneity function, which gives new insights and extends the results in [Britton *et al.*, Science, 369:846–849, 2020]. Moreover, human heterogeneity may decrease the basic reproduction number but increase the level of herd immunity, implying the unreliability of the basic reproduction number in characterizing the spread and control of infectious diseases with human heterogeneity.

AMS subject classifications: 35J55, 35B32

Key words: SEIR model, human heterogeneity, basic reproduction number, herd immunity level, final epidemic size.

1 Introduction

Human heterogeneity is ubiquitous and shares tremendous popularity in the study of social science [39] and epidemiology [4–9, 13–17, 19, 23–25, 28, 30, 37]. There are many heterogeneities in human societies that will influence virus transmission, such as social activity level, age structure, incubation period, individual susceptibility or exposure to infection. Recently, the effects of human heterogeneities on infectious disease are back to spotlight in modeling and precise control of COVID-19 spread. Questions of interest

*Corresponding author. *Email addresses:* song921012@xjtu.edu.cn (P. Song), wujh@yorku.ca (J. Wu), yxiao@mail.xjtu.edu.cn (Y. Xiao)

to mathematical epidemiologists and public health members are how the herd immunity level is affected by human heterogeneities such as susceptibility and infectivity [7,19], the limitations of basic reproduction number and what human heterogeneity can tell [31], why final epidemic size of COVID-19 is so different from the classic SIR model, and why ODE models might fail in modelling COVID-19 and how human heterogeneity shapes the evolution of infectious disease [20]. These questions imply the necessity to incorporate human heterogeneities in disease models to profoundly understand the transmission mechanism of infectious diseases.

Many studies have shown that under certain conditions, human heterogeneities will decrease herd immunity level [6, 7, 19, 25]. It is found in [6, 25] that when susceptibility is the only variable property, the final size of the epidemic is always smaller for a heterogeneous population than for a homogeneous population with the same reproduction number. A recent study on COVID-19 showed that population heterogeneity can significantly impact disease-induced immunity [19]. They estimated that if the basic reproduction number $\mathcal{R}_0 = 2.5$ in an age-structured community with mixing rates adapted to social activity, then the disease-induced herd immunity level can be around 43%, which is substantially less than the classical herd immunity level of 60% obtained through homogeneous immunization of the population. Another study [7] showed that if the coefficient of heterogeneity variation increases from 0 to 4, the herd immunity threshold decreases from greater than 60% to less than 10%.

However, some researchers is probably impossible even with vaccination efforts in full force and the theoretical threshold for vanquishing COVID-19 looks to be out of reach, due to the reasons such as vaccination-induced human behaviour change and mutation of viruses [2]. The effects of human heterogeneity on the spread of infectious disease are comprehensive and repay many more studies from theoretical insights. In this paper, we will consider some SEIR model with mutation and heterogeneous populations, in order to investigate the effect of human heterogeneities such as activity level, susceptibility, exposure to infection on basic reproduction number, final epidemic size and herd immunity.

The rest of the paper is organized as follows. In Section 2, we will introduce the general human heterogeneous SEIR disease model with mutation, define the basic reproduction number, and investigate the threshold dynamics. The effect of human heterogeneities on disease persistence, herd immunity threshold and final epidemic size will be explored in Section 3. Applications to COVID-19 can be found in Section 4. Discussion and conclusion will be in Section 5.

2 A general disease model of human heterogeneities

The population is divided into four classes: susceptible (S), exposed (E, latently infected without symptoms), infected (I, infected with symptoms) and removed (R, recovered or disease-induced death). The susceptible individuals are infected by exposed and infec-

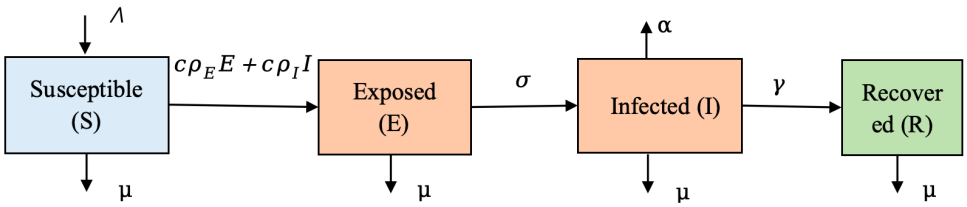

Figure 1: The susceptible individuals are infected by exposed and infected individuals and become exposed, each exposed individual becomes infected individuals with the rate σ, infected individuals are recovered with a rate γ. Λ and μ are natural birth and death rates, respectively. α is disease induced death rate of infected individuals.

tious individuals and become exposed; exposed individuals become infected; infected individuals are recovered or die of disease and become removed, see Fig. 1.

To incorporate human heterogeneities, we assume that the population comprises a continuum family of phenotypes differentiated by some kind of trait $x \in [a,b]$. Human heterogeneities may happen in

- activity level $c(x)$: the number of contacts per unit time of x-type population $N(x)$. Activity level is concerned with genders, age structures, or jobs of the population;
- susceptibility to infection $\rho_S(x)$: the probability that x-type susceptible individuals can be infected. It is concerned with the health condition of the susceptible population;
- exposure to symptomatic $\rho_I(x)$ and asymptomatic infection $\rho_E(x)$: the probability that x-type infected and exposed individuals can infect susceptible individuals, respectively;
- transition rate from E to I $(\sigma(x))$: $1/\sigma(x)$ may satisfy some kind of probability distribution such as lognormal distribution, gamma distribution, or Weibull distribution;
- transition rate from I to R $(\gamma(x))$;
- disease-induced rate $\alpha(x)$;
- natural birth $\Lambda(x)$ and death rate $\mu(x)$.

We call these heterogeneity attributes of population and denote

$$\mathbf{\Theta}(x) := \left\{ c(x), \rho_S(x), \rho_E(x), \rho_I(x), \frac{1}{\sigma(x)}, \frac{1}{\gamma(x)}, \alpha(x), \Lambda(x), \mu(x) \right\}$$

as the attribute values of x-type population. Different phenotypes have different attribute values, i.e.

$$\mathbf{\Theta}(x) \neq \mathbf{\Theta}(y) \quad \text{if} \quad x \neq y.$$

The trait x can be defined as some kind of attribute (heterogeneity) of the population, for example, $c(x)=x$ means that the population is differentiated by the activity level, $\sigma(x)=1/x$ means that the population is differentiated by the transition rate from E to I of x type population. In real applications, one does not need to consider all kinds of heterogeneity, so some attributes can be independent of the trait x.

