A nonlinear relapse model with disaggregated contact rates: Analysis of a forward-backward bifurcation

Jimmy Calvo-Monge [a, *], Fabio Sanchez [a, b], Juan Gabriel Calvo [a, b], Dario Mena [a, b]

[a] Escuela de Matemática, Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San José, 11501, Costa Rica
[b] Centro de Investigación en Matemática Pura y Aplicada, Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San José, 11501, Costa Rica

A B S T R A C T

Throughout the progress of epidemic scenarios, individuals in different health classes are expected to have different average daily contact behavior. This contact heterogeneity has been studied in recent adaptive models and allows us to capture the inherent differences across health statuses better. Diseases with reinfection bring out more complex scenarios and offer an important application to consider contact disaggregation. Therefore, we developed a nonlinear differential equation model to explore the dynamics of relapse phenomena and contact differences across health statuses. Our incidence rate function is formulated, taking inspiration from recent adaptive algorithms. It incorporates contact behavior for individuals in each health class. We use constant contact rates at each health status for our analytical results and prove conditions for different forward-backward bifurcation scenarios. The relationship between the different contact rates heavily influences these conditions. Numerical examples highlight the effect of temporarily recovered individuals and initial conditions on infected population persistence.

© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Epidemiological models serve as an essential tool for understanding disease dynamics. Many historical examples yield insightful results on how initial conditions and parameters alter the progression of an epidemic outbreak (Brauer, 2009); critical concepts developed in this setting, such as the R_0 reproductive number, work as threshold indicators for disease behavior. Modern epidemiological mathematics heavily use bread-and-butter SIR models (Anderson & May 1979), and current research efforts in this area are devoted to modifying the classical models, allowing them to capture all the intricacies of real-world disease dynamics, for example, better representation of social distancing phenomena, compliance conducts, economic conditions and other factors.

One effort in this area is related to studying human contact behavior. Contacts between individuals of different characteristics (health statuses, age groups …) constitute a key factor in disease spread (Zhang et al., 1126; Mossong et al., 1371; Lauro21). The need to study contact differences due to health status requires that classical models be modified. In classical settings, there is an implicit assumption of homogeneous behavior in each compartment (for example, among susceptible and infected individuals) through establishing constant or proportional contact rates. This approach hides different individuals' inherent characteristics and responses toward the disease's progress.

We now have a history of multiple efforts to deal with this problem. A first approach consists of specifying non-linear incidence rate functions by constructing functions that reflect the impact of the state of the model on the contact rates through time; for example (Li and Zhang, 1371; Liu et al., 1987; Xiao & Ruan, 2007; Hu et al., 2012; Hethcote and van den Driessche, 1007; Moghadas and Alexander, 1093), for models without relapse, and (Sanchez et al., 2023; Xiao & Tang, 2010) for models with non-linear relapse rates. In these cases, the general idea is to include functions of the form

$$g_{\kappa,\nu}(\cdot)I = \frac{\kappa I^p}{1 + \nu I^q},$$

as the incidence rate function for the disease, using positive constants κ, ν, p, q and with $I=I(t)$ the infected population size in time. Within relapse phenomena, very similarly, (Sanchez et al., 2023) proposes the function

$$g_{\kappa,\nu}(\cdot) = \frac{\kappa}{1 + \nu \frac{R}{N}}, \tag{1.1}$$

where $R=R(t)$ is the number of (temporarily) recovered individuals at time t, and N is the total population size. As we can see, with this approach, modelers usually specify functions that decrease when the epidemic burden is high. This makes them depend inversely on the sizes of infected or recovered populations in time. The subsequent analysis is commonly focused on the impact of the model hyper-parameters (constants such as κ, ν in the non-linear functions) on the system behavior.

