Western **Graduate&PostdoctoralStudies**

Western University

10-13-2023 10:00 AM

Study of Behaviour Change and Impact on Infectious Disease Dynamics by Mathematical Models

Tianyu Cheng, *The University of Western Ontario*

Supervisor: Zou, Xingfu, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Applied Mathematics
© Tianyu Cheng 2023

Abstract

This thesis uses mathematical models to study human behaviour changes' effects on infectious disease transmission dynamics. It centers on two main topics. The first concerns how behaviour response evolves during epidemics and the effects of adaptive precaution behaviour on epidemics. The second topic is building general framework models incorporating human behaviour response in epidemiological modelling.

In the first project, based on the fact that a fraction of the epidemiologically susceptible population is actually susceptible due to precautions, we present a novel perspective on understanding the infection force, incorporating human protection behaviours. This view explains many existing infection force functions and motivates new forms of infection force functions. We demonstrate an SIRS model with time delay, considering disease surveillance of both the current and the past. This delay model aims to answer a specific question: How do disease surveillance data assigned varying weights affect disease control measures? Is it related to how recent these data are? To do so, we study the local bifurcation caused by delay and the parameter reflecting the weight of past epidemics.

In the second project, motivated by the view of infection force in the first project, we propose a general framework model for the above-mentioned second topic. This framework is based on the Kermack-McKendrick model with infection age and can be reduced to various models by choosing particular kernels. We derive renewal equations for the incidence and the infection force, which are integral equations and focus on the processes. Besides generalizing the infection force function in the first project into renewal equations, another primary objective of this project is to explore how human behaviour changes affect the final epidemic size. We found that non-pharmaceutical precautions adoption can reduce the final epidemic size.

In the last project, we formulated a general framework with evolving precaution levels through some specific SIS type of disease models to the first concerned topic. These models attempt to answer the two questions: (1). How does the behaviour response impact the disease dynamics? (2). How does the response level evolve with the disease? In addition, we consider the time lag in behaviour response since it takes time for humans to process the epidemiological information and plan non-pharmaceutical precautions.

Summary for Lay Audience

Human behaviour plays a key role in infectious disease transmission, especially for those infectious diseases transmitted through close person-to-person contact. Triggered by psychological factors (mainly perception of risk and infection fear), people spontaneously change their behaviour, like adopting various control interventions over time, to reduce their susceptibility in response to an epidemic threat. Roles of human behaviour in transmission motivate the necessity of including behaviour response in mathematical epidemiological models. This thesis mainly studies the interaction between human behaviour change as well as its determinants (psychological factors) and infectious disease transmission by mathematical epidemiological models.

Among epidemiology concepts, infection force relates to the probability of a susceptible getting infected and can describe the impact of individual behaviour. We start by revisiting the *infection force* from a new angle and justifying the notion of infection force from the viewpoint of *susceptibility*. After that, we study behaviour change from two sides in mathematical modelling: (1). the impacts of human behavioural response on epidemics; (2). the adaptive evolution of behavioural response levels with epidemic dynamics. In addition, we answer the main three questions by mathematical theoretical techniques: (1). Is it possible for an "epidemic" to become "endemic" when precautionary intervention is employed? What are the necessary conditions? (2). Which factors affect disease control measures if disease surveillance data of present and past time are weighted differently? How is it related to data recentness and weight? (3). Without demography, will some fraction of the population escape infection over whole epidemics if precautionary behaviour is involved? If so, how large is this fraction? How does behaviour response affect this fraction?

Co-Authorship Statement

This thesis is written by Tianyu Cheng under the supervision of Dr. Xingfu Zou. Chapters 2-4 of this thesis consist of the following papers:

Chapter 2: Tianyu Cheng and Xingfu Zou, A new perspective on infection forces with a demonstration by a DDE infectious disease model. *Math. Biosci. Eng*, 19(5) (2022). 4856-4880.

Chapter 3: Tianyu Cheng and Xingfu Zou, On final and peak sizes of an epidemic with latency under some non-pharmaceutical interventions, in preparation.

Chapter 4: Tianyu Cheng and Xingfu Zou, Modelling the impact of precaution on disease dynamics and its evolution, preprint.

The drafts of the above papers were prepared by Tianyu Cheng and then revised by Tianyu Cheng and Dr. Xingfu Zou.

Acknowlegements

Professor. Xingfu Zou, I do not have enough words to express my gratitude to you! Thank you for your support, excellent intellectual guidance, encouragement, and patience throughout the years. You provided me with a great deal of valuable advice and motivation for my work. Throughout my career, your endless enthusiasm for research has enabled me to pursue my interests while exploring exciting new projects! Your dedication to research, caring attitude, and supportive assistance will significantly enhance my future. Moreover, you helped me improve my invaluable technical and communication skills. Though I was far from an ideal student, you believed in me and patiently fostered my growth amid my slow progress and missteps! You are like a beacon of light for me, illuminating the path ahead and guiding me through not only academic but true life! I am truly grateful and proud to be one of your students! You are the best supervisor!! Thank you for everything!

Next, I would like to thank everyone in the Department of Applied Mathematics at Western, especially members of Dr. Zou's group. Thanks to all the professors, especially Dr. Lindi Wahl, Dr. Colin and Dr. Pei Yu. The administration team is greatly appreciated, particularly Audrey Kager, Adriana and Angela.

Finally, I would like to thank my family for their unconditional love and support. We haven't seen each other for four years, and I miss them so much! I also would like to thank my friends Ao, Cun, Doli, Tedi, Yi and Yang for their encouragement and help.

Contents

List of Figures

Chapter 1

Introduction

Human behaviour plays an important role in the spread of infectious diseases. Individuals may change their behaviours during an epidemic. Such behaviour changes are a type of response to the epidemics. Some changes are actively made by individuals based on their knowledge and understanding of the disease; others may be the consequences of implementing some mandatory intervention/control measures by the government and/or public health agents. Understanding how precautionary behaviours affect disease spread can help the public raise awareness of and precautions for the epidemics and help the government and/or public design and plan better control strategies. Therefore, the interplay between disease spread and precautionary behaviour has sparked mathematical epidemiologists' interest [16, 17, 33] and references therein. However, most of the previous works in the literature have been on how precautionary behaviour influences disease transmission; less attention has been paid to how behaviour response evolves with epidemics and the impact of such evolution of the response. In this thesis, we *systematically* develop mechanistic mathematical models that integrate the behavioural response and disease dynamics, including the effects of adaptive behaviour response on disease dynamics and the adaptive evolution of behavioural levels in response to epidemic dynamics.

Among all kinds of control interventions, non-pharmaceutical interventions (NPIs) are often seen as the first line of defence against disease spread, especially when there are no effective pharmaceutical interventions such as vaccines and medications. This has been exemplified by the COVID-19 pandemic. The effectiveness of NPIs, including mask-wearing, social distancing and hand-sanitizing during the COVID-19 pandemic, has been demonstrated by extensive

work, e.g. [6, 17, 43] and some references therein. Even after pharmaceutical interventions became available, NPIs remained crucial due to their versatility and universal accessibility. In addition, NPIs, being behaviour-based prevention measures, offer direct insights into how protective behaviour affects disease spread and how protective behaviour changes over time. An NPI can typically be related to a behaviour response level, which can reflect the impact of some psychosocial factors on disease control. Thus, the study of NPIs is essential for understanding the interplay between disease transmission and behaviour change.

Human behaviour can be *psychologically* driven by the fear of infectious diseases during (or even after an epidemic), and it, in turn, will affect the disease dynamics. This psychological factor can manifest in various ways, such as the perception of disease risk, fear of infection (illness), peer influence, etc. The fear of infection and the perception of risk are two of the main psychological factors that affect non-pharmaceutical control measures. Those who perceive a high risk and have a high fear level of infection are more likely to practice preventative measures, such as frequent hand washing, wearing masks, social distancing and reducing unnecessary social activities. These behaviours can reduce disease transmission. Fear of infection and perception of risk act as motivators, which can enhance the public's adherence to NPIs and improve the effectiveness of NPIs. For instance, during the flu season or the COVID-19 pandemic, the rapid spread and severe symptoms massively reported by the media and warned by public health agencies led to the public's fear of infection and perception of risk, which drove the public to follow the recommended NPIs, such as mask-wearing, practicing hand hygiene, avoiding large gatherings and social distancing. In this thesis, we explore the potential impact of these psychological factors on the dynamics of an infectious disease that involves precautionary responses of the population.

In general, individuals' precaution levels (level of behaviour change) are mainly adapted to the *present severity* of an infectious disease. For example, if an epidemic is more severe, the individuals' fear level will be higher, will perceive a higher risk and are more likely to adopt a higher precaution level; when severity becomes milder, individuals are less likely to adopt NPIs due to the inconvenience or economic costs, which may help escalate the epidemic. The *momentary severity* of an infectious disease is naturally measured by infection-related variables, often *the prevalence of infection*, i.e. *the number of infected cases at a given time* [14].

In reality, a more reasonable measurement of epidemic severity should include information not only at the current but also over a period of time in the past. This is because such information on the infections during a past period may better reflect the epidemic trend. Moreover, in reality, there is always a delay in the availability of disease surveillance at any given time since diagnosing/testing, reporting cases, and collecting the data all take time. Therefore, in this thesis, we will also take into account the *time delay* in disease surveillance (mainly *prevalence*) in both **Chapters 2** and **3**.

Mathematical models are among the most popular tools for exploring the transmission dynamics of infectious diseases. The application of mathematical models to study disease spread can be traced back to the 17th century. Bernoulli demonstrated the effectiveness of cow-pox inoculation against smallpox by a mathematical model [2]. Then, in 1927, the biochemist Kermack and the physician McKendrick marked the dawn of modern mathematical epidemiology through a series of works [21, 22, 23] by mathematical models. However, most traditional and classical mathematical models do not explicitly incorporate behavioural factors. In the 21st century, as technology (needed for computation) improves and knowledge of disease biology increases, more complex models have been proposed, which explicitly include human behaviour and NPIs. In particular, the recent epidemics/pandemics such as SARS, seasonal influenza epidemics, H1N1 influenza pandemics, and the ongoing COVID-19 pandemic have triggered a flood of research and driven advances in these complex models incorporated with behavioural responses and NPIs, e.g. [13, 14, 16, 17]. These models focus on specific diseases and are much more complex and detailed. This imposes some limitations on these models in their extension and usefulness for theoretical purposes. Thus, these models' conclusions are subject to experimental data, which are often non-reproducible, hence making the conclusion less reliable. In this thesis, we aim to provide a general, natural and fundamental model framework to incorporate behaviour changes. This framework can reconstruct some existing models and motivate new models. Further, it allows valuable theoretical and qualitative analysis for mechanistic study.

Generally, direct contact between infected individuals and susceptible individuals is more likely to carry transmission risk during epidemics. Successful transmission of infection depends on three main factors: the closeness of contacts, the infectivity of infectors and the

susceptibility of the susceptible individuals. NPIs can lower contact rates and block routes of transmission. In mathematical modelling, the incidence rate most directly describes the disease transmission process. Infection force is directly related to the incidence rate in a proportional way and is often used (preferred by many modellers) in many models. NPI-adopting can reduce *infection force*, consequently, incidence. *Infection force* can describe the impacts of individuals' behaviour changes or different NPIs-adopting. To explicitly model adaptive human behavioural response, we adopt a mass action infection mechanism to factor the incidence as a product of the *infection force* and infectiveness parameters.

This thesis' core idea in epidemic-mathematical modelling is to identify the revised *infection force* functions by including the adaptation of behaviour response to epidemic severity. Based on the fact that due to the fear of disease and the resulting behaviour change as a psychological response, only a fraction of the epidemiologically susceptible population is actually practically susceptible, we modify the *infection force* functions. To the author's knowledge, the first mathematical epidemiology paper, including behavioural concepts might be [8] published in 1978 in which nonlinear infection forces resulted from the incorporation of behaviour factors. After that, there have appeared more works proposing nonlinear infection forces by considering psychological factors and human behaviour, e.g. [17, 18, 20, 21] are missing in the literature and references therein. But most of these works focus on pure theoretical research in dynamical systems or just briefly justify from the viewpoint of infectivity. Biological explanations regarding the interplay among human behaviour, psychological factors and epidemics are ambiguous. Infection force functions revised by the notion of "practically susceptible" in this thesis can reconstruct most (if not all) those infection force functions in these previous works.

Among various indicators for assessing epidemic overall severity, basic reproduction number and final size are particularly important. These two notions serve as a kind of infectiveness parameters, which reflect the severity of the disease, not in the instantaneous sense but in the overall sense, and hence, they are often considered as the epidemics' inner properties, in contrast to *prevalence* which is of instantaneous nature only.

Next, we summarize these aforementioned and other concepts in mathematical epidemiology that are applied in this thesis.

1.1 Basic concepts in epidemic modelling

Incidence: the rate at which newly infected individuals arise, i.e. the number of newly infected cases per unit of time [14].

Infection force: the probability per unit of time for a susceptible to become infected [14].

Final size: the fraction of the population that became infected sooner or later [14]. This is often described as the attack ratio. Final size is a popular tool for studying the short-term behaviour of epidemic dynamics,i.e. the epidemic's course is rapid enough to include no demography.

Basic reproduction number R_0: the expected number of secondary infections produced when an infectious individual is introduced into a population where all individuals are susceptible. It is a powerful threshold to know if there is an endemic or epidemic. The threshold condition $R_0 > 1$ is related to not only the condition for disease invasion but also the existence and stability of endemic steady states. The most common method defining R_0 is through the spectral radius of the next-generation operator [2, 9, 13, 44]. For some specific and simple scenarios, we can also obtain R_0 by biological meaning directly.

Infection age: duration since an individual gets infected. Inaba et al. vividly describe that it works as a "clock" to mark the process of infection and recovery [14, 22, 23]. It is often applied to describe the dependence of infectivity on time. An infection age can be in the form of continuous or discrete. The discrete form of infection age induces multiple infective stages and treatment stages.

1.2 Epidemiological models

1.2.1 Basic general model: age-structure model

We start with the original Kermack-McKendrick (KM-27) model given in [23], which has served well for around 100 years! The KM-27 model is the stage/age-structure model without the incorporation of behavioural frameworks. Instead, they assumed the population is well-mixed (often uniformly) and homogeneous in behaviour. Additionally, the assumption for this

model is as follows:

- The population is closed, except through disease-induced death, denoting the total number of the population N.

- Individuals cannot infect others at the moment of being infected. Denote the number of those who have just been infected as I_0.

The content of [23] is far fruitful beyond we imagined, not only just about a well-known compartmental SIR model (shown as an example in [23]) but giving the renewal equation of infection force and the final size relation in implicit ways. This general KM model in [23] provides a common way to compartment models.

We denote[1]

$$i(t, a) := \text{the density of infected individuals at infection-age } a \text{ and time } t,$$

$$\beta(a) := \text{ the rate of infectivity at age } a,$$

$$\gamma(a) := \text{the rate of removal, which is the sum of the recovery and disease-induced death rates,}$$

$$S(t) := \text{the number of susceptibles at time } t,$$

$$I(t) := \text{the number of infected individuals (but not necessarily infective) at time } t$$

$$= \sum_{a=0}^{t} i(t, a) \text{ or } \int_{0}^{\infty} i(t, a) da,$$

$$R(t) := \text{the number of removed individuals by recovery and disease-induced death at time } t$$

$$= N - S(t) - I(t),$$

$$F(t) := \text{the infection force at time } t,$$

and then

$$B(t) := \text{incidence} = F(t)S(t).$$

---

[1] ':=' means defining.

So, without host demography,

$$
\begin{cases}
S'(t) = -B(t), \\
F(t) = \displaystyle\int_0^t A(a)B(t-a)da + A(t)I_0, \\
R'(t) = \displaystyle\int_0^t C(a)B(t-a)da + C(t)I_0,
\end{cases}
$$

and

$$
I(t) = \int_0^t \Phi(a)B(t-a)da + \Phi(t)I_0
$$

where

$$
\Phi(a) = e^{-\int_0^a \gamma(a)da}, \; A(a) = \Phi(a)\beta(a), \; \text{and } C(a) = \gamma(a)\Phi(a).
$$

As shown above, the KM-27 model takes the form of an integral equation for infection force $F(t)$. Inspired by this, we integrated the precaution behaviour change into this integral equation for infection force in **Chapter 2**.

By series of computing, the authors gave

$$
\ln\left(\frac{S(0)}{S(\infty)}\right) = A(N - S(\infty))
$$

with $A = \int_0^\infty A(t)dt$, which is the well-known final size relation, explicitly pointed out later by many researchers, e.g. [3, 13, 26]. And here implicit $R_0 = AN$. See detailed derivations and demonstrations in [23].

In [23], the authors started with the assumption that the density $i(t, a)$ is discrete in infection age, having

$$
i(t, a) - i(t+1, a+1) = \gamma(a)i(t, a).
$$

Indeed, McKendrick's earlier work [34] proposed the continuous one, and details of it are given in [10]. The continuous one is a form of

$$
\frac{\partial i(t, a)}{\partial t} + \frac{\partial i(t, a)}{\partial a} = -\gamma(a)i(t, a).
$$

In **Chapter 2**, we applied this partial differential equation (PDE) to formulate the infection-age

density.

1.2.2   The special case: compartmental models

As aforementioned, [23] gave a compartmental SIR (susceptible-infective-recovered) model as a very special case of the general KM-27 model. It is acknowledged as the earliest compartmental model. Most of the mathematical epidemic models that appeared after this are detailed versions of this model, including more focused information. The compartmental SIR model is of the form:

$$
\begin{cases}
S'(t) = -\beta S(t)I(t), \\
I'(t) = \beta S(t)I(t) - \gamma I(t), \\
R'(t) = \gamma I(t),
\end{cases}
$$

where β represents the transmission coefficient, and γ denotes the recovery/epidemic-induced death rate in the infected class I. The force of infection here is $F(t) = \beta I(t)$. And $R_0 = N\beta/\gamma$. Denote the final size $Z := 1 - S(\infty)/N$. In the case $I(0)/N \to 0$ and $S(0)/N \to 1$, the usual final size relation is

$$
Z = 1 - e^{-R_0 Z}.
$$

For more details of the final size relation, we refer to the book by Diekmann [13], and the work [26, 35].

In the later work [22], Kermack and McKendrick extended the model proposed in [23] by involving births and death. The total population size may change exponentially (grow or die out). The study of the final size becomes meaningless in this model since the population goes through renewal. Later, in 1976, Hethcote suggested the assumption that total population size is constant, with birth and death rates equal in [19]. For ease of theoretical analysis in **Chapters 2 and 3**, we follow this assumption. Such a simple example is as follows.

$$
\begin{cases}
S'(t) = -\beta S(t)I(t) + (dN - dS(t)), \\
I'(t) = \beta S(t)I(t) - \gamma I(t) - dI(t), \\
R'(t) = \gamma I(t) - dR(t),
\end{cases}
$$

where d is the birth/natural death rate, and the constant $\Lambda := dN$ is often called the recruitment rate.

1.3 Thesis outline and objectives

The main goal of this thesis is to explore the interaction between human behaviour change and infectious disease dynamics. The critical point is to build behaviour into the modelling framework given by Kermack and McKendrick [23]. More precisely, as stated earlier, we restructure the expression of the *infection force* by incorporating humans' precautions behaviour and response level. From the view of model forms, we consider both age-structure models (in **Chapter 2**) and compartment models (in **Chapters 1 and 3**), depending on the focus modelling. From different time scales of transmission, we discuss both the final size for short-term dynamics (in **Chapter 2**) and endemics states for long-term dynamics (in **Chapters 1 and 3**). This thesis is structured as follows.

In **Chapter 2**, we revisit the notion of infection force from a new angle, which can offer a new perspective to motivate and justify some infection force functions. Our approach can not only explain many existing infection force functions in the literature, most importantly, but it can also motivate new forms of infection force functions incorporating human protective behaviours, particularly infection forces depending on disease surveillance of the past. As a demonstration, we propose a SIRS model with delay. We comprehensively investigate this model's dynamics, mainly focusing on the local bifurcation caused by the delay and another parameter that reflects the weight of past epidemics in the infection force. The results in this chapter show that, depending on how recent the disease surveillance data are, their assigned weight may have a different impact on disease control measures.

In **Chapter 3**, we explore the final size of epidemics featured by infection-age structure, considering the adoption of NPIs. For the purpose of modelling, we structure a general model based on the age-structure model in [23] and give examples with particular kernels to describe this circumstance. Moreover, we derive a renewal equation for the (both instantaneous and cumulative) force-of-infection, which are integral equations and describe the relationship between the history and future of the force-of-infection. Using this integral equation for the

cumulative force of infection, we get the general formula of the final size relation. Accordingly, we found that NPIs-adopting reduces the population's eventual involvement in the epidemic. Additionally, we theoretically demonstrate that the uptake of NPIs reduces the infection peak and herd immunity threshold in some specific models.

In **Chapter 4**, we explore the features of adaptive behaviour response to epidemics through some specific SIS type of disease models. Triggered by psychological factors, people spontaneously adopt NPIs to reduce their susceptibility in response to an epidemic threat. Our model describes the fact that the behaviour response levels mainly affect the severity of epidemics by directly adjusting the number of *practically susceptible* (the susceptible who have effective contact with infectors). In turn, the severity change of epidemics leads to the evolution of the behaviour response levels. Mathematical analysis shows that when $R_0 > 1$, the endemic equilibrium exists and is subjective to a critical parameter that combines the initial response level and the initial number of the infectious class. In addition, considering the time lag in disease surveillance, we further extend our model, which is a system of delay differential equations (DDE). We figure out the condition that Hopf bifurcations occur for this DDE system by theoretical and numerical techniques.

Chapter 5 concludes this thesis by summarizing the main results in a brief summary. Moreover, we discuss possible future extensions.

1.4   Mathematical theories and methodologies

In this thesis, all the models are in the form of differential equations within dynamical systems(i.e. state-space models). We apply differential equations to describe the dynamics of epidemics in the population. Before studying the behaviour of dynamics, the first step is ensuring the models are well-posed, which means biological state variables remain non-negative and bounded. The natural approach to studying long-term behaviour in a dynamic system is finding the equilibria and determining their stability. The typical technique to determine the local stability of equilibria is linearization. As for the global stability of equilibria, there is no general technique. However, using the Lyapunov function method can realize the determination of global stability in some particular cases [24, 37]. Biologically, individuals within

particular compartments are at equilibrium, meaning the flux of individuals entering and leaving these compartments is balanced. In epidemiology, two important equilibria are focused on: (1) the disease-free equilibrium (DFE) and the endemic equilibrium (EE), corresponding to the individuals in infected states persisting indefinitely.

Since the explicit solution is not obtainable for all dynamic systems involved in my thesis, we focus on the behavioural change of these solutions responding to the adjustment of external parameters. Mathematically, this is called bifurcation theory. Among various types of bifurcations, we pay more attention to Hopf bifurcation, which can generate limit cycles. [24, 37]

However, the models without demographic effects have a line of equilibria: every point with $I = 0$. Linearizing this type of dynamical system about each equilibrium is not applicable because the corresponding linearization matrixes at all equilibria have a zero eigenvalue. The alternative approach to analyzing this type of system is deriving final sizes. We find a relation between $I(t)$ and $S(t)$. Using this relation and $I(\infty) = 0$, we can obtain $S(\infty)$.

Besides compartment models consisting of differential equations, we also obtain the renewal equations of incidence formatted by integral equations. Next, I outline the core mathematical concepts and techniques mentioned briefly.

1.4.1   Stability analysis for equilibria

Equilibria are time-independent solutions of differential equations, revealing the steady-state behaviours of their dynamical systems. An equilibrium is locally stable if the solution trajectories starting nearby eventually converge to it and unstable if it repels solutions. Moreover, If any solution trajectories go to the equilibrium from any initial point in the feasible region, this equilibrium is global asymptotic stable. The dynamic behaviour around the hyperbolic equilibrium of the original nonlinear system is equivalent to the behaviour of the linearization system at this equilibrium. Thus, the local stability of the hyperbolic equilibrium can be obtained by studying the eigenvalues of the corresponding linearization matrix, called the Jacobian matrix, as well. The criteria are

- If all eigenvalues have negative real parts, the equilibrium is locally stable;

- If at least one of the eigenvalues has positive real parts, the equilibrium is unstable.

While if at least a simple pair of purely imaginary eigenvalues of the Jacobian matrix evaluated at the equilibrium exists, this equilibrium is non-hyperbolic. The below criteria are invalid. A non-hyperbolic equilibrium is one of the necessary conditions for the existence of Hopf bifurcation points. The other is the transversal condition, which means the purely imaginary eigenvalue crosses the imaginary axis in a complex plane when external parameters vary [24, 37].

1.4.2 Renewal equation

A renewal equation is an integral equation concerning an unknown function. The classical standard renewal equation is

$$x(t) = g(t) + \int_0^t x(t-s)f(s)ds, \quad t \geq 0 \tag{1.1}$$

where $g(t)$, $f(t)$ are known functions and $x(t)$ is a unknown function [18, 40]. Here, g is a measurable function defined on $[t_0, \infty)$ and called the *force function*. Additionally, g is bounded on any finite interval. If $g = 0$, Eq.(1.1) represents the linear renewal equation. The f is a *kernel* function on $[t_0, \infty)$. The renewal equation as a modelling tool is applied to trace processes of births/incidence versus time in biology [2, 17, 26, 27] and epidemiology [8, 7, 14]. In many cases, asymptotic behaviour as $t \to \infty$ of the renewal equation can provide much information for questions of interest. The renewal theorem is one of the most powerful tools for asymptotic behaviour of $x(t)$. The following variation of (1.1) is usually applied

$$x(t) = \int_0^\infty x(t-s)f(s)ds \quad t \geq 0. \tag{1.2}$$

If $x(t)$ in equation (1.2) express the total reproductive rate, $f(t)$ can be thought of as the product of age-dependent mortality and age-dependent reproduction [17, 26, 27]. In an epidemiological context, if $x(t)$ represents new infectious individuals, $f(t)$ can be interpreted as the product of the transmission rate and the probability of being infectious [8, 7].

By giving the different assumption on the *kernel* function, renewal equations can be sim-

plified as ordinary and delay differential equations. Although analytical methods for solving a renewal equation are generally unavailable, numerical techniques are desirable [5, 42].

This chapter gives basic background, motivation, and a brief introduction to related concepts in epidemic models and mathematical tools.

Bibliography

[1] R. M. Anderson and R. M. May, Infectious diseases of humans: dynamics and control, *Oxford University Press*, Oxford, UK, 1991.

[2] F. Brauer, On a nonlinear integral equation for population growth problems, *SIAM J. Math. Anal.*, **6(2)** (1975) 312-317.

[3] F. Brauer and C. Castillo-Chavez, Mathematical models in population biology and epidemiology, *Springer, New York*, 2001.

[4] D. Breda, O. Diekmann, W. F. De Graaf, A. Pugliese and R. Vermiglio, On the formulation of epidemic models (an appraisal of Kermack and McKendrick), *J. Biol. Dyn.*, **6(sup2)** (2012), 103-117.

[5] D. Breda, O. Diekmann, D.Liessi and F. Scarabel, Numerical bifurcation analysis of a class of nonlinear renewal equations, *Electron. J. Qual. Theory Differ. Equ.*, (2016).

[6] R. E. Baker, S. W. Park, W. Yang, G. A. Vecchi, C. J. E. Metcalf and B. T. Grenfell, The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections, *Proc. Natl. Acad. Sci. U.S.A.*, **117(48)** (2020), 30547-30553.

[7] K. L. Cooke and J. A. Yorke, Some equations modelling growth processes and gonorrhea epidemics, *Math. Biosci.*, **16(1-2)** (1973), 75-101.

[8] V. Capasso and G. Serio, A generalization of the Kermack-McKendrick deterministic epidemic model, *Math. Biosci.*, **42(1-2)** (1978), 43-61.

[9] O. Diekmann, J. A. P. Heesterbeek and J. A. 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** (1990), 365-382.

[10] K. Dietz, Introduction to McKendrick (1926) Applications of mathematics to medical problems, *Breakthroughs in Statistics*, (1997), 17-57.

[11] O. Diekmann and J. A. P. Heesterbeek, Mathematical epidemiology of infectious diseases: model building, analysis and interpretation, *Wiley*, 2000.

[12] O. Diekmann, H. Heesterbeek and T. Britton, Mathematical tools for understanding infectious disease dynamics (Vol. 7), *Princeton University Press*, **7** 2013.

[13] L. Di Domenico, G. Pullano, C. E. Sabbatini, P. Y. Boelle and V. Colizza, Impact of lockdown on COVID-19 epidemic in file-de-France and possible exit strategies, *BMC Medicine*, **18(1)** (2020), 1-13.

[14] J. M. Epstein, J. Parker, D. Cummings and R. A. Hammond, Coupled contagion dynamics of fear and disease: mathematical and computational explorations, *PloS one.*, **3(12)** (2008), e3955.

[15] W. Feller, On the logistic law of growth and its empirical verifications in biology, *Acta Biotheor.*, **5** (1940), 51-66.

[16] W. Feller, On the integral equation of renewal theory, *Ann. Math. Stat*, **12(3)** (1941), 243-267.

[17] S. Funk, M. Salathé and V. A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: a review, *J. R. Soc. Interface.*, **7(50)** (2010), 1247-1256.

[18] J. Guckenheimer and P. Holmes, Nonlinear oscillations, dynamical systems, and bifurcations of vector fields (Vol. 42), *Springer Science Business Media*, 2013.

[19] H.W. Hethcote, Qualitative analysis for communicable disease models, *Math. Biosci.*, **28** (1976), 335-356.

[20] H. Inaba, Age-structured population dynamics in demography and epidemiology, *Springer Singapore*, 2017.

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

[22] W. O. Kermack and A. G. McKendrick, Contributions to the mathematical theory of epidemics. II. The problem of endemicity, *Proc. R. soc. Lond. Ser. A-Contain. Pap. Math. Phys. Character.* , **138(834)** (1932), 55-83.

[23] W. O. Kermack and A. G. McKendrick, Contributions to the mathematical theory of epidemics. III. -Further studies of the problem of endemicity, *Proc. R. soc. Lond. Ser. A.*, **141(843)** (1933), 94-122.

[24] Y. A. Kuznetsov, I. A. Kuznetsov, and Y. Kuznetsov, Elements of applied bifurcation theory (Vol. 112), *New York: Springer*, 1998.

[25] A. J. Lotka, Relation between birth rates and death rates. *Science*, **26(653)** (1907), 21-22.

[26] A. J. Lotka, The stability of the normal age distribution, *Proc. Natl. Acad. Sci. U.S.A.*, **8(11)** (1922), 339-345.

[27] A. J. Lotka, Contact points of population study with related branches of science, *Proc. Am. Philos. Soc.*, **80(4)** (1939), 601-626.

[28] W. Liu, S. A. Levin and Y. Iwasa, Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models, *J. Math. Biol.*, **23** (1986), 187-204.

[29] W. Liu, H. W. Hethcote and S. A. Levin, Dynamical behavior of epidemiological models with nonlinear incidence rates, *J. Math. Biol.*, **25** (1987), 359-380.

[30] M. Lu, J. Huang, S. Ruan and P. Yu, Bifurcation analysis of an SIRS epidemic model with a generalized nonmonotone and saturated incidence rate, *J. Diff. Eqns.*, **267** (2019), 1859-1898.

[31]   M. Lu, J. Huang, S. Ruan and P. Yu,   Global dynamics of a susceptible-infectious-recovered epidemic model with a generalized non-monotone incidence rate, *J. Dyn. Diff. Eqns.*, **33** (2020), 1-37.

[32] J. Ma and D. J. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease, *Bull. Math. Biol.*, **68** (2006), 679-702.

[33] P. Manfredi and A. D'Onofrio, Modelling the interplay between human behaviour and the spread of infectious diseases, *Springer Science Business Media*, Eds, 2013.

[34] A. G. Mkendrick, Applications of mathematics to medical problems, *Proc. Edinb. Math. Soc.*, **44** (1925), 98-130.

[35] J. C. Miller, A note on the derivation of epidemic final sizes, *Bull. Math. Biol.*, **74(9)** (2012), 2125-2141.

[36] B. Morsky, F. Magpantay, T. Day and E. Akcay, The impact of threshold decision mechanism of collective behaviour on disease spread, *Proc. Natl. Acad. Sci. U.S.A.*, **120(19)** (2023), e2221479120.

[37] L. Perko, Nonlinear systems: global theory, *Differ. Equ. Dyn. Syst.*, (1996), 179-310.

[38] N. Perra, D. Balcan, B. Gonçalves and A. Vespignani, Towards a characterization of behaviour-disease models, *PloS one.*, **6(8)** (2011), e23084.

[39] Z. Qiu, B. Espinoza, V. V. Vasconcelos, C. Chen, S. M. Constantino, S. A. Crabtree and M. V. Marathe, Understanding the coevolution of mask-wearing and epidemics: A network perspective, *Proc. Natl. Acad. Sci. U.S.A.*, **119(26)** (2022), e2123355119.

[40] W. L. Smith, Renewal theory and its ramifications, *J. R. Stat. Soc. Ser. B Methodol.*, **20(2)** (1958), 243-284.

[41] C. I. Siettos and L. Russo, Mathematical modelling of infectious disease dynamics, *Virulence.*, **4(4)** (2013), 295-306.

[42] F. Scarabel, O. Diekmann and R. Vermiglio, Numerical bifurcation analysis of renewal equations via pseudospectral approximation, *J. Comput. Appl. Math.*, **397** (2021), 113611.

[43] L.Thunstrom, S. C. Newbold, D. Finnoff, M. Ashworth and J. F. Shogren, The benefits and costs of using social distancing to flatten the curve for COVID-19, *J. Benefit-Cost Anal*, **11(2)** (2020), 179-195.

[44] P. Van den Driessche and J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Math. Biosci.*, **180(1-2)** (2002), 29-48.

Chapter 2

A new perspective on infection forces with demonstration by a DDE infectious disease model

2.1   Introduction

Incidence rate in an infectious disease model plays a crucial role in the disease dynamics. It is the rate (i.e. number of cases per unit time) of incidence of new infections coming from the susceptible population. Expressing the incidence rate in the form fS with S being the susceptible population, then the role of the incidence rate is represented by f, which is often referred to as the infection force. An infection force can be dependent on $I(t)$ or on both $I(t)$ and $S(t)$ where $I(t)$ and $S(t)$ are the populations of the infectious and susceptible classes of the host. Actually, there have been many works in literature that investigate the disease dynamics described by various types of ODE disease models (SI, SIS, SIR, SIRS, etc.) adopting various infection force functions (or incidence rates), see, e.g. [1, 2, 3, 15, 16, 17, 18, 20, 21, 23, 25, 26, 31] and the references therein. For example, in Liu et al. [17, 18] the incidence rate $\beta I^p S^q$ were used for some in SIRS and SEIRS models respectively, and it was later also used by Korobeinikov and Maini [15]. This incidence rate corresponds to an infection force function of the form $f = f(I, S) = \beta I^p S^{q-1}$. In Ruan and Wang [26], a saturated infection force

$f = f(I) = aI^2/(b + I^2)$ was adopted in an SIRS model. In Korobeinikov and Maini [16], general incidence rate of the form $g(S, I, N)$, where N is the total population, was discussed and the global properties of some infectious disease models with various nonlinear g were explored. In Wang [31], some general nonlinear infection forces were discussed for an SIRS ODE model with balanced demography, and these can include the form $f = \lambda I/(1 + pI + qI^2)$ and $f = \lambda I^2/(1 + pI + qI^2)$ which are infection forces discussed in, e.g., [20, 21, 25]. In Liu et al [11] for an SEIHR ODE model, an infection force of the form $\beta I e^{-a_1 E - a_2 I - a_3 H}$ was used where $E(t)$ and $H(t)$ are the numbers of exposed and hospitalized hosts; and along the same line but for an SEI ODE model with logistic demography for the host, an infection force of the form $\beta I e^{-mI}$ was used in [1]. Particularly in McCallum et al. [23], some valuable discussions on transmission functions are presented from biological perspectives, and seven particular transmission functions were collected and compared, including some of the above-mentioned ones.

It has been found that some *nonlinear* infection force functions, particularly those *non-monotone* ones, can lead to very complicated disease dynamics, as opposed to the linear infection force arising from mass action incidence rate. For details in this regard, a reader is referred to, e.g., [1, 11, 15, 20, 21, 25, 26] and the references therein.

We note that the aforementioned literature on nonlinear infection force mainly focuses on pure theoretical research in dynamical systems and, hence, has brief justifications from the viewpoint of *infectivity*, and some are even ambiguous and out of reasonable biological explanations. Here, we first present a novel alternative perspective for considering infection force functions, which can not only reconstruct the infection force function mentioned above but also help motivate our new model. The main idea is that during an epidemic or pandemic, particularly for a newly emerged disease such as the SARS in 2003 and Covid-19 broken out in 2020, due to various control measures including lockdowns, quarantines and isolations, ordered stay-home and even effective use of personal protection equipments (PPEs), only *a proportion of the susceptible individuals* will actually be possibly exposed to the infectious hosts and hence be possibly infected. For convenience of later references, let us call this proportion of susceptible individuals *practically susceptible* individuals. A susceptible host who takes

extreme precautions by either staying home or using full PPEs has zero or little chance of being infected. The practically susceptible population can be expressed as $S_p = PS(t)$, where $P \in [0, 1]$ is a measurement of the proportion; Alternatively, this P can also be explained as the probability that a susceptible individual is a practically susceptible host. In general, such a proportion is non-pharmaceutical and non-biological, and it depends on the level of precaution that is impacted by the level of severity of the epidemics, mediated by, e.g., the control measures implemented by the government, regulations and guidelines imposed by the public health agency, as well as media coverage. Let $L(t)$ denote the measurement of the level of severity, and accordingly, $P = P(L)$. As the severity level increases, people will generally become more precious and more protective by either restricting their activities or using PPEs; moreover, the government may implement more and more restrictive control measures, which will make fewer and fewer people available for infection. These facts justify the following basic assumption for the proportion function $P(L)$:

(H1) $P(L)$ is a non-increasing function, satisfying $P(0) \leq 1$ and $\lim_{L\to\infty} P(L) \geq 0$.

