Journal of Mathematical Biology (2024) 89:1
https://doi.org/10.1007/s00285-024-02100-0

Mathematical Biology

Modelling the impact of precaution on disease dynamics and its evolution

Tianyu Cheng[1] · Xingfu Zou[1]

Received: 6 September 2023 / Revised: 15 February 2024 / Accepted: 18 April 2024

Abstract
In this paper, we introduce the notion of *practically susceptible population*, which is a fraction of the *biologically susceptible* population. Assuming that the fraction depends on the severity of the epidemic and the public's level of precaution (as a response of the public to the epidemic), we propose a general framework model with the response level evolving with the epidemic. We firstly verify the well-posedness and confirm the disease's eventual vanishing for the framework model under the assumption that the basic reproduction number $R_0 < 1$. For $R_0 > 1$, we study how the behavioural response evolves with epidemics and how such an evolution impacts the disease dynamics. More specifically, when the precaution level is taken to be the *instantaneous best response* function in literature, we show that the endemic dynamic is convergence to the endemic equilibrium; while when the precaution level is the *delayed best response*, the endemic dynamic can be either convergence to the endemic equilibrium, or convergence to a positive periodic solution. Our derivation offers a justification/explanation for the best response used in some literature. By replacing "adopting the best response" with "adapting toward the best response", we also explore the adaptive long-term dynamics.

Keywords Infectious disease · SIS model · Non-pharmaceutical interventions · Precaution · Severity of epidemics · Practically susceptible · Stability · Bifurcation

Mathematics Subject Classification 34D20 · 34k18 · 34K20 · 91A22 · 92B20 · 92D30

Dedicated to Professor Glenn Webb in the occasion of his 80th birthday. Partially supported by NSERC of Canada (RGPIN-2022-04744).

✉ Xingfu Zou
xzou@uwo.ca

[1] Department of Mathematics, University of Western Ontario, London, ON N6A 5B7, Canada

Published online: 06 May 2024

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., Gomes et al. (2022), Inaba (2014), Hyman and Li (2005), Katriel (2012), Ketcheson (2021), Lorenzi et al. (2021), May and Anderson (1979), Thieme (1985, 2009) 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 Cheng and Zou (2022), 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

$$\beta I(t)S_p(t) = \beta I(t)PS(t) = [\beta P I(t)] \cdot S(t),$$

leading to an overall infection force $f_m(t) = [\beta P I(t)]$. 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:

$$P(L) \text{ is non-increasing}, \quad P(0) \leq 1 \text{ and } P(\infty) \geq 0. \tag{1.1}$$

Below are some prototypes of such a fraction function satisfying (1.1):

(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, Cui et al. (2008), Liu et al. (2007), McCluskey (2010), Ruan and Wang (2003), Wang (2006), Xiao and Ruan (2007) 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

$$L(t) = \int_0^\tau w(\xi) I(t-\xi)\, d\xi \tag{1.2}$$

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

$$L(t) = \sum_{i=0}^{k} w_i I(t-\tau_i) \text{ with } 0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k. \tag{1.3}$$

Practically, the form (1.3) is more feasible because case reporting is done at discrete times in reality. In Cheng and Zou (2022), through a SIR model incorporated with the exponential decay function $P(L) = P_3(L) = e^{-hL}$ with $L(t)$ given by (1.3) 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 of the practical susceptible population, in addition to the severity of the epidemic, it should also depend on the response level of the public. Given the same severity $L(t)$, the response level may vary for different age groups (e.g., senior people tend to be more precautious than young people during an epidemic), genders, education levels, ethnic groups and regions/countries. Because of the sophistication of human's physiology and societal structures, response levels can be impacted by many factors and can evolve with the severity of the epidemics. This has particularly been demonstrated in the past COVID-19 pandemic. Thus, with the same severity L, different response levels may lead to different fractions P. Such a response level is heterogeneous, differing from individual to individual in a community. 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 (1.1) is revised to

$$\begin{cases} \dfrac{\partial P(X,L)}{\partial X} < 0, \quad \dfrac{\partial P(X,L)}{\partial L} < 0, \\ P(0,L) = 1, \quad P(1,L) \geq 0, \quad P(X,\infty) \geq 0, \quad P(X,0) \leq 1. \end{cases} \tag{1.4}$$

We point out that from their meanings described above, the severity $L(t)$ is an epidemiological notion while the response level is a behavioural factor.

We also point out that the incorporation of such a fraction function $P(X, L)$ is motivated by the recent works Wang et al. (2016) and Wang and Zou (2017) 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., Ketcheson 2021; Li et al. 2021 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. (2022) use the fraction of the mask-wearing population as a measurement of response to an epidemic, and Morsky et al. (2023) 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 Sect. 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 3 and 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 Sect. 2 when formulating the framework model, aiming to demonstrate the feasibility of the general framework. Section 3 adopts an instantaneous form for $M(t)$ but with infection force f_m, resulting in a system of ordinary differential equations (ODEs), while Sect. 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 Sect. 5, we employ some ideas in Morsky et al. (2023), Qiu et al. (2022) to explore the disease dynamics when the response level is assumed to

be adapting to a given response function. Finally, in Sect. 6, we summarize the main results and discuss their implications in epidemiological and social contexts.