Key analytical results can be obtained using this approach. They can tackle the problem of different behaviors among health classes: what we call the *epidemiological heterogeneity* of agents involved in the disease progression. As mentioned in (Funk et al., 2010), although these models are rich in dynamic and analytical properties, the exact contact dynamic behavior sometimes is not emphasized in their formulation. In recent years new interest has been placed in representing more dynamic information on contact rates from the study population. Particularly, there has been interest in the *economical heterogeneity* of individuals involved in epidemics and in the inclusion of utilitarian adaptive decisions individuals make within the development of the epidemic scenario.

The main contribution from (Fenichel et al., 2011) consists of devising a process in which contact rates for individuals in each health class can be updated simultaneously as disease changes. The idea consists of modeling the individuals as decision agents who consider their environment status and utility to decide optimal contact rates throughout time. This approach considers individuals' economic considerations when deciding how many contacts they should engage with at each time period. A detailed review of other proposals under the epi-economical approach is available in (Funk et al., 2010).

This recent technique allows the computation of contact rates alongside the progress of the disease through an optimization decision process performed by each individual. We call this procedure the *adaptive setting*. This has proved helpful in creating epidemiological models closer to the actual decision-making processes made by individuals. It has been applied to create more realistic settings and compare them with the classical formulation. For example, thanks to the use of the adaptive setting, there are novel insights on the true impact of asymptomatic individuals (Espinoza et al., 1974), a more intuitive understanding of final epidemic burden states in contrast to the classical results (Espinoza et al., 1127), and a deeper analysis on social distancing (Springborn et al., 1186). Analytical comparisons and conjectures for the adaptive setting can be found under the non-relapse case in (Morin et al., 1111).

The adaptive setting is not detached from the first approach. To compute contact rates adaptively, we must first define non-linear incidence rate functions that will use these contact rates. The formulation of non-linear incidence rate functions in the adaptive setting is commonly expressed in the form:

$$g(S, I, R) = \frac{C^s C^i N}{SC^s + IC^i + RC^r}, \tag{1.2}$$

where $C^h = C^h(S, I, R)$ (for $h \in \{s, i, r\}$) is the average number of contacts for each health status individual per time period, and N is the total population size. These contact rates might be functions that depend on the status of the disease S, I, R, especially when using the adaptive approach, where they are updated throughout the disease dynamics.

This adaptive setting constitutes a recent effort and offers a promising strategy to capture complex epi-economical phenomena better. Given its novelty, the literature on adaptive behavior has not been applied to non-linear relapse scenarios. Although several references propose non-linear relapse incidence rate functions for epidemiological differential equation models, the formulation (1.2) merits further analytical inspection in the relapse scenario. This paper uses this

formula for incidence rate functions to study a relapse model. We will examine the analytical impact of specifying contact rates using (1.2) and the repercussions on how to interpret these models. Our results will be framed in terms of the relations between the contact rates C^s, C^i, and C^r when they are assumed constant. In Section 2, we propose our model and explore its main analytical properties. We present our main theoretical results in Section 3, where bifurcation plots and local stability are considered; all mathematical proofs can be found in Section 6. Section 4 provides some numerical simulations of sensitivity to contact rates and initial conditions. A discussion of our main results can be found in Section 6.

2. Non-linear relapse rate model

We propose an epidemiological model with the presence of non-linear relapse behavior. Following (Sanchez et al., 2019) and (Sanchez et al., 2007), we consider three compartments of individuals: S (susceptible), I (infected), and R (recovered with the possibility of reinfection) and represent the model dynamics using the following system of equations,

$$
\begin{aligned}
\frac{dS}{dt} &= -g(\cdot)\beta\frac{SI}{N} + \mu N - \mu S, \\
\frac{dI}{dt} &= g(\cdot)\beta\frac{SI}{N} + \varphi\frac{RI}{N} - (\gamma + \mu)I, \\
\frac{dR}{dt} &= \gamma I - \varphi\frac{IR}{N} - \mu R.
\end{aligned}
\tag{2.1}
$$

$N = S + I + R$ is constant. The function $g(\cdot)$ is the incidence rate function of the model. From now on, we will take $g(\cdot)$ given by (1.2). This is the proposed relaxation of the burden of health status homogeneity, present in the classical formulation. The likelihood of infection when there is contact with an infected individual is given by β, the rate of recovery by γ, and the rate of reinfection represented by φ. We have a demographic exit and entrance rate for the system given by μ. The incidence rate function, $g(\cdot)$, represents the contact rate between susceptible and infected individuals, implying that $g(\cdot)\beta$ acts as the rate at which susceptible become infected.