We assume that the population has overlapping generations such that the mutation is modeled by a diffusion process with constant rate $\epsilon\geq 0$ acting on the phenotypic trait variable. Then the SEIR model cooperated with heterogeneity read as follows:

$$\begin{cases} \dfrac{\partial S}{\partial t}=\Lambda(x)-\mathrm{Foi}(x)S-\mu(x)S+\epsilon S_{xx}, & x\in(a,b), \quad t>0, \\ \dfrac{\partial E}{\partial t}=\mathrm{Foi}(x)S-\sigma(x)E-\mu(x)E+\epsilon E_{xx}, & x\in(a,b), \quad t>0, \\ \dfrac{\partial I}{\partial t}=\sigma(x)E-\gamma(x)I-\mu(x)I-\alpha(x)I+\epsilon I_{xx}, & x\in(a,b), \quad t>0, \\ \dfrac{\partial R}{\partial t}=\gamma(x)I+\alpha(x)I-\mu(x)R+\epsilon R_{xx}, & x\in(a,b), \quad t>0, \\ S_x=E_x=I_x=R_x=0, & x=a,b, \quad t>0. \end{cases} \tag{2.1}$$

Here

$$\mathrm{Foi}(x)=\frac{c(x)\rho_S(x)\displaystyle\int_a^b c(y)[\rho_I(y)I(y,t)+\rho_E(y)E(y,t)]dy}{\displaystyle\int_a^b c(y)N(y,t)dy}$$

is the force of infection upon x-type susceptible individuals in the total population, where

$$N(x,t)=S(x,t)+E(x,t)+I(x,t)+R(x,t)$$

denotes the x-type population. Recall that $c(x)$ is the number of contacts per unit time; $\rho_S(x)$ is the probability that x-type susceptible individuals can be infected; $\rho_I(x)$ and $\rho_E(x)$ are the probability of symptomatic (I) and asymptomatic infections (E), respectively; $c(y)I(y,t)/\int_a^b c(y)N(y,t)dy$ and $c(y)E(y,t)/\int_a^b c(y)N(y,t)dy$ are the probabilities that each person meets the population infected and exposed population, respectively.

Throughout this paper, we denote

$$p(x,t)=\frac{N(x,t)}{\displaystyle\int_a^b N(x,t)dx}$$

the frequency of x-type population at time t, and we write $p(x,0)$ as $p(x)$ for the sake of notational simplicity. Define

$$<f>:=\int_a^b f(x)p(x)dx, \quad \bar{x}:=<x>=\int_a^b xp(x)dx,$$

and the variance of heterogeneity

$$\text{Var} := \int_a^b (x-\bar{x})^2 p(x)dx = < x^2 > - < x >^2 . \tag{2.2}$$

2.1 Equilibrium problem

This paper also concerns non-negative equilibrium solutions of (2.1) which satisfy

$$\begin{cases} \Lambda(x) - \text{Foi}(x)\widetilde{S} - \mu(x)\widetilde{S} + \epsilon\widetilde{S}_{xx} = 0, & x \in (a,b), \\ \text{Foi}(x)\widetilde{S} - \sigma(x)\widetilde{E} - \mu(x)\widetilde{E} + \epsilon\widetilde{E}_{xx} = 0, & x \in (a,b), \\ \sigma(x)\widetilde{E} - \gamma(x)\widetilde{I} - \mu(x)\widetilde{I} - \alpha(x)\widetilde{I} + \epsilon\widetilde{I}_{xx} = 0, & x \in (a,b), \\ \gamma(x)\widetilde{I} + \alpha(x)\widetilde{I} - \mu(x)\widetilde{R} + \epsilon\widetilde{R}_{xx} = 0, & x \in (a,b), \\ \widetilde{S}_x = \widetilde{E}_x = \widetilde{I}_x = \widetilde{R}_x = 0, & x = a,b, \\ \text{Foi}(x) = \dfrac{c(x)\rho_S(x)\displaystyle\int_a^b c(y)\left[\rho_I(y)\widetilde{I}(y) + \rho_E(y)\widetilde{E}(y)\right]dy}{\displaystyle\int_a^b c(y)\widetilde{N}(y)dy}, \end{cases} \tag{2.3}$$

where $\widetilde{S}, \widetilde{E}, \widetilde{I}, \widetilde{R}$ denote the density of susceptible, exposed, infected and removed individuals at equilibrium, respectively. A disease-free equilibrium (DFE) is a solution of (2.3) satisfying $\widetilde{I}(x) = 0$ for every $x \in [a,b]$. An endemic equilibrium (EE) is a solution of (2.3) for which $\widetilde{I}(x) > 0$ for some $x \in [a,b]$. It is easy to verify that the disease free equilibrium is unique, given by $E_0 = (N_0(x), 0, 0, 0)$, where $N_0(x)$ is the unique solution of the following system:

$$\Lambda(x) - \mu(x)\widetilde{S} + \epsilon\widetilde{S}_{xx} = 0, \quad x \in (a,b), \quad S_x(a) = S_x(b) = 0.$$

By the strong maximum principle [18], for any endemic equilibrium, $\widetilde{S}(x), \widetilde{E}(x), \widetilde{I}(x), \widetilde{R}(x)$ are positive for any $x \in [a,b]$.

2.2 The basic reproduction number and threshold dynamics

For infectious disease models, the basic reproduction number, defined as the expected number of secondary cases produced in a completely susceptible population by an infective individual, is one of the most significant concepts in studying the transmission of infectious disease [1, 11]. More importantly, it often determines the threshold behavior for many epidemic models. It is often the case that a disease dies out if the basic reproduction number is less than unity and the disease is established in the population if it is greater than unity. More importantly, the basic reproduction number is also used to characterize the final epidemic size and herd immunity threshold $(1 - 1/\mathcal{R}_0)$. We refer to [12] for the approach of next generation operators for the basic reproduction number and to [27, 36, 38, 42] for related studies.

To define the basic reproduction number of system (2.1), we appeal to the next generation operators theory developed in [12]. Note that exposed (E) and infected (I) classes are the infected compartments of the system. Set $X_1=C([a,b];\mathbb{R}^2)$ and $X_1^+=C([a,b];\mathbb{R}^2_+)$. Let $T(t)$ be the solution semigroup on X_1 of the following system:

$$\begin{cases} \dfrac{\partial u_E}{\partial t}=-\sigma(x)u_E-\mu(x)u_E+\epsilon(u_E)_{xx}, & x\in(a,b), \quad t>0, \\ \dfrac{\partial u_I}{\partial t}=\sigma(x)u_E-\gamma(x)u_I-\mu(x)u_I-\alpha(x)u_I+\epsilon(u_I)_{xx}, & x\in(a,b), \quad t>0, \\ (u_I)_x=(u_E)_x=0, & x=a,b, \quad t>0. \end{cases} \tag{2.4}$$

To define the basic reproduction number for model (2.1), we assume that the state variables are near the disease-free steady state E_0. Then we introduce the distribution of initial infection described by $\boldsymbol{\psi}(x)=(\psi_E(x),\psi_I(x))^\top$. Under the synthetic influences of mutation, mortality, and transfer of individuals in infected compartments, the distribution of those infected members as time evolves is $T(t)\boldsymbol{\psi}(x)$. Thus, the distribution of new infection at time t is $F(x)T(t)\boldsymbol{\psi}(x)$, where for any $\mathbf{u}=(u_E,u_I)^\top\in X_1$,

$$F(x)\mathbf{u}:=\frac{p(x)c(x)\rho_S(x)}{<c>}\begin{pmatrix}\displaystyle\int_a^b c(y)\rho_I(y)u_E(y)dy & \displaystyle\int_a^b c(y)\rho_E(y)u_I(y)dy\\ 0 & 0\end{pmatrix}. \tag{2.5}$$

Consequently, the distribution of total new infections is $\int_0^\infty F(x)T(t)\boldsymbol{\psi}(x)dt$. Define

$$L(\boldsymbol{\psi})(x):=\int_0^\infty F(x)T(t)\boldsymbol{\psi}(x)dt=F(x)\int_0^\infty T(t)\boldsymbol{\psi}(x)dt.$$

Then L is a continuous and positive operator which maps the initial infection distribution ψ to the distribution of the total infected members. Following the idea of next generation operators ([12]), we define the spectral radius of L as the basic reproduction number

$$\mathcal{R}_0=\rho(L)$$

for system (2.1).