Below are just three prototypes of such a function satisfying (H1):

(A) $P_1(L) = \frac{m_1 L + 1}{m_2 L + 1}$ where m_1, m_2 are all positive constants satisfying $\frac{m_1}{m_2} < 1$;

(B) $P_2(L) = \frac{m_1 L + b}{cL^2 + m_2 L + b}$, where all parameters are positive constants and satisfy $\frac{m_1}{m_2} < 1$;

(C) $P_3(L) = e^{-hL}$, where $h > 0$.

Note that although $P_j(L)$, $j = 1, 2, 3$ all satisfy (H1), their decay rates are different, with $P_1(L)$ decaying slowest and $P_3(L)$ decaying the fastest. Also, $P_1(L)$ has a positive lower bound $m_1/m_2 < 1$, and this may accommodate the scenario of maintaining essential services open during a severe epidemic, such as COVID-19.

Now let us consider the mass action infection mechanism which has the incidence rate at time t as $\beta I(t)S(t)$. This corresponds to a linear infection force $f(t) = \beta I(t)$, where β is the transmission rate. With the above observation of *practically susceptible* population, the incidence rate under the mass action infection mechanism should be revised to $\beta I(t)S_p(t) = \beta I(t)[P(L(t))S(t)] = [\beta I(t)P(L(t))]S(t)$, resulting in the infection force $f(t) = \beta I(t)P(L(t))$.

This way, the infection force at time t is explicitly related to the proportion function $P(L(t))$ which depends on the severity level $L(t)$ at time t. When $L(t)$ is measured just by the number of infected individuals at time t, i.e. $L(t) = I(t)$, those infection force functions used in the aforementioned literature can be easily obtained by properly chosen proportion function $P(I)$ for the practically susceptible population. In this case, the infection force $f = f(I) = IP(I)$ can be monotone, either increasing and unbounded, or increasing and saturated if $P(I)$ decreases slowly; it can also be non-monotone (one hump shape) if $P(I)$ decreases fast enough, including those used in [1, 11, 15, 20, 21, 25, 26] where complicated dynamics have been observed. We point out that Liu et al [11] proposed and analyzed an SEIR-H model with $E(t)$ and $H(t)$ representing the populations of exposed and hospitalized groups, and the infection force there is a result of adopting $L(t) = a_1 E(t) + a_2 I(t) + a_3 H(t)$ and $P(L) = e^{-L(t)}$.

Note that in practice, a more reasonable and meaningful measurement for the severity of an epidemic should contain information not only at the present time but also certain information over a period of the past time. When determining their precaution and protection levels, people do look at the numbers of infected cases both at the present time and in recent times. This suggests the following form for $L(t)$:

$$L(t) = \int_0^\tau w(\xi) I(t - \xi) \, d\xi \qquad (2.1)$$

where the constant $\tau > 0$ specifies a length of time interval that one wants to look at, and the weight function $w(\xi)$ reflects the variation of the impact of disease surveillance in the past interval $[t - \tau, t]$ on the severity at the present time. Considering the fact that, in reality, the data are typically collected at discrete times (e.g., hourly, daily and weekly), $w(\xi)$ is accordingly taken as $w(\xi) = \sum_{j=0}^n k_j \delta(\tau_j)$, leading to the following discrete form for $L(t)$:

$$L(t) = \sum_{j=0}^n k_j I(t - \tau_j), \qquad (2.2)$$

where $k_j \geq 0, \ j = 0, 1, \cdots, n$ are the discrete weights, satisfying the normal condition $\sum_{j=0}^n k_j = 1$; $\tau_0 = 0$ and $0 < \tau_1 < \tau_2 < \cdot < \tau_n$ are the positive numbers that may account for the past n time points at which the infected cases are reported/released. When the data in these past n

time points and the present time are all given the same weight, there holds $k_j = 1/(n + 1),\ j = 0, 1, \cdots, n$.

To demonstrate the above new perspective for infection force, we incorporate the above new framework for infection force into the classic SIRS model to obtain the following model system for the transmission dynamics of an infectious disease of SIRS type:

$$\begin{cases} S'(t) = \Lambda - dS(t) - f(t)S(t) + \alpha R(t), \\ I'(t) = f(t)S(t) - (d + r + \epsilon)I(t), \\ R'(t) = rI(t) - (d + \alpha)R(t). \end{cases} \tag{2.3}$$

where, as discussed above, the infection force at time t is given by

$$f(t) = \beta I(t)P(L(t)), \quad L(t) = \sum_{j=0}^{n} k_j I(t - \tau_j), \tag{2.4}$$

with the proportion function satisfying (H1). Although $P(L)$ does not have to be smooth and even continuous to reflect the situation when the control measures are often implemented in terms of time intervals, for convenience of theoretical analysis, we assume $P(L)$ is differentiable, in addition to (H1). Notice that all parameters are positive in model (2.3); in addition to the transmission rate β, the others respectively stand for

- Λ: the recruitment rate of the population;

- d: the natural death rate of the population;

- r: the recovery rate of infective individuals;

- ϵ: the disease-induced death rate;

- α: the rate of removed individuals who lose immunity and return to the susceptible class.

We point out that in a very special case: $\epsilon = 0,\ \alpha = 0$ and $n = 1,\ k_0 = 0$ and $k_1 = 1$ (i.e., severity at time t is measured by $L(t) = I(t - \tau)$) and $P(L) = e^{-hL}$, system (2.3) reduces to the

delayed SIR model

$$
\begin{cases}
S'(t) = \Lambda - dS(t) - \beta e^{-hI(t-\tau)}I(t)S(t), \\
I'(t) = \beta e^{-hI(t-\tau)}I(t)S(t) - (d+r)I(t), \\
R'(t) = rI(t) - dR(t).
\end{cases}
\tag{2.5}
$$

which was thoroughly discussed recently by Song and Xiao [29], including a local and global bifurcation analysis by using the delay τ as the bifurcation parameter, which reveals the onset and termination of Hopf bifurcation from a positive equilibrium. The authors attribute the term $e^{-hI(t-\tau)}$ as the impact of media.

In more recent work, in further studying the impact of media on the disease dynamics of a SEIS model, Song and Xiao [30] introduced a variable $M(t)$ to denote the average number of news items related to the disease outbreak and proposed the following SEIS-M model

$$
\begin{cases}
S'(t) = \Lambda - dS(t) - \beta e^{-hM(t-\tau_2)}I(t)S(t) + \gamma I, \\
E'(t) = \beta e^{-hM(t-\tau_2)}I(t)S(t) - (d+\sigma)E(t), \\
I'(t) = \sigma E(t) - (d+\gamma)I(t), \\
M'(t) = \delta I(t-\tau_1) - \mu M(t).
\end{cases}
\tag{2.6}
$$

Threshold dynamics and bifurcation from an endemic equilibrium were explored in [30]. Here, the media coverage variable $M(t)$ plays a similar role to that of the severity variable $L(t)$ in our model (2.3).

The models (2.3)-(2.4) and (2.5) have something in common: the infection forces contain time delay. We would like to draw readers' attention to another scenario for delay−dependent infection forces but for *vector-borne diseases*. The key idea traces back to Cooke [4] where it was assumed that the *population of the infectious vectors at time t* is proportional to the *population of the infectious host in the previous time $t - \tau$* with τ being the latency of disease in the vector. This suggests the incidence rate for the host $F(I_v(t))S_h(t) = F(pI_h(t - \tau))S_h(t)$, and accordingly, reduces a model system for vector-borne disease to a system only containing the respective variables (susceptible and infectious etc) for the host populations, but with delayed infection force function. Along this time, there have been many works with the corresponding models typically demonstrating the global threshold dynamics between the disease-free equi-

librium and endemic equilibrium, thanks to the powerful Lyapunov method. The success of this method is mainly attributed to assuming that F only depends on the delayed term $I_h(t - \tau)$ and $F(\cdot)$ is nondecreasing. For some details, see, e.g., [13, 14, 15, 24, 28, 33, 34] and the references. therein.

In the rest of this paper, we explore the dynamics of the SIRS model (2.3) with the infection force given by (2.4). In Section 2.2, we address the well-posedness of (2.3)-(2.4) with general $P(L)$. In Section 2.3 we identify the basic reproduction number R_0 and discuss the equilibria of the model system in the desired region. In Section 2.4, we consider a particular $P(L(t)) = e^{-hL}$ which leads to the special infection force $f(t) = \beta I(t)e^{-h(k_0 I(t) + k_1 I(t - \tau))}$, resulting to the model system (2.13). For this infection force, and in the case of $R_0 > 1$ so that a unique endemic equilibrium E^* exists, we analyze the stability of E^* and obtain conditions for Hopf bifurcation to occur. In Section 2.5, we compare our results with those in [29] which is a special case of (2.13) ($\alpha = 0, \epsilon = 0, k_0 = 0, k_1 = 1$). Such a comparison particularly reveals the difference in the disease dynamics between using multiple-time disease surveillance (i.e., $I(t)$ and $I(t - \tau)$) and only using single-time disease surveillance (i.e., $I(t - \tau)$ only). In Section 2.6, we present some numerical simulations to illustrate the theoretical results. Finally, we conclude the paper with a brief summary together with some discussion in Section 2.7.

2.2 Well-posedness of the model

As is customary, we denote by $C = C\left([-\tau, 0], \mathcal{R}^3\right)$ the set of all continuous functions defined on $[-\tau, 0]$, and by $C_+ = C\left([-\tau, 0], \mathcal{R}_+^3\right)$ the subset of all of non-negative functions in C. Then C is a Banach space with the maximum norm, and C_+ is the positive cone of C, which induces the natural order in C.

By the fundamental theory of functional differential equations (see, e.g., Hale and Ver-duyn Lunel [12]), for $\phi = (\phi_1, \phi_2, \phi_3)^T \in C$, the model system (2.3) has a unique solution $(S(t), I(t), R(t))$ satisfying

$$(S(\theta), I(\theta), R(\theta)) = (\phi_1(\theta), \phi_2(\theta), \phi_3(\theta)), \quad \theta \in [-\tau, 0], \tag{2.7}$$

which exists in a maximal interval of existence $(0, T_m)$.

By the nature of the variables in the model (2.3)-(2.4), we are only interested in non-negative solutions. Accordingly, we further require the initial functions to be in C_+. For such non-negative initial functions, we show below that the corresponding solution is also non-negative. Firstly, from the second equation in (2.3)-(2.4), we have

$$I(t) = I(0)e^{\int_0^t [\beta P(L(\theta))S(\theta)-(d+r+\epsilon)]d\theta} \geq 0, \text{ for } t \in (0, T_m).$$

Similarly, the third equation in (2.3) lead to

$$R(t) = R(0)e^{-\alpha t} + r \int_0^t I(\theta)e^{-\alpha(t-\theta)}d\theta,$$

and thus, by $R(0) \geq 0$ and the non-negativity of $I(t)$ confirmed above, we conclude $R(t) \geq 0$ for $t \in (0, T_m)$. To verify that $S(t) > 0$ for $t > 0$, we assume the contrary. Let $t_1 > 0$ be the first time such that $S(t_1) = 0$, then by the first equation of (2.3) we have $S'(t_1) = \Lambda + \alpha R(t_1) > 0$ and therefore, $S(t) < 0$ for $t \in (t_1 - \epsilon_1, t_1)$ where $\epsilon_1 > 0$ is sufficiently small. This contradicts $S(t) > 0$ for $t \in [0, t_1)$, and the contradiction proves $S(t) > 0$ for all $t \in (0, T_m)$.

Next, we prove the boundedness of the solution. To this end, we consider $N(t) = S(t) + I(t) + R(t)$. Simple calculations lead to

$$N'(t) = \Lambda - dN - \epsilon I \leq \Lambda - dN.$$

This implies that $\limsup_{t \to \infty} N(t) \leq \Lambda/d$, concluding the boundedness of $N(t)$. This together with the non-negativity of all components in the solution, implies that $S(t)$, $I(t)$ and $R(t)$ are all bounded. By the theory on the continuation of solutions for functional differential equations (see, e.g., [12]), the boundedness of the solution also implies that the $T_m = \infty$, i.e., the solution exists globally.

Combining the above, we have actually established the well-posedness of (2.3)-(2.4) as stated in the following Theorem.

Theorem 2.2.1 *For every $\phi = (\phi_1, \phi_2, \phi_3)^T \in C_+$, the model system (2.3)-(2.4) has a unique solution $(S(t), I(t), R(t))$ satisfying (2.7), which exists globally in $(0, \infty)$, is nonnegative and*

remains bounded. Moreover, the region

$$D = \{(S, I, R) \mid S \geqslant 0, I \geqslant 0, R \geqslant 0, 0 \leqslant S + I + R = N \leqslant \Lambda/d\}$$

is both positively invariant and attractive for (2.3)-(2.4).

2.3 Disease free equilibrium and basic reproduction number

It is easy to see that the model (2.3) has the disease-free equilibrium $E_0 = (S_0, 0, 0)$, where $S_0 = \Lambda/d$. To explore the stability of E_0 and identify the basic reproduction number of the model system (2.3)-(2.4), we linearize the model at E_0 to obtain

$$\begin{cases} S'(t) = -dS(t) - \beta P(0)S_0 I(t) + \alpha R(t) \\ I'(t) = [\beta P(0)S_0 - (d + r + \epsilon)]I(t) \\ R'(t) = rI(t) - (d + \alpha)R(t). \end{cases} \tag{2.8}$$

This is a linear ODE system (although (2.3)-(2.4) is a DDE system) with the coefficient matrix

$$\begin{pmatrix} -d & -\beta P(0)S_0 & \alpha \\ 0 & \beta P(0)S_0 - (d + r + \epsilon) & 0 \\ 0 & r & -(d + \alpha) \end{pmatrix}, \tag{2.9}$$

which has two negative eigenvalues are $-d < 0$, $-(d + \alpha) < 0$, with the third eigenvalue $\beta P(0)S_0 - (d + r + \epsilon)$, $-(d + \alpha)$ being negative if and only if $\beta P(0)S_0 < (d + r + \epsilon)$. Thus E_0 is asymptotically stable if $\beta P(0)S_0 < (d + r + \epsilon)$ and unstable if $\beta P(0)S_0 > (d + r + \epsilon)$.

Actually, we can prove that E_0 is globally asymptotically stable when $\beta P(0)S_0 < (d+r+\epsilon)$. Indeed, when proving the boundedness of solutions, we have seen that

$$\limsup_{t \to \infty} S(t) \leq \limsup_{t \to \infty} N(t) \leq \Lambda/d = S_0.$$

Choose a $\delta > 0$ sufficiently small such that $\beta P(0)(S_0 + \delta) < (d + r + \epsilon)$. Then, there is a $T > 0$

such that

$$\limsup_{t\to\infty} S(t) \le S_0 + \delta, \quad \text{for } t \ge T.$$

By the conditions on $P(L)$, the second equation in (2.3) has the following linear ODE as a comparison equation from large t from above

$$I'(t) = [\beta P(0)(S_0 + \delta) - (d + r + \epsilon)]I(t), \quad \text{for } t \ge T. \tag{2.10}$$

Note that all solutions converge to 0 as $t \to \infty$ (since $\beta P(0)(S_0 + \delta) < (d + r + \epsilon)$). By a comparison argument, the $I(t)$ component of the solution to (2.3)-(2.4) also tends to 0 as $t \to \infty$. This implies that the third equation in (2.3)-(2.4) has a limit equation $R'(t) = -(d+\alpha)R(t)$ which has the global convergence dynamics: $R(t) \to 0$ as $t \to \infty$; and the first equation in (2.3)-(2.4) has also a limit equation: $S'(t) = \Lambda - dS(t)$ which also have the global convergence dynamics: $S(t) \to \Lambda/d = S_0$ as $t \to \infty$. Now, by the theory of asymptotically autonomous systems (see, e.g., Castillo-Chavez and Thieme [5]), every solution with non-negative initial functions satisfies converges to $(S_0, 0, 0) = E_0$ as $t \to \infty$ when $\beta P(0)S_0 < (d + r + \epsilon)$.

The basic reproduction number of the model can be obtained from (4.18) by using the next-generation approach. However, here it is more straightforward by tracking the duration of infection, which is $1/(d+r+\epsilon)$, and transmission rate near the disease-free equilibrium which is $\beta P(0)S_0$. Such tracking immediately leads to the following formula for the basic reproduction number R_0:

$$R_0 = \frac{1}{d + r + \epsilon} \cdot [\beta P(0)S_0] = \frac{\beta P(0)S_0}{d + r + \epsilon}.$$

Note that $\beta P(0)S_0 < (d + r + \epsilon)$ is equivalent to $R_0 < 1$.

When $R_0 > 1$ ($\beta P(0)S_0 > (d + r + \epsilon)$), E_0 becomes unstable, and one would expect an endemic equilibrium. Note that because of the normality condition for the weights k_j, $j = 0, 1, \cdots, n$ in (4.3), possible endemic equilibrium $E^* = (S^*, I^*, R^*)$ is determined by

the following equations

$$
\begin{cases}
\Lambda - dS + \left(\dfrac{r\alpha}{d+\alpha} - (d+r+\epsilon) \right) I = 0, \\[2mm]
\beta P(I)S = d + r + \epsilon, \\[2mm]
R = \dfrac{rI}{\alpha + d}.
\end{cases}
\tag{2.11}
$$

The first two equations in (2.11) can be rewritten as

$$
S = \frac{\Lambda}{d} + \frac{1}{d}\left(\frac{r\alpha}{d+\alpha} - (d+r+\epsilon) \right) I =: F(I) \quad \text{and} \quad S = \frac{d+r+\epsilon}{\beta P(I)} =: G(I).
$$

$F(I)$ is decreasing since $\frac{r\alpha}{d+\alpha} - (d+r+\epsilon) < -(d+\epsilon) < 0$; while $G(I)$ is increasing because $P(L)$ is decreasing. Hence, (2.11) has a positive solution (unique) if and only if $G(0) < F(0)$, that is,

$$
\frac{d+r+\epsilon}{\beta P(0)} < \frac{\Lambda}{d},
\tag{2.12}
$$

which is equivalent to $R_0 > 1$. Thus, we have proved that (2.3)-(2.4) has a unique endemic equilibrium $E^* = (S^*, I^*, R^*)$ if and only if $R_0 > 1$.

Summarizing the above analysis, we have actually proved the following Theorem.

Theorem 2.3.1 *The disease-free equilibrium E_0 for (2.3)-(2.4) is globally asymptotically stable if $R_0 < 1$; and if $R_0 > 1$, E_0 becomes unstable and there occurs a unique endemic equilibrium $E^* = (S^*, I^*, R^*)$ determined by (2.11).*

2.4 Stability of the endemic equilibrium for a particular $P(L(t))$

In the preceding section, we have shown that when $R_0 > 1$, the model (2.3)-(2.4) has a unique endemic equilibrium $E^* = (S^*, I^*, R^*)$. This section is devoted to the stability of E^*. Unfortunately, for general $P(L(t))$, it does not seem to be possible to obtain any meaningful results on the stability of E^*. Thus, in this section, we consider the following particular severity measurement. $L(t) = k_0 I(t) + k_1 I(t-\tau)$, which only looks at the weighed total infected cases at the present

time t and a previous time $t - \tau$, and the exponential decay function for $P(L)$: $P(L) = e^{-hL}$. The above particular choices result in the infection force $f(t) = \beta I(t)e^{-h(k_0 I(t)+k_1 I(t-\tau))}$ and accordingly reduce (2.3)-(2.4) to the following more concrete model system:

$$
\begin{cases}
S'(t) = \Lambda - dS(t) - \beta I(t)e^{-h(k_0 I(t)+k_1 I(t-\tau))}S(t) + \alpha R(t), \\
I'(t) = \beta I(t)e^{-h(k_0 I(t)+k_1 I(t-\tau))}S(t) - (d + r + \epsilon)I(t), \\
R'(t) = rI(t) - (d + \alpha)R(t).
\end{cases}
\tag{2.13}
$$

Now the basic reproduction number reduces to $R_0 = \frac{\beta\Lambda}{d(d+r+\epsilon)}$. Note that when $k_0 = 0$ and $\alpha = 0$, (2.13) further reduces to (2.5).

Assume $R_0 > 1$ so that the unique endemic equilibrium $E^* = (S^*, I^*, R^*)$ exists. The linearization of (2.13) at E^* is given by

$$
\begin{cases}
S'(t) = -\left(d + m\dfrac{I^*}{S^*}\right) S(t) - m(1 - hk_0 I^*)\, I(t) + \alpha R(t) + mhk_1 I^* \, I(t - \tau) \\
I'(t) = m\dfrac{I^*}{S^*} S(t) - mhk_0 I^* \, I(t) - mhk_1 I^* \, I(t - \tau) \\
R'(t) = rI(t) - (d + \alpha)R(t).
\end{cases}
\tag{2.14}
$$

where $m = d + r + \epsilon$. From (2.14), we can derive the characteristic equation (CE) as

$$
F_1(\lambda, \tau) := Q_2(\lambda)e^{-\lambda\tau} + Q_3(\lambda) = 0,
\tag{2.15}
$$

where
$$
Q_2(\lambda) = (d + \lambda)(d + \alpha + \lambda)\, mI^* hk_1,
$$
$$
Q_3(\lambda) = \lambda^3 + r_2\lambda^2 + r_1\lambda + r_0,
$$
$$
r_2 = (mhk_0 I^* + \alpha + 2d) + \frac{mI^*}{S^*} > 0,
$$
$$
r_1 = d(d + \alpha) + mhk_0 I^* (\alpha + 2d) + m(\alpha + d + m)\frac{I^*}{S^*} > 0,
$$
$$
r_0 = dmhk_0 I^* (\alpha + d) + m((\alpha + d)m - \alpha r)\frac{I^*}{S^*} > 0.
$$

When $\tau = 0$, the transcendental equation (2.15) reduces to the following polynomial:

$$
F_1(\lambda, 0) = Q_2(\lambda) + Q_3(\lambda) = \lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = 0,
\tag{2.16}
$$

where

$$a_2 = \frac{(hmI^* + \alpha + 2d)\,S^* + mI^*}{S^*} > 0,$$

$$a_1 = \frac{(mh\,(2d + \alpha)\,I^* + (d + \alpha)\,d)\,S^* + mI^*\,(\alpha + d + m)}{S^*} > 0,$$

$$a_0 = \frac{mI^*\,((d + \alpha)\,m + hd\,(d + \alpha)\,S^* - \alpha r)}{S^*} > 0.$$

Straightforward calculation also shows that

$$a_2 a_1 - a_0 = \frac{1}{S^{*2}}[(hmI^* + \alpha + d)\,(2d + \alpha)\,(hmI^* + d)\,S^{*2}$$

$$+ \Big((2\alpha + 3d + m)\,mhI^* + dm + 3d^2 + 4\alpha d + \alpha\,(\alpha + r)\Big)\,mI^*S^* + m^2 I^{*2}\,(\alpha + d + m)] > 0.$$

By the Routh−Hurwitz criterion, all roots of the above cubic equation $F_1(\lambda, 0) = 0$ have negative real parts, implying that $E^* = (S^*, I^*, R^*)$ is locally asymptotically stable when $\tau = 0$.

Next, we explore the possibility for (2.15) to have a root with a positive real part when τ is increased. This can occur only when a root of (2.15) crosses the purely imaginary axis in the complex plane from the left half to the right half when τ increases from zero. Noting that

$$F_1(0, \tau) = d(\alpha + d)mI^*hk_1 + r_0 > 0$$

for all $\tau \geq 0$, the aforementioned crossing cannot happen at the origin of the complex plane. Thus, a crossing can only possibly occur in pairs of purely imaginary roots when τ is increased to some critical values.

To investigate this possibility, we plug $\lambda = i\omega$ (without loss of generality, we assume $\omega > 0$) into (2.15). Separating the real part and imaginary part of in the resulting equation $F_1(i\omega, \tau) = 0$ leads to

$$\begin{aligned}
\Big[A_1(\omega^2)\cos(\omega\tau) + \omega A_2 \sin(\omega\tau)\Big]mI^*hk_1 &= B_1(\omega^2), \\
\Big[\omega A_2 \cos(\omega\tau) - A_1(\omega^2)\sin(\omega\tau)\Big]mI^*hk_1 &= \omega B_2(\omega^2),
\end{aligned} \tag{2.17}$$

where

$$A_1(x) = d\alpha + d^2 - x, \quad A_2 = \alpha + 2d,$$

$$B_1(x) = r_2 x - r_0, \quad B_2(x) = x - r_1.$$

By eliminating the trigonometric functions, we obtain (4.37) the following equation:

$$B_1^2(\omega^2) + \omega^2 B_2^2(\omega^2) - (mI^* h k_1)^2 \left[A_1^2(\omega^2) + \omega^2 A_2^2 \right] = 0. \tag{2.18}$$

Denoting $x = \omega^2$, we obtain

$$
\begin{aligned}
F_2(x) &= B_1^2(x) + x B_2^2(x) - (mI^* h k_1)^2 \left[A_1^2(x) + x A_2^2 \right] \\
&= x^3 + q_2(k_1) x^2 + q_1(k_1) x + q_0(k_1) = 0,
\end{aligned}
\tag{2.19}
$$

where

$$
\begin{aligned}
q_2(k_1) &= -(mI^* h k_1)^2 + r_2^2 - 2r_1, \\
q_1(k_1) &= -(mI^* h k_1)^2(\alpha^2 + 2d\alpha + 2d^2) - 2r_0 r_2 + r_1{}^2, \\
q_0(k_1) &= r_0{}^2 - (mdI^* h k_1)^2(\alpha + d)^2.
\end{aligned}
\tag{2.20}
$$

Thus, if $F_2(x) = 0$ has no positive root, then E^* remains asymptotically stable for all $\tau > 0$. If $F_2(x) = 0$ has a positive root $x > 0$, then from (4.37), $\omega = \sqrt{x}$ would satisfy

$$
\begin{cases}
\cos(\omega\tau) = \dfrac{A_2 B_2(\omega^2)\omega^2 + A_1(\omega^2)B_1(\omega^2)}{mI^* h k_1 \left(A_2^2 \omega^2 + A_1^2(\omega^2) \right)} =: G, \\[4mm]
\sin(\omega\tau) = -\dfrac{\omega\left(A_1(\omega^2)B_2(\omega^2) - A_2 B_1(\omega^2) \right)}{mI^* h k_1 \left(A_2^2 \omega^2 + A_1^2(\omega^2) \right)} =: N,
\end{cases}
\tag{2.21}
$$

which yields a sequence $\{\tau^n\}$, $n = 0, 1, 2, 3...$, of critical values for the delay parameter τ, given by

$$\tau^n = \tau^0 + \frac{2n\pi}{\omega}, \qquad \tau^0 = \begin{cases} \frac{\arccos G}{\omega}, & N \geq 0; \\[2mm] \frac{2\pi - \arccos G}{\omega}, & N < 0. \end{cases} \tag{2.22}$$

To verify the transversality at these critical values, we first establish the following lemma and the detailed proof is shown in the Appendix.

Lemma 2.4.1 $\mathrm{sign}\left(\mathrm{Re}\left(\frac{d\lambda}{d\tau} \right)\Big|_{\tau=\tau^n} \right) = \mathrm{sign}\left(\frac{dF_2(x)}{dx}\Big|_{x=\omega^2} \right)$, *where τ^n is given by* (4.40).

Proof Take derivative of (2.15) about τ:

$$\frac{dF_1(\lambda, \tau)}{d\tau} = \frac{\partial F_1(\lambda, \tau)}{\partial \tau} + \frac{\partial F_1(\lambda, \tau)}{\partial \lambda} \frac{d\lambda}{d\tau} = 0, \tag{2.23}$$

then

$$\left(\frac{d\lambda}{d\tau}\right)^{-1} = -\frac{\frac{\partial F_1(\lambda,\tau)}{\partial \lambda}}{\frac{\partial F_1(\lambda,\tau)}{\partial \tau}} = \frac{(Q_2'(\lambda) - \tau Q_2(\lambda))e^{-\lambda\tau} + Q_3'(\lambda)}{\lambda Q_2(\lambda)e^{-\lambda\tau}}. \tag{2.24}$$

This together with $F_1(i\omega, \tau^n) = 0$ further leads to

$$\left.\frac{d\lambda}{d\tau}\right|_{\tau=\tau^n}^{-1} = \frac{Q_2'(i\omega)}{i\omega Q_2(i\omega)} - \frac{\tau}{i\omega} - \frac{Q_3'(i\omega)}{i\omega Q_3(i\omega)} \tag{2.25}$$

and

$$
\begin{aligned}
\mathrm{Re}\left(\left.\frac{d\lambda}{d\tau}\right|_{\tau=\tau^n}^{-1}\right) &= \mathrm{Re}\left(\frac{Q_2'(i\omega)}{i\omega Q_2(i\omega)}\right) - \mathrm{Re}\left(\frac{Q_3'(i\omega)}{i\omega Q_3(i\omega)}\right) \\
&= \frac{-\alpha^2 - 2\alpha d - 2d^2 - 2\omega^2}{(d^2 + \omega^2)(\alpha^2 + 2\alpha d + d^2 + \omega^2)} - \frac{-3\omega^4 + \left(-2r_2^2 + 4r_1\right)\omega^2 + 2r_0 r_2 - r_1^2}{(r_2\omega^2 - r_0)^2 + \omega^2(\omega^2 - r_1)^2} \\
&= \left.\frac{F_2'(x)}{B_1^2(x) + xB_2^2(x)}\right|_{x=\omega^2}.
\end{aligned} \tag{2.26}
$$

Therefore,

$$\mathrm{sign}\left(\left.\frac{d\,\mathrm{Re}\,(\lambda)}{d\tau}\right|_{\tau=\tau^n}\right) = \mathrm{sign}\left(\mathrm{Re}\left(\left.\frac{d\lambda}{d\tau}\right|_{\tau=\tau^n}\right)\right) = \mathrm{sign}\left(\left.\frac{dF_2(x)}{dx}\right|_{x=\omega^2}\right).$$

We have seen that when $\tau = 0$, E^* is asymptotically stable. Thus, the first critical value τ^0 is of the most significance, as it is the smallest possible value for τ at which E^* may lose its stability when τ passes it. In other words, we need to be mainly concerned with the case when there is a pair of eigenvalues crossing from the left complex plane to the right complex plane when τ is increased to pass τ^0. Lemma 2.4.1 shows that only when the positive root x of $F_2(x)$ satisfied $F_2'(x) > 0$, can such a crossing occur due to the increase of $\tau > 0$. Noticing that $F_2(x)$ is a cubic polynomial, it can have at most two positive roots at which $F_2'(x) > 0$. The crossing direction at $\pm i\sqrt{x}$ when x is the *largest* positive root of $F_2(x) = 0$ (as τ passes a critical value given in (4.40)) is always from left to right; at $\pm i\sqrt{x}$ when x is the *second largest* positive root of $F_2(x) = 0$ (if any) is from right to left. Thus, we have the following cases.

(C1) If $F_2(x) = 0$ has no positive root, then all roots of (2.15) have negative real parts.

(C2) If there is only one positive root x_1, then the crossing at $\pm \sqrt{x_1}$ rightward when τ increases

to pass τ^0; in this case, all roots of (2.15) have negative real parts for $\tau < \tau^0$, and (2.15) has a pair of root with positive real parts when $\tau > \tau^0$. Here τ^0 is defined by (4.40) with $\omega = \sqrt{x_1}$).

(C3) If there are two positive roots $x_1 > x_- > 0$ at which $F_2'(x_1) > 0$ and $F_2'(x_-) < 0$ respectively. The crossing direction at $\pm i\sqrt{x_1}$ when τ passes $\tau = \tau_1^{n,+}$ is rightward, but at $\pm i\sqrt{x_-}$ the crossing direction as τ passes $\tau = \tau^{n,-}$, $n = 0, 1, 2...$ is leftward. Unlike case (C2), in this case, it is possible for the branch of the root $\lambda(\tau)$ that enters the right half of the complex plane through $\pm i\sqrt{x_1}$ as τ passes $\tau_1^{0,+}$, to return to the left half of the complex plane when τ further increases to $\tau^{0,-}$ (defined by (4.40) with $\omega_- = \sqrt{x_-}$).

(C4) If there are three positive roots: $x_1 > x_- > x_2$ at which $F_2'(x_i) > 0$, $i = 1, 2$ and $F_2'(x_-) < 0$ respectively, then the crossings at $\pm i\sqrt{x_1}$ and $\pm i\sqrt{x_2}$ when τ increases to pass $\tau = \tau_1^{n,+}$ and $\tau = \tau_2^{n,+}$ (defined by (4.40) with $\omega_1 = \sqrt{x_1}$ and $\omega_2 = \sqrt{x_2}$ respectively) are both rightward, but the crossing at $\pm i\sqrt{x_-}$ is leftward when τ increases to pass $\tau = \tau^{n,-}$ (defined by (4.40) with $\omega_- = \sqrt{x_-}$).

Combining the above summarization with the Hopf Bifurcation Theorem for delay differential equations, we accordingly proved the following theorem.

Theorem 2.4.2 *Assuming $R_0 > 1$ so that E^* exists. Then, there can be four cases.*

(i) *If $F_2(x) = 0$ has no positive root, then E^* is locally asymptotically stable for all $\tau \geq 0$.*

(ii) *If $F_2(x) = 0$ has only one positive root x_1 at which $F_2'(x_1) > 0$, then E^* is locally asymptotically stable for $\tau \in [0, \tau^0)$ and unstable for $\tau > \tau^0$, where τ^0 is given in (4.40) with $\omega = \sqrt{x_1}$. Moreover, system (2.13) undergoes Hopf bifurcation around E^* at $\tau = \tau^0$.*

(iii) *If $F_2(x) = 0$ has two positive roots x_1 and x_- at which $F_2'(x_1) > 0$ and $F_2'(x_-) < 0$ respectively, then E^* is locally asymptotically stable for $\tau \in [0, \tau_1^{0,+})$ and becomes unstable when τ passes through $\tau = \tau_1^{0,+}$, where $\tau_1^{0,+}$ is given in (4.40) with $\omega = \sqrt{x_1}$. Moreover, system (2.13) undergoes Hopf bifurcation around E^* at $\tau = \tau_1^{0,+}$.*

(iv) If $F_2(x) = 0$ has three positive roots: $x_1 > x_- > x_2$ at which $F_2'(x_i) > 0$, $i = 1, 2$ and $F_2'(x_-) < 0$ respectively, then E^* is asymptotically stable for $\tau \in [0, \tau^0)$ and becomes unstable when τ passes through $\tau = \tau^0$, where $\tau^0 = \min\{\tau_1^{0,+}, \tau_2^{0,+}\}$ with

$$\tau_i^{0,+} = \begin{cases} \frac{\arccos G_i}{\omega_i}, & N_i \geq 0, \\ \frac{2\pi - \arccos G_i}{\omega_i}, & N_i < 0; \end{cases}$$

and

$$\begin{cases} G_i = \dfrac{A_2 B_2(x_i)x_i + A_1(x_i)B_1(x_i)}{mI^* h k_1 \left(A_2^2 x_i + A_1^2(x_i)\right)} \\ N_i = -\dfrac{\sqrt{x_i}\,[A_1(x_i)B_2(x_i) - A_2 B_1(x_i)]}{mI^* h k_1 \left(A_2^2 x_i + A_1^2(x_i)\right)}, \end{cases} \quad i = 1, 2. \qquad (2.27)$$

Moreover, system (2.13) undergoes Hopf bifurcation at $\tau = \tau_1^{0,+}$ and $\tau = \tau_2^{0,+}$.

We point out that for cases (ii)-(iv) in the above theorem, E^* loses its stability when τ increases to τ^0 or $\tau_1^{0,+}$; however addition, for case (iii) and iv), E^* may regain its stability at some $\tau > \tau^{0,-}$. We will numerically demonstrate this later.

Theorem 2.4.2 is a general conclusion, stated in terms of the delay τ. To better understand the role of allocation of the weights k_0 and k_1 in detail, we can do further specific analysis on positive roots of the cubic equation $F_2(x) = 0$ below, in terms of k_0 and k_1. To this end, we denote

$$\Delta(k_1) = 12q_0(k_1)[q_2(k_1)]^3 - 3[q_1(k_1)]^2[q_2(k_1)]^2 - 54q_0(k_1)q_1(k_1)q_2(k_1)$$
$$+ 12[q_1(k_1)]^3 + 81[q_0(k_1)]^2, \qquad (2.28)$$

where $q_j(k_1)$, $j = 0, 1, 2$, are give in (2.20). It can be readily seen that $q_0(k_1)$, $q_1(k_1)$ and $q_2(k_1)$ are decreasing in k_1 and each of them has a unique positive root $\tilde{k}_{1i}$, $i = 0, 1, 2$. From (2.20), plugging the formulas for r_j, $j = 0, 1, 2$, and after some tedious calculations (omitted here), these roots are given by

$$\tilde{k}_{10} = \frac{1}{2} + \frac{md + (d + \epsilon)\alpha}{2S^* dh(d + \alpha)}, \quad \tilde{k}_{11} = \frac{q_1(0)}{q_1(0) - q_1(1)}, \quad \tilde{k}_{12} = \frac{q_2(0)}{q_2(0) - q_2(1)}$$

with $q_1(0)$, $q_1(1)$, $q_2(0)$ and $q_2(1)$ explicitly given by

$$q_1(0) = -\frac{2mI^*\left((hdS^* + m)(\alpha + d) - \alpha r\right)\left((hmI^* + \alpha + 2d)S^* + I^*m\right)}{S^{*2}}$$
$$+ \frac{\left((hmI^*(\alpha + 2d) + d(\alpha + d))S^* + (\alpha + d + m)I^*m\right)^2}{S^{*2}},$$

$$q_1(1) = \frac{\left(d(\alpha + d)S^* + (\alpha + d + m)mI^*\right)^2}{S^{*2}} - (mhI^*)^2(\alpha^2 + 2d\alpha + 2d^2)$$
$$- \frac{2mI^*(m(\alpha + d) - \alpha r)((\alpha + 2d)S^* + I^*m)}{S^{*2}},$$

$$q_2(0) = \frac{(hmI^*S^*)^2 + 2(hI^* - 1)m^2S^*I^* + \Lambda^2}{S^{*2}} + (\alpha + d)^2,$$

$$q_2(1) = \frac{-(hmI^*S^*)^2 - 2m^2S^*I^* + \Lambda^2}{S^{*2}} + (\alpha + d)^2.$$

It can be readily seen that $q_0(k_1)$, $q_1(k_1)$ and $q_2(k_1)$ are decreasing in k_1. In addition, $q_i(\tilde{k}_{1i}) = 0$ ($i = 0, 1, 2$). Here tedious calculations for the above formulas are omitted.