2 A general framework model and some preliminary results

Recall that a classic SIS model is given

$$\begin{cases} S'(t) = \Lambda - dS(t) - f(t)S(t) + rI(t), \\ I'(t) = f(t)S(t) - (d + r)I(t), \end{cases} \tag{2.1}$$

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

$$f(t) = f_m(t) = \beta I(t) P(X(t), L(t)) \tag{2.2}$$

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

$$\frac{dX}{dt} = \epsilon X(1 - X)M(t). \tag{2.3}$$

Note that the term $X(1 - X)$ ensures that X is enclosed in the interval $[0, 1]$ (Takeuchi et al. 2009), $\epsilon M(t)$ 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,

$$L(t) = I(t), \qquad \text{or} \qquad L(t) = I(t - \tau) \ \text{ with } \ \tau > 0, \tag{2.4}$$

and consider the following two simple choices for $M(t)$

$$M(t) = M_1(t) = I'(t), \qquad \text{or} \qquad M(t) = M_2(t) = I'(t - \tau) \ \text{ with } \ \tau > 0. \tag{2.5}$$

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. Equation (2.3) is formally analogous to the replicator equations for evolutionary population games (Hofbauer and Sigmund 1998), yet the fitness is time-dependent.

Putting (2.1) and (2.3) together results in the following framework model

$$\begin{cases} S'(t) = \Lambda - dS(t) - f(t)S(t) + rI(t), \\ I'(t) = f(t)S(t) - (d+r)I(t), \\ X'(t) = \epsilon X(1-X)M(t). \end{cases} \tag{2.6}$$

where $f(t)$, $M(t)$ and $L(t)$ are given by (2.2), (2.4) and (2.5) with the involving function $P(X, L)$ satisfying (1.4).

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 (2.1). The proof is similar to that of Theorem 2.1 in Cheng and Zou (2022) and is thus omitted here.

From the last equation in (2.6), $X(t)$ can be expressed by $M(t)$ as

$$X(t) = \frac{1}{1 + C_0 e^{-\epsilon \int_0^t M(s)ds}} \tag{2.7}$$

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

$$X(t) = X_1(I(t)) = \frac{1}{1 + C_1 e^{-\epsilon I(t)}} \quad \text{with} \quad C_1 = C_0 e^{\epsilon I(0)} \tag{2.8}$$

or

$$X(t) = X_2(I(t-\tau)) = \frac{1}{1 + C_2 e^{-\epsilon I(t-\tau)}} \quad \text{with} \quad C_2 = C_0 e^{\epsilon I(-\tau)}. \tag{2.9}$$

We point out that the form (2.8) for the average response level function $X(t)$ is referred to as the "best response function" in some literature (see, e.g., Morsky et al. 2023), while (2.9) may be accordingly called the "delayed best response function." Actually, even the general response form (2.7) is similar to the so-called *Smoothed Best Response* (i.e the *Logit dynamic*) for *evolutionary population game* (see, e.g., Fudenberg and Levine 1998). This can be seen by rewriting (2.7) as

$$X(t) = \frac{\exp\left[\epsilon \int_0^t M(s)ds\right]}{\exp\left[\epsilon \int_0^t M(s)ds\right] + C_0} = \frac{X_0 \exp\left[\epsilon \int_0^t M(s)ds\right]}{X_0 \exp\left[\epsilon \int_0^t M(s)ds\right] + (1 - X_0)} \tag{2.10}$$

To some extent, (2.10) can be interpreted in terms of the "perspective" of the evolutionary population game. For example, if the response level X is considered as the fraction of the population "adopting" precaution behaviour, then X_0 is the initial fraction of the population with the "adopting" strategy, while $(1 - X_0)$ is the initial "non-adopting" fraction. Treating $U := \int_0^t M(s)ds$ as the total *relative* benefit (i.e. payoff gain) at time t and splitting it as

$$U = \int_0^t M(s)ds = q_a \int_0^t M(s)ds - \left(-q_n \int_0^t M(s)ds\right) =: U_a - U_n. \quad (2.11)$$

Here $q_a,\ q_n \in [0, 1]$ with $q_a + q_n = 1$ reflects the weights for "adopting" and "non-adopting" in relative benefit; accordingly $U_a = q_a U$ represents the relative benefit of "adopting" and $U_n = -q_n U$ accounts for the relative benefit of "non-adopting". With these notations, (2.10) is then further expressed as

$$\begin{aligned} X(t) &= \frac{X_0 \exp\left[\epsilon q_a \int_0^t M(s)ds\right]}{X_0 \exp\left[\epsilon q_a \int_0^t M(s)ds\right] + (1 - X_0) \exp\left[-\epsilon q_n \int_0^t M(s)ds\right]}. \\ &= \frac{X_0 \exp(\epsilon U_a)}{X_0 \exp(\epsilon U_a) + (1 - X_0) \exp(\epsilon U_n)}. \end{aligned} \quad (2.12)$$

which contains a memory component: the current response is affected by both the current benefits of the strategies and the initial choice/response states. We emphasize that the benefits here, i.e. U_a and U_n, are time-varying. We also note the parameter ϵ serves a similar role to the *sensitivity parameter* in *Logit choice*:

- when $\epsilon = 0$, the response is independent of the epidemics' severity, i.e. $X(t) \equiv X_0$; when $\epsilon \to \infty$, the response is certainly determined: $X(t)$ approaches 0 or 1, relying on the evolution of the epidemics' severity;
- Low (reps. High) ϵ indicates the response is less (highly) sensitive to the epidemic's severity.