We re-scale system (2.1) by substituting $s := \frac{S}{N}$, $i := \frac{I}{N}$ and $r := \frac{R}{N}$, to obtain the equivalent model

$$
\frac{ds}{dt} = -g(\cdot)\beta si + \mu - \mu s.
\tag{2.2a}
$$

$$
\frac{di}{dt} = g(\cdot)\beta si + \varphi ri - (\gamma + \mu)i.
\tag{2.2b}
$$

$$
\frac{dr}{dt} = \gamma i - \varphi ri - \mu r.
\tag{2.2c}
$$

The incidence rate function $g(\cdot)$ can also be re-scaled and substituted by:

$$
g(\cdot) = g(s, i, r) = \frac{C^s C^i}{sC^s + iC^i + rC^r}.
$$

Remark 2.1. In general, contact rates are functions that depend on the status of the disease, that is $C^h = C^h(S, I, R)$ for each $h \in \{s, i, r\}$. For the remainder of this article, we consider these functions constant. We aim to generalize mathematical results obtained in (Sanchez et al., 2023) and elucidate possible analytical properties of the adaptive algorithm in the relapse case. In this case, our mathematical analysis, including the calculation of R_0 and the determination of stable equilibria, will be greatly simplified. As will be seen shortly, all our results are greatly influenced by the ratios between the contact rates C^h.

3. Mathematical analysis

3.1. Basic reproductive number

Using the next generation matrix approach (Hethcote, 1007), we compute a basic reproductive number R_0 for this system. Here, it is simple to see that

$$R_0 = \frac{\beta}{\gamma + \mu} \lim_{(s,i,r) \to (1,0,0)} g(s,i,r) = \frac{\beta}{\gamma + \mu} C_0.$$

Thus, R_0 depends on C_0, the limit of the incidence function value when the system converges to the disease-free state. When all contact coefficients C^h are constant, then $C_0 = C^i$.

3.2. Finding equilibrium points

First, we study the disease-free equilibrium, where $(s(t), i(t), r(t)) = (1, 0, 0)$.

Theorem 3.1. *The disease-free equilibrium is stable if and only if $R_0 < 1$.*

Proof. . Note that the Jacobian matrix of system (2.2) is given by

$$J(s,i,r) = \begin{pmatrix} -\beta i(g_s s + g) - \mu & -\beta s(g_i i + g) & -\beta sig_r \\ \beta i(g_s s + g) & \beta s(g_i i + g) + \varphi r - (\mu + \gamma) & \beta sig_r + \varphi i \\ 0 & \gamma - \varphi r & -\varphi i - \mu \end{pmatrix},$$

where g_h is the partial derivative of g with respect to the variable $h \in \{s, i, r\}$. Taking the limit to the disease-free point, we get

$$\lim_{(s,i,r) \to (1,0,0)} J(s,i,r) = \begin{pmatrix} -\mu & -\beta C_0 & 0 \\ 0 & \beta C_0 - (\mu + \gamma) & 0 \\ 0 & \gamma & -\mu \end{pmatrix},$$

which has eigenvalues $\lambda_1, \lambda_2 = -\mu$ and $\lambda_3 = \beta C_0 - (\mu + \gamma)$. This point is stable if and only if $\lambda_3 < 0$, which is equivalent to $R_0 < 1$.