By the definition of $\tilde{k}_{1i}$, $i = 0, 1, 2$, we immediately have the following.

Lemma 2.4.3 $q_i(k_1) > 0$ iff $k_1 < \tilde{k}_{1i}$, where $i = 0, 1, 2$.

According to [30](Lemma A.2 and A.3) or [11, 27], as well as Lemma 2.4.3 above, we obtain the following results.

Corollary 2.4.4 *Assume that $R_0 > 1$.*

(I) Case (ii) in Theorem 2.4.2 occurs if and only if one of the following conditions is satisfied

   1) $\Delta(k_1) \geq 0$ and $\tilde{k}_{10} < k_1$;

   2) $\Delta(k_1) < 0$, $\tilde{k}_{10} < k_1 < \min\{\tilde{k}_{11}, \tilde{k}_{12}\}$;

   3) $\Delta(k_1) < 0$, $\max\{\tilde{k}_{11}, \tilde{k}_{10}\} < k_1$;

   4) $\tilde{k}_{10} < k_1 = \tilde{k}_{11}$ or $\tilde{k}_{11} < k_1 = \tilde{k}_{10}$;

(II) Case (iii) in Theorem 2.4.2 occurs if and only if one of the following conditions is satisfied

   1) $\Delta(k_1) < 0$, $\tilde{k}_{12} < k_1 < \min\{\tilde{k}_{10}, \tilde{k}_{11}\}$;

   2) $\Delta(k_1) < 0$, $\tilde{k}_{11} < k_1 < \tilde{k}_{10}$;

(III) Case (iv) occurs if and only if there holds (v): $\Delta(k_1) < 0$, $\max\{\tilde{k}_{10}, \tilde{k}_{12}\} < k_1 < \tilde{k}_{11}$.

(IV) Case (i) occurs if and only if none of (I)-(III) is satisfied. Moreover, Case (i) holds if $k_1 \le \min\{\tilde{k}_{10}, \tilde{k}_{11}, \tilde{k}_{12}\}$ (i.e. $q_i(k_1) \ge 0$, $i = 0, 1, 2$).

Particularly, if $k_1 > \max\{\tilde{k}_{10}, \tilde{k}_{11}\}$ holds, then one of the cases (I)-1) and (I)-3) must occur. Based on this observation, we obtain

Corollary 2.4.5 *If $k_1 > \max\{\tilde{k}_{10}, \tilde{k}_{11}\}$ (i.e. $q_i(k_1) < 0$, $i = 0, 1$), then Case (ii) in Theorem 2.4.2 occurs.*

From Corollaries 2.4.4 and 2.4.5, a necessary condition for Case (ii) of Theorem 2.4.2 is $\tilde{k}_{10} \le k_1$, and a necessary condition for Case (iii) of Theorem 2.4.2 is $\tilde{k}_{10} > k_1$. Interestingly, according to (IV) in Corollary 2.4.4, there will be no bifurcation at all provided that the weight k_1 is small ($k_1 \le \min\{\tilde{k}_{10}, \tilde{k}_{11}, \tilde{k}_{12}\}$). In other words, restricting the weight to the previous surveillance $I(t - \tau)$ to be sufficiently small can avoid Hopf bifurcation (sustained oscillation). From Corollary 2.4.5, on the other hand, a relatively large weight k_1 will allow τ to destroy the stability of E^* at some critical value. In Section 6, we will numerically illustrate the impact of $k_1 = 1 - k_0$ in terms of these two corollaries.

2.5 Special case $\alpha = 0$, $\epsilon = 0$: a comparison to related work

In order to compare our results with those in paper [29], let us also consider the infection force $f(t) = \beta I(t)e^{-h(k_0 I(t) + k_1 I(t-\tau))}$ and let $\alpha = 0$, $\epsilon = 0$ in the model (2.13), reducing the model (2.13) to the following model system

$$\begin{cases} \dot{S} = \Lambda - \beta I e^{-h(k_0 I(t) + k_1 I(t-\tau))} S - dS, \\ \dot{I} = \beta I e^{-h(k_0 I(t) + k_1 I(t-\tau))} S - (d + r)I, \\ \dot{R} = rI - dR. \end{cases} \tag{2.29}$$

The case $k_0 = 0$ (hence $k_1 = 1$) leads to the model (1) in [29] (i.e., Eq.(2.5) in the introduction of this paper), which has been thoroughly analyzed. Note that $k_0 = 0$ means that only infected cases in the previous time $t - \tau$ are used to measure the severity L. Here in this section, we

allow $k_0 \in (0, 1]$ (hence $k_1 \in [0, 1)$ and hence, the severity L is measured by the weighted total number of infected cases at the present time t and the past time $t - \tau$. This seems to be a more realistic and reasonable choice as it allows public health agents to put different weights on the disease information at different times; in the meantime, this also constitutes a good demonstration of the flexibility of our model framework in accommodating a wide range of scenarios. By analyzing (2.29) in more detail, we can reveal the impact of the weights k_0 and k_1, together with the time delay τ (see a discussion in Section 2.7).

The equilibria structure and the basic reproduction number R_0 for (2.29) remain the same as in [29] for (2.5), due to the complementary constraint $k_0 + k_1 = 1$. The first two equations of (2.29) form a decoupled sub-system:

$$\begin{cases} \dot{S} = \Lambda - \beta I e^{-h(k_0 I(t) + k_1 I(t-\tau))} S - dS \\ \dot{I} = \beta I e^{-h(k_0 I(t) + k_1 I(t-\tau))} S - (d + r) I. \end{cases} \tag{2.30}$$

Let $m = d + r$. The CE at E^* for (2.30) becomes

$$F_1(\lambda, \tau) = Q_2(\lambda) e^{-\lambda \tau} + Q_3(\lambda), \tag{2.31}$$

in which

$$Q_2(\lambda) = m I^* h k_1 (\lambda + d), \quad Q_3(\lambda) = \lambda^2 + r_1 \lambda + r_0,$$

with

$$r_1 = k_0 h m I^* + \frac{\Lambda}{S^*}, \quad r_0 = m I^* d h k_0 + \frac{m^2 I^*}{S^*}.$$

To be consistent with Section 2.4, we use the notations here:

$$A_1 = d, \quad A_2 = 1, \quad B_1(x) = x - r_0, \quad B_2 = -r_1$$

and

$$G = \frac{d\omega^2 - r_1 \omega^2 - dr_0}{(d^2 + \omega^2) h k_1 m I^*}, \quad N = -\frac{\omega(-dr_1 - \omega^2 + r_0)}{h k_1 m I^* (d^2 + \omega^2)}.$$

Lemma 2.4.1 also holds for this model, but now function $F_2(x)$ becomes a quadratic function

(instead of a cubic function):

$$F_2(x) = x^2 + q_2(k_1)x + q_0(k_1) \tag{2.32}$$

where

$$q_2(k_1) = -(mI^*k_1h)^2 + r_1^2 - 2r_0, \quad q_0(k_1) = r_0^2 - (k_1hmI^*d)^2. \tag{2.33}$$

Let

$$\Delta(k_1) = [q_2(k_1)]^2 - 4q_0(k_1). \tag{2.34}$$

When $q_0(k_1) < 0$, $F_2(x)$ has a unique positive root: $x_1 = -\frac{q_2(k_1)}{2} + \frac{\sqrt{\Delta(k_1)}}{2}$ at which $F_2'(x_1) > 0$. Thus, E^* loses its stability for *all* $\tau > \tau^0$ to a periodic solution through Hopf bifurcation.

When $q_0(k_1) > 0$, $q_2(k_1) < 0$ and $\Delta(k_1) > 0$, there are two positive roots $x_1 = -\frac{q_2(k_1)}{2} + \frac{\sqrt{\Delta(k_1)}}{2}$ and $x_- = -\frac{q_2(k_1)}{2} - \frac{\sqrt{\Delta(k_1)}}{2}$ such that $F_2'(x_1) > 0$ and $F_2'(x_-) < 0$. Then (2.31) has two pairs of simple purely imaginary roots $\pm i\omega_+$, $\pm i\omega_-$ at $\tau^{n,+}$, $\tau^{n,-}$, $n = 0, 1, 2...$, where $\omega_+ = \sqrt{x_1}$, $\omega_- = \sqrt{x_-}$ and $\tau^{n,+}$, $\tau^{n,-}$ are defined by the same form in (4.40) with ω given by $\omega_+ = \sqrt{x_1}$, $\omega_- = \sqrt{x_-}$ respectively. According to Lemma 2.4.1, we have

$$\left. \text{Re}\left(\frac{d\lambda}{d\tau}\right)^{-1} \right|_{\tau=\tau^{n,+}} > 0 > \left. \text{Re}\left(\frac{d\lambda}{d\tau}\right)^{-1} \right|_{\tau=\tau^{n,-}}. \tag{2.35}$$

Hence, at each $\tau = \tau^{n,+}$, a pair of roots cross the imaginary axis at $\pm i\omega_+$ into the right half-plane, and at each $\tau = \tau^{n,-}$, a pair of roots crosses at $\pm i\omega_-$ back into the left half-plane. These observations together with the fact that all roots of (2.31) are on the left half of the complex plane when $\tau = 0$, imply that $\tau^{0,+} < \tau^{0,-}$. However, the orders of other numbers in the two sequences $\tau = \tau^{n,+}$ and $\tau = \tau^{n,-}$ do not seem to have a pattern and seem to be complicated.

For example, for the first couple of critical values in the two sequences, the following two cases are possible:

(A) $\tau^{0,+} < \tau^{0,-} < \tau^{1,+}$. In this case, all roots of (2.31) have negative real parts when $\tau \in [0, \tau^{0,+})$; one pair of roots of (2.31) have positive real parts when $\tau \in (\tau^{0,+}, \tau^{0,-})$ and that pair return to the left complex plane again when $\tau \in (\tau^{0,-}, \tau^{1,+})$.

(B) $\tau^{0,+} < \tau^{1,+} < \tau^{0,-}$. In this case, all roots of (2.31) have negative real parts when $\tau \in$

$[0, \tau^{0,+})$; one pair of roots of (2.31) have positive real parts when $\tau \in (\tau^{0,+}, \tau^{1,+})$ and there will be two pairs of roots on the right complex plane when $\tau \in (\tau^{1,+}, \tau^{0,-})$.

Summarizing the above, we have the following.

Theorem 2.5.1 *Assuming $R_0 > 1$, then*

(i) *If $q_0(k_1) < 0$, or $q_0(k_1) = 0$ but $q_2(k_1) < 0$, the endemic equilibrium E^* of system (2.5.1) is locally asymptotically stable for $\tau \in [0, \tau^0)$ and unstable for all $\tau > \tau^0$; moreover, system (2.5.1) undergoes Hopf bifurcation when τ passes $\tau = \tau^n, n = 0, 1, 2 \ldots$, where τ^n, $n = 0, 1, 2, \cdots$ are defined by (4.40) with $\omega = \sqrt{x_1}$.*

(ii) *If $q_0(k_1) > 0$, $q_2(k_1) < 0$ and $\Delta(k_1) > 0$, then there are two sequences $\{\tau^{n,+}\}$ and $\{\tau^{n,-}\}$, still defined by ((4.40)) but with ω being $\sqrt{x_1}$ and $\sqrt{x_-}$ respectively, such that E^* is locally asymptotically stable when $\tau \in [0, \tau^{0,+})$, and it loses its stability when τ increases to pass $\tau^{0,+}$ through Hopf bifurcation. It is possible for E^* to regain its stability for some $\tau > \tau^{0,-}$. Moreover, system (2.5.1) undergoes Hopf bifurcation at each $\bar{\tau} \in \{\tau^{n,+}\} \cup \{\tau^{n,-}\}$.*

(iii) *For other cases of $q_0(k_1)$ and $q_2(k_1)$, E^* is asymptotically stable for all $\tau \geq 0$.*

To obtain more details on how the weight parameter k_1 affects the conditions in Theorem 2.5.1, we need to explore the signs of $q_0(k_1)$ and $q_2(k_1)$ given by (2.33). Both $q_0(k_1)$ and $q_2(k_1)$ are decreasing in k_1 and each has a unique root $\tilde{k}_{1i} > 0$ which can be expressed in terms of the model parameters by

$$\tilde{k}_{10} = \frac{1}{2} + \frac{m}{2S^*dh}, \quad \tilde{k}_{12} = \frac{q_2(0)}{q_2(0) - q_2(1)},$$

where

$$q_2(0) = \frac{(hmI^*S^*)^2 + 2(hI^* - 1)m^2S^*I^* + \Lambda^2}{S^{*2}},$$

$$q_2(1) = \frac{-(hmI^*S^*)^2 - 2m^2S^*I^* + \Lambda^2}{S^{*2}},$$

$$q_2(0) - q_2(1) = \frac{2h(-dS^* + \Lambda)^2(hS^* + 1)}{S^*} > 0,$$

$$q_0(0) - q_0(1) = \frac{2dh(-dS^* + \Lambda)^2(dhS^* + m)}{S^*} > 0.$$

Thus, a version of Lemma 2.4.3 also holds here: $q_i(k_1) > 0$ iff $k_1 < \tilde{k}_{1i}$, $i = 0, 2$. Moreover, if $\tilde{k}_{12} \leq \tilde{k}_{10}$, then there exists

$$k_{1c} = \tilde{k}_{12} + \frac{2\sqrt{(q_2(1) - q_2(0))^2(q_0(1) - q_0(0))(\tilde{k}_{12} - \tilde{k}_{10}) + (q_0(1) - q_0(0))^2} + 2(q_0(1) - q_0(0))}{(q_2(1) - q_2(0))^2} \geq \tilde{k}_{12},$$

$$k_{2c} = \tilde{k}_{12} - \frac{2\sqrt{(q_2(1) - q_2(0))^2(q_0(1) - q_0(0))(\tilde{k}_{12} - \tilde{k}_{10}) + (q_0(1) - q_0(0))^2} + 2(q_0(1) - q_0(0))}{(q_2(1) - q_2(0))^2} < \tilde{k}_{12},$$

such that $\Delta(k_{ic}) = 0$ ($i = 1, 2$), and $\Delta(k_1) > 0$ when $k_1 > k_{1c}$ or $k_1 < k_{2c}$.

By symmetry, using the same arguments for Theorem 2.4.2 and Corollary 2.4.4, we have the following more specific results for Theorem 2.29, in terms of k_1:

Since Lemma 2.4.3 also holds, then we get more specific results as follows:

Corollary 2.5.2 *Assuming $R_0 > 1$, then*

(i) *If $k_1 > \tilde{k}_{10}$ or $\tilde{k}_{12} < k_1 = \tilde{k}_{10}$, the conclusion of Theorem 2.5.1-(i) holds.*

(ii) *If $\tilde{k}_{12} < k_1 < \tilde{k}_{10}$ and $\Delta(k_1) > 0$ (i.e., $\tilde{k}_{12} < k_{1c} < k_1 < \tilde{k}_{10}$), then the conclusion of Theorem 2.5.1-(ii) holds.*

(iii) *For other range of k_1, particularly when $k_1 \leq \min\{\tilde{k}_{12}, \tilde{k}_{10}\}$, the conclusion of Theorem 2.5.1-(iii) holds.*

As the above results show, when k_1 is small to a certain extent (i.e. $k_1 \leq \min\{\tilde{k}_{12}, \tilde{k}_{10}\}$), the stability of E^* can be ensured whatever $\tau > 0$. When k_1 is large to a certain extent (i.e. $k_1 > \tilde{k}_{10}$), the system (2.29) will undergo Hopf bifurcations at a sequence $\{\tau^n\}$ as τ increases. We point out that when $k_1 = 1$ ($k_0 = 0$), the condition $1 = k_1 > \tilde{k}_{10}$ in (i) of Corollary 2.5.2 reduces to $\frac{m}{2S^*dh} < 1$ which is equivalent to $\frac{dh}{\beta} \exp(\frac{\Lambda h}{m} - 1) < 1$ by the properties of the Lambert W function [6], and thus, is consistent with the conclusion of Case1: (H1) in [29]. Corollary 2.5.2-(ii) corresponds to the conclusion of Case2: (H2) in [29]. Because of the existence of two sequences of critical values $\{\tau^{n,+}\}$ and $\{\tau^{n,-}\}$ which may allow different order patterns (see, e.g., (A) and (B)), the above analysis indicates that the stability switches as τ increase for this case can be more complicated than stated in Proposition 4 in [29]; for example, there may be

multiple switches between the stability and instability of E^* as τ increases, and this will be demonstrated numerically in the next section.

2.6 Numeric exploration of multiple switches of stability

From the results in Sections 4 and 5, we have seen that within certain ranges of the parameter k_1, there may be one, two or even three sequences of critical values for the delay parameter τ at which Hopf bifurcations occur. Among these critical values (if any), the smallest $\tau^{0,+}$ is most important because the endemic equilibrium E^* is deemed to lose its stability when τ increases to it. However, the effect of the remaining critical values (if any) depends on the order they appear. Particularly when the F_2 function in Section 4 or 5 has more than one positive root, it is generally very difficult (if not impossible) to determine whether there is a pattern for the order of those critical values in the two or three sequences of critical values for τ related to the two or three positive roots of F_2. Below, we provide some numeric examples to show that such orders can be different depending on the values of the model parameter. Our numerical diagrams are constructed mainly by using the DDE Biftool package [8, 9, 10].

We begin with (2.29). For convenience of comparison with the results in [29], we choose the same values of the parameters for (2.29) as used in [29] as below:

$$\Lambda = 0.2, \beta = 1, h = 3, d = 0.2, r = 0.1. \tag{2.36}$$

By numerical computations, we obtain $E^* \approx (0.62946, 0.24703, 0.12351)$ and $R_0 = 3.3$, together with the three threshold values of k_1: $\tilde{k}_{12} \approx 0.8735925, \ k_{1c} \approx 0.89673437$ and $k_{2c} < 0$. Then $k_1 > k_{1c}$ (resp. $k_1 \in [0, k_{1c})$) implies $\Delta(k_1) > 0$ (resp. $\Delta(k_1) < 0$).

The related conclusions corresponding to Corollary 2.5.2 are illustrated in the $k_1 - \tau$ plane in Fig. 2.1 and are restated below:

a) When $k_1 < k_{1c}$, no Hopf bifurcation occurs when τ increases, and E^* remains stable for all $\tau \geq 0$ (Fig. 2.1-(a));

b) For $k_{1c} < k_1 < \tilde{k}_{10}$ (see the zoom-in figure in Fig. 2.1-(b)), there are *two sequences* of critical values beginning with $\tau^{0,+}$ and $\tau^{0,-}$ respectively. The stability of E^* switches as τ

increases from zero: E^* first loses stability at $\tau^{0,+}$ and regains the stability at $\tau^{0,-}$;

c) For $\tilde{k}_{10} < k_1 \leq 1$, there is *only one sequence* of critical values beginning with $\tau^{0,+}$; E^*
   loses stability at $\tau = \tau^{0,+}$ and remains unstable for all $\tau > \tau^{0,+}$ (see the zoom-in figure Fig.
   2.1(b));

d) When $\tau < 11.7575$, E^* remains stable regardless of $k_1 \in [0, 1]$ (see Fig. 2.1(a)).

Especially, when $k_1 = 0.8968 \in (k_{1c}, \tilde{k}_{10})$, we can calculate the first few critical values to
obtain the following *alternating order*:

$$\tau^{0,+} = 59.83 < \tau^{0,-} = 92.93 < \tau^{1,+} = 186.782 < \tau^{1,-} = 285.958 < \tau^{2,+} = 313.73 \qquad (2.37)$$

with $\omega_+ = 0.0597$, $\omega_- = 0.0326$. Accordingly, when τ increases from $\tau = 0$ to pass $\tau^{2,+}$, the
stability/instability E^* switches at least five times: losing stability at $\tau^{i,+}$, $i = 0, 1, 2$, and regain
stability at $\tau^{i,-}$, $i = 0, 1$. See Fig. 2.2.

When $k_1 = 1$, bifurcation also occurs at τ^n with $\tau^0 \approx 11.7575$ and no re-stabilization of E^*
occurs when $\tau > \tau^0$, which is in line with the conclusion and Fig.1 in [29].

If we fix τ at a given value, we can also use our results to numerically demonstrate the
impact of k_1. This corresponds to a scenario in which current data and the data of a fixed time
ago are available, and we want to explore how the weight allocation weight (k_0, k_1) will affect
the disease dynamics.

Next, we look at the more general (2.13). Fix the other parameters in (2.13) as below

$$\Lambda = 0.2, \beta = 1, h = 3, d = 0.2, r = 0.05, \epsilon = 0.1, \alpha = 0.06. \qquad (2.38)$$

Then by numeric computation we obtain $E^* \approx (0.65043, 0.20657, 0.03972)$, $R_0 = 2.86$, $\tilde{k}_{10} \approx$
0.93364, $\tilde{k}_{11} \approx 0.88864$, $\tilde{k}_{12} \approx 1.27804$ and $\Delta(k_c) = 0$ with $k_c = 0.929195195$. Note that
$k_1 > k_c$ (resp. $k_1 \in [0, k_c)$) implies $\Delta(k_1) < 0$ (resp. $\Delta(k_1) > 0$). By Corollary 2.4.4, we have
the following conclusions for (2.13) which are similar to a)-d) above for (2.29):

(a When $0 < k_1 < k_c$, no Hopf bifurcation occurs when τ increases, and E^* remains stable for
   all $\tau \geq 0$;
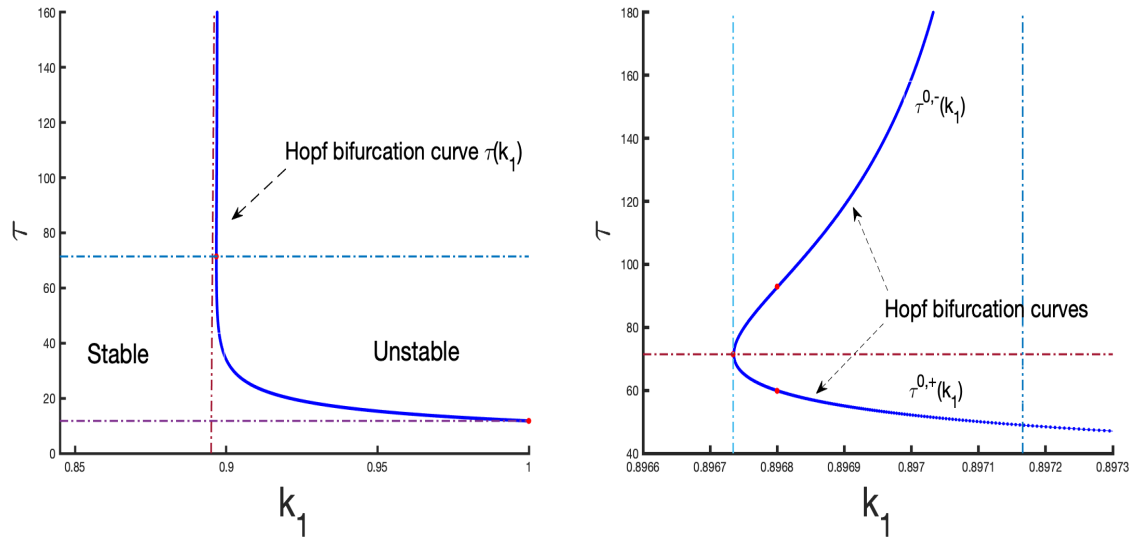
Figure 2.1: The Hopf bifurcation curves $\tau^0(k_1)$ in the $k_1 - \tau$ parameter space, with (b) being a zoom-in of (a) near $k_1 = 0.90$.
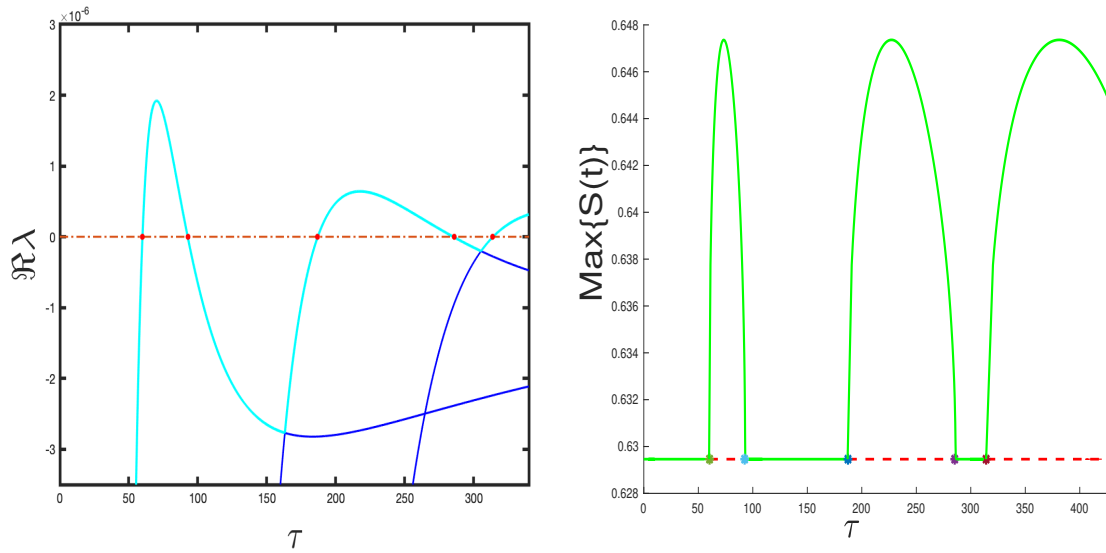
Figure 2.2: (a) Real part of characteristic roots along increasing τ for $k_1 = 0.8968$. There is one curve passing through $\Re\lambda = 0$ at least five times at Hopf points $\tau^{0,+} = 59.83$, $\tau^{0,-} = 92.93$, $\tau^{1,+} = 186.78$, $\tau^{1,-} = 285.96$ and $\tau^{2,+} = 313.73$. (b) The bifurcation diagram with respect to τ with fixed $k_1 = 0.8968$.

(b When $k_c < k_1 < \tilde{k}_{10}$, there are *two sequences* of critical values for τ starting with $\tau^{0,+}$ and

   $\tau^{0,-}$ respectively. There are multiple switches of the stability/instability of E^* as τ increases

   from zero: E^* first loses stability at $\tau^{0,+}$ and regains the stability at $\tau^{0,-}$; and then loses

   stability again at $\tau^{1,+}$;

(c When $\tilde{k}_{10} < k_1 \leq 1$, there is *only one sequence* of critical values for τ starting with $\tau^{0,+}$, and

   E^* loses stability at $\tau = \tau^{0,+}$ remains unstable for all $\tau > \tau^{0,+}$;

(d If $\tau < 12.45 = \tau^{0,+}(1)$, then E^* remains stable for all $k_1 \in [0, 1]$.

The first bifurcation branch in terms of k_1 (i.e., $\tau^{0,+}$ and $\tau^{0,-}$) is demonstrated in Fig. 2.3,

which visually reflects the above conclusions (a-(d.

Although the first bifurcation branches in Fig. 2.3 and the zoom-in Fig. 2.1-(a) are quanti-

tatively the same, the order of subsequential branches can be different. To see this, we choose

$k_1 = 0.931 \in (k_c, \tilde{k}_{10})$. Then there are two positive roots of $F_2(x) = 0$: $x_1 = 0.0099$ with

$F_2'(x_1) > 0$ and $x_- = 0.0021$ with $F_2'(x_-) < 0$; and by numerical calculations, we obtain

$\omega_1 = 0.0997$, $\omega_- = 0.0453$, and accordingly

$$\tau_1^{0,+} = 27.93 < \tau^{0,-} = 66.05 < \tau_1^{1,+} = 90.94 < \tau_1^{2,+} = 153.95 < \tau^{1,-} = 204.67 < \tau_1^{3,+} = 216.96,$$

(2.39)

which differs from the *alternating order* in (2.37). The switches of stability/instability of E^*

is shown in Fig. 2.4-(a), which has a different pattern of zeros from that in Fig. 2.2-(a). For

the above parameter values, Fig. 2.6-(a) and Fig. 2.6-(b) illustrate the solutions of (2.13) with

initial value $(0.7, 0.2, 0.04)$: for $\tau = 15 < \tau_1^{0,+}$, the solution tends to E^*; when $\tau = 30 \in$

$(\tau_1^{0,+}, \tau^{0,-})$, the solution converges to a periodic orbit surrounding E^*.

If fixing $k_1 = 0.94 \in [\tilde{k}_{10}, 1]$, then $F_2(x) = 0$ has a unique positive root $x_1 = 0.0163 > 0$

at which $F_2'(x_1) > 0$. For this case, there is *only one* sequence of critical values $\{\tau^n\}$ with

$\tau^0 = 20.82$, $\tau^1 = 69.97$, $\tau^2 = 119.13$ and $\tau^3 = 168.28$, shown in Fig. 2.4-(b). Noticing that

0.94 and 0.931 are very close, this indicates that the occurrence of Hopf bifurcation from the

endemic equilibrium E^* with respect to the two parameters and the path of bifurcation can be

sensitive and complicated. This phenomenon was not explored in [29] since the model in [29]

has $k_1 = 1$.
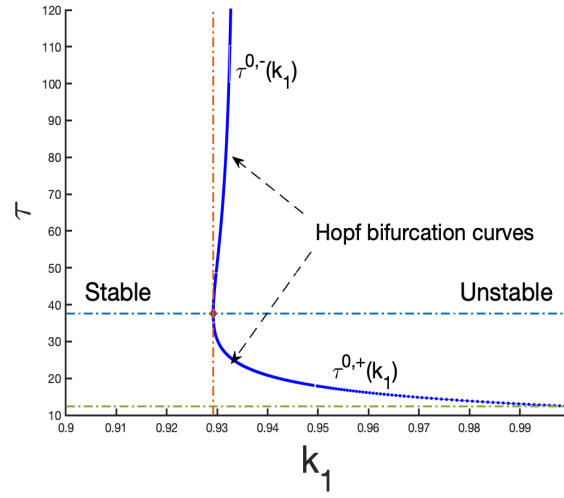
Figure 2.3: The Hopf bifurcation curve $\tau^0(k_1)$ in the k_1 -τ plane.
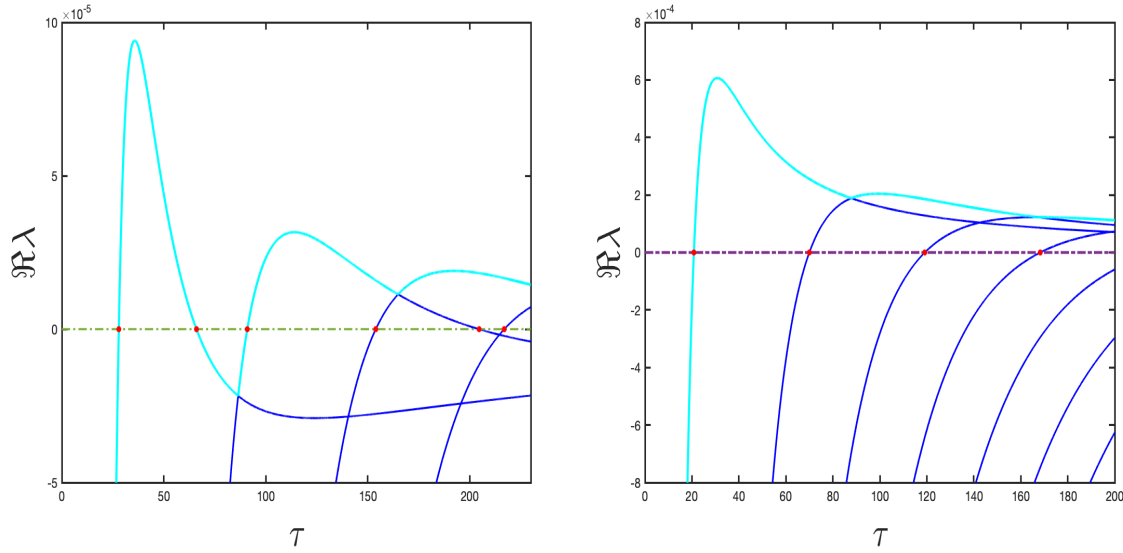
Figure 2.4: Real part of roots of the characteristics equation (2.15) at E^* when increasing τ for fixed k_1: (a). Fixing $k_1 = 0.931$, one curve hit $\Re\lambda = 0$ at $\tau_1^{0,+} = 27.93$, $\tau^{0,-} = 66.05$, $\tau_1^{1,+} = 90.94$, $\tau_1^{2,+} = 153.95$, $\tau_1^{1,-} = 204.67$ and $\tau_1^{3,+} = 216.96$ respectively; (b). Fixing $k_1 = 0.94$, four curves hit $\Re\lambda = 0$ at $\tau^0 = 20.82$, $\tau^1 = 69.97$, $\tau^2 = 119.13$ and $\tau^3 = 168.28$ respectively.
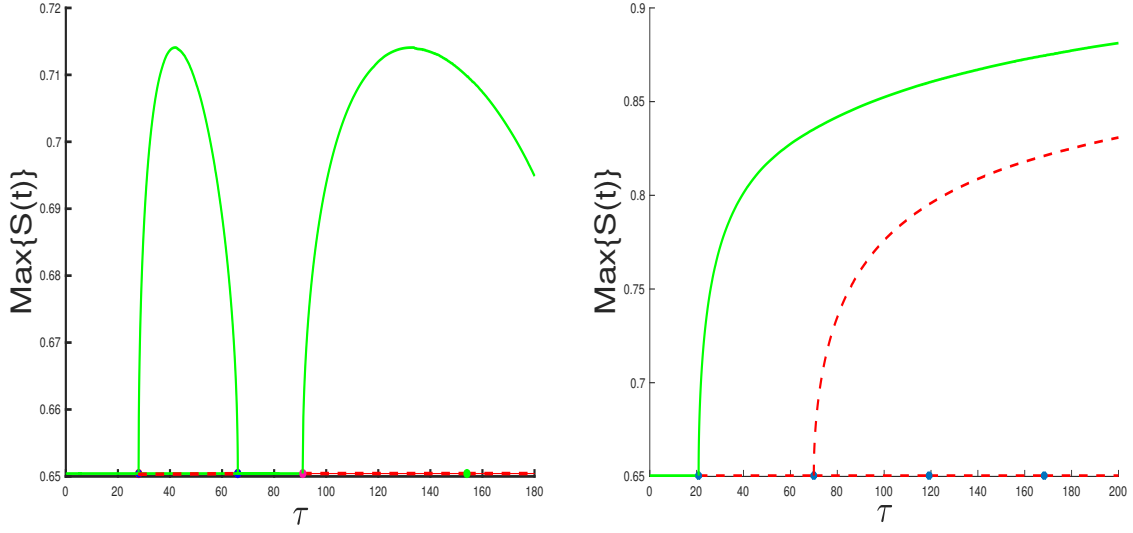
Figure 2.5: The bifurcation diagram for a fixed k_1, choosing the time delay τ as the bifurcation parameter. (a). $k_1 = 0.931$, there are Hopf bifurcation points: $\tau_1^{0,+} = 27.93$, $\tau^{0,-} = 66.05$, $\tau_1^{1,+} = 90.94$ and $\tau_1^{2,+} = 153.95$. (b). $k_1 = 0.94$, there are Hopf bifurcation points: $\tau^0 = 20.82$ and $\tau^1 = 69.97$ $\tau^2 = 119.13$ and $\tau^3 = 168.28$. The stable and unstable parts are denoted by solid green and dashed red lines, respectively; the curve describes the maximum value of the periodic $S(t)$.
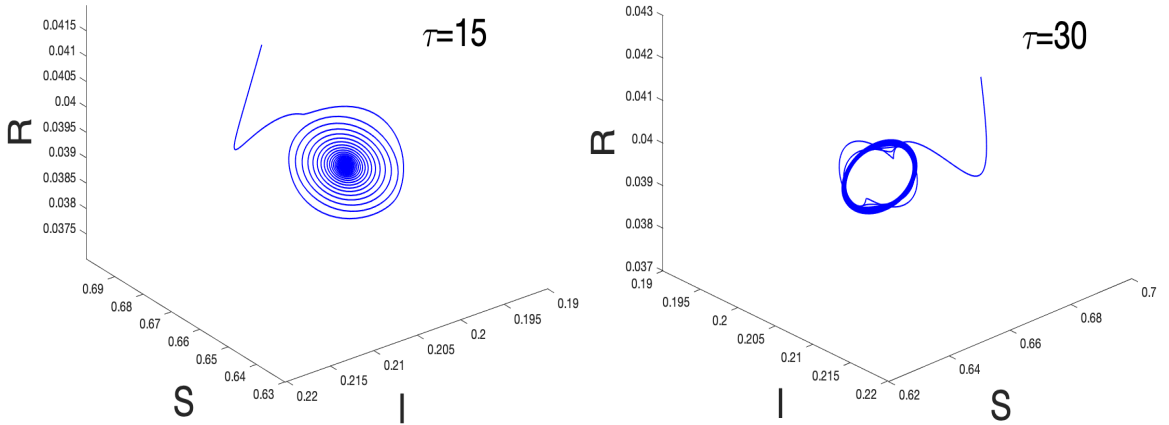
Figure 2.6: The solutions of (2.13) with $k_1 = 0.931$, $(0.7, 0.2, 0.04)$ as the initial value, and other parameters given by (2.39). (a) When $\tau = 15$, the solution converges to $E^* = (0.65043, 0.20657, 0.03972)$. (b) when $\tau = 30$, the solution converges to a periodic solution surrounding E^*.