Let Q be the generator of the continuous semigroup $T(t)$, i.e. for any $\mathbf{u}=(u_E,u_I)^\top\in X_1$,

$$Q\mathbf{u}=\begin{pmatrix}-\sigma(x)u_E-\mu(x)u_E+\epsilon(u_E)_{xx} & 0\\ \sigma(x)u_E & -\gamma(x)u_I-\mu(x)u_I-\alpha(x)u_I+\epsilon(u_I)_{xx}\end{pmatrix}.$$

Note that $T(t)$ is a positive semigroup and Q is resolvent positive, then by [36, Theorem 3.12], we have

$$Q^{-1}\boldsymbol{\psi}=\int_0^\infty T(t)\boldsymbol{\psi}(x)dt.$$

Let $\boldsymbol{\phi}=Q^{-1}\boldsymbol{\psi}$, it can be seen that ϕ is positive and satisfies

$$F(x)\boldsymbol{\phi}=\mathcal{R}_0 Q\boldsymbol{\phi},$$

i.e.

$$\begin{cases} -\epsilon(\phi_E)_{xx}+\sigma(x)\phi_E+\mu(x)\phi_E \\ =\dfrac{1}{\mathcal{R}_0}\left(\dfrac{p(x)c(x)\rho_S(x)}{<c>}\displaystyle\int_a^b c(y)[\rho_I(y)\phi_I+\rho_E(y)\phi_E]dy\right), & x\in(a,b), \\ -\epsilon(\phi_I)_{xx}+\gamma(x)\phi_I+\mu(x)\phi_I+\alpha(x)\phi_I-\sigma(x)\phi_E=0, & x\in(a,b), \\ (\phi_E)_x=(\phi_I)_x=0, & x=a,b. \end{cases} \tag{2.6}$$

By the Krein-Rutman theorem [26], the eigenvalue problem $Q\boldsymbol{\varphi}=\lambda F(x)\boldsymbol{\varphi}$ has a unique principal eigenvalue λ_1, that is, a real and simple eigenvalue with positive eigenfunctions, and it is strictly less than the real parts of all other eigenvalues. Since $\boldsymbol{\phi}$ is positive, it follows that $\lambda_1=1/\mathcal{R}_0$ is the unique principal eigenvalue of the eigenvalue problem $Q\boldsymbol{\varphi}=\lambda F(x)\boldsymbol{\varphi}$.

Lemma 2.1. *System* (2.6) *admits a unique principal eigenvalue, denoted by* λ_1*, with positive eigenfunctions, and the basic reproduction number of system* (2.1) *satisfies*

$$\mathcal{R}_0=\frac{1}{\lambda_1}.$$

Moreover, if $\epsilon=0$*, i.e. mutation of human heterogeneities is not considered, then*

$$\mathcal{R}_0=\int_a^b\frac{p(x)c(x)}{<c>}\mathcal{R}_0(x)dx=\int_a^b\frac{p(x)c(x)}{<c>}\left(\mathcal{R}_0^E(x)+\mathcal{R}_0^I(x)\right)dx, \tag{2.7}$$

where

$$\mathcal{R}_0^E(x):=\frac{c(x)\rho_S(x)\rho_E(x)}{\sigma(x)+\mu(x)},\quad \mathcal{R}_0^I(x):=\frac{c(x)\rho_S(x)\rho_I(x)\sigma(x)}{\left(\sigma(x)+\mu(x)\right)\left(\gamma(x)+\mu(x)+\alpha(x)\right)} \tag{2.8}$$

are the basic reproduction numbers of x-type exposed and infected individuals, respectively, and $\mathcal{R}_0(x):=\mathcal{R}_0^E(x)+\mathcal{R}_0^I(x)$.

Proof. We only need to consider the case $\epsilon=0$. Note that when $\epsilon=0$, the operator $F(x)Q^{-1}$ is still compact. Then by [10, Proposition 2.1], we have

$$\lim_{\epsilon\to0}\lambda_1(\epsilon)=\lambda_1(0),$$

and

$$\sigma(x)\phi+\mu(x)\phi=\lambda_1(0)\left(\frac{p(x)c(x)\rho_S(x)}{<c>}\int_a^b c(y)\left(\frac{\rho_I(y)\sigma(y)}{\gamma(y)+\mu(y)+\alpha(y)}+\rho_E(y)\right)\phi(y)dy\right),$$

which yields that

$$\mathcal{R}_0=\frac{1}{\lambda_0}=\int_a^b \frac{p(x)c(x)}{<c>}\left(\frac{c(x)\rho_S(x)\rho_I(x)\sigma(x)}{\big(\sigma(x)+\mu(x)\big)\big(\gamma(x)+\mu(x)+\alpha(x)\big)}+\frac{c(x)\rho_S(x)\rho_E(x)}{\sigma(x)+\mu(x)}\right)dx.$$

Let

$$\begin{aligned}
&\mathcal{R}_0^E(x)=\frac{c(x)\rho_S(x)\rho_E(x)}{\sigma(x)+\mu(x)},\\
&\mathcal{R}_0^I(x)=\frac{c(x)\rho_S(x)\rho_I(x)\sigma(x)}{\big(\sigma(x)+\mu(x)\big)\big(\gamma(x)+\mu(x)+\alpha(x)\big)},\\
&\mathcal{R}_0(x)=\mathcal{R}_0^E(x)+\mathcal{R}_0^I(x),
\end{aligned}$$

then

$$\mathcal{R}_0=\int_a^b \frac{p(x)c(x)}{<c>}\big(\mathcal{R}_0^E(x)+\mathcal{R}_0^I(x)\big)dx.$$

This completes the proof. □

Now we establish the threshold dynamics of system (2.1). For further purposes to prove the global stability of DFE, we consider the following eigenvalue problem:

$$\begin{cases}
-\epsilon(\phi_E)_{xx}+\sigma(x)\phi_E+\mu(x)\phi_E\\
\quad-\left(\dfrac{p(x)c(x)\rho_S(x)}{<c>}\displaystyle\int_a^b c(y)\big(\rho_I(y)\phi_I+\rho_E(y)\phi_E\big)dy\right)=\lambda_0\phi_E,\\
-\epsilon(\phi_I)_{xx}+\gamma(x)\phi_I+\mu(x)\phi_I+\alpha(x)\phi_I-\sigma(x)\phi_E=\lambda_0\phi_I,\\
(\phi_E)_x=(\phi_I)_x=0,\quad x=a,b,
\end{cases}\tag{2.9}$$

and the corresponding adjoint eigenvalue problem

$$\begin{cases}
-\epsilon(\phi_E^*)_{xx}+\sigma(x)\phi_E^*+\mu(x)\phi_E^*\\
\quad-c(x)\rho_E(x)\displaystyle\int_a^b\frac{p(y)c(y)\rho_S(y)}{<c>}\phi_E^*dy-\sigma(x)\phi_I^*=\lambda_0\phi_E^*,\\
-\epsilon(\phi_I^*)_{xx}+\gamma(x)\phi_I^*+\mu(x)\phi_I^*+\alpha(x)\phi_I^*\\
\quad-c(x)\rho_I(x)\displaystyle\int_a^b\frac{p(y)c(y)\rho_S(y)}{<c>}\phi_E^*dy=\lambda_0\phi_I^*,\\
(\phi_E^*)_x=(\phi_I^*)_x=0,\quad x=a,b.
\end{cases}\tag{2.10}$$

By the Krein-Rutman heorem [26], the eigenvalue problems (2.9) and (2.10) each admit a unique principal eigenvalue λ_0, that is, a real and simple eigenvalue with positive eigenfunctions, and it is strictly less than the real parts of all other eigenvalues. The following lemma is a direct result of [36, Theorem 3.5].