The case of epidemiological interest is when $R_0 > 1$, in which we study the existence of endemic equilibria. Initial calculations show that these points must be in the form $\left(1 - i^* - \frac{\gamma i^*}{\varphi i^* + \mu}, i^*, \frac{\gamma i^*}{\varphi i^* + \mu}\right)$. To find the value of i^* at the equilibrium, we can substitute this point into (2.2b), use that $r = 1 - s - i$ (as N is constant in this model) and obtain that i^* must satisfy the cubic equation $a_3 X^3 + a_2 X^2 + a_1 X + a_0 = 0$, whose coefficients are:

$$
\begin{aligned}
a_3 &= R_\varphi^2 R_0 - R_\mu R_\varphi^2 (1 - \kappa), \\
a_2 &= R_\varphi \left[R_0(1 - R_\varphi) + R_\mu(R_0 + R_\varphi) - R_\mu(1 - R_\mu)(1 - \theta) - R_\mu(1 + R_\mu)(1 - \kappa) \right], \\
a_1 &= R_\mu \left[R_0(1 - R_\varphi) + R_\varphi(1 - R_0) - (1 - R_\mu)(1 - \theta) + R_\mu R_\varphi - R_\mu(1 - \kappa) \right], \\
a_0 &= R_\mu^2 (1 - R_0),
\end{aligned}
\tag{3.1}
$$

where

$$\kappa = \frac{C^i}{C^s}, \quad \theta = \frac{C^r}{C^s}, \quad R_\mu = \frac{\mu}{\mu + \gamma}, \quad R_\varphi = \frac{\varphi}{\mu + \gamma}. \tag{3.2}$$

Mathematically, the model proposed in (Sanchez et al., 2023) can be seen as a special case of our model: if we use $C^s = C^i$ and $C^r = C^i(1 + \nu)$, we obtain $g(\cdot)$ given by (1.1). This cubic equation also becomes the generalization of the corresponding one obtained in (Sanchez et al., 2023). We also point out the biological interpretation of these contact quotients: κ represents the change expected in contacts made by an individual after it becomes infected, and θ compares the difference between the individual contacts before infection and after recovery.

We proceed to examine the behavior and existence of equilibria points based only on the disease parameters of the model (R_φ, R_μ), the infected individual response to the disease (R_0), and the relationship between the average contact rates between compartments ($\frac{C^i}{C^s}$ and $\frac{C^r}{C^s}$). In Figure (1) we use $\mu = 0.00015$, $\gamma = 0.0027$, $\beta = 0.00096$ and $\varphi = 0.044$, taken from simulations made in (Sanchez et al., 2023) and drug epidemic parameter estimation performed in (Sanchez et al., 2007). We create bifurcation plots for each R_0 and varying the quotients $\kappa = \frac{C^i}{C^s}$ and $\theta = \frac{C^r}{C^s}$. First, see Figure (1) for $\theta = 1.7$.

We divided these plots into four regions of interest for the basic reproductive number: $\mathcal{R}_1$ where no endemic equilibrium is attained, $\mathcal{R}_2$ where a stable endemic equilibrium, and another non-stable can be found, $\mathcal{R}_3$ where three possible equilibrium states can be found, one of which is stable, and $\mathcal{R}_4$ where there is just one stable, steady state.