Plugging (2.8) (resp. (2.9)) into the first two equations of (2.6), one be reduced as the following two-dimensional model:

$$\begin{cases} S'(t) = \Lambda - dS(t) - f(t)S(t) + rI(t), \\ I'(t) = f(t)S(t) - (d + r)I(t), \end{cases} \quad (2.13)$$

with $f(t)$, $M(t)$ and $L(t)$ are given by (2.2), (2.4) and (2.5) with the involving function $P(X, L)$ satisfying (1.4), (2.13) is indeed an *autonomous system* for the variables $S(t)$ and $I(t)$ for which

$$D = \{(S, I) \mid S \geq 0, I \geq 0,\ S + I \leq \Lambda/d\}. \quad (2.14)$$

is invariant.

Moreover, it is easy to see that (2.13) 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 (2.13) at E_0 to obtain

$$\begin{cases} S'(t) = -dS(t) - \beta I(t) P_E S_0 + rI(t) \\ I'(t) = [\beta P_E S_0 - (d + r)]I(t). \end{cases} \tag{2.15}$$

where

$$P_E = P(X_0, 0) \quad \text{for the case} \quad f(t) = f_m(t). \tag{2.16}$$

From (2.15), 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 (1.4), there holds $f(t) \le \beta P_E I(t)$. Thus, the second equation in (2.13) has the second equation in the linear system (2.15) as a comparison equation from above. By a comparison argument, we then conclude that E_0 is actually 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

$$R_0 = \frac{1}{d + r} \cdot \beta P_E S_0 = \frac{\beta P_E S_0}{d + r}. \tag{2.17}$$

Obviously, $R_0 < 1$ (resp. $R_0 > 1$) if and only if $\lambda_0 < 0$ (resp. $\lambda_0 > 0$).

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

Theorem 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*

$$\lim_{t\to\infty} X(t) = \frac{1}{1 + C_1} = \frac{1}{1 + C_0 e^{\epsilon I(0)}} \approx \frac{1}{1 + C_0} = X_0$$
$$\textit{(since } I(0) \textit{ is generally very small).}$$

By the above theorem, we just need to discuss the disease dynamics of (2.6) 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) $L(t) = I(t)$, $M(t) = I'(t)$ and $f(t) = f_m(t)$;

(B) $L(t) = I(t - \tau)$, $M(t) = I'(t - \tau)$ and $f(t) = f_m(t)$.

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 (2.13) becomes

$$\begin{cases} S'(t) = \Lambda - dS(t) - \beta I(t)P_1(I(t))S(t) + rI(t) := B_1, \\ I'(t) = \beta I(t)P_1(I(t))S(t) - (d + r)I(t) := B_2, \end{cases} \tag{3.1}$$

where the fraction for this case becomes $P(X, L) = P(X_1(I), I) =: P_1(I)$ with

$$P_1(I(t)) = \exp\left(\frac{-hI(t)}{1 + C_1 e^{-\epsilon I(t)}}\right), \tag{3.2}$$

For this 2-D ODE system, an endemic equilibrium is given by the intersection of the following two curves:

$$\begin{cases} S = \dfrac{\Lambda}{d} - I =: g_1(I) \\ S = \dfrac{d + r}{\beta} \exp\left(\dfrac{hI}{1 + C_1 e^{-\epsilon I}}\right) =: g_2(I) \end{cases} \tag{3.3}$$

Noting that $g_1(I)$ is decreasing $g_2(I)$ is increasing, and hence, (3.3) 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 (3.1) 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:

$$J = \begin{bmatrix} -\frac{mI_1^*}{S_1^*} - d & mhI_1^* X_1^* \left(\epsilon(1 - X_1^*)I_1^* + 1\right) - d \\ \frac{mI_1^*}{S_1^*} & -mhI_1^* X_1^* \left(\epsilon(1 - X_1^*)I_1^* + 1\right) \end{bmatrix}$$

$$\det(J) = \frac{\left(1 + S_1^* X_1^* h(1 + I_1^* \epsilon(1 - X_1^*))\right) mdI_1^*}{S_1^*} > 0, \quad tr(J) = -\frac{\det(J) + d^2}{d} < 0.$$

Thus, E_1^* is asymptotically stable.