2.7 Conclusion and discussion

In this paper, we have provided a framework from a new perspective to look at infection forces in modelling infectious disease transmission dynamics. This new angle was motivated by the fear effect in predator-prey interactions as well as by the effects of all those non-pharmaceutical control measures implemented around the world during the ongoing COVID-19 pandemic. Such a framework cannot only accommodate various infection force functions in existing infectious disease models but can also suggest new forms of infection forces, particularly infection forces that depend on past disease surveillance. We have demonstrated this new framework by incorporating, into a classic SIRS model, an infection force that depends on both the disease surveillance at current time t and the disease surveillance at a past time $t - \tau$, with each being given a weight (k_0 and k_1 respectively). The resulting new model (2.13) is a system of delay differential equations, which includes some existing models as special cases. Here, the two new parameters τ and k_1 ($k_0 = 1 - k_1$) can be considered strategy parameters, reflecting the impact of the current and past disease surveillances on disease control measures and human behaviours.

We have performed a standard analysis on the new model (2.13), including confirming the well-posedness of the model, identifying the basic reproduction number R_0 for the model, establishing the threshold dynamics for the model in terms of R_0, and exploring the stability of and bifurcation from E^* with respect to the two key parameters τ and k_1. Our results show that although the basic reproduction number of the new model (featured by τ and k_1 compared to existing models) does not affect R_0 and the threshold between disease-free and endemic, they do affect the disease dynamics (long-time disease patterns) when it becomes endemic. More specifically, we have found that there are ranges on τ and k_1 for which the model demonstrates convergent (to E^*) endemic pattern and oscillatory endemic pattern (due to Hopf bifurcations), respectively. Most interestingly, we have observed that the *path of Hopf bifurcations* when τ increases can be different when other model parameters are at different values, as illustrated by the first few critical values of τ in (2.37) and (2.39). Note that the model in the recent work [29] is a special case of our reduced model (2.4.2) where k_1 is fixed at 1 (hence $k_0 = 0$). Hence, even our reduced model (2.4.2) offers more flexibility in the use of disease surveillance data:

depending on how recent the datum is (e.g., the value of τ), its impact on the long-term disease dynamics (*through k_1*) could be different. The results in Section 2.5, particularly the numerical results provided in Section 2.6 clearly show the role $k_1 \in [0, 1]$ can play and the difference it can make. These results on the impact of k_1 and τ on the long-term dynamics of our models imply that the long-term effect of control measures changes when the delay in reporting of infection increases; therefore, they actually have demonstrated that timely reporting is important to the management of a disease.

We point out that our results on Hopf bifurcations are only local. It is possible to carry out a global bifurcation analysis in the case that the equation $F_2(x) = 0$ only has one positive root, generating only one sequence of critical values for τ. We choose not to do such a global bifurcation in this already lengthy paper since the main purpose is to demonstrate our new framework of modelling infection forces in transmission dynamics when non-pharmaceutical effects are considered. When $F_2(x) = 0$ has more than one positive root, the dynamics are even richer since multiple switches of stability/instability of endemic equilibrium may happen.

In reality, unavoidably there is a delay in the availability of disease surveillance at any given time. This is because it takes time to diagnose/test report cases and collect the data during an epidemic. This fact forces one to use past disease surveillance in measuring the severity of the epidemic and hence justifies the form (4.2) and (4.3) for the severity function $L(t)$. Moreover, in addition to the number of infected cases $I(t)$, the rate of change of $I(t)$ (i.e., $I'(t)$) will also have an impact on the practically susceptible population (i.e., $S_p = PS$) through the proportion function $P(L)$: positive $I'(t)$ would increase L (decrease $P(L)$) and negative $I'(t)$ would decrease L (include $P(L)$). This idea was explored in Xiao et al [35] where the authors proposed a media-impact function of the form $\beta_0 e^{M(t)}$, where

$$M(t) = \max \left\{ 0, p_1 I(t) + q_1 I_q(t) + p_2 \frac{dI(t)}{dt} + q_2 \frac{dI_q(t)}{dt} \right\}.$$

Incorporating such a dependence on $I'(t)$ would significantly make the model more challenging to analyze, as seen in [35]. A natural alternative way to measure the rate of change is to look at an approximation of $I'(t)$ by the difference $\Delta I(t) = I(t) - I(t - \sigma)$ which has recently been used in Li et al. [22]. Note that the dependence of $L(t)$ on $\Delta I(t)$ can be absorbed into the form of

(4.3). Therefore, the form (4.3) and the more general for (4.2) deserve more investigations in association with various proportion functions $P(L)$ (e.g., those in (A)-(B)) for infection disease models of various types.

Bibliography

[1] R. M. Anderson and R. M. May, Infectious diseases of humans: dynamics and control, *Oxford University Press, Oxford,* 1992.

[2] M. Alexander and S. Moghadas, Periodicity in an epidemic model with a generalized non-linear incidence, *Math. Biosci*, **189** (2004), 75-96.

[3] M. Alexander and S. Moghadas, Bifurcation analysis of an SIRS epidemic model with generalized incidence, *SIAM J. Appl. Math.*, **65** (2005), 1794-1816.

[4] K. L. Cooke, Stability analysis of a vector disease model, *Rocky Mt. J. Math.*, **5**(1979), 31-42.

[5] C. Castillo-Chavez and H. R. Thieme, Asymptotically autonomous epidemic models. In: O. Arinoet al.: Mathematical population dynamics: analysis of heterogeneity, *I. Theory of Epidemics.*, Wuerz, Canada, 1995.

[6] R. M. Corless, G. H. Gonnet, D. E. Hare, D. Jeffrey and D.E. Knuth, On the Lambert W function, *Adv Comput Math.*, **5(1)** (1996), 329-359.

[7] J. Cui, Y. Song and H. Zhu, The impact of media on the control of infectious diseases, *J Dyn. Diff. Eqns.*, **20** (2008), 31-53.

[8] J. A. Collera, Numerical continuation and bifurcation Analysis in a Harvested Predator-prey Model with Time Delay using DDE-Biftool, *SEAMS School on Dynamical Systems and Bifurcation Analysis*, (2018), 225-241.

[9] K. Engelborghs, T. Luzyanina and G. Samaey, A Matlab package for bifurcation analysis of delay differential equations, *DDE-BIFTOOL.*, **2** (2001).

[10] K. Engelborghs, T. Luzyanina and D. Roose, Numerical bifurcation analysis of delay differential equations using DDE-BIFTOOL, *ACM Trans Math Softw (TOMS).*, **28(1)** (2002), 1-21.

[11] S. Fan, A new extracting formula and a new distinguishing means on the one variable cubic equation, *Nat Sci J Hainan Teach Coll*, **2** (1989), 91-98.

[12] J. K. Hale and S. M. Verduyn Lunel, Introduction to functional differential equations, *Springer, New York*, 1993.

[13] G. Huang, Y. Takeuchi, W. Ma and D. Wei, Global stability for delay SIR and SEIR epidemic models with nonlinear incidence rate, *Bull. Math. Biol.*, **72(5)** (2010), 1192-1207.

[14] G. Huang and Y. Takeuchi, Global analysis on delay epidemiological dynamic models with nonlinear incidence, *J. Math. Biol.*, **63(1)** (2010), 125-139.

[15] A. Korobeinikov and P. K. Maini, A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence, *Math. Biosci. Eng.*, **1** (2004), 57-60.

[16] A. Korobeinikov and P. K. Maini, Non-linear incidence and stability of infectious disease models, *Math. Med. Biol.*, **22** (2005), 113-128.

[17] W. Liu, S. A. Levin and Y. Iwasa, Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models, *J. Math. Biol.*, **23** (1986), 187-204.

[18] W. Liu, H. W. Hethcote and S. A. Levin, Dynamical behavior of epidemiological models with nonlinear incidence rates, *J. Math. Biol.*, **25** (1987), 359-380.

[19] R. Liu, J. Wu and H. Zhu, Media/psychological impact on multiple outbreaks of emerging infections diseases, *Comp. Math. Meth, Medic.*, **8** (2007), 153-164.

[20] M. Lu, J. Huang, S. Ruan and P. Yu, Bifurcation analysis of an SIRS epidemic model with a generalized nonmonotone and saturated incidence rate, *J. Diff. Eqns.*, **267** (2019), 1859-1898.

[21] M. Lu, J. Huang, S. Ruan and P. Yu, Global dynamics of a susceptible-infectious-recovered epidemic model with a generalized non-monotone incidence rate, *J. Dyn. Diff. Eqns.*, **33** (2020), 1-37.

[22] A. Li, Y. Wang, P. Cong, and X. Zou, Re-examination of the impact of some non-pharmaceutical interventions and media coverage on the COVID-19 outbreak in Wuhan, *Infect. Dis. Model.*, **6** (2021), 975-987.

[23] H. McCallum, N. Barlow and J. Hone, How should pathogen transmission be modelled? *Trends Ecol. Evol.*, **6(16)** (2001), 295-300.

[24] C. C. McCluskey, Complete global stability for an SIR epidemic model with delay—distributed or discrete, *Nonlin. Anal. RWA.*, **11** (2010), 55-59.

[25] C. C. McCluskey, Global stability for an SIR epidemic model with delay and nonlinear incidence, *Nonlin. Anal. RWA.*, **11** (2010), 3106-3109.

[26] S. Ruan and W. Wang, Dynamical behavior of an epidemic model with a nonlinear incidence rate, *J. Diff. Eqns.*, **188** (2003), 135-163.

[27] S. Ruan and J. Wei, On the zeros of transcendental functions with applications to stability of delay differential equations with two delays, *Dyn Contin Discrete Impuls Syst Ser A Math Anal.*, **10** (2003), 863-874.

[28] A. M. Rahman and X. Zou, Global dynamics of a two-strain disease model with latency and saturating incidence rate, *Can. Appl. Math. Quat.*, **20** (2012), 51-73.

[29] P. Song and Y. Xiao, Global hopf bifurcation of a delayed equation describing the lag effect of media impact on the spread of infectious disease, *J. Math. Biol.*, **76** (2018), 1249-1267.

[30] P. Song and Y. Xiao, Analysis of an epidemic system with two response delays in media impact function, *Bull. Math. Biol.*, **81** (2019), 1582-1612.

[31] W. Wang, Epidemic models with nonlinear infection forces, *Math. Biosci. Eng.*, **3(1)** (2006), 267.

[32] D. Xiao and S. Ruan, Global analysis of an epidemic model with nonmonotone incidence rate, *Math. Biosci.*, **20** (2007), 419-429.

[33] R. Xu and Z. Ma, Global stability of a SIR epidemic model with nonlinear incidence rate and time delay, *Nonlin. Anal. RWA.*, **10** (2009), 3175-3189.

[34] R. Xu and Z. Ma, Global dynamics of a vector disease model with saturation incidence and time delay, *IMA J. Appl. Math.*, **76** (2011), 919-937.

[35] Y. Xiao, S. Tang and J. Wu, Media impact switching surface during an infectious disease outbreak, *Sci. Rep.*, **5(1)** (2015), 7838.

Chapter 3

On final and peak sizes of an epidemic with latency under some non-pharmaceutical interventions

3.1 Introduction

Many new emerging diseases have much shorter durations than the life span of the human host. Thus, in modelling the dynamics of such diseases, one typically ignores the demography of the host and just focuses on the transmission dynamics. For such a model, the diseases eventually die out, and hence, the long-term disease dynamics are no longer a major concern. Instead, the scale of the epidemic is generally the major concern, which can be reflected by the final size and the peak size. Taking the classic Kermack-McKendrick (K-M) disease model in [23] as an example, the final size r_∞ is determined in [23] by the equation

$$r_\infty = 1 - e^{-\mathcal{R}_0 r_\infty} \tag{3.1}$$

where $r(t)$ represents the fraction of recovery in a wholly closed population denoted by N_0 at time t and $r_\infty := r(\infty)$. Similar formulas have been obtained for final sizes of some other circumstances of epidemics (some expressed in terms of the susceptible population rather than the recovered population), see, e.g., [1, 3, 7, 13, 26] and the references therein.

For a special and the simplest case of the K-M model described by the following ODE system

$$\begin{cases} S'(t) = -\beta I(t)S(t), \\[2mm] I'(t) = \beta I(t)S(t) - \gamma I(t), \\[2mm] R'(t) = \gamma I(t). \end{cases} \tag{3.2}$$

The final size equation can be easily obtained by analyzing the trajectories of the $I-S$ equations and using the relation $S(\infty) + R(\infty) = S(0) =: S_0$. For the general case of the K-M model and some other expansions of the K-M model without demography, the *renewal equation (RE) approach* is often employed, and it turns out to be powerful. For instance, Diekmann *et al.* [8] reformulated the K-M model in terms of the RE for the force-of-infection (FOI) $F(t)$

$$F(t) = \int_0^\infty F(t - \tau)S(t - \tau)A(\tau)d\tau, \tag{3.3}$$

and the cumulative force-of-infection (CFOI) $y(t)$

$$y(t) = \int_0^\infty (1 - e^{-y(t-\tau)})S(-\infty)A(\tau)d\tau. \tag{3.4}$$

Here $A(\tau)$ describes the expected contribution to the force-of-infection. Indeed, by the RE approach, the authors established the following equation for $y(\infty)$

$$y(\infty) = \mathcal{R}_0(1 - e^{-y(\infty)}), \tag{3.5}$$

and identified the relation

$$r_\infty = \frac{y(\infty)}{\mathcal{R}_0} \tag{3.6}$$

which transforms the final size equation (3.5) for the CFOI to the final size equation (3.1) for the recovered fraction.

The above brief illustration of the RE approach in terms of FOI shows that FOI is indeed an important notion and offers an alternative candidate for employing the RE approach. On the other hand, depending on specific scenarios of epidemics, various forms of FOI (leading to various incidence rate functions) have been used in disease models in the literature. In

this context, in a recent work [2], motivated by the various non-pharmaceutical interventions implemented during the COVID-19 pandemic, we provided a new framework and insights into FOI by introducing the notion of "practically susceptible population", which is a fraction of the "biologically susceptible population," consisting of those who are not influenced by any NPIs and behave as there was no epidemic at all. This new viewpoint on FOI allows us to explore the effects of non-biological and non-pharmaceutical interventions and psychological responses (precautions) on disease dynamics.

More precisely, practically susceptible population, denoted by $S_p(t)$, is a fraction of the biologically susceptible population denoted by $S(t)$, that is

$$S_P(t) := P(t)S(t) \tag{3.7}$$

where the fraction $P(t) = P(L(t))$ naturally depends on the severity of the epidemics denoted by $L(t)$ in a way reflected by the following properties (see, [2]):

$$\frac{dP(L)}{dL} \leq 0 \quad \text{with} \quad P(0) \leq 1 \quad \text{and} \quad \lim_{L \to \infty} P(L) \geq 0. \tag{3.8}$$

In general, $P(L)$ reflects the effect of any NPIs and the host's precaution level (e.g., behaviour changes and use of EPPs). The severity level function L can be described by some disease variables like the infected population $I(t)$ or the weighted sum of these disease variables and reflect the information on the disease severity. Some general forms of L based on the infected population have been given in [2]. Correspondingly, replacing $S(t)$ by $S_p(t)$ in the mass action incidence rate $\beta I(t)S(t)$ and and standard incidence rate $\beta I(t)S(t)/[I(t) + S(t)]$ respectively, we obtain the respective new incidence rates

$$\beta I(t)S_p(t) = [\beta P(t)I(t)] \cdot S(t) \text{ and } \frac{\beta I(t)S_p(t)}{I(t) + S_p(t)} = \left[\frac{\beta P(t)I(t)}{I(t) + P(t)S(t)}\right] \cdot S(t),$$

which leads to the respective new FOIs as

$$\beta P(t)I(t) \text{ and } \frac{\beta P(t)I(t)}{I(t) + P(t)S(t)}.$$

In [2] and a follow-up work [11], using the revised FOIs that reflect the effects of NPIs as explained above, we explored the impact of the corresponding NPIs on the long-term disease dynamics for some disease models *with demography*. In addition, [2, 11] focused on compartmental models that are simply described by differential equations (such as ordinary differential equations(ODEs) or delay differential equations(DDEs)). Now, as we mentioned in the beginning, our interest in this work is in the final size and peak size of an epidemic that is deemed to die out (reflected by the nature of "without demography"). Then there arises the question:

(Q1) What would be the impact of NPIs and the host's response/precaution (affecting $P(L)$) on the final size and peak size of a disease model without demography?

We have seen earlier that for some models, the final size equation can be derived by renewal equations for the force of infection. For the above-revised force of infection reflecting the host's fear of diseases, we may ask:

(Q2) Can we apply the REs approach for the revised FOI to address (Q1)?

It is well-known that most compartmental epidemic models can be described by REs, and hence, we wish to employ the RE theory to address these issues. and conversely, choosing some particular kernels in REs yields compartmental models [9, 16](references therein). Biologically, REs give a reasonable way to trace renewals/reproductions of the given class/population over time. The idea was initially developed by Lotka [24, 31] and studied by many researchers, e.g., [17, 18, 22, 25] *et al.* in modelling demography and species reproduction. Particularly, Diekmann *et al.* re-drew researchers' attention to REs and highlighted REs' power in mathematical epidemiology-modelling in his several works (e.g., [8, 9, 13, 15, 16]).

Encouraged by these works, in this paper, we will tackle the above two questions. To this end, in Section 2, we will first present some preliminaries that involve infection age. Going through the references [8, 23] intensively, we find that age/stage density is the key to associating REs and compartment systems — age density deduces a RE and a compartment system can be obtained through the age density. In Section 3, we derive the renewal equations for the revised force of infection function and the corresponding incidence rate function. Then, in Section 4, we use the REs for the FOI and the incidence rate to derive equations that govern

the final size. Section 5 is dedicated to some special cases of the general model represented by the REs model in Section 3, for which we are able to obtain more detailed and more explicit results, particularly some results on the peak size which is, in general, very challenging (if not impossible) for general case. We conclude the paper with a discussion section (Section 6), in which we discuss the implications of our results and some possible extensions of this work.

3.2 Preliminaries on infection age structure and related notions

3.2.1 Age density of the infected class

Denote by $a \in [0, \infty)$ the infection age which defines the length of time an individual has been infected, and let $u(t, a)$ be the density (with respect to a) of the infected population at time t is $u(t, a)$. Let $\gamma(a)$ denote the recovery rate at age a satisfying $\gamma(\infty) < \infty$. For the age density of infected individuals, there is the well-known as the McKendrick/von Foerster equation.

$$\frac{\partial u(t, a)}{\partial a} + \frac{\partial u(t, a)}{\partial t} = -\gamma(a)u(t, a), \tag{3.9}$$

where $\gamma(a)$ is death rate infected population with infection age a.

Firstly, we note that the characteristics of (3.9) define a family of lines with equation $t = t_0 + a$, where t_0 is considered the "birthdate" (infection time) of an infected individual. Then, the infection age density along the characteristic is $u(t, a) = u(t_0 + a, a) := u_{t_0}(a)$ which solves

$$\frac{du_{t_0}(a)}{da} = -\gamma(a)u_{t_0}(a). \tag{3.10}$$

The solution of (3.10) is:

$$u(t, a) = u_{t_0}(a) = u_{t_0}(0)e^{-\int_0^a \gamma(x)dx} = u(t_0, 0)e^{-\int_0^a \gamma(x)dx} = u(t - a, 0)e^{-\int_0^a \gamma(x)dx}. \tag{3.11}$$

Let

$$\sigma(a) := e^{-\int_0^a \gamma(x)dx},$$

which expresses the probability of a newly infected individual remaining infected at infection age a. That means a fraction $\sigma(a)$ of the population infected at time $t - a$ will still remain infected at time t with age a. Simple calculation yields

$$\sigma'(a) = -\gamma(a)\sigma(a) < 0.$$

Naturally, the following assumption on $\sigma(a)$

$$\sigma(\infty) = e^{-\int_0^\infty \gamma(x)dx} = 0, \tag{3.12}$$

should be imposed since no individual will "survive" to infinity age. Accordingly $u(t, \infty) = 0$ for any $t \in \mathbb{R}$.

Next, we need to track those infected individuals who are already infected at $t = 0$ with age a_0 and still remain infected at t with age $a = a_0 + t$. Let $u_0(a_0) = u(0, a_0)$, which is the initial age distribution of the infected class. Solving (3.9) for $a > t$, one obtains

$$u(t, a) = u_0(a - t)e^{-\int_{a-t}^a \gamma(x)dx} = u_0(a - t)\frac{\sigma(a)}{\sigma(a - t)} \tag{3.13}$$

Here $\sigma(a)/\sigma(a - t)$ describes the probability of those already infected initially at $t = 0$ with age $a_0 = a - t$ still remaining in the infected class at time t (with infection age a).

Summarizing the above, the solution of (3.9) is expressed piecewise as below

$$u(t, a) = \begin{cases} u(t - a, 0)\sigma(a), & a < t \\ \\ u_0(a - t)\frac{\sigma(a)}{\sigma(a-t)}, & a > t. \end{cases} \tag{3.14}$$

The piecewise nature accounts for the two sources for $u(t, a)$: one is from the initially infected individuals for $a > t$ represented by $u_0(a - t) = u(0, a - t)$, and the other is from the new infections ("birth" or "production") during $[0, t-a]$ for $a < t$ represented by $B(t-a) = u(t-a, 0)$.

Many diseases have latency, and the latent period can vary from individual to individual. For some diseases, the individual latencies vary slightly, and in such a case, one may assume

the disease to have a fixed constant latency. If considering a fixed latency $\tau \geq 0$,

$$E(t) := \int_0^\tau u(t,a)da, \quad I(t) := \int_\tau^\infty u(t,a)da,$$

represent the total numbers of latent class and infectious class respectively. Accordingly, $U(t) := E(t) + I(t)$ is the total number of infected individuals and

$$R(t) := \int_0^\infty \gamma(a)u(t,a)da,$$

sums up all those who recover at different infection ages and, hence, gives the total number of recovered classes at time t.

Throughout this work, we use the following classical/standard notations:

$$s(t) := S(t)/N_0, \quad s(\infty) := s_\infty \quad \text{and} \quad r(t) := R(t)/N_0, \quad r(\infty) := r_\infty. \tag{3.15}$$

3.2.2 The related mean period

Consider the recovery age as a random variable ξ. Then, at infection age a, an infected individual either has recovered or becomes infectious. Thus, the probability of a latent individual becoming infectious at age a is

$$P(0 < \xi \leq a) = 1 - \sigma(a) = \int_0^a \gamma(x)\sigma(x)dx$$

Accordingly, the probability density of ξ is $f(a) = \gamma(a)\sigma(a)$. Assuming fixed constant latency $\tau > 0$ as stated above, then

$$
\begin{aligned}
T_E &= \frac{\int_0^\tau x\gamma(x)\sigma(x)dx}{\int_0^\tau \gamma(x)\sigma(x)dx} = \frac{\int_0^\tau \sigma(x)dx}{1 - \sigma(\tau)} - \frac{\tau\sigma(\tau)}{1 - \sigma(\tau)}, \\
T_I &= \frac{\int_\tau^{+\infty} x\gamma(x)\sigma(x)dx}{\int_\tau^\infty \gamma(x)\sigma(x)dx} = \frac{\int_\tau^\infty \sigma(x)dx}{\sigma(\tau)} + \tau
\end{aligned}
\tag{3.16}
$$

respectively, give the mean time an infected individual stays in the latent class and the mean-time of an infected individual stays in the infectious class. Consequently, $T_I - \tau$ is the mean

infectious period and

$$
\begin{aligned}
T_{tol} &= \int_0^\infty x\gamma(x)\sigma(x)dx = -\int_0^\infty xd\sigma(x) = \int_0^\infty \sigma(x)dx - x\sigma(x)|_{x=0}^{x=\infty} = \int_0^\infty \sigma(x)dx \\
&= \int_0^\tau \sigma(x)dx + \int_\tau^\infty \sigma(x)dx \\
&= \sigma(\tau)T_I + (1 - \sigma(\tau))T_E
\end{aligned}
\tag{3.17}
$$

is the mean infected period, which is the weighted sum of T_E and T_I.

3.3 Derivation of the model

3.3.1 The derivation of the REs for $B(t)$ and FOI $F(t)$

Assume that the infectivity(or weight) of infected individuals with infection age a is $\beta(a) \in [0, 1]$ for $a \in [\tau, \infty)$, and then the total infectivity of the ill population at time t is,

$$
C(t) := \int_\tau^\infty \beta(a)u(t, a)da.
$$

As is known, this is traditionally defined as the force-of-infection when focusing on physio-logical infectivity (see, e.g., Anderson and May [2]). Particularly, if $\beta(a)$ is independent on a, $C(t) = \beta I(t)$.

Now, consider the effect of NPIs and behaviour changes during an epidemic. As mentioned in the introduction, due to fear of the disease and/or possible NPIS imposed by public health agents or government, people will typically reduce their social activities, and thus, only a fraction $P = P(L)$ of the biologically susceptible population $S(t)$ is practically susceptible. Then, the incidence rate is accordingly revised to

$$
B(t) := \int_\tau^\infty \{\beta(a)P_l(t)S(t)u(t, a)\}\, da = S_P(t)C(t) = P(L)S(t)C(t)
\tag{3.18}
$$

and accordingly, the force-of-infection is revised to

$$
F(t) := C(t)P(L),
\tag{3.19}
$$

in which $C(t)$ and $P_l(t)$ measure the efficiency of physiological infection (focusing more on features of epidemics) and non-biological/non-pharmaceutical interventions (reflecting the features of human behaviour changes), respectively. In other words, F combines the efficacy of epidemic infection transmission and non-biological/non-pharmaceutical interventions.

Obviously, $u(t, 0) = B(t)$ holds. From (3.18) it follows that, for $t \geq \tau$,

$$
\begin{aligned}
B(t) &= \int_\tau^\infty \beta(a) S_P(t) u(t, a) da \\
&= S_P(t) \left(\int_\tau^t \beta(a) u(t, a) da + \int_t^\infty \beta(a) u(t, a) da \right) \\
&= S_P(t) \left(\int_\tau^t \varphi(a) B(t - a) da + G(t) \right)
\end{aligned}
$$
(3.20)

where

$$
\varphi(a) := \begin{cases} 0, & a \in [0, \tau) \\ \beta(a)\sigma(a), & a \in [\tau, \infty) \end{cases}
$$
(3.21)

and

$$
G(t) := \int_t^\infty \beta(a) u_0(a - t) \frac{\sigma(a)}{\sigma(a - t)} da = \int_t^\infty u_0(a - t) \frac{\varphi(a)}{\sigma(a - t)} da = \int_0^\infty u_0(\eta) \frac{\varphi(\eta + t)}{\sigma(\eta)} d\eta.
$$
(3.22)

Here $\varphi(a)$ is continuous on $a \in [\tau, \infty)$, and it is the expected contribution to the force of infection by per illed individual at age a according to [8]. Note that $T_{tol} < \infty$ ensures $\int_\tau^\infty \varphi(a) da < \infty$.

For $0 \leq t < \tau$, $a > \tau$ implies $a > t$ and hence

$$
B_\tau(t) := S_P(t) C(t) = S_P(t) \int_\tau^\infty \beta(a) u(t, a) da = S_P(t) G_\tau(t)
$$
(3.23)

where

$$
\begin{aligned}
G_\tau(t) &= \int_\tau^\infty \beta(a) u(t, a) da = \int_\tau^\infty \beta(a) u_0(a - t) \frac{\sigma(a)}{\sigma(a - t)} da \\
&= \int_\tau^\infty u_0(a - t) \frac{\varphi(a)}{\sigma(a - t)} da = \int_{\tau - t}^\infty u_0(\eta) \frac{\varphi(\eta + t)}{\sigma(\eta)} d\eta.
\end{aligned}
$$
(3.24)

Denote

$$G^\diamond(t) = \begin{cases} G(t) & t \geq \tau, \\ G_\tau(t) & t < \tau. \end{cases} \tag{3.25}$$

$G^\diamond(t)$ reflects the contribution to new infection at t by those already infective at $t = 0$. $G^\diamond(t)$ is continuous at $t = \tau$. Because $\varphi(a)$ goes to zero when the infection age a of an individual tends to infinity, $G(t)$ also tends to zero, i.e. $\lim_{t \to \infty} G^\diamond(t) = \lim_{t \to \infty} G(t) = 0$.

Actually, if *extending the domain of $B(t)$ to the negative axis*, then we actually have

$$G(t) = \int_t^\infty \varphi(a)B(t - a)da \quad \text{and} \quad G_\tau(t) = \int_\tau^\infty \varphi(a)B(t - a)da,$$

due to the fact that $u_0(a - t) = B(t - a)\sigma(a - t)$. Combining the above, we then have obtained the REs for $B(t)$ and $F(t)$:

$$B(t) = S_P(t) \int_\tau^\infty \varphi(a)B(t - a)da \tag{3.26}$$

and

$$F(t) = P_l(t)) \int_\tau^\infty \varphi(a)B(t - a)da = P_l(t) \int_\tau^\infty \phi(a)F(t - a)S(t - a)\,da \tag{3.27}$$

on the *whole time axis* instead of the positive axis, where $P_l(t) = P(L(t))$. This allows us to consider the initial time to be at $-\infty$ for both (3.26) and (3.27), as is done in Diekmann et al [8, 15]. We point out that in contrast to (3.3), the "renewal equation" (3.27) is no longer the traditional scalar RE due to the involvement of the fraction function $P_l(t) = P(L(t))$ and latency. Practically, we need knowledge/information of not only the current *'practically susceptible'* which includes the current information of the epidemic severity but also the entire history of B over the interval $[-\infty, t]$, even if for the current value $B(t)$ or $F(t)$.

We point out that the REs (3.26) and (3.27) naturally lead to the iterated transition property (semi-group property) and allow us to track how the current value depends on the past values. For example, for any $a_1, a_2 \geq 0$, from (3.26), we can obtain

$$B(t) = S_P(t) \int_\tau^\infty \varphi(a_2) \left\{ S_P(t - a_2) \int_\tau^\infty \varphi(a_1)B(t - a_2 - a_1)da_1 \right\} da_2 \tag{3.28}$$

which is a result of two iterative trackings. See Figure 3.1 for an illustration of (3.28).
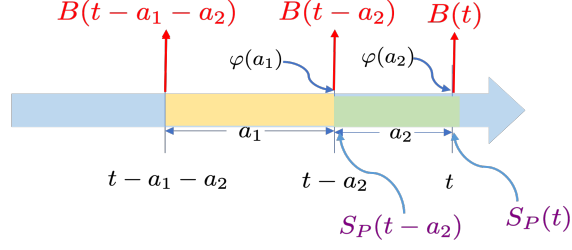
Figure 3.1: Depiction of the RE (3.28) of $B(t)$

3.3.2 Full general model

With the above preparation, a general model without demography can be formulated by RE as

$$
\begin{cases}
S'(t) = -B(t) = -F(t)S(t), \\
F(t) = P_l(t) \int_\tau^\infty \varphi(a)F(t-a)S(t-a)da.
\end{cases}
\tag{3.29}
$$

Corresponding to the model (3.29) with $\varphi(a)$ given in (3.21), we have the following compartmental model (see details in Appendix):

$$
\begin{cases}
S'(t) = -F(t)S(t), \\
E'(t) = -\int_0^\tau \gamma(a)u(t,a)da - u(t,\tau) + F(t)S(t), \\
I'(t) = -\int_\tau^\infty \gamma(a)u(t,a)da + u(t,\tau), \\
R'(t) = \int_0^\infty \gamma(a)u(t,a)da.
\end{cases}
\tag{3.30}
$$

We assume there is no recovery individual at the epidemic's beginning time $t = t_0$ (i.e., $R(t_0) = 0$) throughout this paper. In the sequel, as we proceed further, we will take $t_0 = -\infty$ or $t_0 = 0$, depending on the purpose.

Remark 3.3.1 *It worth noting that (3.29) and (3.30) can both be obtained from the equation (3.9), respectively. Age density of the infected population $u(t,a)$ establishes a bridge between the REs model (3.29) and the compartmental model (3.30).*

Observe that in (3.30), the equations of E' and I' couple via the term $u(t,\tau)$, which is the rate at which The infectious class gains from the latent class. Now we determine $u(t,\tau)$ by (3.14) as below.

For $t \geq \tau$ (long-time), by (3.14),

$$u(t, \tau) = u(t - \tau, 0)\sigma(\tau) = B(t - \tau)\sigma(\tau).$$

That means that, with probability $\sigma(\tau)$, those individuals who get infected at time $t - \tau$ will enter the infectious class from the latent class at time t.

If $t < \tau$ (short-time), again by (3.14),

$$u(t, \tau) = u_0(\tau - t)\frac{\sigma(\tau)}{\sigma(\tau - t)}.$$

At such a time t, only those who are already infected at time $t = 0$ are possible to enter the infectious class (with probability $\sigma(\tau)/\sigma(\tau - t)$), because at time t, all newly infected individuals during $[0, t]$ have no chance to become infectious (their infection ages are all less than the latency τ. Moreover, noting that

$$u_0(\tau - t) = u(-(\tau - t), 0)\sigma(\tau - t) = B(t - \tau)\sigma(\tau - t),$$

we have actually also have $u(t, \tau) = B(t - \tau)\sigma(\tau)$.

Summarizing the above, we conclude that on the whole time axis, there holds

$$u(t, \tau) = B(t - \tau)\sigma(\tau). \tag{3.31}$$

From the first equation of (3.29), we obtain

$$S(t) = S(t_0)e^{-\int_{t_0}^{t} F(\xi)d\xi} = S(t_0)\mathcal{A}(t; t_0) \tag{3.32}$$

with

$$\mathcal{A}(t; t_0) := e^{-\int_{t_0}^{t} F(\xi)d\xi}$$

representing the probability for an uninfected individual at time t_0 to escape from becoming infected at least until time t. Furthermore, the cumulative force of infection (CFOI) during the

period $[t_0, t]$ is given as

$$Y(t; t_0) := \int_{t_0}^{t} F(\xi) d\xi. \tag{3.33}$$

According to (3.32), given that $S(t_0) \approx N_0$ (the total population), we indeed have the following relation among the CFI $Y(t; t_0)$, "escaping" probability $\mathcal{A}(t; t_0)$ and $s_\infty = s(\infty) = S(\infty)/N_0$:

$$s_\infty = \mathcal{A}(\infty; t_0) = e^{-Y(\infty; t_0)}. \tag{3.34}$$

Obviously, divergence of $Y(\infty; t_0)$ is equivalent to $s_\infty = 0$ from (3.34).

Denote

$$\hat{\mathcal{R}}_o(t) := P(0) N_0 \int_{\tau}^{t} \varphi(a) da, \quad t \geq \tau. \tag{3.35}$$

Then the basic reproduction number of (3.29) is given by

$$\mathcal{R}_o := \hat{\mathcal{R}}_o(\infty).$$

The following theorem simply means that the disease will eventually die out. This is not surprising because the model does not have demographic structure, and hence, there is no recruitment for the susceptible class.

Theorem 3.3.2 *For model (3.30), the total infected population $U(\infty) = 0$ holds. Accordingly, $E(\infty) = 0, I(\infty) = 0$.*

Proof: We leave the proof in the Appendix.

Because of the results in Theorem (3.3.2), it is reasonable to assume the following for the severity variable:

Assumption 3.3.3 $L(\infty) = 0$ *and* $L'(\infty) = 0$; *and* $U = 0 \leftrightharpoons L = 0$.

3.4   Final epidemic sizes for the model (3.30)

3.4.1   The case that the initial condition is given by $S(-\infty)$:

To make mathematical analysis easier, Breda and Dikemann *et al.* [8, 15](references therein) assume time starts from $-\infty$. This assumption gives some convenience in the analysis due to the occurrence of the related integrals in the REs. In this subsection, we adopt this assumption on the starting time, i.e. $t_0 = -\infty$, so that $B(t)$ and $F(t)$ have the explicit expressions (3.26) and (3.27), respectively. Further, we also accordingly impose the following initial conditions:

Assumption 3.4.1 $\lim_{t \to -\infty} S(t) = N_0$ *and* $\lim_{t \to -\infty} U(t) = 0$.