Lemma 2.2. $\text{sign}\{\lambda_0\} = \text{sign}\{1-\mathcal{R}_0\}$.

Our first main result is stated as follows.

Theorem 2.1. *(i) If $\mathcal{R}_0 \le 1$, then DFE is globally asymptotically stable.*

(ii) If $\mathcal{R}_0 > 1$, there exists some positive constant ϵ_0 such that any positive solution of system (2.1) *satisfies*

$$\liminf_{t\to\infty} \big\| \big(S(\cdot,t),E(\cdot,t),I(\cdot,t),R(\cdot,t)\big) - \big(N_0(x),0,0,0\big) \big\| \ge \epsilon_0.$$

In particular, system (2.1) *admits at least one endemic equilibrium.*

Proof. We prove (i) by constructing a Lyapunov functional and applying LaSalle's invariance principle [21, Theorem 1] for infinite dimensional dynamical systems. Let $X = C([a,b];\mathbb{R}^4)$ with the supremum norm $\|\cdot\|_\infty$, then X is an ordered Banach space with the cone P consisting of all nonnegative functions in X, and X has a nonempty interior, denoted by $\text{int}(P)$. It is easy to verify that (2.1) defines a dynamic system on X. Denote the unique solution of system (2.1) with initial value $(s_0,e_0,i_0,r_0) \in P$ by $\Phi_t(s_0,e_0,i_0,r_0) = (S(\cdot,t),E(\cdot,t),I(\cdot,t),R(\cdot,t))$ for any $t>0$. It can be verified that Φ_t is compact and for each $\mathbf{u}_0 \in U$, the orbit of $\mathbf{u}_0$ under the dynamical system generated by (2.1) has compact closure in U.

Define the functional

$$L(\mathbf{u}) = \int_a^b (u_e\phi_E^* + u_i\phi_I^*)dx$$

for $\mathbf{u}\in U$, where (ϕ_E^*,ϕ_I^*) is the eigenfunction corresponding to the principal eigenvalue λ_1 associated with the eigenvalue problem (2.10). Now we prove $L(\mathbf{u})$ is a Lyapunov functional for system (2.1). For an arbitrary solution $\mathbf{u} = (S,E,I,R)$ of system (2.1), we have

$$\begin{aligned}
\frac{d}{dt}L(\mathbf{u}(\cdot,t)) &= \int_a^b \big(E_t\phi_E^* + I_t\phi_I^*\big)dx \\
&= \int_a^b \Big(\big(\text{Foi}(x)S - \sigma(x)E - \mu(x)E + \epsilon E_{xx}\big)\phi_E^* \\
&\qquad + \big(\sigma(x)E - \gamma(x)I - \mu(x)I - \alpha(x)I + \epsilon I_{xx}\big)\phi_I^* \Big)dx \\
&= -\int_a^b \text{Foi}(x)(E+I+R)\phi_E^* dx - \lambda_0 \int_a^b (E\phi_E^* + I\phi_I^*)dx. \qquad (2.11)
\end{aligned}$$

By Lemma 2.2, $R_0 \le 1$ yields that $\lambda_1 \ge 0$. Furthermore, S,E,I,R are non-negative and $\text{Foi}(x)$ are nonnegative. Hence, $d(L(\mathbf{u}(\cdot,t)))/dt \le 0$, which implies $L(\mathbf{u})$ is a Lyapunov functional of system (2.1).

Next, define

$$\dot{L}(\mathbf{u}_0) := \frac{d}{dt}L\big(\mathbf{u}(\cdot,t)\big)\big|_{t=0}, \quad M = \{\mathbf{u}_0 \in P \,|\, \dot{L}(\mathbf{u}_0) = 0\},$$

where $\mathbf{u}=(S,E,I,R)$ is the unique solution of (2.1) with initial condition $\mathbf{u}_0=(s_0,e_0,i_0,r_0)\in P$. By (2.11), we have $M=\{\mathbf{u}_0=(s_0,e_0,i_0,r_0)\in U|e_0=i_0=0\}$ if $\lambda_1\geq 0$. It follows from (2.1) that for $\lambda_1\geq 0$, the maximal invariant set in M is given by

$$\hat{M}:=\{\mathbf{u}_0=(s_0,e_0,i_0,r_0)\in U|e_0=i_0=0\}.$$

Therefore, by the LaSalle invariant principle [21, Theorem 1], we obtain

$$\big(E(x,t),I(x,t)\big)\ \to\ (0,0)\quad \text{in}\ \ [L^\infty(\Omega)]^2\quad \text{as}\quad t\ \to\ \infty,$$

which together with (2.1) imply $R(x,t)\to 0$ uniformly in $\overline{\Omega}$ as $t\to\infty$. Therefore, we obtain $S(x,t)\to N_0$ in $L^\infty(\Omega)$ as $t\to\infty$.

The proof of (ii) is similar to that of [33, Theorem 1.1(ii)], so we omit the proof here. The proof of theorem is complete. □

3 The effect of human heterogeneities on disease persistence, herd immunity threshold and final epidemic size

In this section, we investigate the effect of human heterogeneities on disease persistence and herd immunity threshold. In Theorem 2.1, we show that disease persistence is determined by the basic reproduction number $\mathcal{R}_0$. Therefore, we will compare the basic reproduction number between homogeneous and heterogeneous populations. Throughout this section, we assume that $\epsilon=0$ since in this part, our focus is not on mutation.

Before proceeding, we calculate the basic reproduction number of homogeneous population. Consider the following well-mixed homogeneous SEIR model:

$$\begin{cases}\dfrac{\mathrm{d}S}{\mathrm{d}t}=\Lambda(\bar{x})-c(\bar{x})\rho_S(\bar{x})\big(\rho_I(\bar{x})I+\rho_E(\bar{x})E\big)S/N-\mu(\bar{x})S,\\ \dfrac{\mathrm{d}E}{\mathrm{d}t}=c(\bar{x})\rho_S(\bar{x})\big(\rho_I(\bar{x})I+\rho_E(\bar{x})E\big)S/N-\sigma(\bar{x})E-\mu(\bar{x})E,\\ \dfrac{\mathrm{d}I}{\mathrm{d}t}=\sigma(\bar{x})E-\gamma(\bar{x})I-\mu(\bar{x})I-\alpha(\bar{x})I,\\ \dfrac{\mathrm{d}R}{\mathrm{d}t}=\gamma(\bar{x})I-\mu(\bar{x})R.\end{cases}\tag{3.1}$$

The basic reproduction number $\overline{\mathcal{R}}_0$ of the homogeneous SEIR model is calculated as

$$\overline{\mathcal{R}}_0=\frac{c(\bar{x})\rho_S(\bar{x})\rho_I(\bar{x})\sigma(\bar{x})}{\big(\sigma(\bar{x})+\mu(\bar{x})\big)\big(\gamma(\bar{x})+\mu(\bar{x})+\alpha(\bar{x})\big)}+\frac{c(\bar{x})\rho_S(\bar{x})\rho_E(\bar{x})}{\sigma(\bar{x})+\mu(\bar{x})},$$

where $\bar{x}=\int_a^b xp(x)dx$. Define

$$\mathcal{R}_0^E(\bar{x})=\frac{c(\bar{x})\rho_S(\bar{x})\rho_E(\bar{x})}{\sigma(\bar{x})+\mu(\bar{x})},\quad \mathcal{R}_0^I(\bar{x})=\frac{c(\bar{x})\rho_S(\bar{x})\rho_I(\bar{x})\sigma(\bar{x})}{\big(\sigma(\bar{x})+\mu(\bar{x})\big)\big(\gamma(\bar{x})+\mu(\bar{x})+\alpha(\bar{x})\big)},$$

then

$$\overline{\mathcal{R}}_0 = \mathcal{R}_0^E(\bar{x}) + \mathcal{R}_0^I(\bar{x}). \tag{3.2}$$

Note that $\mathcal{R}_0^E(\bar{x}), \mathcal{R}_0^I(\bar{x})$ are the basic reproduction numbers of exposed and infected individuals, respectively.