We can actually prove that E_1^* is globally asymptotically stable. To this end, we consider the following Dulac function. $Q(S, I) = 1/[I\, P_1(I)],\ \ (S, I) \in D$. Then we have

$$\begin{aligned}
&\frac{\partial(QB_1)}{\partial S} + \frac{\partial(QB_2)}{\partial I} \\
&= \frac{\partial}{\partial S}\left(\frac{\Lambda - dS}{P_1(I)I} - \beta S + \frac{r}{P_1(I)}\right) + \frac{\partial}{\partial I}\left(\beta S - \frac{d+r}{P_1(I)}\right) \\
&= -\left(\frac{d}{P_1(I)I} + \beta\right) + \frac{(d+r)P_1'(I)}{P_1(I)^2}.
\end{aligned}$$

Differentiating $P_1(I) = P(X_1(I), I)$ with respect to I shows

$$\begin{aligned}
P_1'(I) = \frac{dP_1(I)}{dI} &= \frac{\partial P(X_1, I)}{\partial X_1}\frac{dX_1}{dI} + \frac{\partial P(X_1, I)}{\partial I} \\
&= \epsilon(1 - X_1)X_1\frac{\partial P(X_1, I)}{\partial X_1} + \frac{\partial P(X_1, I)}{\partial I} \\
&= \epsilon(1 - X_1)X_1 hP_1(I)[-I - X_1] = -P_1(I)hX_1(1 - X_1)(I + X_1) < 0.
\end{aligned}$$

Thus

$$\frac{\partial(QB_1)}{\partial S} + \frac{\partial(QB_2)}{\partial I} < 0, \quad \text{for } (S, I) \in D.$$

According to the Bendixson Dulac theorem, System (3.1) does not have endemic periodic orbits, implying that E_1^* is globally asymptotically stable. Hence, we have proved the following theorem.

Theorem 3.1 *Assume that $R_0 > 1$. Then (3.1) 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*

$$X_1^* = \lim_{t\to\infty} X(t) = \frac{1}{1 + C_1 e^{-\epsilon I_1^*}} \tag{3.4}$$

where

$$C_1 = C_0 e^{\epsilon I(0)} = \left(\frac{1}{X_0} - 1\right) e^{\epsilon I_0}. \tag{3.5}$$

From this theorem and Theorem 2.1, we can see that the specific disease model (3.1) 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 (3.3).

We highlight a new and interesting phenomenon for the above-stated global threshold dynamics: the globally stable endemic equilibrium E_1^* depends on the initial

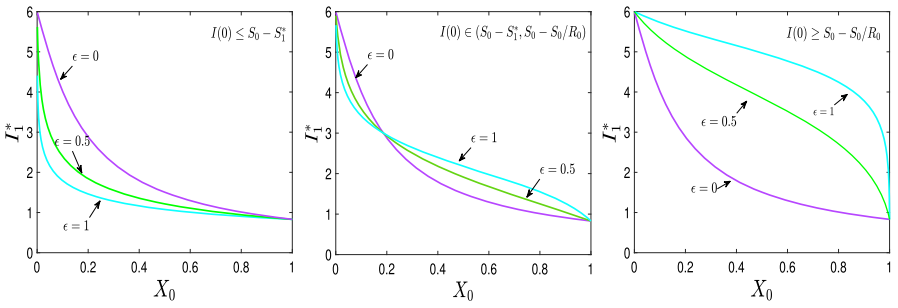

Fig. 1 While the endemic level I_1^* is always decreasing in X_0, its monotonicity on ϵ various with I_0 and X_0: **a** $I(0) \approx 0$; **b** $I(0) = 3$ and **c** $I(0) = 8$. Other parameters are chosen to be $\Lambda = 0.12$, $\beta = 0.1$, $h = 1$, $d = 0.012$, $r = 0.388$ and then $S_0 = 10$

prevalence $I_0 = I(0)$ and the initial response level $X_0 = X(0)$. This is because when reducing the 3-D system (2.6) to the 2-D system (3.1) through the best response function (2.8), the parameter C_1 in (2.8) depends on $X_0 = X(0)$ and $I_0 = I(0)$ as well as the evolution speed ϵ, so does I_1^*. Note from (3.3) and (3.5) 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 (3.3), 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

$$g_2(I_0) = \frac{d+r}{\beta} e^{hI_0X_0} < \frac{\Lambda}{d} - I_0 = g_1(I_0) \text{ for small } I_0.$$

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. 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. 1b, c. In Fig. 1a, I_0 is small and I_1^* is decreasing in ϵ. In Fig. 1b, 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. Figure 1c is with large value I_0 for which, I_1^* is increasing in ϵ.