Remark 3.2. The presence of this cubic phenomenon was first observed in (Sanchez et al., 2023). This case is of particular interest when analyzing the effect of the value of R_0 in epidemics with relapse. The small region in the $[1, 1 + \varepsilon(\kappa, \theta)]$ interval represents the possibility of having a very small stable equilibrium state of the disease even when the reproductive number is higher than 1. These simulations suggest the importance of the contact rate C^r in creating such a scenario.
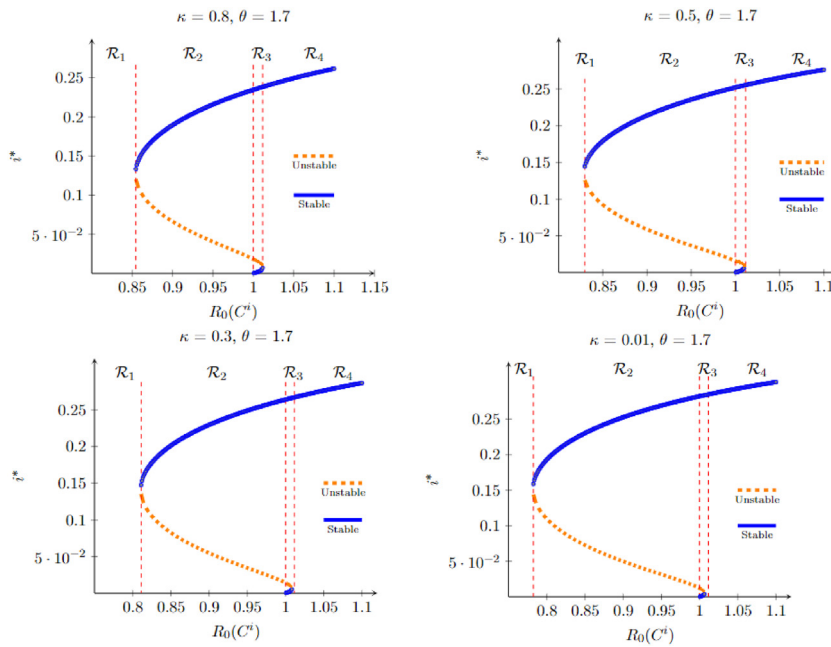
Fig. 1. Equilibria points computed using $\theta = 1.7$ and varying $\kappa = 0.8, 0.5, 0.3$ and 0.01. A cubic bifurcation plot can be found, and three equilibria points occur within an interval $R_0 \in [1, 1 + \varepsilon(\kappa, \theta)]$. We note that decreasing the value of κ diminishes the window $\varepsilon(\kappa, \theta)$, and it decreases the minimal R_0 value for which we find stable equilibria, for example, for $\kappa = 0.8$ this value is at $R_0 \approx 0.85$, but for $\kappa = 0.01$ it is at $R_0 \approx 0.8$. For a discussion on the length of the window $\varepsilon(\kappa, \theta)$, refer to Fig. 5 below. Stable and unstable equilibria regions are highlighted in these plots.
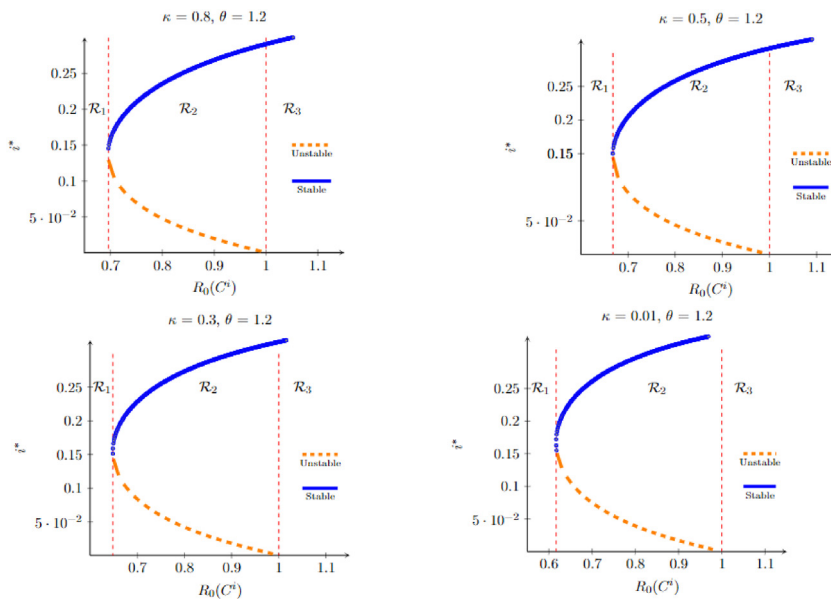

Fig. 2. Equilibria points computed using $\theta = 1.2$ and we varying $\kappa = 0.8, 0.5, 0.3$ and 0. We can see that no R_0 allows us to obtain three possible equilibria points.