3.1 The basic reproduction number: Homogeneous vs heterogeneous

Recall that $p(x)$ is the frequency of x-type population and the variance

$$\text{Var} = \int_a^b (x-\bar{x})^2 p(x)dx = <x^2> - <x>^2$$

is an important indicator characterizing the human heterogeneity variation. Now we will use Var to describe the difference between the basic reproduction number of homogeneous and heterogeneous populations.

Theorem 3.1. *Assume that $c(x)$ is independent of x, i.e. activity level heterogeneity is not considered. If $\mathcal{R}_0(x)$ is twice continuously differentiable and there are finite bounds m and M such that*

$$m \le \big(\mathcal{R}_0(x)\big)'' \le M,$$

then

$$\mathcal{R}_0 = \overline{\mathcal{R}}_0 + \frac{\xi}{2}\text{Var}, \tag{3.3}$$

where ξ is some constant in $[m,M]$. In particular, if $\mathcal{R}_0(x)$ is convex (concave) on $x \in [a,b]$, then $\mathcal{R}_0 \ge (\le)\overline{\mathcal{R}}_0$.

Proof. Assume that $c(x)$ is independent of x, then

$$\mathcal{R}_0 = \int_a^b p(x)\mathcal{R}_0(x)dx.$$

By Jessen's inequality, if $\mathcal{R}_0(x)$ is convex(concave) on $x \in [a,b]$, then $\mathcal{R}_0 \ge (\le)\overline{\mathcal{R}}_0$. Moreover, if $\mathcal{R}_0(x)$ is twice continuously differentiable, then it follows Holder's defect formula [35, Problem 6.5, p.94] that

$$\mathcal{R}_0 = \overline{\mathcal{R}}_0 + \frac{\xi}{2}\text{Var},$$

where ξ is some constant in $[m,M]$. □

Here we mention that Theorem 3.1 still holds for small ϵ by implicit function theorem and perturbation theory. Since it is not focus here, we omit the proof.

Direct calculation yields the following result.

Corollary 3.1. *Assume that the trait x is defined as activity level, i.e. $c(x)=x$, and other attributes (heterogeneities) have no relations to activity level, then*

$$\mathcal{R}_0=\frac{<x^2>}{<x>^2}\overline{\mathcal{R}}_0=\left(1+\frac{\mathrm{Var}}{\bar{x}^2}\right)\overline{\mathcal{R}}_0.$$

Corollary 3.1 shows that if only activity level is considered, then heterogeneous activity level will increase the basic reproduction number and enhance the disease persistence.

3.2 Herd immunity threshold

Assume that $R_0(x)>1$ for any $x\in[a,b]$ and the purpose of herd immunity is to ensure that the basic reproduction number of any type of population should be smaller than one. Thus, for x-type population, the minimum population number with immunity is $(1-1/\mathcal{R}_0(x))N(x)$, the herd immunity threshold in heterogeneous populations is

$$\mathcal{H}^{he}=\frac{\int_a^b\left(1-1/\mathcal{R}_0(x)\right)N(x)dx}{\int_a^b N(x)dx}=1-\int_a^b\frac{p(x)}{\mathcal{R}_0(x)}dx, \tag{3.4}$$

and the herd immunity threshold in a homogeneous population is

$$\mathcal{H}^{ho}=1-\frac{1}{\mathcal{R}_0(\bar{x})}. \tag{3.5}$$

Theorem 3.2. *If $1/\mathcal{R}_0(x)$ is twice continuously differentiable and there are finite bounds m and M such that*

$$m\le\left(\frac{1}{\mathcal{R}_0(x)}\right)''\le M,\quad x\in[a,b],$$

then

$$\mathcal{H}^{he}=\mathcal{H}^{ho}-\frac{\xi}{2}\mathrm{Var}, \tag{3.6}$$

where ξ is some constant in $[m,M]$. In particular, if $1/\mathcal{R}_0(x)$ is convex (concave) on $x\in[a,b]$, then $\mathcal{H}^{he}\le(\ge)\mathcal{H}^{ho}$.

Proof. The results can be directly derived from Jensen's inequality and Holder's defect formula [35, Problem 6.5, p.94]. □

Here we mention that Theorem 3.2 still holds for small ϵ by implicit function theorem and perturbation theory. Since it is not focus here, we omit the details.

Corollary 3.2. *Assume that $\mathcal{R}_0(x)=kx$. Then*

$$\mathcal{H}^{he}=\mathcal{H}^{ho}-\frac{\xi}{2k}\mathrm{Var}, \tag{3.7}$$

where ξ is some constant in $[1/b^2,1/a^2]$.

If the trait x is defined as activity level, i.e. $c(x)=x$, and other attributes (heterogeneities) has no relations to activity level, then basic reproduction number of x-type population satisfies $\mathcal{R}_0(x)=kx$. Corollary 3.2 shows that heterogeneous activity level will decrease the herd immunity level.

3.3 Final epidemic size

In this part, we assume that $\Lambda(x)=\mu(x)=0$ and investigate the effect of human heterogeneities on the final epidemic size. Note that

$$\begin{cases} \dfrac{\partial \ln S}{\partial t}=-\mathrm{Foi}(x), & x\in(a,b), \quad t>0, \\ \dfrac{\partial (I+R)}{\partial t}=\sigma(x)E, & x\in(a,b), \quad t>0, \\ \dfrac{\partial R}{\partial t}=\gamma(x)I+\alpha(x)I, & x\in(a,b), \quad t>0. \end{cases} \tag{3.8}$$

Here

$$\mathrm{Foi}(x)=\frac{c(x)\rho_S(x)\displaystyle\int_a^b c(y)\big(\rho_I(y)I(y,t)+\rho_E(y)E(y,t)\big)dy}{\displaystyle\int_a^b c(y)N(y,t)dy}.$$

(3.8) implies that

$$I=\frac{R_t}{\gamma(x)+\alpha(x)}, \quad E=\frac{(I+R)_t}{\sigma(x)},$$

and then

$$-(\ln S)_t=\frac{c(x)\rho_S(x)\displaystyle\int_a^b c(y)\big(\rho_I(y)R_t/\big(\gamma(y)+\alpha(y)\big)+\rho_E(y)(I+R)_t/\sigma(y)\big)dy}{\displaystyle\int_a^b c(y)N(y,t)dy},$$

i.e.

$$-(\ln S)_t=\frac{c(x)\rho_S(x)}{<c>N}\int_a^b \frac{\mathcal{R}_0^I(y)R_t+\mathcal{R}_0^E(y)(I+R)_t}{\rho_S(y)}dy \tag{3.9}$$

with N denotes the total population.