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 delayed responses, which is common in reality. For this case, the general framework model (2.6) reduces to the following specific system of delay differential equations:

$$\begin{cases} S'(t) = \Lambda - dS(t) - \beta I(t)\left(e^{-hI(t-\tau)X_2(t)}S(t)\right) + rI(t), \\ I'(t) = \beta I(t)\left(e^{-hI(t-\tau)X_2(t)}S(t)\right) - (d+r)I(t), \end{cases} \tag{4.1}$$

where $X_2(t)$ is given by (2.9):

$$X_2(t) = \frac{1}{1 + C_2 e^{-\epsilon I(t-\tau)}} \quad \text{with} \quad C_2 = C_2(\tau) = C_0 e^{\epsilon I(-\tau)} = \left(\frac{1}{X_0} - 1\right) e^{\epsilon I(-\tau)}. \tag{4.2}$$

Rewrite (4.1) with (4.2) as

$$\begin{cases} S'(t) = \Lambda - dS(t) - \beta I(t)\left(P_2(I_\tau(t))S(t)\right) + rI(t), \\ I'(t) = \beta I(t)\left(P_2(I_\tau(t))S(t)\right) - (d+r)I(t), \end{cases} \tag{4.3}$$

with

$$P_2(I_\tau(t)) = P_2(I(t-\tau)) = \exp\left(\frac{-hI(t-\tau)}{1 + C_2 e^{-\epsilon I(t-\tau)}}\right). \tag{4.4}$$

When $\tau = 0$, (4.3)–(4.4) simply reduces to (3.1)–(3.2).

As for (3.1)–(3.2), an endemic equilibrium is an intersection of the following two curves:

$$\begin{cases} S = \dfrac{\Lambda}{d} - I =: \hat{g}_1(I) \\ S = \dfrac{d+r}{\beta} \exp\left(\dfrac{hI}{1 + C_2 e^{-\epsilon I}}\right) =: \hat{g}_2(I) \end{cases} \tag{4.5}$$

Note that (4.5) is the same as (3.3) 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 Sect. 3, we conclude that (4.3)–(4.4) 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

$$X_2^* = \lim_{t\to\infty} X(t) = \frac{1}{1 + C_2 e^{-\epsilon I_2^*}}. \tag{4.6}$$

In the sequel, we will analyze the stability of the endemic equilibrium E_2^*. Note that $C_2 = C_0 e^{\epsilon I(-\tau)}$ 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 (4.3) at E_2^* is given by

$$\begin{cases} S'(t) = -\left(d + \dfrac{mI_2^*}{S_2^*}\right) S(t) - d\, I(t) - \dfrac{mI_2^* P_2'(I_\tau)}{P_2(I_\tau)}\Big|_{I_\tau = I_2^*} I(t-\tau) \\ I'(t) = \dfrac{mI_2^*}{S_2^*} S(t) + \dfrac{mI_2^* P_2'(I_\tau)}{P_2(I_\tau)}\Big|_{I_\tau = I_2^*} I(t-\tau). \end{cases} \tag{4.7}$$

The corresponding characteristic equation(CE) can be derived as

$$F(\lambda, \tau) := (d + \lambda)\left(a_0 e^{-\lambda\tau} + \lambda + b_0\right) = 0, \tag{4.8}$$

where

$$\begin{aligned} a_0 &= mI_2^* h X_2^* (\epsilon(1 - X_2^*) I_2^* + 1) \\ &= \frac{mI_2^* h \left(1 + C_2\left(\epsilon I_2^* + 1\right) \mathrm{e}^{-\epsilon I_2^*}\right)}{\left(1 + C_2\, \mathrm{e}^{-\epsilon I_2^*}\right)^2} \quad \text{and} \quad b_0 = \frac{mI_2^*}{S_2^*}. \end{aligned} \tag{4.9}$$

When $\tau = 0$, the transcendental equation (4.8) reduces to the following polynomial:

$$F(\lambda, 0) = (d + \lambda)(\lambda + a_0 + b_0) = 0. \tag{4.10}$$

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

$$G(\lambda, \tau) = a_0 e^{-\lambda\tau} + \lambda + b_0. \tag{4.11}$$

Next, we discuss if there is a root of (4.11) 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 (4.11) and separate the real part and imaginary part of $G(i\omega, \tau) = 0$ to obtain

$$a_0 \cos(\omega\tau) = -b_0, \quad a_0 \sin(\omega\tau) = \omega. \tag{4.12}$$

Equivalently,

$$\cos(\omega\tau) = -\frac{b_0}{a_0}, \quad \sin(\omega\tau) = \frac{\omega}{a_0}. \tag{4.13}$$

From (4.13), $\frac{\omega}{a_0} = \sin(\omega\tau) \leq \omega\tau$, and thus $\tau \geq \frac{1}{a_0}$. This implies that when τ is small (e.g., $\tau < 1/a_0$), roots of (4.11) cannot cross the pure imaginary axis in the complex plane and, hence all roots still remain in the left half plane, ensuring the endemic equilibrium remains asymptotically stable when τ is small. To further investigate the critical value(s) of τ that result in roots of (4.11) crossing the pure imaginary axis, we square both equations in (4.13) and add the two resulting equations to obtain

$$\omega^2 = (a_0^2 - b_0^2) \tag{4.14}$$

which is valid if and only if $a_0 > b_0$.