On the other hand, decreasing the value of θ leads to a different behavior, as shown in Figure (2). In this case, we count three regions of interest exhibiting a typical backward quadratic bifurcation plot.

Remark 3.3. From the previous numerical results, we can derive the following conjectures:

a) A cubic bifurcation plot can be found for sufficiently high values of $\theta = \frac{C_s^i}{C_s}$, independently of κ. We observe an interval for R_0 in which those three equilibria points can be found, and it depends on (θ, κ).

b) For cubic bifurcation plots, a sensible region $[1, 1 + \varepsilon]$ could attain an endemic stable or a small non-stable equilibrium.
c) When θ is small, there is no cubic behavior for any κ, and for all $R_0 > 1$, there is only one possibility for an endemic equilibrium.

3.3. Theoretical results

Based on our previous simulations, we would like to formalize the conditions for the existence of regions that provide us with such cubic behavior. For that, we propose the following result.

Theorem 3.4. *Let μ, γ, φ be positive real numbers. Define R_μ and R_φ as in (3.2) and suppose that*

$$R_\varphi > \frac{1 + R_\mu^2}{(1 - R_\mu)^2}. \tag{3.3}$$

Then, there exist $0 < \theta_1 < \theta_2$ such that for every $\theta \in [\theta_1, \theta_2]$ and every $\kappa \in [0, 1]$, there is an $R_0 > 0$ such that the polynomial equation $a_0 + a_1 X + a_2 X^2 + a_3 X^3 = 0$ where $a_0, \dots, a_3$ are defined by (3.1), has three distinct real roots in the interval $[0, 1]$. Moreover, for each pair (κ, θ), this hold for all R_0 in a neighborhood of the form $[1, 1 + \varepsilon(\theta)]$.

In other words, there is a range of the fraction $\theta = \frac{C_s^3}{C_s^i}$ which yields a cubic bifurcation plot, under our condition (3.3), independently of the value of $\kappa = \frac{C^i}{C^s}$. We present the proof of this theorem, which uses the algebraic theory of Sturm chains. A preamble for this theory can be found in the Appendix.

Proof. Let us assume that the polynomial $f(X)$ has three different real roots. In this case, as discussed in proposition (6.6), the sequence of higher derivatives of $f(X)$ forms a Sturm sequence on any interval. This sequence is then

$$[a_0 + a_1 X + a_2 X^2 + a_3 X^3, a_1 + 2a_2 X + 3a_3 X^2, 2a_2 + 6a_3 X, 6a_3].$$

Its values at $x = 0$ and $x = 1$ are respectively $(a_0, a_1, 2a_2, 6a_3)$ and $(a_0 + a_1 + a_2 + a_3, a_1 + 2a_2 + 3a_3, 2a_2 + 6a_3, 6a_3)$. Our goal will be to find values for R_0 for which the signs of these sequences are $-, +, -, +$ at $x = 0$ and $+, +, +, +$ at $x = 1$. By Proposition (6.6), this would prove the existence of three different roots in the interval $[0, 1]$, since in this case $V_S(0) = 3$ and $V_S(1) = 0$. The reader can verify that for this to happen it is enough to have:

$$a_0 < 0, \quad a_2 < 0, \quad a_3 > 0, \quad a_0 + a_1 > 0, \quad \text{and } a_2 + a_3 > 0. \tag{3.4}$$

Note that if $R_0 > 1$ then $a_0 < 0$ and if $R_0 > R_\mu$, then for all $\kappa \in [0, 1]$ we have

$$a_3 = R_\varphi^2 [R_0 - R_\mu(1 - \kappa)] > R_\varphi^2 [R_0 - R_\mu] > 0.$$

Because $R_\mu < 1$, then $R_0 > 1$ is sufficient to ensure that $a_0 < 0$ and $a_3 > 0$.