This assumption means the total infected population U was negligible in the infinite past. Under the Assumptions (3.3.3) and (3.4.1), there are hold

$$\lim_{t \to \pm\infty} P_l(t) = P(L(\pm\infty)) = P(0), \ \lim_{t \to \pm\infty} P'_l(t) = P'(L)L' = 0 \ \text{ and } \ \lim_{t \to \pm\infty} F(t) = 0. \qquad (3.36)$$

It follows from the expression (3.26) that

$$
\begin{aligned}
F(t) &= P_l(t) \int_\tau^\infty \varphi(a)B(t-a)da \\
&= P_l(t) \int_\tau^\infty F(t-a)S(t-a)\varphi(a)da \\
&= P_l(t) \int_{-\infty}^{t-\tau} F(\eta)S(\eta)\varphi(t-\eta)d\eta.
\end{aligned}
\qquad (3.37)
$$

The cumulative force-of-infection over $(-\infty, T)$ is

$$
\begin{aligned}
Y(T; -\infty) &= \int_{-\infty}^T F(t)dt \\
&= -\int_\tau^{+\infty} \varphi(a)\left\{\int_{-\infty}^T \frac{dS(t-a)}{dt}P_l(t)dt\right\}da \\
&= -\int_\tau^{+\infty} \varphi(a)\left\{P_l(t)S(t-a)|_{t=-\infty}^{t=T} - \int_{-\infty}^T S(t-a)\frac{dP_l(t)}{dt}dt\right\}da \qquad (3.38) \\
&= \hat{\mathcal{R}}_p(T) + S(-\infty)\int_\tau^{+\infty} \varphi(a)\left\{P_l(-\infty) - P_l(T)e^{-Y(T-a;-\infty)}\right\}da \\
&= \hat{\mathcal{R}}_p(T) + \mathcal{R}_o - P_l(T)\hat{\mathcal{H}}(T)
\end{aligned}
$$

where

$$\hat{\mathcal{R}}_p(t) := \int_{-\infty}^t \hat{\mathcal{H}}(\xi)P_l'(\xi)d\xi, \quad \mathcal{R}_p := \hat{\mathcal{R}}_p(\infty) \tag{3.39}$$

with

$$\hat{\mathcal{H}}(t) := \int_\tau^{+\infty} \varphi(a)S(t-a)da = \int_{-\infty}^{t-\tau} \varphi(t-\eta)S(\eta)d\eta. \tag{3.40}$$

Thus,

$$\hat{\mathcal{R}}_P'(t) = P_l'(t)\hat{\mathcal{H}}(t), \tag{3.41}$$

which indicates $P_l(t)$ and $\hat{\mathcal{R}}_p(t)$ have same direction of change.

Differentiating $\hat{\mathcal{H}}(t)$ gives

$$\hat{\mathcal{H}}'(t) = \int_\tau^\infty \frac{dS(t-a)}{dt}\varphi(a)da = -\int_\tau^\infty B(t-a)\varphi(a)da = -C(t) < 0; \tag{3.42}$$

We may also use the alternative formula in (3.40) to obtain

$$\hat{\mathcal{H}}'(t) = -\int_\tau^\infty \varphi(a)\frac{dS(t-a)}{da}da = S(t-\tau)\varphi(\tau) + \int_\tau^\infty \varphi'(a)S(t-a)da. \tag{3.43}$$

Here, we should mention that only for very special $\varphi(a)$ do the (3.41) and (3.43) reduce to explicit differential equations (DEs). For some details on this, see Section 3.5.

Also, simple calculation leads to

$$P_l(-\infty)\hat{\mathcal{H}}(-\infty) = P_l(-\infty)\int_\tau^{+\infty} \varphi(a)S(-\infty)da = \mathcal{R}_o;$$
$$P_l(\infty)\hat{\mathcal{H}}(\infty) = \mathcal{R}_o s_\infty. \tag{3.44}$$

Summarizing the above, the properties of $\hat{\mathcal{H}}(t)$ are given in the following

Lemma 3.4.2 $\hat{\mathcal{H}}(t)$ *given in (3.40) is strictly decreasing on* $t \in (-\infty, \infty)$ *with* $\hat{\mathcal{H}}(-\infty) = \frac{\mathcal{R}_o}{P(0)}$ *and* $\hat{\mathcal{H}}(\infty) = \frac{\mathcal{R}_o s_\infty}{P(0)}$.

Before discussing the property of $\mathcal{R}_p$, we give the following Lemma.

Lemma 3.4.3 *Assume that f and g are functions defined on $\mathbb{R}$, satisfying*

(i) $g : \mathbb{R} \to (g_m, g_M) \subseteq (0, \infty)$ *be a differentiable, bounded and monotonic function with*

$$\inf_{t \in \mathbb{R}} g(t) := g_m, \ \sup_{t \in \mathbb{R}} g(t) := g_M;$$

(ii) $f : \mathbb{R} \to (0, f_M)$ *be a differentiable and bounded function satisfying*

$$f(\pm\infty) = \sup_{t \in \mathbb{R}} f(t) := f_M < \infty.$$

Then

$$\int_{-\infty}^{\infty} g(t) f'(t) dt$$

converges. Moreover, if g is strictly decreasing, then

$$(g_m - g_M) f_M \leq \int_{-\infty}^{+\infty} g(t) f'(t) dt \leq 0. \tag{3.45}$$

Proof: See Appendix.

By the property of function $P_l(t)$ (i.e. $P(L)$), Lemma 3.4.2 and Lemma 3.4.3, we immediately have the following theorem.

Theorem 3.4.4 *For $\mathcal{R}_p$ is defined in (3.39), there holds*

$$-\mathcal{R}_o < \mathcal{R}_p \leq 0. \tag{3.46}$$

By (3.38), the CFOI at $t = \infty$ with $t_0 = -\infty$ is

$$Y(\infty; -\infty) = \ln\left(\frac{1}{s_\infty}\right) = \mathcal{R}_p + \mathcal{R}_o\left(1 - e^{-Y(\infty;-\infty)}\right). \tag{3.47}$$

That is

$$-\ln(s_\infty) = \mathcal{R}_p + \mathcal{R}_o(1 - s_\infty); \tag{3.48}$$

equivalently,

$$r_\infty = \frac{Y(\infty; -\infty)}{\mathcal{R}_o} - \frac{\mathcal{R}_p}{\mathcal{R}_o}. \tag{3.49}$$

Remark 3.4.5 *Compared with (3.5) and (3.6) which depend on a single parameter $\mathcal{R}_0$, (3.47) and (3.49) depend only only on parameter $\mathcal{R}_o$ but also $\mathcal{R}_p$. Note that $\mathcal{R}_p$ accounts for the impact of the fraction function $P_l(t) = P(L(t))$ which reflects the behaviour changes during an epidemic.*

Rearranging the final size equation (3.48) yields

Theorem 3.4.6 *Given that $S(-\infty) \approx N_0$, the final size relation of model (3.30) is explicitly presented in the form*

$$s_\infty = e^{-\mathcal{R}_o(1-s_\infty)-\mathcal{R}_p}; \tag{3.50}$$

or equivalently,

$$r_\infty = 1 - e^{-\mathcal{R}_p}e^{-\mathcal{R}_o r_\infty} \tag{3.51}$$

where $\hat{\mathcal{R}}_p(t)$ and $\mathcal{R}_p$ are defined in (3.39) and s_∞ and r_∞ are as in (3.15).

From the formula (3.50), we can see the fraction s_∞ is necessarily the intersection point of the following functions

$$q_1(x) = e^{-\mathcal{R}_o(1-x)}, \quad q_2(x) = e^{\mathcal{R}_p}x \tag{3.52}$$

in the open interval $(0, 1)$. Note that $q_1(x)$ is an exponentially growing function and $q_2(x)$ is an linearly growing function with slope $e^{-\mathcal{R}_p}$ depending on $\mathcal{R}_p$.

Observe that $q_2(x)$ is increasing increasing in the parameter $\mathcal{R}_p$, and $q_2(x) = x$ when $\mathcal{R}_p = 0$. This together with the fact that $\mathcal{R}_0 \leq 0$ implies the following

(C1) If $\mathcal{R}_o \leq 1$, $q_1(x) > x$ on $x \in [0, 1)$, that is $q_1(x) > q_2(x)$ for any $\mathcal{R}_p \leq 0$. It means there is only one $s_\infty = 1 \in [0, 1]$. Therefore, in this case, the final size r_∞ is close to zero. The same conclusion is also obtained in Brauer, Breda and Diekmann et al. [6, 8] without $\mathcal{R}_p$ term.

(C2) If $\mathcal{R}_o > 1$, there exists a critical value $\mathcal{R}_p^{cr}$ for critical value $\mathcal{R}_p$ such that

    (i) when $\mathcal{R}_p \in (\mathcal{R}_p^{cr}, 0)$, $q_1(x)$ and $q_2(x)$ have two intersect points $x_1^* \in (0, 1/\mathcal{R}_o)$ and $x_2^* \in (1/\mathcal{R}_o, 1)$;

    (ii) when $\mathcal{R}_p = \mathcal{R}_p^{cr}$, there is only one intersect point $x^* = 1/\mathcal{R}_o \in (0, 1)$;

(iii) when $\mathcal{R}_p = 0$, there is only one intersect point $x^* \in (0, 1/\mathcal{R}_o)$.

The critical value $\mathcal{R}_p^{cr}$ is indeed determined by the two tangential conditions $q_1(x) = q_2(x)$ and $q_1'(x) = q_2'(x)$ which turns out to be

$$\mathcal{R}_p^{cr} := \ln(\mathcal{R}_o) - \mathcal{R}_o + 1$$

which satisfies

$$-\frac{(1 - \mathcal{R}_o)^2}{\mathcal{R}_o} < \mathcal{R}_p^{cr} \leq 0$$

for any $\mathcal{R}_o > 1$. And

$$q_2^{cr}(x) := e^{\mathcal{R}_p^{cr}} x \tag{3.53}$$

is the tangent line of $q_1(x)$ with the tangent point $(1/\mathcal{R}_o, e^{1-\mathcal{R}_o})$. See Fig. 3.2 for a illustration.

Remark 3.4.7 $P_l(t) \equiv 1$ *indicates* $\mathcal{R}_p = 0$. *In this special case, C2-(ii))-(iii) shows* s_t *will eventually fall down to a level* s_∞ *below* $1/\mathcal{R}_o$, *provided that* $\mathcal{R}_o > 1$.
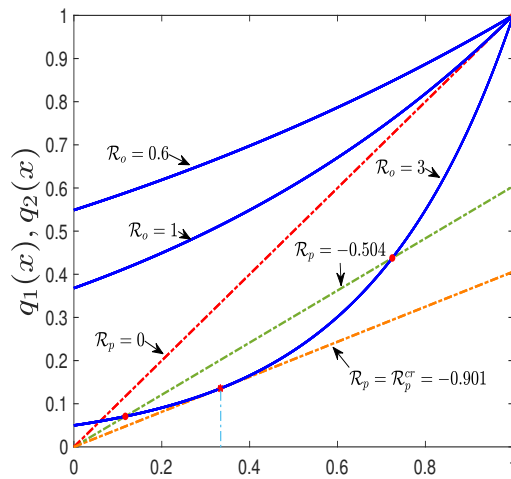
Figure 3.2: Demonstration of $q_1(x)$ and $q_2(x)$.

Remark 3.4.8 *For the case (C2)-(i), there are two intersection values* x_1^* *and* x_2^*. *Remark 3.4.7 seems to suggest that* x_2^* *should be excluded and* x_1^* *is the true value of* s_∞. *Another intuition to support this suggestion is that* x_2^* *is not* s_∞ *is that* $\mathcal{R}_p$ *impact* x_1^* *and* x_2^* *in totally opposite*

direction: the larger $|\mathcal{R}_p|$ is (note that $\mathcal{R}_p \leq 0$), the larger x_1^* is and the smaller x_2^* is. Noting that $\mathcal{R}_p$ accounts for the effect of the intervention or human behaviour changes and precaution induce the fraction $P_l(t)$ should help mitigate an epidemic and hence increase the final since s_∞. Unfortunately, we cannot exclude x_2^* at present and can only leave it as a conjecture, as stated below.

Conjecture 3.4.9 *If $\mathcal{R}_o > 1$, the final size s_∞ satisfies $s_\infty \leq \frac{1}{\mathcal{R}_o}$.*

It is difficult to prove the conjecture 3.4.9 here for a general case which is plausible. Fortunately, we can prove this conjecture for some special cases, as is done in Theorem 3.5.6 in the next Section.

Assume that the above conjecture is true, then the dependence of $s_\infty = x_1^*$ on $\mathcal{R}_p$ leads to the following lemma.

Theorem 3.4.10 *Suppose $\mathcal{R}_o > 1$ is fixed and $s_\infty \leq \frac{1}{\mathcal{R}_o}$, then the decrement of $\mathcal{R}_p$ leads to the increment of s_∞, i.e. the decrement of the final size r_∞. Indeed, the decrement of $\mathcal{R}_p$ from 0 to a negative value leads to the increment of s_∞ with the rate $1/|\mathcal{R}_o - \frac{1}{s_\infty}|$.*

Proof: See the Appendix.

We have mentioned that adopting the initial time to be at $t_0 = -\infty$ brings in some convenience in analysis. However, it has an inconvenience for numerical computation. For convenience simulations that will be done in Sect.3.5, we explore the scenario of adopting initial time at $t_0 = 0$ in the next subsection.

3.4.2   The case with initial condition given by $S(0)$:

Denote $L(0) := L_0$. According to the property of $P(L)$ described in the Assumption 3.8, we have

$$P_l(0) = P(L_0), \quad P_l(\infty) = P(L(\infty)) = P(0).$$

For $t \geq \tau$,

$$
\begin{aligned}
F_1(t) &= P_l(t)\left(\int_\tau^t \varphi(a)B(t-a)da + G(t)\right) \\
&= P_l(t)\left(\int_\tau^t -\frac{dS(t-a)}{dt}\varphi(a)da + G(t)\right) = P_l(t)\left(\int_\tau^t \frac{dS(t-a)}{da}\varphi(a)da + G(t)\right) \quad (3.54) \\
&= P_l(t)\left(S(0)\varphi(t) - S(t-\tau)\varphi(\tau) - \int_\tau^t S(t-a)\varphi'(a)da + G(t)\right)
\end{aligned}
$$

where

$$
\frac{dS(t-a)}{dt} = -\frac{dS(t-a)}{da} = -F(t-a)S(t-a).
$$

For $0 \leq t < \tau$

$$
F_2(t) = P_l(t)\left(\int_\tau^\infty \beta(a)u(0,a-t)\frac{\sigma(a)}{\sigma(a-t)}da\right) = P_l(t)G_\tau(t). \quad (3.55)
$$

In summary,

$$
F(t) = \begin{cases} F_1(t) & t \geq \tau; \\[2mm] F_2(t) & 0 \leq t < \tau. \end{cases} \quad (3.56)
$$

In addition

$$
\lim_{t\to\tau} F_2(t) = P_l(\tau)\left(\int_\tau^\infty \beta(a)u(0,a-\tau)\frac{\sigma(a)}{\sigma(a-\tau)}da\right) = P_l(\tau)G(\tau) = F_1(\tau), \quad (3.57)
$$

Thus, $F(t)$ is continuous.

Analogously to (3.40) and (3.39), define the following function

$$
\hat{\mathcal{H}}_0(t) := \begin{cases} \int_\tau^t \varphi(a)S(t-a)da = \int_0^{t-\tau} \varphi(t-\eta)S(\eta)d\eta & t \geq \tau \\[2mm] 0 & t < \tau \end{cases} \quad (3.58)
$$

$$
\check{\mathcal{R}}_p(t) := \int_\tau^t \hat{\mathcal{H}}_0(a)P_l'(a)da, \quad t \geq \tau \quad (3.59)
$$

$$
\hat{\mathcal{R}}_c(t) := N_0\int_\tau^t P_l(a)\varphi(a)da, \quad t \geq \tau \quad (3.60)
$$

$$\hat{\mathcal{R}}_g(t) := \int_\tau^t P_l(a)G(a)da, \quad t \geq \tau \tag{3.61}$$

$$\hat{\mathcal{R}}_{op}(t) := \frac{S_0}{N_0}\hat{\mathcal{R}}_c(t) + \check{\mathcal{R}}_p(t) + \hat{\mathcal{R}}_g(t), \quad t \geq \tau \tag{3.62}$$

and use the notations

$$\mathcal{H}_0 := \hat{\mathcal{H}}_0(\infty), \; \mathcal{R}_{pp} := \check{\mathcal{R}}_p(\infty), \; \mathcal{R}_c := \hat{\mathcal{R}}_c(\infty), \; \mathcal{R}_g := \hat{\mathcal{R}}_g(\infty), \; \mathcal{R}_{op} := \hat{\mathcal{R}}_{op}(\infty). \tag{3.63}$$

The following properties hold:

$$\hat{\mathcal{H}}_0(\tau) = 0, \quad \mathcal{H}_0 = \frac{\mathcal{R}_o s_\infty}{P(0)}, \quad P_l(\infty)\mathcal{H}_0 = \mathcal{R}_o s_\infty,$$

$$\hat{\mathcal{R}}_c(\tau) = 0, \; \check{\mathcal{R}}_p(\tau) = 0, \; \hat{\mathcal{R}}_g(\tau) = 0, \; \hat{\mathcal{R}}_{op}(\tau) = 0,$$

and

$$\frac{\mathcal{R}_c}{\mathcal{R}_o} = \frac{\int_\tau^\infty P_l(a)\varphi(a)da}{P(0)\int_\tau^\infty \varphi(a)da} \leq 1.$$

Differentiating $\mathcal{H}_0(t)$ leads to

$$\mathcal{H}_0'(t) = \varphi(t)S(0) + \int_\tau^t \varphi(a)\frac{dS(t-a)}{dt}da = \varphi(t)S(0) - \left(\frac{F_1(t)}{P(t)} - G(t)\right)$$
$$= \varphi(\tau)S(t-\tau) + \int_\tau^t \varphi'(a)S(t-a)da. \tag{3.64}$$

The sign of $\mathcal{H}_0'(t)$ is not clear. Unfortunately, it seems not direct to get the boundary of $\mathcal{R}_{pp}$, which is different from the $\mathcal{R}_p$ defined by (3.39) in Section 3.4.1.

Assume that $T > \tau > 0$,

$$Y(T; 0) := \ln\left(\frac{S(0)}{S(T)}\right) = \int_0^T F(t)dt = \left(\int_0^\tau F_2(t)dt + \int_\tau^T F_1(t)dt\right)$$
$$= Y(\tau; 0) + Y(T; \tau) = Y_0(T; 0) + \frac{S_0}{N_0}\hat{\mathcal{R}}_c(T) - P_l(T)\hat{\mathcal{H}}_0(T) + \check{\mathcal{R}}_p(t) \tag{3.65}$$

where

$$
\begin{aligned}
Y_0(T;0) : &= \int_0^T P_l(t)G^\circ(t)dt = \int_0^\tau P_l(t)G_\tau(t)dt + \int_\tau^T P_l(t)G(t)dt = Y(\tau;0) + \hat{\mathcal{R}}_g(t) \\
&= \int_\tau^T \varphi(a) \left\{ \int_0^a \frac{P_l(t)u_0(a-t)}{\sigma(a-t)} dt \right\} da + \int_T^\infty \varphi(a) \left\{ \int_0^T \frac{P_l(t)u_0(a-t)}{\sigma(a-t)} dt \right\} da \quad (3.66) \\
&= \int_\tau^\infty \varphi(a) \left\{ \int_0^T \frac{P_l(t)u_0(a-t)}{\sigma(a-t)} dt \right\} da - \int_\tau^T \varphi(a) \left\{ \int_a^T \frac{P_l(t)u_0(a-t)}{\sigma(a-t)} dt \right\} da
\end{aligned}
$$

with

$$
Y(\tau;0) = \int_0^\tau P_l(t)G_\tau(t)dt = \int_0^\tau P_l(t) \left\{ \int_{\tau-t}^\infty u_0(\eta) \frac{\varphi(\eta+t)}{\sigma(\eta)} d\eta \right\} dt. \quad (3.67)
$$

$Y_0(T;0)$ is the something that occurred before time $t = 0$.

Letting T tend to infinity, we get

$$
\begin{aligned}
Y_0(\infty;0) &= \int_\tau^\infty \varphi(a) \left\{ \int_0^\infty \frac{P_l(t)u_0(a-t)}{\sigma(a-t)} dt \right\} da - \int_\tau^\infty \varphi(a) \left\{ \int_a^\infty \frac{P_l(t)u_0(a-t)}{\sigma(a-t)} dt \right\} da \\
&= \int_\tau^\infty \varphi(a) \left\{ \int_0^a \frac{P_l(a-\eta)u_0(\eta)}{\sigma(\eta)} d\eta \right\} da. \quad (3.68)
\end{aligned}
$$

Furthermore,

$$
Y(t;\tau) = \hat{\mathcal{R}}_g(t) + \frac{S_0}{N_0}\hat{\mathcal{R}}_c(t) - P_l(t)\hat{\mathcal{H}}_0(t) + \check{\mathcal{R}}_p(t) = \hat{\mathcal{R}}_{op}(t) - P_l(t)\hat{\mathcal{H}}_0(t) \quad (3.69)
$$

and

$$
\begin{aligned}
Y(\infty;0) &= Y_0(\infty;0) + S(0) \int_\tau^\infty \varphi(a)P_l(a) \left\{ 1 - \frac{P_l(\infty)}{P_l(a)} e^{-Y(\infty;0)} \right\} da + \int_\tau^\infty \left\{ \int_\tau^t \varphi(a)S(t-a)da \right\} P_l'(t)dt \\
&= Y_0(\infty;0) + \mathcal{R}_{pp} + \frac{S_0}{N_0}\mathcal{R}_c - P_l(\infty)\mathcal{H}_0 \\
&= \mathcal{R}_{pp} + \left(1 - \frac{S_\infty}{N_0} \right)\mathcal{R}_o - \int_\tau^\infty \varphi(a) \left\{ P(0)N_0 - S(0)P(I(a)) - \int_0^a \frac{P_l(t)u_0(a-t)}{\sigma(a-t)} dt \right\} da \\
&= \mathcal{R}_{pp} + \frac{S_0}{N_0} \left(\frac{\mathcal{R}_c}{\mathcal{R}_o} - e^{-Y(\infty;0)} \right)\mathcal{R}_o + Y_0(\infty;0).
\end{aligned}
$$
$$(3.70)$$

Correspondingly,

$$
Y(\infty;\tau) = \mathcal{R}_{op} - \mathcal{R}_o \frac{S(\tau)}{N_0} e^{-Y(\infty;\tau)}. \quad (3.71)
$$

We collect our findings into the following theorem.

Theorem 3.4.11 *Given the initial condition $S(0) = S_0$, the final size relation of the model (3.30) is*

$$Y(\infty; 0) = \mathcal{R}_{pp} + \frac{S_0}{N_0}\left(\frac{\mathcal{R}_c}{\mathcal{R}_o} - e^{-Y(\infty;0)}\right)\mathcal{R}_o + Y_0(\infty; 0) \tag{3.72}$$

where $Y_0(\infty; 0)$ is given in (3.68). Moreover, the following form also holds:

$$Y(\infty; \tau) = \ln\left(\frac{s(\tau)}{s_\infty}\right) = \mathcal{R}_{op} - \mathcal{R}_o\frac{S(\tau)}{N_0}e^{-Y(\infty;\tau)};$$

equivalently,

$$r_\infty = \frac{Y(\infty; \tau)}{\mathcal{R}_o} - \left(\frac{\mathcal{R}_{op}}{\mathcal{R}_o} - 1\right)$$

where $\hat{\mathcal{R}}_{op}(t)$ and $\mathcal{R}_{op}$ are given in (3.62) and (3.63).

Remark 3.4.12 *In works of Brauer [7] and Breda, Diekmann et at [8], $Y_0(\infty; 0)$ is set to 0. Letting $Y_0(\infty; 0) \downarrow 0$ in (3.70), we obtain*

$$Y(\infty; 0) = \mathcal{R}_{pp} + \frac{S_0}{N_0}\left(\frac{\mathcal{R}_c}{\mathcal{R}_o} - e^{-Y(\infty;0)}\right)\mathcal{R}_o \leq \mathcal{R}_{pp} + \frac{S_0}{N_0}\left(1 - e^{-Y(\infty;0)}\right)\mathcal{R}_o, \tag{3.73}$$

Note that $\mathcal{R}_g$ reflects the contribution from those already in the infected class at time $t = 0$ to cumulative force-of-infection through the epidemic. When the initial fraction of infectors $\frac{N_0 - S_0}{N_0}$ is small, then $\mathcal{R}_g$ is small and hence, can be safely omitted. Thus,

$$\hat{\mathcal{R}}_{op}(t) - \check{\mathcal{R}}_p(t) \approx \hat{\mathcal{R}}_c(t) \tag{3.74}$$

which leads to the following result:

Corollary 3.4.13 *If $\frac{N_0 - S_0}{N_0}$ is small i.e. $\frac{N_0 - S_0}{N_0} = \epsilon$ for $0 < \epsilon \ll 1$, then*

$$Y(\infty; 0) = \mathcal{R}_{pp} + \left(\frac{\mathcal{R}_c}{\mathcal{R}_o} - e^{-Y(\infty;0)}\right)\mathcal{R}_o$$

or

$$-\ln(s_\infty) \approx \mathcal{R}_{pp} + \mathcal{R}_o\left[\frac{\mathcal{R}_c}{\mathcal{R}_o} - s_\infty\right] < \mathcal{R}_{pp} + \mathcal{R}_o r_\infty.$$

In what follows, we consider a special case of "survival" probability (probability of staying in the infected class). For such a special case, we can obtain more explicit and possibly more

useful results.

Consider

$$\gamma(a) = \begin{cases} \gamma_1, & a \in [0, \tau), \\ \gamma_2, & a \in [\tau, \infty). \end{cases} \tag{3.75}$$

With this piecewise constant recovery rate function $\gamma(a)$, the "survival" probability function $\sigma(a)$ becomes

$$\sigma(a) = \begin{cases} e^{-\gamma_1 a}, & a \in [0, \tau) \\ V_\tau e^{-\gamma_2 a}, & a \in [\tau, \infty) \end{cases} \tag{3.76}$$

where

$$V_\tau := V(\tau) = e^{(\gamma_2 - \gamma_1)\tau}$$

with

$$V(t) = e^{(\gamma_2 - \gamma_1)t}, \quad t \in [0, \tau]. \tag{3.77}$$

Accordingly,

$$T_E = \frac{1}{\gamma_1} - \frac{e^{-\gamma_1 \tau} \tau}{1 - e^{-\gamma_1 \tau}} \geq 0, \quad T_I = \frac{V_\tau(\gamma_2 \tau + 1)e^{-\gamma_2 \tau}}{\gamma_2 e^{-\gamma_1 \tau}} = \tau + \frac{1}{\gamma_2}.$$

Obviously,

$$\lim_{\gamma_1 \to 0} T_E = \frac{\tau}{2}, \quad \lim_{\gamma_1 \to \infty} T_E = 0.$$

Accordingly,

$$\varphi(a) = \begin{cases} 0, & a \in [0, \tau), \\ \beta(a)V_\tau e^{-\gamma_2 a}, & a \in [\tau, \infty). \end{cases} \tag{3.78}$$

By computing, we get

$$G_\tau(t) = \sigma(t) \int_{\tau-t}^{\tau} u_0(\eta)\beta(\eta + t)\frac{1}{V(\eta)}d\eta + \frac{\sigma(t)}{V_\tau} \int_{\tau}^{\infty} u_0(\eta)\beta(\eta + t)d\eta \quad t < \tau,$$
$$G(t) = \sigma(t) \int_{0}^{\tau} u_0(\eta)\beta(\eta + t)\frac{1}{V(\eta)}d\eta + \frac{\sigma(t)}{V_\tau} \int_{\tau}^{\infty} u_0(\eta)\beta(\eta + t)d\eta \quad t \geq \tau. \tag{3.79}$$

In reality, the initial age density function $u_0(a)$ can have various situations. In the sequel, we will discuss two situations: continuous and discretely centred.

Assumption 3.4.14 *The initial infection-age distribution $u_0(a)$ is continuous and satisfies*

$$\int_0^\tau u_0(a)da = E_0, \quad \int_\tau^\infty u_0(a)da = I_0. \tag{3.80}$$

For the continuous initial density function, we have

Lemma 3.4.15 *If $u_0(a)$ satisfies the Assumption 3.4.14, and*

$$\varphi(a) = \begin{cases} 0, & a \in [0, \tau), \\ \\ \beta V_\tau e^{-\gamma_2 a}, & a \in [\tau, \infty), \end{cases} \tag{3.81}$$

then the following final size inequality holds:

$$Y(\infty; \tau) \le \mathcal{R}_o \left(\max\left\{ \frac{1}{V_\tau}, 1 \right\} - \frac{S(\tau)}{N_0} e^{-Y(\infty;\tau)} \right) + \mathcal{R}_{pp}. \tag{3.82}$$

Especially, if $\gamma_1 = 0$,

$$Y(\infty; \tau) \le \mathcal{R}_o \left(1 - \frac{S(\tau)}{N_0} e^{-Y(\infty;\tau)} \right) + \mathcal{R}_{pp}. \tag{3.83}$$

Proof. See the Appendix.

Assumption 3.4.16 *The initial infection-age distribution $u_0(a)$ is discretely centred, meaning that $u_0(a) = 0$ for all a except for finite many values of a.*

For example, if all initial infectives have infection-age τ at $t = 0$, and all the infected but not infective have infection-age $\bar{a}_l \in (0, \tau)$ at $t = 0$ then the initial infection-age distribution is

$$u_0(a) = \begin{cases} I_0, & \text{for } a = \tau, \\ E_0, & \text{for } a = \bar{a}_l, \\ 0, & \text{Otherwise.} \end{cases} \tag{3.84}$$

The Assumption 3.4.16 on $u_0(a)$ indicates $G^\diamond(t) \equiv 0$, $\hat{\mathcal{R}}_g(t) \equiv 0$ and $Y_0(t; 0) \equiv 0$, $Y(\tau; 0) \equiv 0$. Further,

$$\hat{\mathcal{R}}_{op}(t) = \frac{S_0}{N_0} \hat{\mathcal{R}}_c(t) + \check{\mathcal{R}}_p(t), \tag{3.85}$$

Therefore, we have

$$Y(t; 0) = Y(t; \tau) = \hat{\mathcal{R}}_{op}(t) - P_l(t)\hat{\mathcal{H}}_0(t), \tag{3.86}$$

and thus, the limit of $Y(t, \tau)$ as $t \to \infty$ reduces to

$$
\begin{aligned}
Y(\infty; \tau) &= \mathcal{R}_{pp} + \left(1 - \frac{S(\tau)}{N_0}e^{-Y(\infty;\tau)}\right)\mathcal{R}_o - \int_\tau^\infty \varphi(a)\left\{P(0)N_0 - S(0)P(I(a))\right\}da \\
&= \mathcal{R}_{pp} + \left(\frac{S_0}{N_0}\frac{\mathcal{R}_c}{\mathcal{R}_o} - \frac{S(\tau)}{N_0}e^{-Y(\infty;\tau)}\right)\mathcal{R}_o \le \mathcal{R}_{pp} + \mathcal{R}_o\left(s_0 - s_\infty\right).
\end{aligned}
\tag{3.87}
$$

Summarizing the above, we have the following conclusion.

Lemma 3.4.17 *If $u_0(a)$ satisfies the Assumption 3.4.16, then the final size relation is,*

$$-\ln\left(s_\infty\right) = \mathcal{R}_{pp} + \frac{S_0}{N_0}\mathcal{R}_c - \mathcal{R}_o s_\infty - \ln\left(\frac{S(\tau)}{N_0}\right). \tag{3.88}$$

Remark 3.4.18 *If $P_l(t) \equiv 1$, then*

$$\check{\mathcal{R}}_p(t) \equiv 0, \ \hat{\mathcal{R}}_c(t) \equiv \hat{\mathcal{R}}_o(t).$$

Moreover,

$$\mathcal{R}_c = \mathcal{R}_o, \ \mathcal{R}_{op} = \frac{S_0}{N_0}\mathcal{R}_o,$$

and the final size equation (3.88) *for s_∞ is transformed to the formula* (3.1) *for r_∞.*

We point out that when $P_l(t) \le 1$, in our numerical simulations for the special models presented in the next section, we observe that the inequality $\mathcal{R}_{op} \le \mathcal{R}_o$ (corresponding to Theorem 3.4.4 in Subsect.3.4.1) always holds. See, Fig.3.3(b), Fig.3.7. Unfortunately, we are unable to prove this inequality for the general model.

3.5   Some applications and numerical simulations: reduction to DEs

In this section, we demonstrate the theoretical results established in the preceding sections by some particular simple examples that can reduce the model to DE models. We will also present

some numeric simulation results to illustrate the theoretical results.

We begin with, as a preparation, the following Lemma, which is parallel to the Lemma 3.4.3 but is only on the half line $[0, \infty)$.

Lemma 3.5.1 *Assume that*

(i) $g : [0, \infty) \to [0, \infty)$ *is differentiable, bounded and strictly increasing with $g(0) = 0$ (*

  hence $g(t) \in [0, g_M)$ where $g_M := \sup_{t \in [0, \infty)} g(t) = g(\infty)$);

(ii) $f : [0, \infty) \to [0, \infty)$ *is differentiable and bounded, and satisfies $f(t) \le f(\infty) =: f_M < \infty$.*

Then,

$$0 \le \int_0^\infty f'(t)g(t)dt \le g_M f_M. \tag{3.89}$$

Proof: See the Appendix.

We now choose two particular simple forms for the kernel function $\phi(a)$ to explore the impact of $P_l(t)$ on the final size, reflected by the relation between s_∞ and $\mathcal{R}_P$ or $\mathcal{R}_{op}$.

The first one is the one given in (3.78) with $\beta(a) = \beta$ (constant), which is rewritten below as

$$\varphi_1(a) := \begin{cases} 0, & a \in [0, \tau) \\ \beta e^{(\gamma_2 - \gamma_1)\tau} e^{-\gamma_2 a}, & a \in [\tau, \infty), \end{cases} \tag{3.90}$$

For this kernel, the corresponding basic reproductive number is

$$\mathcal{R}_o = P(0)N_0 \int_\tau^\infty \varphi_1(a)da = \frac{\varphi_1(\tau)P(0)N_0}{\gamma_2} = \{P_L(0)N_0\beta\sigma(\tau)\}(T_I - \tau). \tag{3.91}$$

Remark 3.5.2 *The special case of $\gamma_1 \to 0$ corresponds to the scenario that a pre-infectious individual cannot recover before getting infectious. In this case, the latent period τ makes no difference in $\mathcal{R}_o$.*

The second form is taken from [9] and is given by

$$\varphi_2(a) := \begin{cases} \beta \frac{\epsilon}{\epsilon - \bar{\gamma}}(e^{-\bar{\gamma}a} - e^{-\epsilon a}), & \text{if} \quad \bar{\gamma} \ne \epsilon, \\ \beta\bar{\gamma}ae^{-\bar{\gamma}a}, & \text{if} \quad \bar{\gamma} = \epsilon, \end{cases} \quad a \in (0, \infty). \tag{3.92}$$

With this kernel, the basic reproduction number of the model is

$$\mathcal{R}_o = \frac{\beta P(0) N_0}{\bar{\gamma}}.$$

(3.93)

3.5.1 The case $\varphi(a) = \varphi_1(a)$

For this kernel, $u(t, \tau)$ is determined to be

$$u(t, \tau) = B(t - \tau)\sigma(\tau) = \varphi(\tau) S_P(t - \tau) I(t - \tau).$$

(3.94)

The formulas (3.42) and (3.43) lead to

$$-\beta I(t) = \varphi(\tau) S(t - \tau) - \gamma_2 \hat{\mathcal{H}}(t)$$

(3.95)

which establishes an explicit expression for $\hat{\mathcal{H}}(t)$ as

$$\hat{\mathcal{H}}(t) = \frac{\varphi(\tau)}{\gamma_2} S(t - \tau) + \frac{\beta}{\gamma_2} I(t)$$

(3.96)

Also, the model (3.30) for this kernel reduces to

$$\begin{cases} S'(t) = -\beta I(t) S_P(t), \\[2mm] E'(t) = \beta I(t) S_P(t) - \gamma_1 E - \varphi(\tau) I(t - \tau) S_P(t - \tau), \\[2mm] I'(t) = -\gamma_2 I + \varphi(\tau) I(t - \tau) S_P(t - \tau), \\[2mm] R'(t) = \gamma_1 E + \gamma_2 I. \end{cases}$$

(3.97)

In summary, the integrated model reads

$$\begin{cases} S'(t) = -\beta I P_l(t) S, \\[2mm] E'(t) = \beta I(t) S_P(t) - \gamma_1 E - \varphi(\tau) I(t - \tau) S_P(t - \tau), \\[2mm] I'(t) = -\gamma_2 I + \beta e^{-\gamma_1 \tau} I(t - \tau) P(I(t - \tau)) S(t - \tau), \\[2mm] \hat{\mathcal{R}}'_P(t) = P'_l(t) \hat{\mathcal{H}}(t). \end{cases}$$

(3.98)

For the scenario of $t_0 = -\infty$, the initial condition is

$$(S(-\infty), E(-\infty), I(-\infty), \hat{\mathcal{H}}(-\infty), \hat{\mathcal{R}}_p(-\infty)) = (N_0, 0, 0, \mathcal{R}_o/P(0), 0).$$

and for the scenario $t_0 = 0$, the initial condition is

$$S(0) = S_0, \quad E(0) = E_0 = \int_{-\infty}^{0} E'(t)dt, \quad I(0) = I_0 = \int_{-\infty}^{0} I'(t)dt \tag{3.99}$$

and

$$\hat{\mathcal{R}}_p(0) = \int_{-\infty}^{0} \hat{\mathcal{H}}(t)P_l'(t)dt \tag{3.100}$$

which is difficult to determine.

Although the theoretical results for the scenario that t starts from the infinite past described in Subsection 3.4.1 are more brief and straightforward, it is not easy to verify these conclusions by numerical simulation.

In the rest of this section, we will choose $P(L) = e^{-hL}$ with $h \geq 0$ and use the results in subsection 3.4.2 to show the numerical results.