Lemma 3.1. *Let $\Lambda(x)=\mu(x)=\epsilon=0$ in system* (2.1)*. Let $s_0(x)$ denote the initial susceptible proportion. In the limit $s_0(x)\to 1$, the final outbreak size of x-type population in system* (2.1) *approaches the solution $r_\infty(x)$ to the following final outbreak size relation:*

$$-\ln\big(1-r_\infty(x)\big)=\frac{\rho_S(x)c(x)}{<c>}\int_a^b \frac{p(y)\mathcal{R}_0(y)r_\infty(y)}{\rho_S(y)}dy. \tag{3.10}$$

Moreover, the final epidemic size of the total population is $<r_\infty>$.

Proof. Denote

$$s(x,t)=\frac{S(x,t)}{N(x)},\quad e(x,t)=\frac{E(x,t)}{N(x)},$$
$$i(x,t)=\frac{I(x,t)}{N(x)},\quad r(x,t)=\frac{R(x,t)}{N(x)}.$$

It follows from (2.1) that

$$\big(s(x)+e(x)\big)_t=-\sigma(x)e(x,t),\quad (s+e+i)_t=-\big(\gamma(x)+\alpha(x)\big)i(x,t),$$

which by $s(x,t)+e(x,t)+i(x,t)+r(x,t)=i$ implies

$$e(x,t),\ i(x,t)\ \to\ 0.$$

Thus, integrating (3.9) by parts, we have

$$-\ln\big(1-r_\infty(x)\big)=\frac{\rho_S(x)c(x)}{<c>}\int_a^b\frac{p(y)\mathcal{R}_0(y)r_\infty(y)}{\rho_S(y)}dy,$$

where $\lim_{t\to\infty}r(x,t)=r_\infty(x)$. □

Denote

$$\mathcal{R}_0^{ho}=\overline{\mathcal{R}_0},\quad \mathcal{R}_0^{he}=\mathcal{R}_0.$$

For homogeneous population, the final epidemic size r^{ho} satisfies

$$-\ln(1-r^{ho})=r^{ho}\mathcal{R}_0^{ho}. \tag{3.11}$$

Now we compare the final epidemic size of the heterogeneous population

$$r^{he}:=<r_\infty>$$

and homogeneous population r^{ho}. The following result is a direct consequence of Lemma 3.1 and Theorem 3.1.

Theorem 3.3. *Assume that $c(x)$ and $\rho_S(x)$ are independent of x, i.e. activity level and susceptibility heterogeneities are not considered. Then*

$$-\ln(1-r^{he})=r^{he}\mathcal{R}_0^{he}.$$

Moreover, if $\mathcal{R}_0(x)$ is convex (concave) on $x\in[a,b]$, then $r_0^{he}\geq(\leq)r^{ho}$.

Theorem 3.4. *Assume that the trait x is defined as susceptibility, i.e. $\rho_S(x)=x$, and other attributes (heterogeneities) has no relations to susceptibility, then*

$$\mathcal{R}_0^{he}=\mathcal{R}_0^{ho},\quad r^{he}<r^{ho}.$$

Proof. Rewrite (3.10) as

$$-\ln\big(1-r_\infty(x)\big)=\mathcal{R}_0(x)r^{he},$$

which yields that

$$1-r^{he}=\int_a^b p(x)e^{-\mathcal{R}_0(x)r^{he}}.$$

By Jensen's inequality, we have

$$1-r^{he}=e^{-\mathcal{R}_0^{ho}r^{he}}+C,\quad C>0,$$

which together with $1-r^{ho}=e^{-\mathcal{R}_0^{ho}r^{ho}}$ imply that $r^{he}<r^{ho}$. □

It follows from Theorem 2.1 that heterogeneous susceptibility does not affect the disease persistence, but lower the final epidemic size, which agrees with the results in [6,25].

Proposition 3.1. *Assume that the trait x is defined as activity level, i.e. $c(x)=x$, and other attributes (heterogeneities) have no relations to activity level, then*

$$1-r^{he}=e^{\mathcal{R}_0^{ho}(r^{he}+C_1)}+C_2$$

for some positive constants C_1, C_2.

Proof. If $c(x)=x$, then $\mathcal{R}_0(x)=(x/\bar{x})\mathcal{R}_0^{ho}$, which implies that $\mathcal{R}_0^{he}=(1+\mathrm{Var}/\bar{x}^2)\mathcal{R}_0^{ho}$ and

$$-\ln\big(1-r_\infty(x)\big)=\frac{x\mathcal{R}_0^{ho}}{\bar{x}^2}\int_a^b yr_\infty(y)p(y)=\frac{x\mathcal{R}_0^{ho}}{\bar{x}^2}<r_\infty(x)x>.$$

Thus,

$$1-r_\infty(x)=e^{-Ax},\quad A=\frac{\mathcal{R}_0^{ho}}{\bar{x}^2}<r_\infty(x)x>,$$

and

$$1-r^{he}=\int_a^b p(x)e^{-Ax}dx,$$

which yields by Jenson's inequality that

$$1-r^{he}=e^{-A\bar{x}}+C_2.$$

By Chebyshev inequality and monotonicity of $r_\infty(x)$, we have

$$<r_\infty(x)x>=(r^{he}+C_1)\bar{x}.$$

Thus, $1-r^{he}=e^{\mathcal{R}_0^{ho}(r^{he}+C_1)}+C_2$. □

In this section, theoretical results depend on convexity of $\mathcal{R}_0(x)$. This kind of convexity can be found in epidemiology. For example, if we consider the effect of heterogeneous activity level on disease spreading, the basic reproduction number can be assumed to be linear function of activity level when activity level is very small, or saturated function of activity level (e.g. $ax/(x+b)$) when activity level is large.

4 Applications on the COVID-19

In this section, we will apply the results developed in Section 3 to explore the effect of heterogeneous transition rate from E to I ($\sigma(x)$), activity levels ($c(x)$), susceptibility ($\rho_S(x)$), exposure to symptomatic ($\rho_I(x)$) and asymptomatic infection ($\rho_E(x)$) on the basic reproduction number $\mathcal{R}_0$ and herd immunity threshold $\mathcal{H}^{he}$. In this section, mutation, natural birth and death are not included, and disease-induced death is omitted, thus we have

$$\epsilon=0,\quad \Lambda(x)=\mu(x)=\alpha(x)=0.$$

Then model (2.1) becomes

$$\begin{cases}\dfrac{\partial S}{\partial t}=-\mathrm{Foi}(x)S,\\ \dfrac{\partial E}{\partial t}=\mathrm{Foi}(x)S-\sigma(x)E,\\ \dfrac{\partial I}{\partial t}=\sigma(x)E-\gamma(x)I,\\ \dfrac{\partial R}{\partial t}=\gamma(x)I,\end{cases}\tag{4.1}$$

and

$$\mathrm{Foi}(x)=\frac{c(x)\rho_S(x)\displaystyle\int_a^b c(y)\big(\rho_I(y)I(y,t)+\rho_E(y)E(y,t)\big)dy}{\displaystyle\int_a^b c(y)N(y,t)dy}.$$

Moreover, the basic reproduction number

$$\mathcal{R}_0=\int_a^b\frac{p(x)c(x)}{<c>}\mathcal{R}_0(x)dx=\int_a^b\frac{p(x)c(x)}{<c>}\big(\mathcal{R}_0^E(x)+\mathcal{R}_0^I(x)\big)dx,$$

where

$$\mathcal{R}_0^E(x):=\frac{c(x)\rho_S(x)\rho_E(x)}{\sigma(x)},\quad \mathcal{R}_0^I(x):=\frac{c(x)\rho_S(x)\rho_I(x)}{\gamma(x)},$$

and the herd immunity threshold

$$\mathcal{H}^{he}=1-\int_a^b\frac{p(x)}{\mathcal{R}_0(x)}dx.$$

Here $p(x)$ is the initial frequency of x-type population.