Assuming $a_0 > b_0$, (4.13) and (4.14) defines a sequence of critical values for the delay parameter τ given by

$$\tau_n = \tau_0 + \frac{2n\pi}{\omega}, \quad \tau_0 = \frac{1}{\omega}\arccos\left(-\frac{b_0}{a_0}\right) \tag{4.15}$$

Taking derivative in $f_1(\lambda, \tau) = 0$ with respect to τ, we obtain

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = -\frac{\frac{\partial f_1(\lambda,\tau)}{\partial \tau}}{\frac{\partial f_1(\lambda,\tau)}{\partial \lambda}} = -\frac{\lambda(\lambda + b_0)}{1 + (\lambda + b_0)\tau}. \tag{4.16}$$

Since $f_1(i\omega, \tau_n) = 0$, then

$$\left.\frac{\mathrm{d}\,\mathrm{Re}\,(\lambda)}{\mathrm{d}\tau}\right|_{\tau=\tau_n} = \mathrm{Re}\left(\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\tau=\tau_n}\right) = \mathrm{Re}\left(-\frac{i\omega(i\omega + b_0)}{1 + (i\omega + b_0)\tau_n}\right) = b_0 + \omega^2 > 0. \tag{4.17}$$

This verifies the transversality condition at critical value τ_n, $n = 0, 1, 2, \ldots$.

Combining the above analysis, we have proved the following theorem based on the Hopf bifurcation Theorem for DDE.

Theorem 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 (4.9) and τ_0 satisfying $\tau_0 > 1/a_0$ is given in (4.15). Furthermore, system (4.1) undergoes Hopf bifurcation around E_2^* at $\tau = \tau_n$, $n = 0, 1, 2 \ldots$ where τ_n is given in (4.15).*

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 starts, 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

$$\begin{aligned}\frac{a_0}{b_0} &= -\left.\frac{S_2^* P_2'(I_\tau)}{P_2(I_\tau)}\right|_{I_\tau = I_2^*} = S_2^* h\left(\epsilon(1 - X_2^*)X_2^* I_2^* + X_2^*\right) \\ &= \frac{S_2^* h\left(1 + C_0\left(\epsilon I_2^* + 1\right)\mathrm{e}^{-\epsilon I_2^*}\right)}{\left(1 + C_0\,\mathrm{e}^{-\epsilon I_2^*}\right)^2} := f_{ab}(C_0, h).\end{aligned} \tag{4.18}$$

Note that $X_0 \to 0^+ \Longleftrightarrow C_0 \to \infty$ and $X_0 \to 1^- \Longleftrightarrow C_0 \to 0^+$; 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

$$h_{cr} = \frac{1}{S_0} \cdot W\left(R_0 e\right).$$

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 4.1.

Note that 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 $a_0/b_0 > 1$ can hold and hence, Hopf bifurcation can occur as τ passes those critical values. To illustrate this, we choose $\Lambda = 0.06$, $\beta = 0.2$, $d = 0.012$, $r = 0.388$ and $I(-\tau) = 0$. For these values $h_{cr} = W(R_0 e)/S_0 \approx 0.3014$. Observe that a_0/b_0 depends on ϵ as is seen directly from (4.18), while h_{cr} is independent of ϵ because both S_0 and R_0 are not affected by ϵ. This observation suggests us to numerically explore different values of ϵ, at which X_0 may affect a_0/b_0 in different ways. Figure 2 presents the numerical results for $\epsilon = 0.5$ and $\epsilon = 2$ respectively. Figure 2a with $\epsilon = 0.5$ 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 $a_0/b_0 > 1$. However, Fig. 2b with $\epsilon = 2$ demonstrates a totally different pattern of dependence of a_0/b_0 on X_0; particularly,

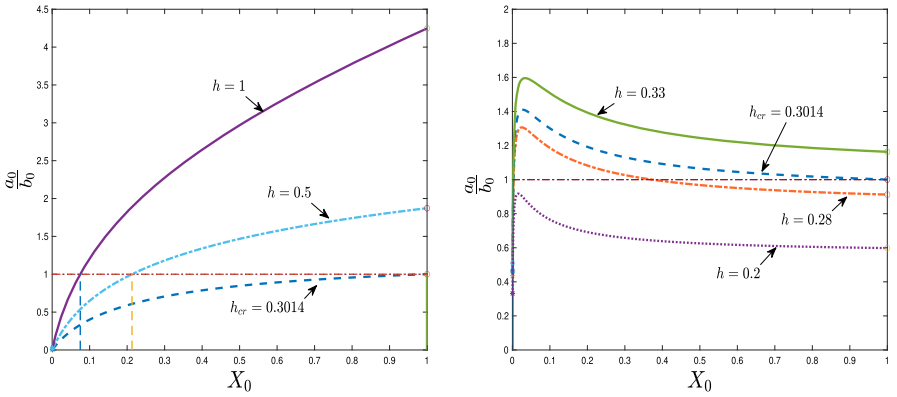

Fig. 2 Impact of X_0 on a_0/b_0 with different given values of h and ϵ: **a** $\epsilon = 0.5$; **b** $\epsilon = 2$. Other parameters are chosen to be $\Lambda = 0.06$, $\beta = 0.2$, $d = 0.012$, $r = 0.388$ and $I(-\tau) = 0$, leading to $h_{cr} = W(R_0 e)/S_0 \approx 0.3014$

when $h = 0.28 < 0.3410 = h_{cr}$, there is an intermediate range of X_0 in which $a_0/b_0 > 1$.