3.5.1.1 The case $\tau > 0$ with $t_0 = 0$

We consider a particular initial density specified below.

$$u_0(a) = \begin{cases} \dfrac{E_0}{\tau} & a \in [0, \tau), \ a \in [0, \tau), \\ \text{integral on } [\tau, \infty) \text{ with contraint } \displaystyle\int_{\tau}^{\infty} u_0(\eta)d\eta = I_0. \end{cases} \tag{3.101}$$

Then for $t \in [0, \tau)$,

$$u(t, \tau) = \frac{E_0}{\tau}\sigma(t) = \frac{E_0}{\tau}e^{-\gamma_1 t}, \tag{3.102}$$

and, accordingly

$$G_\tau(t) = \begin{cases} \dfrac{\beta\sigma(t)}{V_\tau}\left(\dfrac{1-V(t)}{\gamma_1-\gamma_2}\dfrac{E_0}{\tau} + I_0\right) & \gamma_1 \neq \gamma_2, \\ \beta\sigma(t)\left(\dfrac{E_0 t}{\tau} + I_0\right) & \gamma_1 = \gamma_2. \end{cases} \tag{3.103}$$

$$G(t) = \begin{cases} \frac{\beta\sigma(t)}{V_\tau}\left(\frac{1-V_\tau}{\gamma_1-\gamma_2}\frac{E_0}{\tau} + I_0\right) & \gamma_1 \neq \gamma_2, \\[2mm] \beta\sigma(t)(E_0 + I_0) & \gamma_1 = \gamma_2. \end{cases} \tag{3.104}$$

$$\hat{\mathcal{R}}_g(t) = \begin{cases} \frac{1}{N_0 V_\tau}\left(\frac{1-V_\tau}{\gamma_1-\gamma_2}\frac{E_0}{\tau} + I_0\right)\hat{\mathcal{R}}_c(t) & \gamma_1 \neq \gamma_2, \\[2mm] \frac{E_0+I_0}{N}\hat{\mathcal{R}}_c(t) & \gamma_1 = \gamma_2. \end{cases} \tag{3.105}$$

The disease dynamics for $t \in [0, \tau]$ is then governed by the ODE system

$$\begin{cases} S'(t) = -\beta I(t)S_P(t), \\[2mm] E'(t) = \beta I(t)S_P(t) - \gamma_1 E - \frac{E_0}{\tau}e^{-\gamma_1 t}, \\[2mm] I'(t) = -\gamma_2 I + \frac{E_0}{\tau}e^{-\gamma_1 t}. \end{cases} \tag{3.106}$$

For $t \geq \tau$,

$$u(t, \tau) = \sigma(\tau)B(t - \tau) = \beta e^{-\gamma_1 \tau}I(t - \tau)S_P(t - \tau).$$

Plugging the above calculations into (3.97), we obtain

$$\begin{cases} S'(t) = -\beta I(t)S_P(t), \\[2mm] E'(t) = \beta I(t)S_P(t) - \gamma_1 E - \beta e^{-\gamma_1 \tau}I(t - \tau)S_P(t - \tau), \\[2mm] I'(t) = -\gamma_2 I + \beta e^{-\gamma_1 \tau}I(t - \tau)S_P(t - \tau), & t \in [\tau, \infty). \\[2mm] \hat{\mathcal{H}}_0'(t) = \varphi(\tau)S(t - \tau) - \gamma_2 \hat{\mathcal{H}}_0(t), \\[2mm] \mathcal{R}_{op}'(t) = \left(\hat{\mathcal{H}}_0'(t) + \hat{\mathcal{H}}_0(t)\frac{P_l'(t)}{P_l(t)} + \beta I(t)\right)P_l(t). \end{cases} \tag{3.107}$$

with the new and translated initial condition $(S(\tau), E(\tau), I(\tau), \hat{\mathcal{H}}_0(\tau), \hat{\mathcal{R}}_{op}(\tau)) = (S_\tau, E_\tau, I_\tau, 0, 0)$ with $(S(\tau), E(\tau), I(\tau)) = (S_\tau, E_\tau, I_\tau)$ determined by solving the ODE system (3.106) on $[0, \tau]$.

By simulating the system (3.107) with the particular $L(t) = I(t)$, it is easy to verify the FSR given in the Theorem numerically 3.4.11. Fig. 3.3 are the numeric results with the parameter values specified in the caption. From Fig.3.3, we observe the following.

(a). If $\mathcal{R}_o$ is fixed, an increase in h from 0 increases s_∞ (i.e. decreases the final size r_∞); for

fixed $h \geq 0$, s_∞ decrease as $\mathcal{R}_o$ increase; Further, $s_\infty < 1/\mathcal{R}_o$ holds, which confirms the Conjecture 3.4.9.

(b). $\mathcal{R}_o - \mathcal{R}_{op} > 0$. This inequality can be theoretically proved for the SIR model (3.108) (the case of $\tau = 0$), see Lemma 3.5.4 below. In addition, s_∞ decreases (or the final size r_∞ increases) as $\mathcal{R}_{op}$ increases.

Unfortunately, all the rigorous proof of these aforementioned numerical results (shown in Fig.3.3) is not easy for the model (3.107).
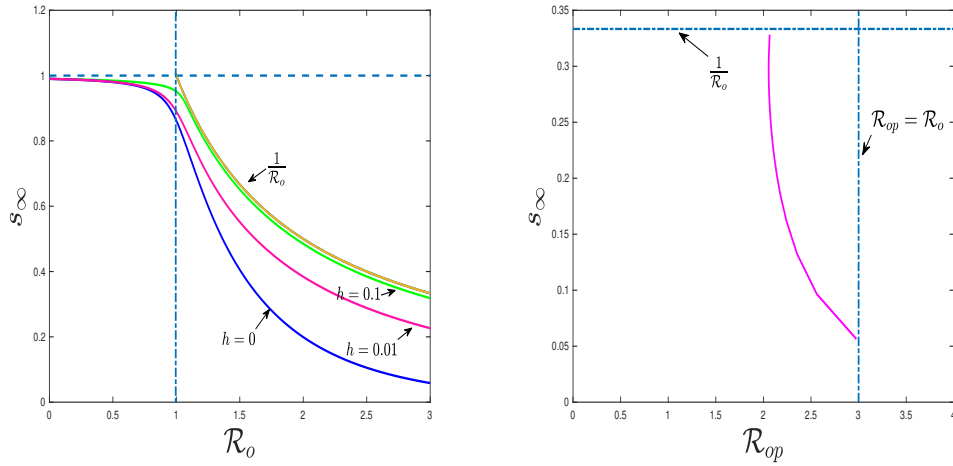
Figure 3.3: (a). $\mathcal{R}_o \in [0, 3]$; (b). $h \in [0, 1]$, $\mathcal{R}_o = 3$. Other parameters are $\tau = 4, \gamma_1 = 0, \gamma_2 = \frac{1}{7}, N_0 = 300, E_0 = 5, I_0 = 5, R(0) = 0$.

3.5.1.2 The special case $\tau \equiv 0$ and $t_0 = 0$

If $\tau \equiv 0$, the model (3.30) reduces to the SIR model

$$
\begin{cases}
S'(t) = -\beta I(t) S_P(t), \\[2mm]
I'(t) = -\bar{\gamma} I + \beta I(t) S_P(t), \\[2mm]
R'(t) = \bar{\gamma} I.
\end{cases}
\tag{3.108}
$$

for which the basic reproduction number is

$$
\mathcal{R}_o = \frac{\beta P(0) N_0}{\bar{\gamma}}.
\tag{3.109}
$$

By simple calculations, we can obtain the explicit formulas for functions $\hat{\mathcal{H}}_0(t)$, $\hat{\mathcal{R}}_c(t)$ and $\check{\mathcal{R}}_p(t)$ as

$$\begin{aligned}
\hat{\mathcal{H}}_0(t) &= \frac{\beta}{\gamma_2} S(t) + \frac{\beta}{\gamma_2} I(t) - \frac{\varphi(t)N_0}{\gamma_2} \\
&= \frac{\mathcal{R}_o}{P(0)}(1 - \sigma(t)) - \frac{\mathcal{R}_o}{P(0)} r(t)
\end{aligned} \tag{3.110}$$

with $\hat{\mathcal{H}}_0(0) = 0$ and $\hat{\mathcal{H}}_0(\infty) = \mathcal{R}_o s_\infty / P(0)$;

$$\begin{aligned}
\hat{\mathcal{R}}_c(t) &= \beta N_0 \int_0^t \sigma(a)P_l(a)da \\
&= -\frac{\mathcal{R}_o}{P(0)} \int_0^t \sigma'(a)P_l(a)da;
\end{aligned} \tag{3.111}$$

and

$$\check{\mathcal{R}}_p(t) = -\hat{\mathcal{R}}_c(t) + \frac{\mathcal{R}_o}{P(0)}P_l(t)(1 - \sigma(t)) - \frac{\mathcal{R}_o}{P(0)} \int_0^t r(t)P'_l(t)dt. \tag{3.112}$$

Corresponding to the Assumption 3.4.14, we give the following assumption:

Assumption 3.5.3 $\int_0^\infty u(0, \eta)d\eta = I_0$.

Under Assumption 3.5.3, we obtain

$$G^\diamond(t) = G(t) = \beta \int_t^\infty u(0, a - t)e^{-\gamma_2 t}da = \beta e^{-\gamma_2 t}\int_0^\infty u_0(\eta)d\eta = \beta e^{-\gamma_2 t}I_0 = \varphi(t)I_0, \tag{3.113}$$

where $G(t)$ is the total infectivity of the infected population at $t = 0$ at time t.

$$Y_0(t; 0) = \hat{\mathcal{R}}_g(t) = \frac{I_0}{N_0}\hat{\mathcal{R}}_c(t). \tag{3.114}$$

Then

$$\begin{aligned}
\hat{\mathcal{R}}_{op}(t) &= \hat{\mathcal{R}}_c(t) + \check{\mathcal{R}}_p(t) \\
&= \frac{\mathcal{R}_o}{P(0)}P_l(t)(1 - \sigma(t)) - \frac{\mathcal{R}_o}{P(0)} \int_0^t r(t)P'_l(t)dt
\end{aligned} \tag{3.115}$$

Using the Lemma 3.5.1, we get

$$\int_0^t r(t)P'_l(t)dt > 0.$$

Then the following result holds

Lemma 3.5.4 *For model (3.108) with $P_l(t) = P(I(t))$ under the assumption 3.5.3, $\mathcal{R}_o - \mathcal{R}_{op} > 0$ holds, see Fig.3.7.*

According to the theorem (3.4.11), we get the following inequality of the FSR of the model (3.108)

$$- \ln(s_\infty) = \mathcal{R}_o r(\infty) + (\mathcal{R}_{op} - \mathcal{R}_o) - \ln(s_0) < \mathcal{R}_o r(\infty) - \ln(s_0), \tag{3.116}$$

which includes

$$\frac{Y(\infty, 0)}{\mathcal{R}_o} < r_\infty. \tag{3.117}$$

Based on numerical simulation, system (3.108) in this Subsect.3.5.1.2 exhibits similar qualitative features to system (3.108) in Subsect.3.5.2. For brevity, the related figures and results are omitted here and included in the Appendix.

3.5.2 The case $\varphi(a) = \varphi_2(a)$

If choosing the kernel $\varphi(a) = \varphi_2(a)$ and letting $\tau = 0$, the model (3.29) reduces to (Details in the Appendix)

$$\begin{cases} S'(t) = -\beta S_P I, \\ \\ E'(t) = -\epsilon E + \beta S_P I, \\ \\ I'(t) = -\bar{\gamma} I + \epsilon E, \\ \\ R'(t) = \bar{\gamma} I. \end{cases} \tag{3.118}$$

For this kernel, the basic reproduction number $\mathcal{R}_o$ of (3.30) has the same expression as (3.91) with $\tau = 0$.

Functions $\hat{\mathcal{H}}(t)$ and $\hat{\mathcal{R}}_p(t)$ have the following forms:

$$\hat{\mathcal{H}}(t) = -\frac{\beta}{\bar{\gamma}} R(t), \quad \hat{\mathcal{R}}_p(t) = \int_0^t P_l'(a) \hat{\mathcal{H}}(a) da. \tag{3.119}$$

Since

$$S_P(t) = P_l(t) S \leq S(t) \leq S(0), \quad t \geq 0,$$

and accordingly,

$$s_P(t) := \frac{S_P(t)}{N_0} \le s(t) \le s(0), \quad t \ge 0.$$

Remark 3.5.5 *It needs to be emphasized that we choose $L(t) = U(t)$ in this Subsect. 3.5.2 to make the following theoretical analysis much easier. However, in practice, U_0 is not easy to assess due to the difficulty of obtaining and tracking E_0 in the early stages of epidemics.*

Calculations lead to

$$U'(t) = P_l(t)S\beta I - \bar{\gamma}I = \mathcal{R}_o\left(\frac{s_P(t)}{P(0)} - \frac{1}{\mathcal{R}_o}\right)\bar{\gamma}I \le \mathcal{R}_o\left(\frac{s_P(t)}{P(0)} - \frac{1}{\mathcal{R}_o}\right)\bar{\gamma}U, \tag{3.120}$$

$$U''(t) = \mathcal{R}_o\left(\frac{s_P'(t)}{P(0)}\right)\bar{\gamma}I + \mathcal{R}_o\left(\frac{s_P(t)}{P(0)} - \frac{1}{\mathcal{R}_o}\right)\bar{\gamma}I'. \tag{3.121}$$

With the above preparation for this kernel, we then the following results for the reduced model (3.118).

Theorem 3.5.6 *For system (3.118), s_∞ satisfies $s_\infty \le \frac{1}{\mathcal{R}_o}$. Moreover, if $P_l(t) = P(U)$, $s_\infty < \frac{1}{\mathcal{R}_o}$.*

Proof: See the Appendix.

Moreover,

Lemma 3.5.7 *If $U'(t_m) = 0$, then $U''(t_m) \le 0$.*

Proof: See the Appendix.

Remark 3.5.8 *Lemma 3.5.7 indicates that if U has a critical point, it must be a local maximum. Further, the continuous differentiability of U implies that $U(t)$ has, at most, a single peak $U(t_m)$. In other words, if $U'(0) < 0$ then $U_{max} = U(0)$; if $U'(0) > 0$, then there exits $t_m > 0$ such that $U'(t_m) = 0$ and $U_{max} = U(t_m)$ and t_m is the precisely the time at which the fraction of infected individuals reaches the peak $U_{\max}$. Unfortunately, we are unable to determine the peak size analytically.*

In summary, we have the following Theorem.

Theorem 3.5.9 (Peak time and size) *Consider* $P_l(t) = P(U)$. *Then,* $U'(0) \geq 0$ *(which is equivalent to* $P'_l(0) \geq 0$*) if and only if* $s_P(0) \geq \frac{P(0)}{\mathcal{R}_o}$. *Moreover,*

(I) *When* $\mathcal{R}_o < 1$, *then* $U'(t) < 0$ *for all* $t > 0$, *the maximum value of* $U(t)$ *is attained at* $t_0 = 0$, *that is,* $U_{max} = I(0) + E(0)$;

(II) *When* $\mathcal{R}_o > 1$, *then*

    (a) *if* $s_P(0) > \frac{P(0)}{\mathcal{R}_o}$, *there exists a* $t_m > 0$ *such that* $U'(t_m) = 0$. *Accordingly,* $s(t_m) = \frac{1}{\mathcal{R}_o} \frac{P(0)}{P(U_{max})}$ *where* $U_{max} = U(t_m)$;

    (b) *if* $s_P(0) < \frac{P(0)}{\mathcal{R}_o}$, *then* $U'(t) < 0$ *for all* $t > 0$, *the maximum value occurs at* $t_0 = 0$, *that is,* $U_{max} = I_0 + E_0$.

Remark 3.5.10 *(See Fig.3.5) From Theorem* (3.5.9)-*(II), since* t_m *is the threshold for* $U'(t)$, *the value*

$$S(t_m) = \frac{P(0)}{P(U_{max})} \frac{N_0}{\mathcal{R}_o}$$

is indeed the threshold for herd immunity (HIT), i.e. the value of S under which U can no longer increase [20]. Thus, the involvement of $P_l(t)$ *leads to the increment of herd immunity threshold compared with the classical herd immunity level* $N_0/\mathcal{R}_o$, *since* $P(0)/P(U_{max}) > 1$.

The following is a direct result of Lemma 3.5.1.

Theorem 3.5.11 *(See. Fig.3.4-(b)) For* (3.118), $\mathcal{R}_p = \hat{\mathcal{R}}_p(\infty) < 0$ *always holds.*

Recall that the CFI over period $[0, t]$ of the model (3.118) is

$$Y(t; 0) = \frac{\beta P_l(t) R(t)}{\bar{\gamma}} + \hat{\mathcal{R}}_p(t). \tag{3.122}$$

Moreover, pluging $t = \infty$ into the above fomula (3.122) deduces the FSR of the model (3.118) to

$$\ln \frac{s_0}{s_\infty} = \mathcal{R}_o r(\infty) + \mathcal{R}_p.$$

which can be rewritten as

$$\ln s_\infty = \mathcal{R}_0 s_\infty + [\ln s_0 - \mathcal{R}_o - \mathcal{R}_p].$$

The equation clearly shows that s_∞ is decreasing in $\mathcal{R}_p$, implying the NPIs or behaviour changes reflected by the fraction function $P_l(t) = P(L(t))$ can increase the final size s_∞ provided that $\mathcal{R}_p$ is no less than the minimal (critical) value $\mathcal{R}_p^{cr}$, where

$$\mathcal{R}_p^{cr} = 1 + \ln s_0 - \mathcal{R}_0 + \ln \mathcal{R}_0.$$

The above conclusion is obtained for (3.118) by analyzing its CFI, similar results have also been discussed in Theorem 3.5.6 and Theorem 3.4.10 for other cases of the model. In summary, we have the following theorem.

Theorem 3.5.12 *(See Fig. 3.4-(b)) For any fixed $\mathcal{R}_o > 1$ is fixed, $\mathcal{R}_p$ decrease if and only if s_∞ increase (i.e. r_∞ decrease).*

We point out that, by its definition, $\mathcal{R}_p$ (hence s_∞) actually depends on several things, including $\mathcal{R}_0$ through the kernel $\phi(a)$ and $P_l(t) = P(L(t))$. However, the dependence is generally very difficult to analyze. In Fig.3.4 and Fig.3.5, we present some numerical results to demonstrate such dependence. To this end, we choose $L(t) = I(t)$ and $P(L) = e^{-hL} = e^{-hI}$. Fig.3.4(a) shows the relationships between h and $\mathcal{R}_p$. By altering the non-biological parameter h, one can adjust the efficacy of NPIS (i.e. the value of $\mathcal{R}_p$). Fig.3.5 focuses the impact of the basic reproduction number $\mathcal{R}_o$ on both $\mathcal{R}_p$ and s_∞. The fig.3.5(b) parallel and also is similar to fig.3.3(a) and the explanation of fig.3.5(b) is omitted here. In the Theorem 3.5.9 and Remark 3.5.10, the theoretical results for the peak epidemic size of the case $L = U$ are presented. In practice, the number of the exposed class E is sometimes more difficult to collect precisely, whereas the mathematical analysis of the case $L = I$ is relatively complicated.

Fig. 3.6 aims to present some numeric simulations on how the final size and peak size are impacted by the choice of $L(t)$ and h in the fraction function $P(L) = e^{-hL}$. The comparison is between $L = I$ and $L = U$ with different values of h. The numeric results seem to suggest that, in controlling peak size and final size, the effectiveness of Non-pharmaceutical and Non-biological precautions when applying only the number of $I(t)$ to measure the severity level $L(t)$ is not as good as when considering all infected numbers (i.e. $U(t)$), but it is still better than no precaution (i.e. $h = 0$).
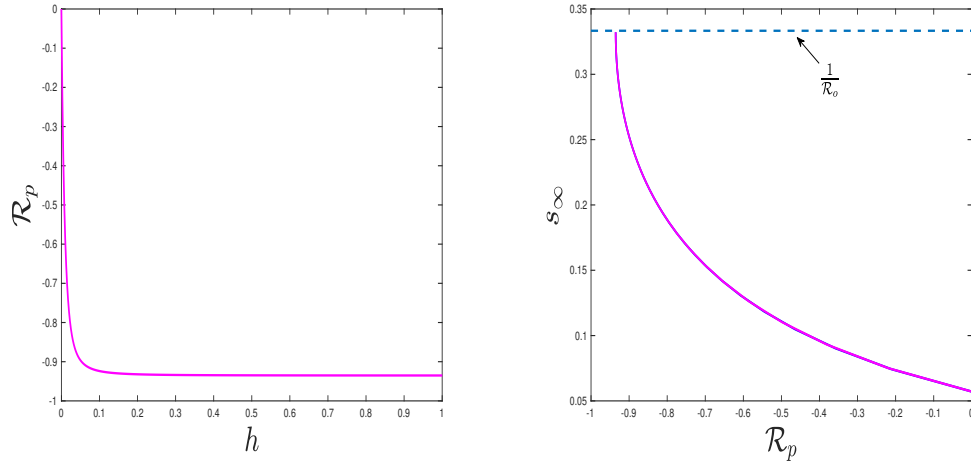
Figure 3.4: $\mathcal{R}_o = 3, \bar{\gamma} = \frac{1}{20}, \epsilon = \frac{1}{14}, h \in [0, 5]$ and $N_0 = 300, E_0 = 5, I_0 = 5, R(0) = 0$.
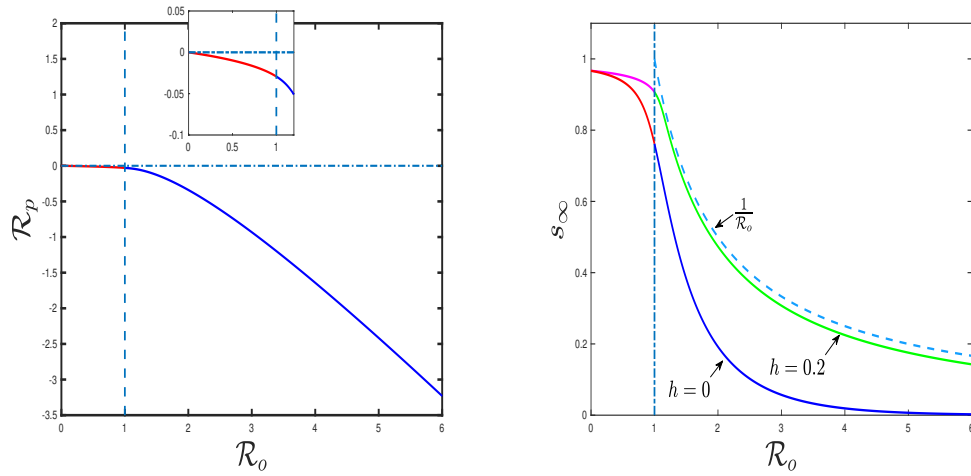

Figure 3.5: $\mathcal{R}_o \in [0, 6], \bar{\gamma} = \frac{1}{20}, \epsilon = \frac{1}{14}, h = 0.2$ and $N_0 = 300, E_0 = 5, I_0 = 5, R(0) = 0$.
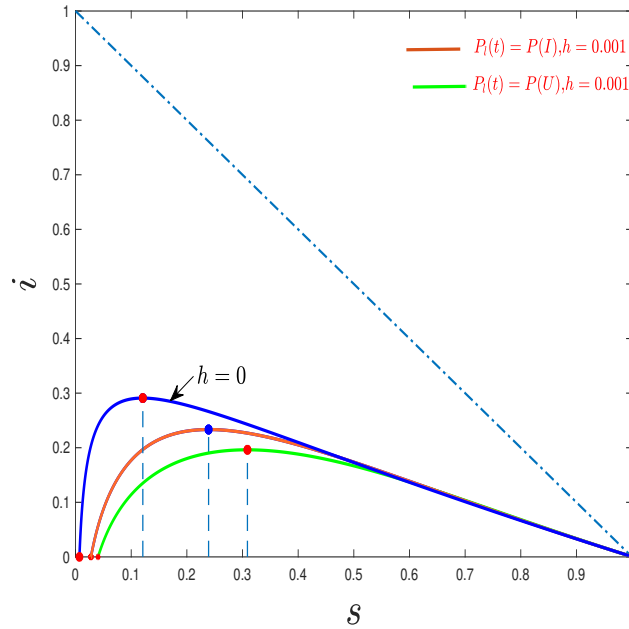
Figure 3.6: $\mathcal{R}_o = 5, \bar{\gamma} = \frac{1}{20}, \epsilon = \frac{1}{14}$ and $N_0 = 2000, E_0 = 5, I_0 = 5, R(0) = 0$.

3.6 Discussion

In this paper, we aim to explore the final epidemic size for general epidemics characterized by infection age, focusing on how NPIs and behaviour changes influence the epidemic size. We employ the renewal equation approach, the power of which seems to be under-estimated in literature. Our results show that the final size relation depends on both $\mathcal{R}_o$ and $\mathcal{R}_p$ (or $\mathcal{R}_{op}$) with $\mathcal{R}_p$ (or $\mathcal{R}_{op}$) accounting for the effect of the NPIs and/or behaviour changes. In addition, by the approach in [8], we figure out the relation of the final size to the CFOI. As described in Remark 3.4.5, NPIs and/or behaviour change alters this relation via introducing the parameter $\mathcal{R}_p$, in contrast to the results in [8].

The assumption that time extends along the entire axis ($t_0 = -\infty$) simplifies mathematical analysis but makes numerical simulations impractical. By adopting this extension, some researchers are able to establish relatively more theoretical conclusions, but they can obtain less quantitative simulating results [7, 8, 16]. Limiting the time starting from a finite value ($t_0 > -\infty$), we repeat, in Subsection 3.4.2, the derivations and inductions in great detail to overcome the stumbling block in numerical simulations. Subsection 3.4.2 places more empha-

sis on the contribution of initial states than Subsection 3.4.1, through incorporating the term $\hat{\mathcal{R}}_g(t)$ or $\mathcal{R}_g$.

The theorems on the FSR in Section 3.4 provide theoretical support for estimating the magnitude of the epidemic through the estimate of the basic reproduction number $\mathcal{R}_o$ as well as the intervention effect term $\mathcal{R}_p$ which is not easy to determine. In Section 3.5, by choosing two particular and simple forms for kernel function $\varphi(a)$, we are led to a differential equation for $\hat{\mathcal{R}}_p(t)$ or $\hat{\mathcal{R}}_{op}(t)$ (see the last equation of (3.98) or 3.107), which can be analyzed and simulated easily. Further research is needed to analyze $\mathcal{R}_p$ so that more information can be gathered on the impact of fraction function $P_l(t) = P(L(t))$ on the final size.

As mentioned in the previous sections, the Conjecture 3.4.9 is analytically challenging, although its conclusion is essential and biologically meaningful. It is certainly worth working on. Thus, an immediate research project is to establish sufficient conditions under which the Conjecture 3.4.9 holds.

Section 3.5 shows that for some special cases as dealt with in Subsections 3.108 and 3.5.2, the framework of the general infection-age model (3.30) reduced to simpler forms that can offer some information about the herd immunity threshold (HIT), the peak size (PZ) and the calendar time t_m. However, for the general case of the general model (3.30), it is very challenging to analyze the HIT and PZ. More mathematical techniques are needed to theoretically obtain information on the HIT and PZ for the general model (3.30).

In summary, this work provides a general and plausible framework for estimating the final epidemic size when considering Non-pharmaceutical interventions and/or psychological effects. From a mathematical point of view, this work generalizes the force of infection described in [2] to reflect the above-mentioned effects and derives the corresponding REs. As is seen, incorporating Non-pharmaceutical interventions and/or psychological effects results in REs with *time-varying kernels*. It is the time-varying nature of the kernel that makes the model novel, the analysis more difficult, and the results more interesting both mathematically and biologically.

Appendix

The derivation of model(3.30)

$$
\begin{aligned}
E'(t) = \int_0^\tau \frac{\partial u(t,a)}{\partial t} da &= \int_0^\tau -\gamma(a)u(t,a) - \frac{\partial u(t,a)}{\partial a} da \\
&= -\int_0^\tau \gamma(a)u(t,a)da - u(t,\tau) + u(t,0) \\
&= -\int_0^\tau \gamma(a)u(t,a)da - u(t,\tau) + B(t) \\
I'(t) = \int_\tau^\infty \frac{\partial u(t,a)}{\partial t} da &= -\int_\tau^\infty \gamma(a)u(t,a)da - \int_\tau^\infty \frac{\partial u(t,a)}{\partial a} da \\
&= -\int_\tau^\infty \gamma(a)u(t,a)da - u(t,\infty) + u(t,\tau) \\
&= -\int_\tau^\infty \gamma(a)u(t,a)da + u(t,\tau).
\end{aligned}
\tag{3.123}
$$

Proof of Theorem 3.3.2

Proof

$$
\frac{d(S+U)}{dt} = -\int_0^\infty \gamma(a)u(t,a)da = -\int_0^\infty B(t-a)\sigma(a)\gamma(a)da
\tag{3.124}
$$

which is strictly negative, $S+U$ is strictly decreasing and positive. Therefore, $\lim_{t\to\infty} \frac{d(S+U)}{dt} = 0$, that is, $U(\infty) = 0$. Further, since $I(t), E(t) > 0$ for all $t < \infty$, $E(\infty) = 0$, $I(\infty) = 0$.

Proof of Lemma 3.4.15

Proof Under the Assumption 3.4.14 of $u_0(a)$, then

$$
\min\left\{\frac{1}{V_\tau}, 1\right\} G_M(t) \le G(t) \le \max\left\{\frac{1}{V_\tau}, 1\right\} G_M(t)
$$

where

$$
G_M(t) = (E_0 + I_0)\varphi(t)\frac{\int_0^\infty u_0(\eta)\frac{\beta(\eta+t)}{\beta(t)} d\eta}{\int_0^\infty u_0(\eta)d\eta}, \quad t \ge \tau.
\tag{3.125}
$$

Moreover, particularly, if $\beta(a) \equiv \beta$,

$$
\int_\tau^t P_l(a)G_M(a)da = \frac{E_0 + I_0}{N_0}\hat{\mathcal{R}}_c(t),
\tag{3.126}
$$

$$\min\left\{\frac{1}{V_\tau},1\right\}\frac{N_0-S_0}{N_0}\hat{\mathcal{R}}_c(t) \le \hat{\mathcal{R}}_g(t) \le \max\left\{\frac{1}{V_\tau},1\right\}\frac{N_0-S_0}{N_0}\hat{\mathcal{R}}_c(t) \tag{3.127}$$

and

$$\begin{aligned}
\min\left\{\frac{1}{V_\tau},1\right\}\hat{\mathcal{R}}_c(t) &\le \left(\min\left\{\frac{1}{V_\tau},1\right\}\frac{E_0+I_0}{N_0}+\frac{S_0}{N_0}\right)\hat{\mathcal{R}}_c(t) \le \hat{\mathcal{R}}_{op}(t)-\hat{\mathcal{R}}_p(t) \\
&\le \left(\max\left\{\frac{1}{V_\tau},1\right\}\frac{E_0+I_0}{N_0}+\frac{S_0}{N_0}\right)\hat{\mathcal{R}}_c(t) \le \max\left\{\frac{1}{V_\tau},1\right\}\hat{\mathcal{R}}_c(t) \le \max\left\{\frac{1}{V_\tau},1\right\}\hat{\mathcal{R}}_o(t).
\end{aligned} \tag{3.128}$$

This inequality implies (3.82).

Proof of Lemma 3.4.3

Proof For any $a < b$, with the help of the second mean value theorem for integrals[12, 29, 30], there exists $\xi_{a,b} \in [a,b]$, which depend on the choice of a,b, such that

$$\begin{aligned}
\int_a^b g(t)f'(t)dt &= g(a)\int_a^{\xi_{a,b}} f'(t)dt + g(b)\int_{\xi_{a,b}}^b f'(t)dt \\
&= g(a)\left(f(\xi_{a,b})-f(a)\right) + g(b)\left(f(b)-f(\xi_{a,b})\right).
\end{aligned} \tag{3.129}$$

Further, we get the following inequalities:

$$\begin{aligned}
&g(a)\left(f(\xi_{a,b})-f(a)\right) + g(b)\left(f(b)-f(\xi_{a,b})\right) \\
&\quad < g(a)\left(f(\xi_{a,b})-f(a)\right) + g(b)\left(f_M-f(\xi_{a,b})\right) \\
&\quad < g_M\left(f_M-f(a)\right) + g_M\left(f_M-f(\xi_{a,b})\right) \\
&\quad < g_M\left(f_M-f(a)\right) < g_M f_M
\end{aligned} \tag{3.130}$$

and

$$\begin{aligned}
&g(a)\left(f(\xi_{a,b})-f(a)\right) + g(b)\left(f(b)-f(\xi_{a,b})\right) \\
&\quad = f(\xi_{a,b})\left(g(a)-g(b)\right) + g(b)f(b)-g(a)f(a) \\
&\quad > f_M(g_m-g_M) + g(b)f(b)-g(a)f(a) \\
&\quad > f_M(g_m-g_M) - g(a)f(a) \\
&\quad > f_M(g_m-g_M) - g_M f_M = f_M(g_m-2g_M).
\end{aligned} \tag{3.131}$$

Let $a \to -\infty$ and $b \to \infty$,

$$f_M(g_m - 2g_M) < \int_{-\infty}^{+\infty} g(t)f'(t)dt < g_M f_M.$$

Moreover, if $g : \mathbb{R} \to (g_m, g_M)$ is strictly decreasing,

$$
\begin{aligned}
& g(a)\left(f(\xi_{a,b}) - f(a)\right) + g(b)\left(f(b) - f(\xi_{a,b})\right) \\
& = f(\xi_{a,b})\left(g(a) - g(b)\right) + g(b)f(b) - g(a)f(a) \\
& > g(b)f(b) - g(a)f(a)
\end{aligned}
\tag{3.132}
$$

holds. Hence,

$$(g_m - g_M)f_M \leq \int_{-\infty}^{+\infty} g(t)f'(t)dt.$$

Choose $a = -b < 0 < b$ and there exists $\xi_b \in [-b, b]$ such that

$$
\begin{aligned}
& \int_{-b}^{b} g(t)f'(t)dt \\
& \leq g(-b)\left(f(\xi_b) - f(-b)\right) + g(-b)\left(f_M - f(\xi_b)\right) \\
& = g(-b)\left(f_M - f(-b)\right)
\end{aligned}
\tag{3.133}
$$

Let $b \to +\infty$,

$$\int_{-\infty}^{\infty} g(t)f'(t)dt \leq 0. \tag{3.134}$$

Proof of Lemma 3.4.10

Proof If choosing another set of parameters and keeping $\mathcal{R}_o$ the same, we get

$$
\begin{aligned}
-\Delta\mathcal{R}_p &= \ln(s_\infty + \Delta s_\infty) + \mathcal{R}_o(1 - (s_\infty + \Delta s_\infty)) + \mathcal{R}_p \\
&= \ln\left(1 + \frac{\Delta s_\infty}{s_\infty}\right) - \mathcal{R}_o \Delta s_\infty.
\end{aligned}
\tag{3.135}
$$

Moreover, if $\frac{\Delta s_\infty}{s_\infty}$ is small, we find by straightforward Taylor expansion that

$$
\begin{aligned}
\Delta\mathcal{R}_p &= -\left(\frac{\Delta s_\infty}{s_\infty} + O\left(\frac{\Delta s_\infty}{s_\infty}\right) - \mathcal{R}_o \Delta s_\infty\right) \\
&\approx \left(\mathcal{R}_o - \frac{1}{s_\infty}\right)\Delta s_\infty
\end{aligned}
\tag{3.136}
$$

or

$$\lim_{\Delta s_\infty \to 0} \frac{\Delta \mathcal{R}_p}{\Delta s_\infty} = \mathcal{R}_o - \frac{1}{s_\infty}. \tag{3.137}$$

Proof of Lemma 3.5.1

Proof

$$\int_0^b f'(t)g(t)dt = g(0)\int_0^{\xi_b} f'(t)dt + g(b)\int_{\xi_b}^b f'(t)dt$$

$$= g(0)\left[f(\xi_b) - f(0)\right] + g(b)[f(b) - f(\xi_b)] \tag{3.138}$$

$$= g(b)[f(b) - f(\xi_b)]$$

$$g(b)f(b) > \int_0^b f'(t)g(t)dt > g(b)f(b) - f_M g(b). \tag{3.139}$$

Let $b \to \infty$,

$$g_M f_M \geq \int_0^\infty f'(t)g(t)dt \geq g_M f_M - f_M g_M = 0. \tag{3.140}$$

Proof of the Theorem 3.5.6

Proof Suppose to the contrary $s_\infty > \frac{1}{\mathcal{R}_o}$, which indicates that there exists a $\delta > 0$ such that $s_\infty = \frac{1}{\mathcal{R}_o} + \delta$ and then $s(t) \geq \frac{1}{\mathcal{R}_o} + \delta$ for all $t \geq 0$ owing to the decrement of $s(t)$. Since $\lim_{t \to \infty} s_P(t) = P_l(\infty)s_\infty = P(0)s_\infty$ and $\delta > 0$, for given $(P(0)\delta)/2$ there exists $\bar{t}_0 \in [0, \infty)$ such satisfying

$$|s_P(t) - P(0)s_\infty| < \frac{P(0)\delta}{2}$$

for all $t \geq \bar{t}_0$. Thus, we have

$$s_P(t) > \frac{P(0)}{\mathcal{R}_o} + \frac{P(0)\delta}{2}$$

and

$$U'(t) = \mathcal{R}_o\left(\frac{s_P(t)}{P(0)} - \frac{1}{\mathcal{R}_o}\right)\bar{\gamma}I > \frac{\mathcal{R}_o\bar{\gamma}I\delta}{2}$$

for all $t \geq \bar{t}_0$. Then, for all $t \geq \bar{t}_0$, we have

$$U(t) > U(t_0)e^{\frac{\mathcal{R}_o\bar{\gamma}\delta}{2}\int_{t_0}^t \frac{I(t)}{U(t)}dt} > U(t_0)$$

which means $U(\infty) = +\infty$, it contradicts with $U(\infty) = 0$. Thus, $s_\infty \leq \frac{1}{\mathcal{R}_o}$ holds.