We now move to a coordinate plane with R_0 in the x-axis and θ in the y-axis. The coefficients $a_0, \dots, a_3$ can be seen as linear equations in this plane, given by

$$a_3 = AR_0 - B(1 - \kappa)$$
$$a_2 = C + DR_0 - E(1 - \kappa) - F(1 - \theta)$$
$$a_1 = G + HR_0 - I(1 - \kappa) - J(1 - \theta)$$
$$a_0 = K(1 - R_0),$$

where the constants $A, B, \dots, K$ depend only on R_φ and R_μ and are given by

$$A := R_\varphi^2, \quad B := R_\mu R_\varphi^2, \quad C := R_\varphi^2 R_\mu, \quad D := R_\varphi(1 - R_\varphi) + R_\varphi R_\mu,$$
$$E := R_\varphi R_\mu(1 + R_\mu), \quad F := R_\varphi R_\mu(1 - R_\mu), \quad G := R_\varphi R_\mu(1 + R_\mu)$$
$$H := R_\mu(1 - R_\varphi) - R_\mu R_\varphi, \quad I := R_\mu^2, \quad J := R_\mu(1 - R_\mu) \quad K := R_\mu^2.$$

Note that all constants are positive, except perhaps D and H (we don't know the sign of $1 - R_\varphi$).

Geometrically, the inequalities in (3.4) refer to an area that is below the line $\ell_1 := \{a_2 = 0\}$ and above the lines $\ell_2 := \{a_2 + a_3 = 0\}$ and $\ell_3 := \{a_0 + a_1 = 0\}$ in the (R_0, θ) plane. Consider the intersections of these three lines with the $R_0 = 1$ vertical line. If we prove that the θ coordinate of the intersection of the ℓ_1 with this vertical axis is bigger than the θ coordinates of intersections of the other two (ℓ_2 and ℓ_3), then there would be an interval to the right of $R_0 = 1$ which is below the line ℓ_1 and above both ℓ_1 and ℓ_3. See the next figure for a visual intuition.
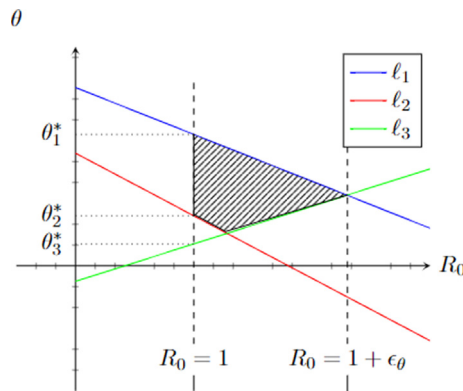
This is a specific example, and we don't know the signs of the slopes of lines ℓ_1, ℓ_2, and ℓ_3. However, their intersections at $R_0 = 1$ define the existence of a region below ℓ_1 and above both ℓ_2 and ℓ_3 in a neighborhood of $R_0 = 1$.

The intersection of line ℓ_1 at $R_0 = 1$ gives us

$$\theta_1^* = 1 - \frac{C+D}{F} + \frac{E}{F}(1-\kappa),$$

and the intersection of line ℓ_2 at $R_0 = 1$ gives us

$$\theta_2^* = 1 - \frac{C+D}{F} + \frac{E}{F}(1-\kappa) + \frac{B(1-\kappa)-A}{F} < \theta_1^* + \frac{B-A}{F},$$

we note that $B - A = R_\mu R_\varphi^2 - R_\varphi^2 = R_\varphi^2(R_\mu - 1) < 0$, so we indeed have $\theta_2^* < \theta_1^*$.