4.1 The effect of heterogeneous $\sigma(x)$

Now we use the transition rate from E to I of x type population as the trait x, i.e. $\sigma(x)=1/x$. Other attribute values in $\mathbf{\Theta}(x)$ are independent of the transition rate from E to I. Thus,

$$\mathcal{R}_0=c\rho_S\rho_E\int_a^b p(x)xdx+\frac{1}{\gamma}c\rho_S\rho_I,\quad \overline{\mathcal{R}}_0=c\rho_S\rho_E\bar{x}+\frac{1}{\gamma}c\rho_S\rho_I.$$

Now we compare the basic reproduction number of COVID-19 with homogeneous population ($\overline{\mathcal{R}}_0$) and population with heterogeneous $\sigma(x)$ ($\mathcal{R}_0$). Here we fixed $\overline{\mathcal{R}}_0=3$ [22] and similar to [40], we assume that $\rho_E=\rho_I$. Note that for COVID-19, individuals who do not have symptom are still infective. Thus, we assume that exposed individuals are infective and the number of exposed individuals can be used to characterize the number of individuals who do not have symptom but are still infective. The transition rate from I to R (γ) of COVID-19 is fixed as 0.1 [34]. $c\rho_S\rho_E$ can be directly derived by the values of $\overline{\mathcal{R}}_0$,$\bar{x}$ and $1/\gamma$. For $p(x)$, we consider three commonly used distributions (Weibull, Gamma and Lognormal) [3, 29]. The parameters, means and standard variances in three distributions are listed in Table 1 [3].

Table 1: Distributions associated with the transition rate from E to I.

Distribution	Mean (day)	SD (day)	Parameters
Weibull$(\lambda;k)$	6.4	2.3	$\lambda=7.16, k=3.05$
Gamma$(\lambda;k)$	6.5	2.6	$\lambda=6.25, k=1.04$
Lognormal$(\lambda;k)$	6.8	2.4	$\lambda=1.85, k=0.34$

By direct calculation, we have $\mathcal{R}_0=\overline{\mathcal{R}}_0=3$, and

$$\mathcal{H}^{ho}=67\%, \qquad \mathcal{H}^{he}=37\%\,(\text{Weibull}),$$
$$\mathcal{H}^{he}=43\%\,(\text{Gamma}), \quad \mathcal{H}^{he}=28\%\,(\text{Lognormal}).$$

4.2 The effect of heterogeneous activity levels

Now we use the activity level as the trait x, i.e. $c(x)=x$, and assume that other attributes are not concerned with the activity level. Then

$$\mathcal{R}_0=\frac{<x^2>}{<x>^2}\overline{\mathcal{R}}_0=\left(1+\frac{\text{Var}}{\bar{x}^2}\right)\overline{\mathcal{R}}_0,$$

and by Corollary 3.1,

$$\mathcal{H}^{he}=\mathcal{H}^{ho}-\frac{\xi}{2k}\text{Var},$$

where ξ is some positive constant. An interesting finding here is that heterogeneous activity levels will enhance the disease persistence, but decrease the herd immunity threshold.

For COVID-19, we fixed $\overline{\mathcal{R}}_0=3$ [22]. Now we use the data of contact numbers in [41] to quantitatively compare the basic reproduction number between homogeneous population ($\overline{\mathcal{R}}_0$) and population with heterogenous activity levels ($\mathcal{R}_0$) of COVID-19 in Shanghai and Wuhan. The used data is from [41, Table S6], where heterogeneous activity levels are induced by gender, age structure, type of profession (preschool student, employed,

unemployed, retired). By direct calculation, we obtain

$$\begin{aligned}&\text{(Shanghai)} \quad \bar{x}=18.7, \quad \mathcal{R}_0=1.08\overline{\mathcal{R}}_0=3.24, \quad \mathcal{H}^{he}=63.7\% \ (\mathcal{H}^{ho}=66.7\%),\\ &\text{(Wuhan)} \quad \bar{x}=18.7, \quad \mathcal{R}_0=1.07\overline{\mathcal{R}}_0=3.21, \quad \mathcal{H}^{he}=64.1\% \ (\mathcal{H}^{ho}=66.7\%).\end{aligned}$$

Finally, we consider a special case that the activity level satisfies a power law distribution, i.e. $p(x)\propto x^{-\lambda}$ with $x>x_{\min},\lambda>3$. Here, the technical condition $\lambda>3$ is assumed to ensure the existence of mean $\bar{x}$ and Var. By direct calculations,

$$\mathcal{R}_0=\left(1+x_{\min}^{\lambda-1}\frac{(\lambda-2)^2}{\lambda-3}\right)\overline{\mathcal{R}}_0,$$

and

$$\mathcal{H}_{he}=1-\frac{\lambda-1}{kx_{\min}\lambda}, \quad \mathcal{H}_{ho}=1-\frac{\lambda-2}{kx_{\min}(\lambda-1)}.$$

5 Conclusions and discussions

Based on the results in Sections 3 and 4, we give some answers to the following problems in Section 1 from the perspective of human heterogeneity:

1. Will human heterogeneity increase or decrease the basic reproduction number?

If multiple types of heterogeneity are considered, human heterogeneity may increase or decrease the basic reproduction number, and it strongly depends on the convexity of the heterogeneity function (see Theorem 3.1). If only the activity level (contact number variation) is considered, then heterogeneous activity level will increase the basic reproduction number and enhance the disease persistence (see Corollary 3.1).

2. Is the statement that human heterogeneity decreases the herd immunity level always right [7,19]?

Human heterogeneity may increase or decrease herd immunity level if multiple types of heterogeneity are considered, and it strongly depends on the convexity of the heterogeneity function (see Theorem 3.2). If only activity level (contact number variation) is considered, then heterogeneous activity level will decrease the herd immunity level (see Corollary 3.2).

3. How does human heterogeneity shape the final epidemic size?

The effect of human heterogeneity on the final size of the epidemic remains a complicated issue even in the case where only the heterogeneous activity level is considered (see Theorem 3.1), and the results are obtained only under certain conditions. If activity level and susceptibility heterogeneity are not considered, then the convexity of heterogeneity function will determine the effects (see Theorem 3.3). If only susceptibility heterogeneity is considered, then heterogeneity will decrease the final epidemic size (see Theorem 3.4).

4. The challenges of human heterogeneity on modelling infectious disease.

The basic reproduction number is still an indicator of disease persistence especially for longtime diseases, but it has limitations in indicating epidemic severity and being used as an indicator for disease control. Results from Corollaries 3.1 and 3.2 show that human heterogeneity may increase the basic reproduction number but decrease herd immunity level. The existence of human heterogeneity makes many simple ODE models fail and gives a lot of difficulties and limitations in modelling emerging infectious diseases. Trade-off between the first principle simple epidemic models and complicated models considering human heterogeneity becomes a critical issue in modelling emerging infectious diseases.