Moreover, if $P_l(t) = P(U)$, $s_\infty < \frac{1}{\mathcal{R}_o}$. Suppose to the contrary $s_\infty = \frac{1}{\mathcal{R}_o}$ i.e. $S_\infty = \frac{N_0}{\mathcal{R}_o}$. Then

$$
\begin{aligned}
\frac{dU}{dS} &= -\frac{\mathcal{R}_o \left(\frac{s_P(t)}{P(0)} - \frac{1}{\mathcal{R}_o} \right) \bar{\gamma}}{\beta N_0 s_P(t)} = -\frac{\left(s_P(t) - \frac{P(0)}{\mathcal{R}_o} \right)}{s_P(t)} \\
\frac{d}{dS} \frac{dU}{dS} &= -\frac{d}{ds_P} \left[\frac{\left(s_P(t) - \frac{P(0)}{\mathcal{R}_o} \right)}{s_P(t)} \right] \frac{ds_P}{dS} \\
&= -\frac{P(0)}{\mathcal{R}_o} \frac{1}{s_P^2} \frac{ds_P}{dS} = -\frac{P(0)}{\mathcal{R}_o} \frac{1}{s_P^2} \left[\frac{dP_l}{dU} \frac{dU}{dS} s + \frac{P_l}{N_0} \right]
\end{aligned}
\tag{3.141}
$$

which means $\frac{d^2U}{dS^2} < 0$ if $\frac{dU}{dS} \le 0$.

$$
\left. \frac{dU}{dS} \right|_{S=S_\infty} = \frac{\left(P(U(S_\infty)) s_\infty - \frac{P(0)}{\mathcal{R}_o} \right)}{P(U(S_\infty)) s_\infty} = \frac{\mathcal{R}_o \left(\frac{P(U(S_\infty))}{\mathcal{R}_o} - \frac{P(0)}{\mathcal{R}_o} \right)}{P(U(S_\infty))} = 0.
$$

Thus, there is a $\delta > 0$ such that for any $S(t) \in [S_\infty, S_\infty + \delta)$, $\frac{dU}{dS} < 0$. And hence, $U(S_\infty) = 0 > U(S_\infty + \delta/2)$, a contradiction.

Proof of Lemma 3.5.7

Proof Since $U(t) > 0$ for all $t \in [0, \infty)$, $U'(t_m) = 0$ tells

$$
\frac{s_P(t_m)}{P(0)} - \frac{1}{\mathcal{R}_o} = 0.
$$

Then, we have

$$
\begin{aligned}
U''(t_m) &= \mathcal{R}_o \left(\frac{s'_P(t_m)}{P(0)} \right) \bar{\gamma} I(t_m) \\
&= \mathcal{R}_o \left(\frac{P'(U)U'(t_m)s(t_m) + P(U(t_m))s'(t_m)}{P(0)} \right) \bar{\gamma} I(t_m) \\
&= \mathcal{R}_o \left(\frac{P(U(t_m))s'(t_m)}{P(0)} \right) \bar{\gamma} I(t_m) \le 0.
\end{aligned}
\tag{3.142}
$$

The details in reduction of the model (3.118)

Inspired by [16], we rewrite $\varphi_2(a)$ as a vector form

$$
\varphi_2(a) = \beta(a)^T e^{\mathcal{B}a} \mathbf{Id}_{2 \times 1}
\tag{3.143}
$$
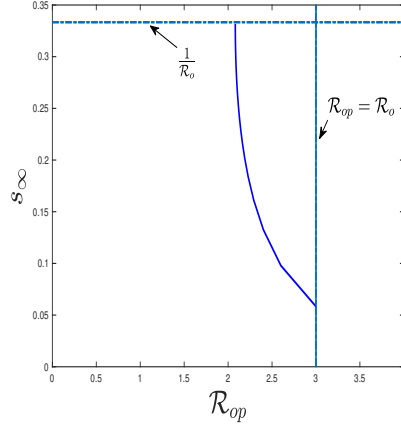
Figure 3.7: $h \in [0, 2], \mathcal{R}_o = 3$. Other parameters are $\gamma_2 = \frac{1}{7}, N_0 = 300, E_0 = 5, I_0 = 5, R(0) = 0$.
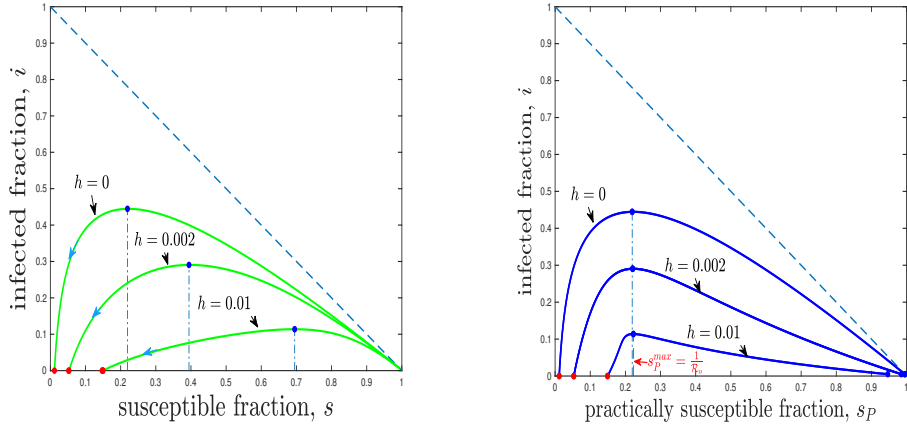
Figure 3.8: Parameters are $\mathcal{R}_o = 4.5, \gamma_2 = \frac{1}{7}, N_0 = 1000, I_0 = 5, R(0) = 0$.

where

$$\beta(a) := \begin{pmatrix} 0 \\ \beta \end{pmatrix}, \; \mathbf{Id}_{2\times 1} := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; \gamma(a) := \mathcal{B} = \begin{bmatrix} -\epsilon & 0 \\ \epsilon & -\bar{\gamma} \end{bmatrix}.$$

The infected density $\mathbf{u}(t, a)$ on infection-age a corresponding to the kernel $\varphi(a) = \varphi_2(a)$ is a two-dimensional vector:

$$\mathbf{u}(t, a) = \begin{pmatrix} u_e(t, a) \\ u_i(t, a) \end{pmatrix},$$

which is different from the equation (3.9):

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)\mathbf{u}(t, a) = \mathcal{B}\mathbf{u}(t, a) \tag{3.144}$$

with boundary condition $\mathbf{u}(t, 0) = [u_e^0(t), 0]^T$. The age densities of total infected individuals $u(t, a) := \|\mathbf{u}(t, a)\|_1 = u_e(t, a) + u_i(t, a)$. Further, denote

$$E(t) = \int_0^{+\infty} u_e(t, a)da, \ I(t) = \int_0^{+\infty} u_i(t, a)da$$

exposed and infectious individuals, respectively. Further, simple calculations lead to the model (3.118).

Bibliography

[1] D. Anderson and R. Watson, On the spread of a disease with gamma-distributed latent and infectious periods, *Biometrika*, **67(1)** (1980), 191-198.

[2] R. M. Anderson and R. M. May, Infectious diseases of humans: dynamics and control, *Oxford University Press*, Oxford, UK, 1991.

[3] J. Arino, F. Brauer, P. van den Driessche, J. Watmough and J. Wu, A final size relation for epidemic models, *Math. Biosci. Eng.*, **4(2)** (2007), 159.

[4] V. Andreasen, The final size of an epidemic and its relation to the basic reproduction number, *Bull. Math. Biol.*, **73(10)** (2011), 2305-2321.

[5] L. Almeida, P. A. Bliman, G. Nadin, B. Perthame and N. Vauchelet, Final size and convergence rate for an epidemic in heterogeneous populations, *Math. Models Methods Appl. Sci.*, **31(05)** (2021), 1021-1051.

[6] F. Brauer, The Kermack–McKendrick epidemic model revisited, *Math. Biosci.*, **198(2)** (2005), 119-131

[7] F. Brauer, Age of infection and the final size relation, *Math. Biosci. Eng.*, **5(4)** (2008), 681.

[8] D. Breda, O. Diekmann, W. F. De Graaf, A. Pugliese and R. Vermiglio, On the formulation of epidemic models (an appraisal of Kermack and McKendrick), *J. Biol. Dyn.*, **6(sup2)** (2012), 103-117.

[9] D. Champredon, J. Dushoff and D. J. Earn, Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation, *SIAM J. Appl. Math.*, **78(6)** (2018), 3258-3278.

[10] T. Cheng and X. Zou, A new perspective on infection forces with the demonstration by a DDE infectious disease model, *Math. Biosci. Eng.*, **19(5)** (2022), 4856-4880.

[11] T. Cheng and X. Zou, Modelling the impact of precaution on disease dynamics and its evolution, *in preprint.*

[12] A. C. Dixon, The second mean value theorem in the integral calculus, *Math. Proc. Camb. Philos. Soc.*, **25(3)** (1929), 282-284.

[13] O. Diekmann and J. A. P. Heesterbeek, Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation, *Wiley*, 2000

[14] O. Diekmann, H. Hans and B. Tom, Mathematical tools for understanding infectious disease dynamics (vol. 7), *Princeton University Press*, 2013.

[15] O. Diekmann, H. G. Othmer, R. Planque and M. C. Bootsma, The discrete-time Kermack-McKendrick model: A versatile and computationally attractive framework for modelling epidemics, *Proc. Natl. Acad. Sci. U.S.A.*, **118(39)** (2021), e2106332118.

[16] O. Diekmann and H. Inaba, A systematic procedure for incorporating separable static heterogeneity into compartmental epidemic models, *J. Math. Biol.*, **86(2)** (2023), 1-19.

[17] W. Feller, On the logistic law of growth and its empirical verifications in biology, *Acta Biotheor.*, **5** (1940), 51-66.

[18] W. Feller, On the integral equation of renewal theory, *Ann. Math. Stat.*, **12(3)** (1941), 243-267.

[19] Z. Feng, Final and peak epidemic sizes for SEIR models with quarantine and isolation, *Math. Biosci. Eng.*, **4(4)** (2007), 675.

[20] P. Fine, K. Eames, and D. L. Heymann, "Herd immunity": a rough guide, *Clin. Infect. Dis.*, **52(7)** (2011), 911-916.

[21] K. Heng and C. L. Althaus, The approximately universal shapes of epidemic curves in the Susceptible-Exposed-Infectious-Recovered (SEIR) model, *Sci. Rep.*, **10(1)** (2020), 1-6.

[22] H. Inaba, Age-structured population dynamics in demography and epidemiology, *Springer Singapore*, 2017.

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

[24] A. J. Lotka, Relation between birth rates and death rates. *Science*, **26(653)** (1907), 21-22.

[25] J. D. Lebreton, Demographic models for subdivided populations: the renewal equation approach, *Theor. Popul. Biol.*, **49(3)** (1996), 291-313.

[26] J. Ma and D. J. Earn, Generality of the final size formula for an epidemic of a newly invading infectious disease, *Bull. Math. Biol.*, **68** (2006), 679-702.

[27] P. Magal, C. C. McCluskey and G. F. Webb, Lyapunov functional and global asymptotic stability for an infection-age model, *J. Appl. Anal.* **89(7)** (2010), 1109-1140.

[28] P. Magal, O. Seydi, and G. Webb, Final size of an epidemic for a two-group SIR model, *SIAM J. Appl. Math.*, **76(5)** (2016), 2042-2059.

[29] W. Roman, E. Hetmaniok and S. Damian, A stronger version of the second mean value theorem for integrals, *Comput. Math. Appl.* **64(6)** (2012), 1612-1615.

[30] W. Rudin, Principles of mathematical analysis, *New York, McGraw-Hill,* 1953.

[31] F. R. Sharpe, A. J. Lotka and J. L. Alfred, A problem in age-distribution, *Lond. Edinb. Dublin philos. mag. j. sci.*, **21(124)** (1911), 435-438.

Chapter 4

Modelling the impact of precaution on disease dynamics and its evolution

4.1 Introduction

In modelling infectious disease dynamics, most models assume a homogeneous susceptibility for the susceptible population, and such an assumption can make the models more mathematically tractable. However, in reality, susceptibility may differ from individual to individual. Physiologically or immunologically, this is due to the differences in, e.g., responses of individuals' immune systems to different pathogens, as well as in the effect of vaccination (taking vaccine or not, the efficacy of the vaccine taken, etc.). There have been some studies that use structured models to reflect such heterogeneity of susceptibility and explore its impact on disease dynamics. See, e.g., [5, 7, 8, 9, 10, 12, 14, 19, 20] and the references therein.

In addition to the physiological/immunological factors, there are also social/behavioural factors that may affect susceptibility of a population. For instance, during the pandemic covid-19 from 2020-2023, in addition to the development and wide use of vaccines and drugs for treatment, various non-pharmaceutical interventions (NPIs) in all countries in the world have also played an essential role in controlling the transmission/spread of covid-19. Although specific forms of NPIs differ from country to country, region to region, and city to city, and although such NPIs varied as the disease evolved, such NPIs, together with massive media coverage and education, typically raised the awareness of the public about this disease, making

them more precautious and less social. As a consequence of such precautions and reduced sociality, some epidemically susceptible people are practically non-susceptible or less susceptible. In other words, during a pandemic or epidemic, due to the precaution, only a fraction $P \in [0, 1]$ of the *epidemiologically susceptible* population $S(t)$, denoted by $S_p(t) = PS(t)$, is actually *susceptible* due to precaution caused by NPIs and media coverage. This would, of course, impact the disease dynamics (transmission dynamics in the population).

In a recent work [2], based on the above observation, we proposed a new perspective for understanding the notion of the force of infection (or infection force), which can not only explain many existing infection force functions used in the literature but also motivate new forms of infection force functions. To be more specific, if the mass action infection mechanism $\beta I(t)S(t)$ is adopted, then replacing the population of the epidemiologically susceptible population $S(t)$ with the practically susceptible population $S_p(t)$ would revise the incidence rate $\beta I(t)S(t)$ to

$$\beta I(t)S_p(t) = \beta I(t)PS(t) = [\beta PI(t)] \cdot S(t),$$

leading to an overall infection force $f_m(t) = [\beta PI(t)]$. Here, the fraction P naturally depends on the severity of the epidemic, denoted by $L(t)$, in such a way that the more severe the epidemic is, the smaller the fraction P is. Accordingly it is reasonable to assume $P = P(L(t))$ satisfies the following condition:

$$P(L) \text{ is non-increasing, } P(0) \le 1 \text{ and } P(\infty) \ge 0. \tag{4.1}$$

Below are some prototypes of such a fraction function satisfying (4.1):

(A) $P_1(L) = \frac{m_1 L + 1}{m_2 L + 1}$ where m_1, m_2 are all positive constants satisfying $\frac{m_1}{m_2} < 1$;

(B) $P_2(L) = \frac{m_1 L + b}{cL^2 + m_2 L + b}$, where all parameters are positive constants and satisfy $\frac{m_1}{m_2} < 1$;

(C) $P_3(L) = e^{-hL}$, where $h > 0$.

There are various ways to measure the severity $L(t)$ at time t. In the simplest case, the severity is measured by the current prevalence of the disease, i.e., $L(t) = I(t)$, the above three forms with this severity lead to the infection force the function used many models in the literature, such as [1, 11, 15, 18, 22, 25] and some references therein. We point out that the severity

measurement may also consider the infections in some past times, meaning that $L(t)$ has the form

$$L(t) = \int_0^\tau w(\xi)I(t - \xi)\,d\xi \tag{4.2}$$

where the constant $\tau > 0$ represents a length of time interval and $w(\xi)$ is the weight function that reflects the variation of the impact of disease surveillance in the past interval $[t - \tau, t]$ on the severity at the present time, or its discrete version

$$L(t) = \sum_{i=0}^k w_i I(t - \tau_i) \ \text{ with } \ 0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k. \tag{4.3}$$

Practically, the form (4.3) is more feasible because case reporting is done at discrete times in reality. In [2], through a SIR model incorporated with the exponential decay function $P(L) = P_3(L) = e^{-hL}$ with $L(t)$ given by (4.3) for $k = 1$, we demonstrated the impacts of the information delay $\tau = \tau_1 > 0$ and information weights k_0 and k_1 on the disease dynamics.

For the fraction P for the practical susceptible population, in addition to the severity of the epidemic, it should also depend on the response level of the public. With the same severity L, different response levels may lead to different fractions P. In general, such a response level is heterogeneous, differing from individual to individual; moreover, it may evolve with the disease dynamics. For simplicity, we just use the *average response level* denoted by $X \in [0, 1]$ to avoid the heterogeneity in the response level. With all the above considerations, P is now denoted $P = P(X, L)$, and accordingly, the properties given in (4.1) is revised to

$$\begin{cases} \dfrac{\partial P(X, L)}{\partial X} < 0, \quad \dfrac{\partial P(X, L)}{\partial L} < 0, \\[2mm] P(0, L) = 1, \quad P(1, L) \geq 0, \quad P(X, \infty) \geq 0, \quad P(X, 0) \leq 1. \end{cases} \tag{4.4}$$

We point out that the incorporation of such a fraction function $P(X, L)$ is motivated by the recent works [23, 24] that investigate the fear effect in predator-pray systems. There is a similarity here: in a predator-prey system, when a prey perceives the risk from the predator as an anti-predation response, the prey will typically adjust their behaviour to reduce the risk of being caught by the predator; during an epidemic, aware of the severity of the epidemics, susceptible individuals will also typically change their behaviours (reducing social activities,

using EPPs, or even locked down by the government's mandatory rules), either actively (voluntarily) or passively (forcedly) to reduce the chances of being infected. We note that there are many research works on modelling non-pharmaceutical interventions (NPIs) (see, e.g., [10, 13] and the references therein), but here in this paper, we focus on the impact of the precaution on infectious disease dynamics. Such precaution can be attributed to the various interventions from governments/public health agencies, media coverage, or increased knowledge of the public on the disease. The level of such precaution is, in general, not easy to quantify, and neither is its impact. As such, this study is mainly of a mechanistic nature. However, we believe that different NPIs may serve different purposes and have different effects; accordingly, the associated response of the public may be reflected in different ways. For example, Qiu et al. [17] use the fraction of the mask-wearing population as a measurement of response to an epidemic, and [16] et al. use the fraction of the population adopting general NPIs.

Two questions naturally arise:

(Q1) How does the response level (together with the severity) impact the disease dynamics?

(Q2) How does the response level evolve with the disease dynamics?

This paper aims to explore these two questions through some specific SIS type of disease models. In Section 4.2, we formulate a general framework for a class of SIS models with evolving precaution levels. With the aforementioned infection-force functions f_m, we establish the well-posedness of the general framework model, discuss the stability of the disease-free equilibrium, identify the basic reproduction number R_0 and discuss its relation to the stability of the disease-free equilibrium. Sections 4.3 and 4.4 are devoted to the endemic dynamics, i.e., the long-term dynamics when $R_0 > 1$, with infection force function specified to $f = f_m$ and assuming some specific forms for the precaution evolution rate $M(t)$ that will be explained in Section 4.2 when formulating the framework model, aiming to demonstrate the feasibility of the general framework. Sections 4.3 adopts an instantaneous form for $M(t)$ but with infection force f_m, resulting in a system of ordinary differential equations (ODEs), while Section 4.4 adopts a form for $M(t)$ with a time delay, leading to a system of delay differential equations (DDEs). We analyze the stability of the endemic equilibrium and explore the possibility for Hopf bifurcation to occur.

We also present some numerical simulations to illustrate the theoretical results. In Section 4.5, we employ some ideas in [16, 17] to explore the disease dynamics when the response level is assumed to be adapting to a given response function. Finally, in Section 4.6, we summarize the main results and discuss their implications in epidemiological and social contexts.

4.2   A general framework model and some preliminary results

Recall that a classic SIS model is given

$$
\begin{cases}
S'(t) = \Lambda - dS(t) - f(t)S(t) + rI(t), \\
I'(t) = f(t)S(t) - (d + r)I(t).
\end{cases}
\tag{4.5}
$$

Here $S(t)$ and $I(t)$ are the epidemiologically susceptible and infectious populations, Λ is the recruitment rate of the population, d is the natural death rate of the population, r is the recovery rate of infective individuals.

Now, for the infection force $f(t)$, we will adopt

$$
f(t) = f_m(t) = \beta I(t) P(X(t), L(t))
\tag{4.6}
$$

where β is the transmission rate, and as discussed in the Introduction, $L(t)$ is a measurement of disease severity and $X(t)$ is the average precaution level of the susceptible population. $X(t)$ is assumed to evolve continuously with time, depending on the severity or trend of the epidemics, and thus, X can be modelled by

$$
\frac{dX}{dt} = \epsilon X(1 - X)M(t).
\tag{4.7}
$$

Note that the term $X(1 - X)$ ensures that X is enclosed in the interval $[0, 1]$ ([21]), $\epsilon M(t)$ then reflects the direction and speed of the evolution of $X(t)$, with $\epsilon > 0$ being a positive constant and $M(t)$ is dependent on the severity and/or trend of the epidemics. A general consideration is that $X(t)$ should evolve in the same direction of the epidemics: when the epidemic is mitigating

(resp. escalating), the average protection level $X(t)$ should be accordingly decreasing (resp. increasing). To avoid complexity but demonstrate this co-evolving idea, we will simply use the prevalence at the present or an earlier time to measure the severity, that is,

$$L(t) = I(t) \qquad \text{or} \qquad L(t) = I(t - \tau) \text{ with } \tau > 0, \tag{4.8}$$

and consider the following two simple choices for $M(t)$

$$M(t) = M_1(t) = I'(t) \qquad \text{or} \qquad M(t) = M_2(t) = I'(t - \tau) \text{ with } \tau > 0. \tag{4.9}$$

Here, $M_1(t)$ accounts for a scenario using the current rate of change of the prevalence to represent the trend of the epidemic evolution, while $M_2(t)$ is based on the same logic as for $M_1(t)$, but takes into consideration the fact that there is usually a delay in reality in obtaining and analyzing data that reflect the prevalence and its change; it also takes time for the public healthy agents and governments to plan and implement various NPIs. Eq.(4.7) is formally analogous to the replicator equations for evolutionary population games [6], yet the benefits are time-dependent.

Putting (4.5) and (4.7) together results in the following framework model

$$\begin{cases} S'(t) = \Lambda - dS(t) - f(t)S(t) + rI(t), \\ I'(t) = f(t)S(t) - (d + r)I(t), \\ X'(t) = \epsilon X(1 - X)M(t). \end{cases} \tag{4.10}$$

where $f(t)$, $M(t)$ and $L(t)$ are given by (4.6), (4.8) and (4.9) with the involving function $P(X, L)$ satisfying (4.4).

Firstly, it is easy to show that the set.

$$\Gamma = \{(S, I, X) \mid S \geq 0, I \geq 0, 0 \leq X \leq 1 \text{ and } S + I \leq \Lambda/d\}$$

is invariant for (4.5). The proof is similar to that of Theorem 2.1 in [2] and is thus omitted here.

From the last equation in (4.10), $X(t)$ can be expressed by $M(t)$ as

$$X(t) = \frac{1}{1 + C_0 e^{-\epsilon \int_0^t M(s)ds}} \tag{4.11}$$

where $C_0 = 1/X_0 - 1 \geq 0$ for $X_0 = X(0) \in (0, 1]$. With $M(t)$ specified to $I'(t)$ or $I'(t - \tau)$, $X(t)$ is given either by

$$X(t) = X_1(I(t)) = \frac{1}{1 + C_1 e^{-\epsilon I(t)}} \quad \text{with} \quad C_1 = C_0 e^{\epsilon I(0)} \tag{4.12}$$

or

$$X(t) = X_2(I(t - \tau)) = \frac{1}{1 + C_2 e^{-\epsilon I(t - \tau)}} \quad \text{with} \quad C_2 = C_0 e^{\epsilon I(-\tau)}. \tag{4.13}$$

We point out that the form (4.12) for the average response level function $X(t)$ is referred to as the "best response function" in some literature (see, e.g., [16]), while (4.13) may be called the "delayed best response function."

The general response function (4.11) is of smooth and sigmoidal form, which is formally similar to the *Smooth Best Response* (i.e. *Logit dynamic*) for *evolutionary population game*[4] after re-organizing (4.11):

$$
\begin{aligned}
X(t) &= \frac{\exp\left[\epsilon \int_0^t M(s)ds\right]}{\exp\left[\epsilon \int_0^t M(s)ds\right] + C_0} = \frac{X_0 \exp\left[\epsilon \int_0^t M(s)ds\right]}{X_0 \exp\left[\epsilon \int_0^t M(s)ds\right] + (1 - X_0)} \\
&= \frac{X_0 \exp\left[\frac{\epsilon}{2} \int_0^t M(s)ds\right]}{X_0 \exp\left[\frac{\epsilon}{2} \int_0^t M(s)ds\right] + (1 - X_0) \exp\left[-\frac{\epsilon}{2} \int_0^t M(s)ds\right]}.
\end{aligned}
\tag{4.14}
$$

To some extent, we can interpret our response level function $X(t)$ in the "perspective" of *evolutionary population game*—for example, measure the response level by the proportion of the population "adopting" precautious behaviour. Then X_0 is the initial proportion of the population with the "adopting" strategy, with $(1 - X_0)$ being the "non-adopting." The relative benefits of "adopting" at time t is $\int_0^t M(s)ds$.

Denote

$$\text{benefits of "adopting"} := U_a, \quad \text{benefits of "non-adopting"} := U_n.$$

We can choose

$$U_a = \frac{1}{2} \int_0^t M(s)ds, \text{ (or } \int_0^t M(s)ds), \quad U_n = -\frac{1}{2} \int_0^t M(s)ds \text{ (resp. or 0)},$$

which means the higher the severity, the higher the benefits of "adopting." Correspondingly, the lower benefits of "non-adopting." Further, (4.14) is expressed as

$$X(t) = \frac{X_0 \exp(\epsilon U_a)}{X_0 \exp(\epsilon U_a) + (1 - X_0) \exp(\epsilon U_n)}. \tag{4.15}$$

This form (4.15) suggests a memory component for *evolutionary population game*: The current decision-making is affected by both the current benefits of the strategies and the initial choices/states. Significantly, the benefits here, i.e. U_a and U_n, are time-varying. ϵ serves a similar function as the *sensitivity parameter* in *Logit choice*:

- $\epsilon = 0$, the response is independent of the epidemics' severity, i.e. $X(t) \equiv X_0$; when $\epsilon \to \infty$, the response is determined: 0 or 1, relying on the epidemics' severity;

- Low (reps. High) ϵ indicates the response is less (highly) sensitive to the epidemic's severity.

Plugging (4.12) (resp. (4.13)) into the first two equations of (4.10), one be reduced as the following two-dimensional model:

$$\begin{cases} S'(t) = \Lambda - dS(t) - f(t)S(t) + rI(t), \\ I'(t) = f(t)S(t) - (d+r)I(t), \end{cases} \tag{4.16}$$

with $f(t)$, $M(t)$ and $L(t)$ are given by (4.6), (4.8) and (4.9) with the involving function $P(X, L)$ satisfying (4.4), (4.16) is indeed an *autonomous system* for the variables $S(t)$ and $I(t)$ for which

$$D = \{(S, I) \mid S \geq 0, I \geq 0, S + I \leq \Lambda/d\} \tag{4.17}$$

is invariant.

Moreover, it is easy to see that (4.16) has a disease free equilibrium $E_0 = (S_0, 0)$ where

$S_0 = \Lambda/d$. To investigate the stability of E_0, we linearize the model (4.16) at E_0 to obtain

$$
\begin{cases}
S'(t) = -dS(t) - \beta I(t)P_E S_0 + rI(t), \\
I'(t) = [\beta P_E S_0 - (d+r)]I(t),
\end{cases}
\tag{4.18}
$$

where

$$
P_E = P(X_0, 0) \quad \text{for the case } f(t) = f_m(t).
\tag{4.19}
$$

From (4.18), we can see that the stability of E_0 is determined by the sign of the principal eigenvalue $\lambda_0 = P_E S_0 - (d+r)$: E_0 is asymptotically stable if $\lambda_0 < 0$ and it is unstable if $\lambda_0 > 0$. Moreover, under the condition (4.4), there holds $f(t) \le \beta P_E I(t)$. Thus, the second equation in (4.16) has the second equation in the linear system (4.18) as a comparison equation from above. By a comparison argument, we then conclude that E_0 is globally asymptotically stable if $\lambda_0 < 0$.

By tracking the average infection time and infection rate in the fully susceptible population, we can easily identify the basic reproduction number R_0 as

$$
R_0 = \frac{1}{d+r} \cdot \beta P_E S_0 = \frac{\beta P_E S_0}{d+r}.
\tag{4.20}
$$

Obviously, $R_0 < 1$ (resp. $R_0 > 1$) if and only if $\lambda_0 < 0$ (resp. $\lambda_0 > 0$).

Summarizing the above, we have proved the following Theorem.

Theorem 4.2.1 E_0 *is globally asymptotically stable if* $R_0 < 1$ *and unstable if* $R_0 > 1$. *In the case* $R_0 < 1$, *there holds*

$$
\lim_{t \to \infty} X(t) = \frac{1}{1+C_1} = \frac{1}{1+C_0 e^{\epsilon I(0)}} \approx \frac{1}{1+C_0} = X_0 \quad (\textit{since } I(0) \textit{ is generally very small}).
$$

By the above theorem, we just need to discuss the disease dynamics of (4.10) under the endemic condition $R_0 > 1$. To make the demonstration convenient, we specify $P(X, L)$ as the exponential decay function $P(X, L) = e^{-hLX}$ in the subsequent sections and explore the following two cases:

(A) $L(t) = I(t), \quad M(t) = I'(t)$ and $f(t) = f_m(t)$;

(B) $L(t) = I(t-\tau), \quad M(t) = I'(t-\tau)$ and $f(t) = f_m(t)$.

4.3 The endemic dynamics for case (A): $L(t) = I(t)$, $M(t) = I'(t)$ and $f(t) = f_m(t)$

For this case with $P(X, L) = e^{-hLX}$, $P_E = 1$ and the model system (4.16) becomes

$$\begin{cases} S'(t) = \Lambda - dS(t) - \beta I(t)P_1(I(t))S(t) + rI(t) := B_1, \\ I'(t) = \beta I(t)P_1(I(t))S(t) - (d + r)I(t) := B_2, \end{cases}$$

(4.21)

where

$$P_1(I(t)) = \exp\left(\frac{-hI(t)}{1 + C_1 e^{-\epsilon I(t)}}\right).$$

(4.22)

For this 2-D ODE system, an endemic equilibrium is given by the intersection of the following two curves:

$$\begin{cases} S = \dfrac{\Lambda}{d} - I =: g_1(I) \\ S = \dfrac{d + r}{\beta} \exp\left(\dfrac{hI}{1 + C_1 e^{-\epsilon I}}\right) =: g_2(I) \end{cases}$$

(4.23)

Noting that $g_1(I)$ is decreasing $g_2(I)$ is increasing, and hence, (4.23) has a positive solution (unique) if and only if $g_1(0) > g_2(0)$, that is

$$\frac{\Lambda}{d} > \frac{d + r}{\beta}$$

which is equivalent to

$$R_0 = \frac{\beta \Lambda P_E}{d(d + r)} = \frac{\beta \Lambda}{d(d + r)} > 1.$$

Assume $R_0 > 1$ so that (4.21) has a unique endemic equilibrium $E_1^* = (S_1^*, I_1^*)$. Denote

$$m = d + r \quad \text{and} \quad X_1^* = \frac{1}{1 + C_1 e^{-\epsilon I_1^*}}.$$

Then, the Jacobian matrix at E_1^* is:

$$J = \begin{bmatrix} -\dfrac{mI_1^*}{S_1^*} - d & mhI_1^*X_1^*\left(\epsilon(1 - X_1^*)I_1^* + 1\right) - d \\ \dfrac{mI_1^*}{S_1^*} & -mhI_1^*X_1^*\left(\epsilon(1 - X_1^*)I_1^* + 1\right) \end{bmatrix}$$

$$\det(J) = \frac{\left(1 + S_1 X_1^* h(1 + I_1^* \epsilon(1 - X_1^*))\right) m d I_1^*}{S_1^*} > 0, \quad tr(J) = -\frac{\det(J) + d^2}{d} < 0.$$

Thus, E_1^* is asymptotically stable.

We can prove that E_1 is globally asymptotically stable. To this end, we consider the following Dulac function. $Q(S, I) = 1/P_1(I), \ (S, I) \in D$. Then we have

$$\begin{aligned}
\frac{\partial(QB_1)}{\partial S} &+ \frac{\partial(QB_2)}{\partial I} \\
&= \frac{\partial}{\partial S}\left(\frac{\Lambda - dS}{P_1(I)I} - \beta S + \frac{r}{P_1(I)}\right) + \frac{\partial}{\partial I}\left(\beta S - \frac{d + r}{P_1(I)}\right) \\
&= -\left(\frac{d}{P_1(I)I} + \beta\right) + \frac{(d + r)P_1'(I)}{P_1(I)^2}.
\end{aligned}$$

Direct calculation shows

$$\begin{aligned}
P_1'(I) &= \frac{\partial P(X, I)}{\partial X}\frac{dX}{dI} + \frac{\partial P(X, I)}{\partial I} = \epsilon(1 - X)X\frac{\partial P(X, I)}{\partial X} + \frac{\partial P(X, I)}{\partial I} \\
&= \epsilon(1 - X)XhP_1[-I - X] = -P_1(I)hX(1 - X)(I + X) < 0.
\end{aligned}$$

Thus

$$\frac{\partial(QB_1)}{\partial S} + \frac{\partial(QB_2)}{\partial I} < 0, \ \text{ for } (S, I) \in D.$$

According to the Bendixson Dulac theorem, System (4.21) does not have endemic periodic orbits, implying that E_1 is globally asymptotically stable. Hence, we have proved the following theorem.

Theorem 4.3.1 *Assume that $R_0 > 1$. Then* (4.21) *has an unique endemic disease equilibrium* $E_1^* = (S_1^*, I_1^*)$ *which is globally asymptotically stable with $X(t)$ evolving toward a steady level*

$$X_1^* = \lim_{t \to \infty} X(t) = \frac{1}{1 + C_1 e^{-\epsilon I_1^*}} \tag{4.24}$$

where

$$C_1 = C_0 e^{\epsilon I(0)} = \left(\frac{1}{X_0} - 1\right)e^{\epsilon I_0}. \tag{4.25}$$

From this theorem and Theorem 4.2.1, we can see that the specific disease model (4.21)

actually has a global threshold dynamics: when $R_0 < 1$ all feasible solutions approach the disease free equilibrium; when $R_0 > 1$ all feasible solution approaches to an endemic equilibrium $E_1^* = (S_1^*, I_1^*)$ defined by (4.23).

We highlight a new and interesting phenomenon for the above-stated global threshold dynamics: the globally stable endemic equilibrium E^* depends on the initial prevalence $I_0 = I(0)$ and the initial response level $X_0 = X(0)$. This is because when reducing the 3-D system (4.10) to the 2-D system (4.21) through the best response function (4.12), the parameter C_1 in (4.12) depends on $X_0 = X(0)$ and $I_0 = I(0)$ as well as the evolution speed ϵ, so does I_1^*. Note from (4.23) and (4.25) that $g_1(I)$ and $g_2(0)$ are independent of the these values, while $g_2(I)$ for $I > 0$ can be rewriting as

$$g_2(I) = \frac{d + r}{\beta} \exp\left(\frac{hI}{1 + (1/X_0 - 1)e^{-\epsilon(I - I_0)}}\right).$$

From this formula and (4.23), we can observe how I_0, X_0 and ϵ affect the final endemic level I_1^*, as summarized below:

(O1) $g_2(I)$ is *increasing* in X_0, implying that I_1^* (the I component of the endemic equilibrium E_1^*) is *decreasing in X_0*.

(O2) $g_2(I)$ is *decreasing* in I_0, implying that I_1^* is *increasing in I_0*.

(O3) $g_2(I)$ is increasing in ϵ when $I > I_0$; and it is decreasing in ϵ when $I < I_0$. On the other hand, I_0 is generally very small, and thus, under the endemic condition $R_0 > 1$, there holds

$$g_2(I_0) = \frac{d + r}{\beta} e^{hI_0 X_0} < \frac{\Lambda}{d} - I_0 = g_1(I_0) \text{ for small } I_0.$$

This means that the unique solution I_1^* is larger than I_0. Noting that $I(t) \to I_1^*$ as $t \to \infty$, we then have $I(t) > I_0$ for large t. Thus, we just need to consider the scenario of $I > I_0$, under which $g_2(I)$ is increasing in I provided that I_0 is small, and therefore, implying that I_1^* is indeed *decreasing in ϵ*.

The above observations are demonstrated in Fig. 4.1. Particularly we note that (O3) is only for small I_0; for larger I_0 it may not be valid, as illustrated in Fig. 4.1-(b)-(c). In Fig. 4.1-(a), I_0 is small and I_1^* is decreasing in ϵ. In Fig. 4.1-(b), I_0 is in an intermediate range for which, while I_1^* is decreasing in ϵ for small X_0 but it is increasing in ϵ for large X_0. Fig. 4.1-(c) is with large value I_0 for which, I_1^* is increasing in ϵ.
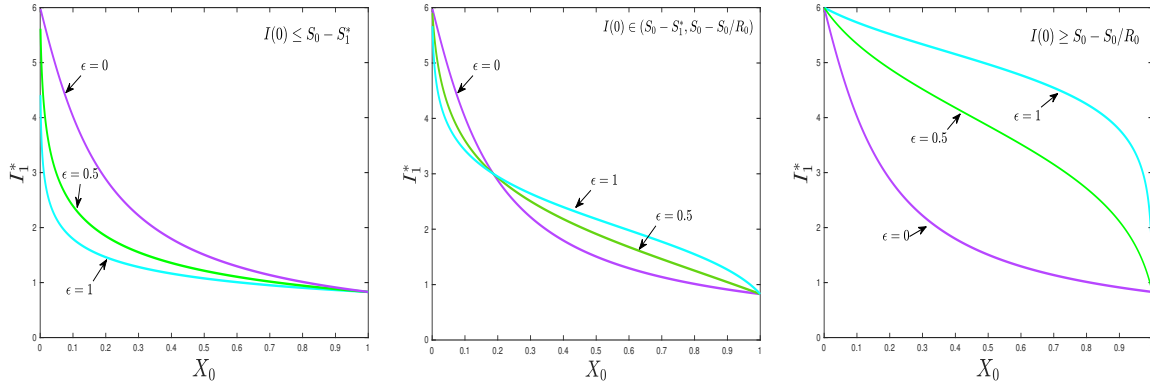
Figure 4.1: While the endemic level I_1^* is always decreasing in X_0, its monotonicity on ϵ various with I_0 and X_0: (a) $I(0) \approx 0$; (b) $I(0) = 3$ (c) $I(0) = 8$. Other parameters are chosen to be $\Lambda = 0.12, \beta = 0.1, h = 1, d = 0.012, r = 0.388$ and then $S_0 = 10$.