5 Adaptation toward the best response

In Sect. 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 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 (2.8), which is referred to as the best response function in some literature (e.g., Morsky et al. 2023; Qiu et al. 2022). Thus, the above-governing equation provides one way to explain/justify the best response strategy in Qiu et al. (2022) where the fraction of mask-wearing susceptible population is used as a type of precaution measurement, and the strategy in Morsky et al. (2023) 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 (2.6) from the 3rd equation, leading to the reduced model system (2.13) with two unknowns only. As is seen in Sects. 3 and 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

$$X_p'(t) = \eta[B(t) - X_p(t)] \tag{5.1}$$

where $\eta > 0$ represents the speed of adaptation. Motivated by (2.8) and (2.9), 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, (5.1) leads to $-\eta X_p \le X_p'(t) \le \eta(1-X_p)$, 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 5.1 *The precaution level function $X_p(t)$ defined by (5.1) satisfies $X_p(t) \in [0, 1]$ if $X_p(0) \in [0, 1]$, and it is also differentiable.*

With (5.1) governing the current strategy adapting toward the given strategy $B(t) = B(I(t))$, the general framework model (2.6) is replaced by

$$\begin{cases} S'(t) = \Lambda - dS(t) - f_p(t)S(t) + rI(t), \\ I'(t) = f_p(t)S(t) - (d + r)I(t), \\ X_p'(t) = \eta\left(B(I(t)) - X_p(t)\right). \end{cases} \tag{5.2}$$

where

$$f_p(t) = \beta I(t) P(X_p(t), L(t)). \tag{5.3}$$

For (5.2), 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 5.1 The governing equation (5.1) 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 that $X_p(t) - B(t) \to 0$ 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 (5.1), under the assumption that $B(t)$ is differentiable, to obtain

$$\begin{aligned} X_p(t) &= \left[\eta \int_0^t B(a)e^{\eta a}da + X_p(0)\right] e^{-\eta t} \\ &= B(t) + [X_p(0) - B(0)]e^{-\eta t} - e^{-\eta t}\int_0^t B'(a)e^{\eta a}da. \end{aligned} \tag{5.4}$$

Thus, $X_p(t) - B(t) \to 0$ as $t \to \infty$, provided that $B'(t) \to 0$ as $t \to \infty$ (by applying the L'Hôpital's rule, the term $e^{-\eta t}\int_0^t B'(a)e^{\eta a}da \to 0$ as $t \to \infty$). When $B(t) = X_1(t) = X_1(I(t))$ given in (2.8), 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 (5.2) 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 (2.17), this is because $P_E = P(X_P(0), 0) = P(X_0, 0)$ which is the same as in (2.16). The Jacobian matrix of (5.2) at $\hat{E}_0$ is calculated as

$$\begin{bmatrix} -d & -\beta P_E S_0 + r & 0 \\ 0 & \beta P_E S_0 - d - r & 0 \\ 0 & 0 & -\eta \end{bmatrix}, \tag{5.5}$$

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 (2.17) for R_0, leads to the following theorem.

Theorem 5.1 *For (5.2), $\hat{E}_0$ is locally asymptotically stable if $R_0 < 1$, and it is unstable if $R_0 > 1$.*

Next, we discuss the endemic dynamics of (5.2) when $R_0 > 1$. As in Sects. 3 and 4, we specify

$$P(X_p, L) = e^{-hLX_p}, \ \ L = I. \tag{5.6}$$

Furthermore, for convenience of demonstration, in what follows, we choose $B(t)$ to be the best response given in (2.8), 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 (5.2), then (S_3^*, I_3^*) solves the exactly the same system as for (S_1^*, I_1^*) (i.e, (3.3)), 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 Sect. 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 (5.2) 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

$$\begin{aligned} u_1 &= hX_p^* mI_3^* + \eta + \frac{mI_3^*}{S_3^*} > 0, \\ u_0 &= \frac{mI_3^*\eta \left(1 + X_p^* h \left(1 + I_3^* \left(1 - X_p^*\right)\epsilon\right) S_3^*\right)}{S_3^*} > 0. \end{aligned} \tag{5.7}$$

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 (5.2) as long as it exists (i.e., if $R_0 > 1$).

Theorem 5.2 *When $R_0 > 1$, the system (5.2) has a unique endemic equilibrium E_3^*, which is locally asymptotically stable.*

Remark 5.2 Now applying the results in Theorems 5.1 and 5.2 to the second equation in (5.2), 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 (5.2) tends to 0 as $t \to \infty$ and hence $I'(t) \to 0$ as $t \to \infty$; and by Remark 5.1, 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 5.1 and 5.2 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 (3.1) with the response $X(t)$ being the best response $B(t) = X_1(t)$ and that of model (5.2) with the response adapting toward $X_1(t)$ are essentially the same.