Our results reveal comprehensive relationships between human heterogeneities and transmission dynamics and also raise some interesting problems. For example, the effect of human heterogeneities on time to extinction and epidemic peak of infectious disease, the study of which will give insights into vaccination priority and ICU beds surveillance. How to incorporate variants data in GISAID [32] to the generalized mutation SEIR model, and study the evolution of influenza virus or SARS-COV2? It is a meaningful project which will be investigated in the future. Note that we consider mutation rate ϵ in the general mutation selection model (2.1), however we did not give much attention on it, coevolution of disease and human heterogeneities at this time is theoretically a difficult issue and will be our focus in sequential works.

Acknowledgments

Research partially supported by the National Natural Science Foundation of China (Grant No. 12101487(PS)) and by the China Postdoctoral Science Foundation (Grant Nos. 2023T160504, 2020M683445(PS)), by the Canada Research Chairs Program (Grant No. 230720) (J. Wu), by the Major International (Regional) Joint Research Project of the National Natural Science Foundation of China (Grant No. 12220101001) (Y. Xiao).

References

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.
[2] C. Aschwanden, *Five reasons why COVID herd immunity is probably impossible*, Nature, 591: 520–522, 2021.
[3] J. A. Backer, D. Klinkenberg, and J. Wallinga, *Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020*, Euro Surveill., 25:2000062, 2020.
[4] F. Bai, *Effect of population heterogeneity on herd immunity and on vaccination decision making process*, J. Theor. Biol., 526:110795, 2021.
[5] H. Berestycki, B. Desjardins, J. S. Weitz, and J.-M. Oury, *Epidemic modeling with heterogeneity and social diffusion*, J. Math. Biol., 86:60, 2023.

[6] R. D. Boylan, *A note on epidemics in heterogeneous populations*, Math. Biosci., 105:133–137, 1991.
[7] T. Britton, F. Ball, and P. Trapman, *A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2*, Science, 369:846–849, 2020.
[8] C. Castillo-Chavez, W. Huang, and J. Li, *The effects of females' susceptibility on the coexistence of multiple pathogen strains of sexually transmitted diseases*, J. Math. Biol., 35:503–522, 1997.
[9] J. Cui, Y. Zhang, and Z. Feng, *Influence of non-homogeneous mixing on final epidemic size in a meta-population model*, J. Biol. Dyn., 13:31–46, 2019.
[10] G. Degla, *An overview of semi-continuity results on the spectral radius and positivity*, J. Math. Anal. Appl., 338:101–110, 2008.
[11] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*, John Wiley and Sons, 2000.
[12] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*, J. Math. Biol., 28:365–382, 1990.
[13] J. G. Dushoff, *Modeling the Effects of Host Heterogeneity on the Spread of Human Diseases*, PhD Thesis, Princeton University, 1997.
[14] J. Dushoff, *Host heterogeneity and disease endemicity: A moment-based approach*, Theor. Popul. Biol., 56:325–335, 1999.
[15] J. Dushoff and S. Levin, *The effects of population heterogeneity on disease invasion*, Math. Biosci., 128:25–40, 1995.
[16] D. M. Ediev, *Population heterogeneity is a critical factor of the kinetics of the COVID-19 epidemics*, medRxiv 2020.06.25.20140442, 2020.
[17] M. El Khalifi and T. Britton, *SIRS epidemics with individual heterogeneity of immunity waning*, J. Theor. Biol., 587:111815, 2024.
[18] D. Gilbarg, N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, in: Classics in Mathematics, Vol. 224, Springer, 1977.
[19] M. G. M. Gomes et al., *Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold*, J. Theor. Biol., 540:111063, 2022.
[20] G. Großmann, M. Backenköhler, and V. Wolf, *Why ODE models for COVID-19 fail: Heterogeneity shapes epidemic dynamics*, medRxiv 2021.03.25.21254292, 2021.
[21] J. K. Hale, *Dynamical systems and stability*, J. Math. Anal. Appl., 26:39–59, 1969.
[22] W. He, G. Y. Yi, and Y. Zhu, *Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID-19: Meta-analysis and sensitivity analysis*, J. Med. Virol., 92(11):2543–255, 2020.
[23] R. I. Hickson and M. G. Roberts, *How population heterogeneity in susceptibility and infectivity influences epidemic dynamics*, J. Theor. Biol., 350:70–80, 2014.
[24] J. A. Jacquez, C. P. Simon, J. Koopman, L. Sattenspiel, and T. Perry, *Modeling and analyzing HIV transmission: The effect of contact patterns*, Math. Biosci., 92:119–199, 1988.
[25] G. Katriel, *The size of epidemics in populations with heterogeneous susceptibility*, J. Math. Biol., 65:237–262, 2012.
[26] M. G. Krein and M. A. Rutman, *Linear operators leaving invariant a cone in a Banach space*, Amer. Math. Soc. Transl., 10:3–95, 1962.
[27] X. Liang, L. Zhang, and X.-Q. Zhao, *Basic reproduction ratios for periodic abstract functional differential equations (with application to a spatial model for lyme disease)*, J. Dynam. Differential Equations, 31:1247–1278, 2019.
[28] S. Merler and M. Ajelli, *The role of population heterogeneity and human mobility in the spread of pandemic influenza*, Proc. R. Soc. B, 277:557–565, 2010.

[29] C. J. Mode, C. K. Sleeman, *Stochastic Processes in Epidemiology: HIV/AIDS, Other Infectious Diseases and Computers*, World Scientific, 2000.
[30] Y. Oz, I. Rubinstein, and M. Safra, *Heterogeneity and superspreading effect on herd immunity*, J. Stat. Mech. Theory Exp., 2021:033405, 2021.
[31] C. L. Shaw and D. A. Kennedy, *What the reproductive number R_0 can and cannot tell us about COVID-19 dynamics*, Theor. Popul. Biol., 137:2–9, 2021.
[32] Y. Shu and J. McCauley, *GISAID: Global initiative on sharing all influenza data – from vision to reality*, Euro Surveill., 22:30494, 2017.
[33] P. Song, Y. Lou, and Y. Xiao, *A spatial SEIRS reaction-diffusion model in heterogeneous environment*, J. Differential Equations, 267:5084–5114, 2019.
[34] Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine Association, *An update on the epidemiological characteristics of novel coronavirus pneumonia (COVID-19)*, Chinese Journal of Epidemiology, 41:139–144, 2020.
[35] J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge University Press, 2004.
[36] H. R. Thieme, *Spectral bound and reproduction number for infinite-dimensional population structure and time heterogeneity*, SIAM J. Appl. Math., 70:188–211, 2009.
[37] A. V. Tkachenko, S. Maslov, A. Elbanna, G. N. Wong, Z. J. Weiner, and N. Goldenfeld, *Time-dependent heterogeneity leads to transient suppression of the COVID-19 epidemic, not herd immunity*, Proc. Natl. Acad. Sci. USA, 118:e2015972118, 2021.
[38] W. Wang and X.-Q. Zhao, *Basic reproduction numbers for reaction-diffusion epidemic models*, SIAM J. Appl. Dyn. Syst., 11:1652–1673, 2012.
[39] Y. Xie, *Population heterogeneity and causal inference*, Proc. Natl. Acad. Sci. USA, 110:6262–6268, 2013.
[40] Z. Yang et al., *Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions*, J. Thorac. Dis., 12(3):165–174, 2020.
[41] J. Zhang et al., *Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China*, Science, 368:1481–1486, 2020.
[42] X.-Q. Zhao, *Dynamical Systems in Population Biology*, Springer, 2017.