4.4 Endemic dynamics for case (B): $L(t) = I(t - \tau)$, $M(t) = I'(t - \tau)$ and $f(t) = f_m(t)$

This case corresponds to the scenario of responses having delay which is common in reality. For this case, the general framework model (4.10) reduces to the following specific system of delay differential equations:

$$\begin{cases} S'(t) = \Lambda - dS(t) - \beta I(t) \left(e^{-hI(t-\tau)X_2(t)} S(t) \right) + rI(t), \\ I'(t) = \beta I(t) \left(e^{-hI(t-\tau)X_2(t)} S(t) \right) - (d + r)I(t), \end{cases} \tag{4.26}$$

where $X_2(t)$ is given by (4.13):

$$X_2(t) = \frac{1}{1 + C_2 e^{-\epsilon I(t-\tau)}} \quad \text{with} \quad C_2 = C_2(\tau) = C_0 e^{\epsilon I(-\tau)} = \left(\frac{1}{X_0} - 1 \right) e^{\epsilon I(-\tau)}. \tag{4.27}$$

Rewrite (4.26) with (4.27) as

$$\begin{cases} S'(t) = \Lambda - dS(t) - \beta I(t) \left(P_2(I_\tau(t))S(t) \right) + rI(t), \\ I'(t) = \beta I(t) \left(P_2(I_\tau(t))S(t) \right) - (d + r)I(t), \end{cases} \tag{4.28}$$

with

$$P_2(I_\tau(t)) = P_2(I(t - \tau)) = \exp \left(\frac{-hI(t - \tau)}{1 + C_2 e^{-\epsilon I(t-\tau)}} \right). \tag{4.29}$$

When $\tau = 0$, (4.28)-(4.29) simply reduces to (4.21)-(4.22).

As for (4.21)-(4.22), an endemic equilibrium is an intersection of the following two curves:

$$
\begin{cases}
S = \dfrac{\Lambda}{d} - I =: \hat{g}_1(I), \\[2mm]
S = \dfrac{d + r}{\beta} \exp\left(\dfrac{hI}{1 + C_2 e^{-\epsilon I}}\right) =: \hat{g}_2(I).
\end{cases}
\tag{4.30}
$$

Note that (4.30) is the same as (4.23) except that $C_1 = (1/X_0 - 1)\exp(\epsilon I_0)$ is now replaced by $C_2 = (1/X_0 - 1)\exp(\epsilon I(-\tau))$. Therefore, by the analysis in Section 3, we conclude that (4.28)-(4.29) has a unique endemic equilibrium if and only if $R_0 = \beta\Lambda/d(d + r) > 1$.

Assume $R_0 > 1$, so that the unique endemic equilibrium $E_2^* = (S_2^*, I_2^*)$ exists and denote

$$
X_2^* = \lim_{t \to \infty} X(t) = \frac{1}{1 + C_2 e^{-\epsilon I_2^*}}.
\tag{4.31}
$$

In the sequel, we will analyze the stability of the endemic equilibrium E_2^*. Note that $C_2 = C_2 e^{\epsilon I(-\tau)}$ is dependent on the delay τ through $I(-\tau)$. However, if we assume the epidemics stars at $t = 0$, meaning that $I(-\tau) = 0$, then $C_2 = C_0$ no longer depends on τ. The analysis below is based on such an assumption: epidemics start at $t = 0$.

The linearization of (4.28) at E_2^* is given by

$$
\begin{cases}
S'(t) = -\left(d + \dfrac{mI_2^*}{S_2^*}\right) S(t) - d\, I(t) - \dfrac{mI_2^* P_2'(I_\tau)}{P_2(I_\tau)}\bigg|_{I_\tau = I_2^*} I(t - \tau), \\[3mm]
I'(t) = \dfrac{mI_2^*}{S_2^*} S(t) + \dfrac{mI_2^* P_2'(I_\tau)}{P_2(I_\tau)}\bigg|_{I_\tau = I_2^*} I(t - \tau).
\end{cases}
\tag{4.32}
$$

The corresponding characteristic equation(CE) can be derived as

$$
F(\lambda, \tau) := (d + \lambda)\left(a_0 e^{-\lambda\tau} + \lambda + b_0\right) = 0,
\tag{4.33}
$$

where

$$a_0 = mI_2^* hX_2^*(\epsilon(1 - X_2^*)I_2^* + 1) = \frac{mI_2^* h\left(1 + C_2\left(\epsilon I_2^* + 1\right)e^{-\epsilon I_2^*}\right)}{\left(1 + C_2\,e^{-\epsilon I_2^*}\right)^2},$$

$$b_0 = \frac{mI_2^*}{S_2^*}.$$

(4.34)

When $\tau = 0$, the transcendental equation (4.33) reduces to the following polynomial:

$$F(\lambda, 0) = (d + \lambda)(\lambda + a_0 + b_0) = 0.$$

(4.35)

Obviously, since the roots of $F(\lambda, 0) = 0$ are $-d$ and $-(a_0 + b_0)$, $E_2^* = (S_2^*, I_2^*)$ is locally asymptotically stable when $\tau = 0$.

We denote

$$G(\lambda, \tau) = a_0 e^{-\lambda\tau} + \lambda + b_0.$$

(4.36)

Next, we discuss if there is a root of (4.36) that crosses the pure imaginary axis in the complex plane from the left half to the right half when τ increases from zero. Since $G(0, \tau) = a_0 + b_0 > 0$, a crossing can only possibly occur in pairs of purely imaginary roots $\pm i\omega$ when τ increase and pass through some critical values.

We plug $\lambda = i\omega$ (assuming $\omega > 0$ without loss of generality) into (4.36) and separate the real part and imaginary part of $G(i\omega, \tau) = 0$ to obtain

$$a_0 \cos(\omega\tau) = -b_0, \quad a_0 \sin(\omega\tau) = \omega.$$

(4.37)

Equivalently,

$$\cos(\omega\tau) = -\frac{b_0}{a_0}, \quad \sin(\omega\tau) = \frac{\omega}{a_0}.$$

(4.38)

Notice that $\sin(\omega\tau) = \frac{\omega}{a_0} \le \omega\tau$, we have $\tau \ge \frac{1}{a_0}$. It essentially means that for small τ, the endemic equilibrium remains asymptotically stable.

Squaring and adding both equations of (4.38) lead to

$$\omega^2 = (a_0^2 - b_0^2)$$

(4.39)

Therefore, Eq. (4.39) has a positive root if and only if $a_0 > b_0$.

Assuming $a_0 > b_0$, (4.38) and (4.39) defines a sequence of critical values for the delay parameter τ given by

$$\tau_n = \tau_0 + \frac{2n\pi}{\omega}, \quad \tau_0 = \frac{1}{\omega} \arccos\left(-\frac{b_0}{a_0}\right). \tag{4.40}$$

Taking derivative in $f_1(\lambda, \tau) = 0$ with respect to τ, we obtain

$$\frac{d\lambda}{d\tau} = -\frac{\frac{\partial f_1(\lambda,\tau)}{\partial \tau}}{\frac{\partial f_1(\lambda,\tau)}{\partial \lambda}} = -\frac{\lambda(\lambda + b_0)}{1 + (\lambda + b_0)\tau}. \tag{4.41}$$

Since $f_1(i\omega, \tau_n) = 0$, then

$$\frac{d\,\mathrm{Re}\,(\lambda)}{d\tau}\bigg|_{\tau=\tau_n} = \mathrm{Re}\left(\frac{d\lambda}{d\tau}\bigg|_{\tau=\tau_n}\right) = \mathrm{Re}\left(-\frac{i\omega(i\omega + b_0)}{1 + (i\omega + b_0)\tau_n}\right) = b_0 + \omega^2 > 0. \tag{4.42}$$

This verifies the transversality condition at critical value τ_n, $n = 0, 1, 2, \cdots$.

Combining the above analysis, we have proved the following theorem based on the Hopf bifurcation Theorem for DDE.

Theorem 4.4.1 *Assuming $R_0 > 1$ so that E_2^* exists. Then, there can be two cases.*

 (i) *If $a_0/b_0 \leq 1$, then E_2^* is locally asymptotically stable for all $\tau \geq 0$.*

 (ii) *If $a_0/b_0 > 1$, then E_2^* is locally asymptotically stable for $\tau \in (0, \tau_0)$ and unstable for $\tau > \tau_0$, where a_0, b_0 are given in (4.34) and τ_0 satisfying $\tau_0 > 1/a_0$ is given in (4.40). Furthermore, system (4.26) undergoes Hopf bifurcation around E_2^* at $\tau = \tau_n$, $n = 0, 1, 2...$ where τ_n is given in (4.40).*

From this theorem, we see that the ratio a_0/b_0 plays a decisive role in determining whether or not there will be Hopf bifurcation. Let us explore a bit more about this ratio in terms of the initial precaution level X_0. Firstly, as mentioned before, we assume that $t = 0$ is the time when the epidemic start, and hence $I(-\tau) = 0$, and accordingly $C_2 = C_0 e^{\epsilon I(-\tau)} = C_0 = 1/X_0 - 1$. Secondly, by calculation, we obtain

$$\frac{a_0}{b_0} = -\frac{S_2^* P_2'(I_\tau)}{P_2(I_\tau)}\bigg|_{I_\tau=I_2^*} = S_2^* h\left(\epsilon(1 - X_2^*)X_2^* I_2^* + X_2^*\right) = \frac{S_2^* h\left(1 + C_0\left(\epsilon I_2^* + 1\right)e^{-\epsilon I_2^*}\right)}{\left(1 + C_0\,e^{-\epsilon I_2^*}\right)^2} := f_{ab}(C_0, h). \tag{4.43}$$

Note that $X_0 \to 0^+ \iff C_0 \to \infty$ and $X_0 \to 1^- \iff C_0 \to 0^+$; moreover

$$\lim_{X_0 \to 1^-} \frac{a_0}{b_0} = f_{ab}(0, h) = hS_2^* = W\left(\frac{hS_0 \exp(hS_0)}{R_0}\right)$$

and

$$\lim_{X_0 \to 0^+} \frac{a_0}{b_0} = f_{ab}(\infty, h) = 0 < 1,$$

where W is the LambertW function.

Note that $f_{ab}(0, h)$ is increasing in $h \geq 0$. Solving $f_{ab}(0, h) = 1$ for h leads to a unique solution

$$h_{cr} = \frac{1}{S_0} \cdot W(R_0 e).$$

By the property of the LambertW function, $R_0 \geq 1$ implies $h_{cr} \geq 1/S_0$. Thus, $f_{ab}(0, h) > 1$ provided $h > h_{cr}$. Therefore, for every $h > h_{cr}$, there is a $C_0^{cr} > 0$ such that $a_0/b_0 = f_{ab}(C_0, h) > 1$ for $C_0 < C_0^{cr}$. By the relation between C_0 and X_0, there is $X_0^{cr} = 1/(1+C_0^{cr})$ such that $a_0/b_0 > 1$ for $X_0 > X_0^{cr}$. Thus, the conditions $h > h_{cr}$ and $X_0 > X_0^{cr}$ give a range of parameters for Hopf bifurcation to occur at those critical values of τ, according to Theorem 4.4.1.

From the above analysis, $h > h_{cr}$ is only a sufficient condition for $a_0/b_0 > 1$. Numerical investigations show that even if $h < h_{cr}$, there still can be values of X_0 for which Hopf bifurcation can occur as τ passes those critical values. To see this, we choose $\Lambda = 0.06, \beta = 0.2, d = 0.012, r = 0.388$ and $I(-\tau) = 0$. For these values $h_{cr} = W(R_0 e)/S_0 \approx 0.3014$. Note that h_{cr} is independent of ϵ. As shown in Fig. 4.2, when ϵ is set at different values, X_0 would impact a_0/b_0 in different ways. Fig. 4.2-(a) with ϵ illustrates the scenario analyzed above: when $h > h_{cr}$, there is an X_0^{cr} such that for $X_0 > X_0^{cr}$, there holds a_0/b_0. However, Fig. 4.2-(b) with $\epsilon = 2$ demonstrates a totally different pattern of dependence of a_0/b_0 on X_0; particularly, when $h = 0.28 < 0.3410 = h_{cr}$, there is an intermediate range of X_0 in which $a_0/b_0 > 1$.

4.5   Adaptation toward the best response

In Section 4.2, we have seen that the equation $X'(t) = \epsilon X(t)[1 - X(t)]M(t)$ governing the change of the response level can lead to various precaution level functions when $M(t)$ is chosen under
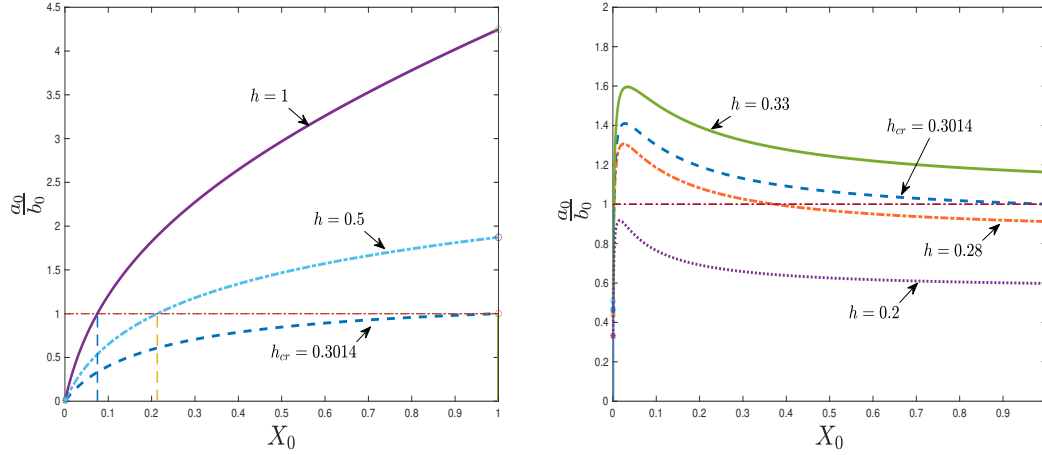
Figure 4.2: Impact of X_0 on a_0/b_0 with different given values of h and ϵ with $h_{cr} = 0.3014$: (a). $h = 0.3014, 0.5, 1, \epsilon = 0.5$; (b) $h = 0.33, 0.3014, 0.28, 0.2, \epsilon = 2$. Other parameters are chosen to be $\Lambda = 0.06, \beta = 0.2, d = 0.012, r = 0.388$ and $I(-\tau) = 0$.

various scenarios. In particular, when choosing $M(t) = I'(t)$, a scenario that the precaution level evolves synchronously with the current disease prevalence, it results in the response level function $X(t) = X_1(t) = X_1(I(t))$ given in (4.12), which is referred to as the best response function in some literature (e.g., [16, 17]). Thus, the above-governing equation provides one way to explain/justify the best response strategy in [17] where the fraction of mask-wearing susceptible population is used as a type of precaution measurement, and the strategy in [16] where a general NPIs strategy is considered. Such a best response strategy or any other response function generated from $X'(t) = \epsilon X(t)[1 - X(t)]M(t)$ by choosing different $M(t)$, directly evolves with the severity of epidemics and hence, it decouples the first two equations in the framework model (4.10) from the 3rd equation, leading to the reduced model system (4.16) with two unknowns only. As is seen in Sections 4.3 and 4.4, this simplifies the analysis of the model to a certain extent. However, the above-mentioned "direct evolution" ignores the possible impact of the individuals' interactions, such as learning from each other's successes or peer pressure. In this section, we briefly discuss the issue of adaptation to a given strategy that is "best response" in some sense.

Let $B(t)$ be a given "best response" (in some sense), and denote the current precaution level (strategy) by $X_p(t)$. An ideal situation is that the precaution level *instantaneously* adapts to the best strategy $B(t)$, that is, $X_p(t) = B(t)$. In reality, however, some transition time is involved,

which means individuals adapt their current response $X_p(t)$ *towards* the best response $B(t)$. Such an adaptation can be described by

$$X'_p(t) = \eta[B(t) - X_p(t)] \tag{4.44}$$

where $\eta > 0$ represents the speed of adaptation. Motivated by (4.12) and (4.13), we assume that the best response $B(t)$ satisfies the following condition:

(H) $B(t) \in [0, 1]$ and it is differentiable for $t \in R^+$.

With this condition, (4.44) leads to $-\eta X_p \leq X'_p(t) \leq \eta(1 - X_p)$, which further implies

$$X_p(0) \exp(-\eta t) < X_p(t) < 1 - (1 - X_p(0)) \exp(-\eta t).$$

Thus, we immediately obtain the following lemma, confirming that $X_p(t)$ preserves the properties stated in (H).

Lemma 4.5.1 *The precaution level function $X_p(t)$ defined by (4.44) satisfies $X_p(t) \in [0, 1]$ if $X_p(0) \in [0, 1]$, and it is also differentiable.*

With (4.44) governing the current strategy adapting toward the given strategy $B(t) = B(I(t))$, the general framework model (4.10) is replaced by

$$\begin{cases} S'(t) = \Lambda - dS(t) - f_p(t)S(t) + rI(t), \\ I'(t) = f_p(t)S(t) - (d + r)I(t), \\ X'_p(t) = \eta\left(B(I(t)) - X_p(t)\right) \end{cases} \tag{4.45}$$

where

$$f_p(t) = \beta I(t)P(X_p(t), L(t)). \tag{4.46}$$

For (4.45), it is reasonable to assume

$$X_p(0) = B(I(0)) = X_0,$$

since there is no pressure to adjust the actual behaviour $X_p(0)$ at the 'germination' stage of the epidemics, which can be considered the best response level X_0.

Remark 4.5.2 *The governing equation (4.44) gives a "scenario of chasing the strategy $B(t)$":
at any $t_0 \in [0, \infty)$, if $B(t_0) < X_p(t_0)$ (resp. $B(t_0) > X_p(t_0)$), then $X_p'(t) < 0$ (resp. $X_p'(t) > 0$) for $t \in (t_0, t_0 + \delta)$ with $\delta \ll 1$, which means $X_p(t)$ is decreasing (resp. increasing) when $t \in (t_0, t_0 + \delta)$; that is, when the actual level is higher (lower) than the best response level at any fixed time point, the actual level will have a downward (upward) trend in a while (can be very short time though). This leads one to expect $X_p(t) - B(t) \to 0$ as $t \to \infty$. However, this expectation may not be true in general unless some extra condition is imposed on $B(t)$. To see this, we can solve (4.44), under the assumption that $B(t)$ is differentiable, to obtain*

$$X_p(t) = \left[\eta \int_0^t B(a)e^{\eta a} da + X_p(0) \right] e^{-\eta t}$$

$$= B(t) + [X_p(0) - B(0)]e^{-\eta t} - e^{-\eta t} \int_0^t B'(a)e^{\eta a} da. \tag{4.47}$$

Thus, $X(t) - B(t) \to 0$ as $t \to \infty$, provided that $B'(t) \to 0$ as $t \to \infty$ ((by applying the L'Hôpital's rule). When $B(t) = X_1(t) = X_1(I(t))$ given in (4.12), the requirement that $B'(t) \to 0$ as $t \to \infty$ is ensured by the condition $I'(t) \to 0$ as $t \to \infty$.

System (4.45) has a unique disease free equilibrium $\hat{E}_0 = (S_0, 0, X_0)$ with $S_0 = \Lambda/d$ and its reproduction number R_0 remains the same as in (4.20), this is because $P_E = P(X_P(0), 0) = P(X_0, 0)$ which is the same as in (4.19). The Jacobian matrix of (4.45) at $\hat{E}_0$ is calculated as

$$\begin{bmatrix} -d & -\beta P_E S_0 + r & 0 \\ 0 & \beta P_E S_0 - d - r & 0 \\ 0 & 0 & -\eta \end{bmatrix}, \tag{4.48}$$

from which we can conclude that $\hat{E}_0$ is locally asymptotically stable if $\beta P_E S_0 - (d + r) < 0$, and is unstable if $\beta P_E S_0 - (d + r) > 0$. This, together with the formula (4.20) for R_0 leads to the following theorem.

Theorem 4.5.3 *For (4.45), $\hat{E}_0$ is locally asymptotically stable if $R_0 < 1$, and it is unstable if $R_0 > 1$.*

Next, we discuss the endemic dynamics of (4.45) when $R_0 > 1$. As in Sections 4.3 and 4.4, we specify

$$P(X_p, L) = e^{-hLX_p}, \quad L = I. \tag{4.49}$$

Furthermore, for convenience of demonstration, in what follows, we choose $B(t)$ to be the best response given in (4.12), that is

$$B(t) = B(I(t)) = X_1(I(t)) = \frac{1}{1 + C_1 e^{-\epsilon I(t)}},$$

and if considering the individual's interaction, we get the adjusted version $B(t) = B(I, X_p)$.

It is easy to see that if $E_3^* = (S_3^*, I_3^*, X_p^*)$ is an endemic equilibrium of (4.45), then (S_3^*, I_3^*) solves the exactly the same system as for (S_1^*, I_1^*) (i.e, (4.23)), with

$$X_p^* = \frac{1}{1 + C_1 e^{-\epsilon I_3^*}} = \frac{1}{1 + C_1 e^{-\epsilon I_1^*}} = X_1^*.$$

As is shown in Section 4.3, $E_3^* = (S_3^*, I_3^*, X_p^*)$ with $(S_3^*, I_3^*) = (S_1^*, I_1^*)$ exists if $R_0 > 1$. For its stability, we calculate the Jacobian matrix of (4.45) at E_3^* as

$$\begin{bmatrix} -d - \frac{mI_3^*}{S_3^*} & hX_p^* mI_3^* - d & hm(I_3^*)^2 \\ \frac{mI_3^*}{S_3^*} & -hX_p^* mI_3^* & -hm(I_3^*)^2 \\ 0 & \eta\epsilon X_p^*\left(1 - X_p^*\right) & -\eta \end{bmatrix}.$$

From this matrix, the characteristic equation is calculated as

$$(d + \lambda)f_p(\lambda) = 0$$

where

$$f_p(\lambda) = \lambda^2 + u_1\lambda + u_0$$

with

$$u_1 = hX_p^* mI_3^* + \eta + \frac{mI_3^*}{S_3^*} > 0,$$

$$u_0 = \frac{mI_3^*\eta\left(1 + X_p^* h\left(1 + I_3^*\left(1 - X_p^*\right)\epsilon\right)S_3^*\right)}{S_3^*} > 0. \tag{4.50}$$

Hence, all roots of the characteristic equation have negative real parts. This leads to the following theorem confirming the stability of the endemic equilibrium E_3^* for (4.45) as long as it exists (i.e., if $R_0 > 1$).

Theorem 4.5.4 *When $R_0 > 1$, the system (4.45) has a unique endemic equilibrium E_3^*, which is locally asymptotically stable.*

Remark 4.5.5 *Now applying the results in Theorems 4.5.3 and 4.5.4 to the second equation in (4.45), we concluded that when $R_0 < 1$ and $I_0 = I(0)$ is small, or when $R_0 > 1$ and I_0 is close to I_1^*, then the right hand side of the second equation in (4.45) tends to 0 as $t \to \infty$ and hence $I'(t) \to 0$ as $t \to \infty$; and by Remark 4.5.2, this implies $X_p(t) - B(t) \to 0$ as $t \to \infty$, meaning that $X_p(t)$ not only adapts toward $B(t)$ but actually approaches $B(t)$. Unfortunately, we are unable (as of now) to expand the local stability of $\hat{E}_0$ and E_3^* in Theorems 4.5.3 and 4.5.4 to global stability, and hence, the convergence of $X_p(t) - B(t) \to 0$ as $t \to \infty$ is only in the local sense.*

The above results show that the long-time (asymptotic) disease dynamics for model (4.21) with the response $X(t)$ being the best response $B(t) = X_1(t)$ and that of model (4.45) with the response adapting toward $X_1(t)$ are essentially the same.

We point out that in a recent work, [17], Qiu et al. used the fraction $m(t)$ of the mask-wearing population to measure the collective (average) response in behaviour and discuss its evolution and impact by assuming the following adapting rule:

$$m(t)' = \underbrace{r}_{\text{Tracking rate}} (\underbrace{F(I, m)}_{\text{Want to wear a mask}} -m), \qquad (4.51)$$

with $F(I, m)$ being interpreted as the best behaviour strategy. The ideal scenario is that all of those who want to will wear masks immediately, without a transitional period, i.e.,

$$m = rF(I, m).$$

In the same line of tracking the best strategy, Morsky et al. [16] considered a general behaviour: NPIs-adopting. Denoting by p the fraction of individuals adopting NPIs (which serves as

another indicator of average behaviour response), the authors propose the following tracking rule for p:

$$p' = \underbrace{\epsilon}_{\text{Behavioral change rate}} (\underbrace{BR(I, p)}_{\text{Best response}} - p) \tag{4.52}$$

with $BR(I, p)$ being the best response given by

$$BR(I, p) = \frac{1}{1 + k \exp(f(I, p))} \tag{4.53}$$

which has a similar form to the expression (4.12) for $X(t)$. Accordingly, $p = BR(I, p)$ may describe the ideal response. Both works consider the adaptation to best response with formulas (4.51) and (4.52) holding a similar form as (4.44).

4.6   Conclusion and discussion

In this paper, by introducing the notion of *practically susceptible*, which is a fraction P of the *biologically susceptible* population and assuming that the fraction P depends on the severity L of the epidemic and the precaution level X of the public, we proposed a general framework model with the response level X involving the epidemic. We verified the well-posedness and confirmed the disease's dying out for the framework model under the assumption that the basic reproduction number $R_0 < 1$. For $R_0 > 1$, when the precaution level X is taken to be the instantaneous best response function X_1, the endemic dynamic is shown to be the dynamic of converging to the endemic equilibrium; while when the precaution level $X(t)$ is the delayed best response X_2; the endemic dynamic can be either convergence to the endemic equilibrium, or convergence to a periodic solution.

Generally, the basic reproduction number R_0 depends on the initial precaution level X_0. We point out that X_0 may be a reflection of many factors (such as cultures, traditions, ethics, lifestyle, genders, ages, professions, and even politics), which is a result of long-term evolution from the past. Hence, such a dependence of R_0 on X_0 is reasonable. This may explain why some infectious diseases can spread in one community (areas, regions, countries, etc.) but cannot spread in another. An interesting and novel finding is that, although R_0 does not depend on the initial disease prevalence $I_0 = I(0)$ and the threshold disease dynamics in terms of R_0 is

confirmed, the endemic equilibrium or endemic periodic solution under $R_0 > 1$ depends on I_0 (and X_0 as well).

In addition to the instantaneous adoption of the best response (with delay or without delay), we also employed some ideas in the two recent works [16, 17] to explore the adaptive disease dynamics, meaning that instead of *adopting the best response*, we assume the precaution level *adapts toward the best response*. Our analysis shows that there is no difference in the threshold long-term dynamics of the disease between the "adopting" and "adapting" if the target strategy $B(t)$ is the instantaneous best response X_1. For a general target strategy $B(t)$ satisfying condition (H), the difference in the endemic dynamics between "adopting" and "adapting" is not clear and remains an open problem.

We remark that here in this paper, we have chosen the "average precaution level $X(t)$" to avoid heterogeneity in precaution level, and the word "average" can also be termed "collective" as in some works (e.g., [16, 17]). In reality, responses to epidemics can be significantly different from individual to individual, from society to society and from government to government. The COVID-19 epidemic/pandemic has clearly demonstrated such big differences. Therefore, it would be more desirable to incorporate the heterogeneity in precaution level to disease models using response-structured equations. We leave this as future research projects.

Bibliography

[1] J. Cui, Y. Song and H. Zhu, The impact of media on the control of infectious diseases, *J Dyn. Diff. Eqns.*, **20(1)** (2008), 31-53.

[2] T. Cheng and X. Zou, A new perspective on infection forces with the demonstration by a DDE infectious disease model, *Math. Biosci. Eng.*, **19(5)** (2022), 4856-4880.

[3] O. Diekmann, M. C. M. De Jong, A. A. De Koeijer and R. Reijnders, The force of infection in populations of varying size: a modelling problem, *J. Biol. Syst.*, **3(2)** (1995), 519-529.

[4] D. Fudenberg and K. D. Levine, The theory of learning in games, *MIT press* **2**, 1998.

[5] M. G. M. Gomes, M. U. Ferreira, R. M. Corder, J. G. King, C. Souto-Maior, C. Penha-Goncalves, G. Goncalves, M. Chikina, W. Pegden and R. Aguas, Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold, *J. Theor. Biol.*, **540** (2022), 111063.

[6] J. Hofbauer and K. Sigmund, Evolutionary games and population dynamics, *Cambridge University Press*, 1998.

[7] H. Inaba, On a pandemic threshold theorem of the early Kermack-McKendrick model with individual heterogeneity, *Math. Popul. Stud.*, **21(2)** (2014), 95-111.

[8] J. M. Hyman and J. Li, Differential susceptibility epidemic models, *J. Math. Biol.*, **50(6)** (2005), 626-644.

[9] G. Katriel, The size of epidemics in populations with heterogeneous susceptibility, *J. Math. Biol.*, **65(2)** (2012), 237-262.

[10] D. I. Ketcheson, Optimal control of a SIR epidemic through finite-time non-pharmaceutical intervention, *J. Math. Biol.*, **83** (2021), 1-21.

[11] R. Liu, J. Wu and H. Zhu, Media/psychological impact on multiple outbreaks of emerging infections diseases, *Comput. Math. Methods Med.*, **8(3)** (2007), 153-164.

[12] T. Lorenzi, A. Pugliese, M. Sensi and A. Zardini, Evolutionary dynamics in an SI epidemic model with phenotype-structured susceptible compartment, *J. Math. Biol.*, **83(6-7)** (2021), 72.

[13] A. Li, Y. Wang, P. Cong, and X. Zou, Re-examination of the impact of some non-pharmaceutical interventions and media coverage on the COVID-19 outbreak in Wuhan, *Infect. Dis. Model.*, **6** (2021), 975-987.

[14] R. M. May and R. M. Anderson, Population biology of infectious diseases, *Nature*, **280(5722)** (1979), 455-461.

[15] C. C. McCluskey, Global stability for an SIR epidemic model with delay and nonlinear incidence, *Nonlin. Anal. RWA.*, **11** (2010), 3106-3109.

[16] B. Morsky, F. Magpantay, T. Day and E. Akcay, The impact of threshold decision mechanism of collective behaviour on disease spread, *Proc. Natl. Acad. Sci. U.S.A.*, **120(19)** (2023), e2221479120.

[17] Z. Qiu, B. Espinoza, V. V. Vasconcelos, C. Chen, S. M. Constantino, S. A. Crabtree and M. V. Marathe, Understanding the coevolution of mask-wearing and epidemics: A network perspective, *Proc. Natl. Acad. Sci. U.S.A.*, **119(26)** (2022), e2123355119.

[18] S. Ruan and W. Wang, *Dynamical behaviour of an epidemic model with a nonlinear incidence rate*, *J. Differ. Equ.*, **188(1)** (2003), 135-163.

[19] H. R. Thieme, Renewal theorems for some mathematical models in epidemiology, *J. Integral Equ.*, **8** (1985), 185-216.

[20] H. R. Thieme, Distributed susceptibility: A challenge to persistence theory in infectious disease models, *Discrete Contin. Dyn. Syst.*, **12(4)** (2009), 865-882.

[21] Y. Takeuchi, W. Wang, S. Nakaoka and I. Shingo, Dynamical adaptation of parental care, *Bull. Math. Biol.*, **71(4)** (2009), 931-951.

[22] W. Wang, Epidemic models with nonlinear infection forces, *Math. Biosci. Eng.*, **3** (2006), 267-279.

[23] X. Wang, L. Y. Zanette and X. Zou, Modelling the fear effect in predator-prey interactions, *J. Math. Biol.*, **73(5)** (2016), 1179-1204.

[24] X. Wang and X. Zou, Modeling the fear effect in predator-prey interactions with adaptive avoidance of predators, *Bull. Math. Biol.*, **79** (2017), 1325-1359.

[25] D. Xiao and S. Ruan, Global analysis of an epidemic model with non-monotone incidence rate, *Math. Biosci.*, **20** (2007), 419-429.

Chapter 5

Conclusions and Future Work

5.1 Conclusion

Triggered by psychological factors, people respond to changes in infectious disease. In this thesis, we incorporated human behaviour changes in mathematical-epidemiological models. We study feedback between the severity of infectious diseases and individuals' protective behaviours. These protective behaviours can lower individuals' susceptibility, hence a lower infection force. In addition, fluctuations in the epidemic severity, often measured by *prevalence*, can alter response levels over time. We present three general model frameworks with different focuses. The types of models consist of systems of ordinary, delay, partial, and integral differential equations. The mathematical results of these models have answered the aforementioned biological questions for which they were initially motivated. In addition, using mathematical simulations, we demonstrated how the human protective response affects disease dynamics in both the long term and short term and how it adaptively evolves.

We started by providing a new perspective on the infection force, which explains the existing notions of infection force in a biological sense and motivates new forms of infection force. Using this revised framework, we later incorporated into a classic SIRS model an infection force that depends on disease surveillance at the current time t and past time $t - \tau$, each given a weight (k_0 and k_1). Consequently, we obtain a new model consisting of delay differential equations. We identify the basic reproduction number R_0 and its threshold principles and study the bifurcation from endemic steady state E^* in relation to the two information-related

131

parameters: the "time delay" τ, the "weight" k_1. We identify the specific ranges of τ and k_1 for the following patterns, respectively: (1). endemic stability; (2). endemic oscillation. Most surprisingly, we found that *path of Hopf bifurcations* can be different with parameters involved in the model under different conditions. To place it commonly, the long-term effects of control measures change when infection reporting delay increases; therefore, timely reporting seems vital to disease management.

In **Chapter 3**, following the viewpoint of infection force in **Chapter 2**, we structure a general model with infection age and give examples with particular kernels. In this **Chapter**, we study the final size of epidemics featured by infection-age structure, considering the uptake of society precautions. Our result shows precautions reduce the population's eventual involvement in the epidemic.

In **Chapter 4**, we propose a general framework model with the response level involving the epidemic by introducing the notion of practically susceptible (i.e. a fraction of the biologically susceptible) given in the **Chapter 2**, and assuming that the fraction depends on the severity of the epidemic and the level of precaution of the public.

5.2   Future work

Hypotheses of the models in this thesis have been simplified and idealized for roughly three purposes: (1). to focus on the study of behaviour response and amplify the impacts of behaviour response; (2). to obtain more comprehensive mathematical results from these models and (3). to make these models easily extended or modified by adding additional mechanistic details to address different biological needs. In future, we can apply these model frameworks to specific epidemics or potentially infectious diseases to provide valuable information to policymakers or verify our theoretical and qualitative results.

As mentioned in **Chapter 2**, we only analyzed the local Hopf bifurcations. The dynamics of the model are much richer than we showed. In the mathematical insight, there is still a lot of work that can be done in global bifurcation analysis.

In **Chapter 3**, we suggest that the initial population of the infected tends to zero. The initial state of an epidemic is more complicated than is conjectured by our model. For instance, the

quantity of origin cases is not easy to track during pandemics.

As we pointed out in **Chapter 4**, to avoid analytical complexity, we assume that the population is homogenous in behaviour. In reality, they differ in response levels when adopting the NPI due to different cultures, societal norms, economic costs and regions, etc. Additionally, **Chapter 4** focuses on studying the mechanistic nature of the precaution level and does not provide the *systematical* approach to quantifying the level, which can be left as future work.

We leave these aforementioned as our future work.

Curriculum Vitae

Name:	Tianyu Cheng
Post-Secondary Education and Degrees:	University of Western Ontario London, ON, Canada 2019 - Present Ph.D. Candidate Shaanxi Normal University Xi'an, Shananxi, China 2016 - 2019 M.Sc. South China Normal University Guangzhou, Guangdong, China 2012 - 2016 B.Sc.
Honours and Awards:	Western Graduate Research Scholarships (2019-2023) Excellent graduate Student Scholarship(2016-2019)
Related Work Experience:	Teaching Assistant, University of Western Ontario (2019-2023) Research Assistant, University of Western Ontario (2019-2023)

Publications:

1. T. Cheng & X. Zou, A new perspective on infection forces with a demonstration by a DDE infectious disease model. *Math. Biosci. Eng* ., 19(5), (2022). 4856-4880.

2. T. Cheng, S. Tang, & R. A. Cheke, Threshold Dynamics and Bifurcation of a State Dependent Feedback Nonlinear Control Susceptible-Infected-Recovered Model, *J. Comput. Nonlinear Dyn.,* 14.7 (2019).

3. Q. Zhang, B. Tang, T. Cheng, & S. Tang, Bifurcation Analysis of a Generalized Impulsive Kolmogorov Model With Applications to Pest and Disease Control, *SIAM J. Appl. Math*, 80.4 (2020), 1796–1819.

4. T. Cheng, & X. Zou, On final and peak sizes of an epidemic with latency under some non-pharmaceutical interventions, in preparation.

5. T. Cheng, & X. Zou, Modelling the impact of precaution on disease dynamics and its evolution, submitted.