We point out that in a recent work (Qiu et al. 2022), Qiu et al. used the fraction $m(t)$ of mask-wearing population measure the collective (average) response in behaviour

and discuss its evolution and impact by assuming the following adapting rule:

$$m(t)' = \underbrace{r}_{\text{Tracking rate}} \left(\underbrace{F(I,m)}_{\text{Want to wear a mask}} - m \right). \tag{5.8}$$

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. (2023) 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:

$$p' = \underbrace{\epsilon}_{\text{Behavioral change rate}} \left(\underbrace{BR(I,p)}_{\text{Best response}} - p \right) \tag{5.9}$$

with $BR(I,p)$ being the best response given by

$$BR(I,p) = \frac{1}{1 + k\exp(f(I,p))} \tag{5.10}$$

which has a similar form to the expression (2.8) for $X(t)$. Accordingly, $p = BR(I,p)$ may describe the ideal response. Both works consider the adaptation to best response with formulas (5.8) and (5.9) holding a similar form as (5.1).

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. In addition, our model framework (2.7) suggests that time-varying and initial-state/choices-dependent strategy should be more thoroughly characterized for application to *evolutionary population game*.

Observe that, in general, 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 and 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 (Morsky et al. 2023; Qiu et al. 2022) 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., Morsky et al. 2023; Qiu et al. 2022). 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 by using response-structured equations. We leave this as future research projects.

Acknowledgements The authors would like to thank (a) The Mathematical Research Center (MRC) of the University of Pittsburgh, which is funded by the Dietrich School of Arts and Sciences and (b) NSF conference Grant DMS- 2304988 for the support to the authors for their participation in the 2023 Mathematical Ecology Conference in which the authors presented this work and received feedback, which helped improve the paper.

References

Cheng T, Zou X (2022) A new perspective on infection forces with the demonstration by a DDE infectious disease model. Math Biosci Eng 19(5):4856–4880

Cui J, Sun Y, Zhu H (2008) The impact of media on the control of infectious diseases. J Dyn Differ Equ 20(1):31–53

Diekmann O, De Jong MCM, De Koeijer AA, Reijnders R (1995) The force of infection in populations of varying size: a modelling problem. J Biol Syst 3(2):519–529

Fudenberg D, Levine KD (1998) The Theory of Learning in Games, vol 2. MIT Press, Cambridge

Gomes MGM, Ferreira MU, Corder RM, King JG, Souto-Maior C, Penha-Goncalves C, Goncalves G, Chikina M, Pegden W, Aguas R (2022) Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. J Theor Biol 540:111063

Hofbauer J, Sigmund K (1998) Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge

Hyman JM, Li J (2005) Differential susceptibility epidemic models. J Math Biol 50(6):626–644

Inaba H (2014) On a pandemic threshold theorem of the early Kermack–McKendrick model with individual heterogeneity. Math Popul Stud 21(2):95–111

Katriel G (2012) The size of epidemics in populations with heterogeneous susceptibility. J Math Biol 65(2):237–262
Ketcheson DI (2021) Optimal control of a SIR epidemic through finite-time non-pharmaceutical intervention. J Math Biol 83:1–21
Li A, Wang Y, Cong P, Zou X (2021) Re-examination of the impact of some non-pharmaceutical interventions and media coverage on the COVID-19 outbreak in Wuhan. Infect Dis Model 6:975–987
Liu R, Wu J, Zhu H (2007) Media/psychological impact on multiple outbreaks of emerging infections diseases. Comput Math Methods Med 8(3):153–164
Lorenzi T, Pugliese A, Sensi M, Zardini A (2021) Evolutionary dynamics in an SI epidemic model with phenotype-structured susceptible compartment. J Math Biol 83(6–7):72
May RM, Anderson RM (1979) Population biology of infectious diseases. Nature 280(5722):455–461
McCluskey CC (2010) Global stability for an SIR epidemic model with delay and nonlinear incidence. Nonlinear Anal RWA 11:3106–3109
Morsky B, Magpantay F, Day T, Akcay E (2023) The impact of threshold decision mechanism of collective behaviour on disease spread. Proc Natl Acad Sci USA 120(19):e2221479120
Qiu Z, Espinoza B, Vasconcelos VV, Chen C, Constantino SM, Crabtree SA, Marathe MV (2022) Understanding the coevolution of mask-wearing and epidemics: a network perspective. Proc Natl Acad Sci USA 119(26):e2123355119
Ruan S, Wang W (2003) Dynamical behaviour of an epidemic model with a nonlinear incidence rate. J Differ Equ 188(1):135–163
Takeuchi Y, Wang W, Nakaoka S, Shingo I (2009) Dynamical adaptation of parental care. Bull Math Biol 71(4):931–951
Thieme HR (1985) Renewal theorems for some mathematical models in epidemiology. J Integral Equ 8:185–216
Thieme HR (2009) Distributed susceptibility: a challenge to persistence theory in infectious disease models. Discrete Contin Dyn Syst 12(4):865–882
Wang W (2006) Epidemic models with nonlinear infection forces. Math Biosci Eng 3:267–279
Wang X, Zou X (2017) Modeling the fear effect in predator–prey interactions with adaptive avoidance of predators. Bull Math Biol 79:1325–1359
Wang X, Zanette LY, Zou X (2016) Modelling the fear effect in predator–prey interactions. J Math Biol 73(5):1179–1204
Xiao D, Ruan S (2007) Global analysis of an epidemic model with non-monotone incidence rate. Math Biosci 20:419–429