The θ coordinate of the intersection between ℓ_3 and $R_0 = 1$ is

$$\theta_3^* = 1 - \frac{G+H}{J} + \frac{I}{J}(1-\kappa) < 1 - \frac{G+H}{J} + \frac{I}{J}.$$

We observe that $\theta_1^* > 1 - \frac{C+D}{F}$ for all $\kappa \in [0, 1]$, so we would need

$$1 - \frac{C+D}{F} > 1 - \frac{G+H}{J} + \frac{I}{J},$$

to guarantee $\theta_3^* < \theta_1^*$ for all κ. This is equivalent to $(G + H - I)F > (C + D)J$, and by expanding this expression, we get the inequality (3.3) from our statement. Note that this region can be obtained independently of κ. Inequality (3.3) will imply that $\frac{C+D}{F} < 1$ and $\frac{G+H-I}{J} < 1$, and these are readily seen to imply that $\theta_i^* > 0$ for $i = 1, 3$. Therefore the desired interval for θ can be taken between $\max\{\theta_3^*, \theta_2^*\}$ and θ_1^*.

This proves that when (3.3) is true, and our polynomial $f(X)$ doesn't have repeated roots, there is a region in the (R_0, θ) plane for which $f(X)$ has three real distinct roots in the interval $[0, 1]$.

Remark 3.5. We note that both R_0 and the coefficients $a_0, \ldots, a_3$ are dependent, simultaneously, on the values of C^h. If we fix the values of κ and θ, all results will depend only on one of the C^h's.

In the next section, we will explore numerical results related to this theorem. In our examples, we care about the $C^i < C^s$ scenario ($\kappa < 1$). In this case, we assume that infection decreases contacts, which is intuitive. The other option, $\kappa > 1$, makes biological sense when infected populations are large because, under those circumstances, susceptible individuals may be overly cautious about engaging in contact with others (Morin et al., 1111). The case $\kappa > 1$ is more stable, as evidenced by the following theorem. Because of this, we focus on the $\kappa < 1$ scenario from now on.

Theorem 3.6. *If $C^i > C^s$ then (2.1) has no limit cycles in the region $\{(i, \tilde{s}) \in \mathbb{R}^2 : i > 0, \tilde{s} > 0\}$.*

proof. *We use a similar technique as in (Blythe et al., 1048). Writing $s = 1 - i - \tilde{s}$, we obtain the two-variable system:*

$$\frac{di}{dt} = g(1 - i - \tilde{s}, i, \tilde{s})\beta i(1 - i - \tilde{s}) + \varphi \tilde{s}i - (\gamma + \mu)i = g_1(i, \tilde{s})$$

$$\frac{d\tilde{s}}{dt} = \gamma i - \varphi \tilde{s}i - \mu \tilde{s} = g_2(i, \tilde{s}).$$
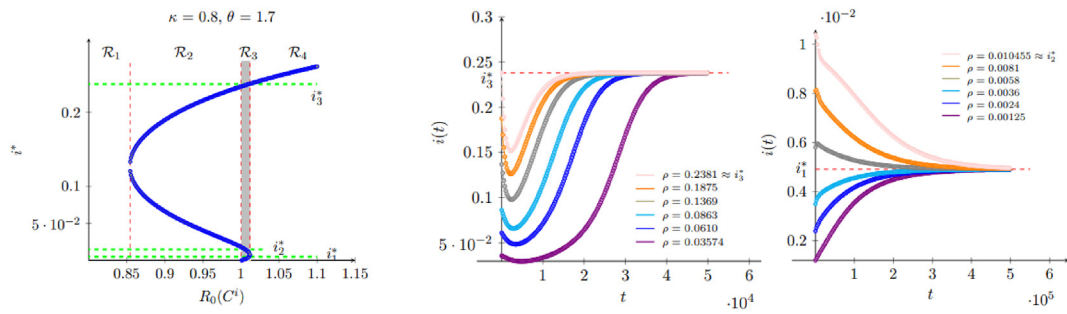
$$(3.5)$$

Fig. 3. Convergence of $i(t)$ to the equilibrium point $i*$. On the left, the bifurcation plot was obtained for this case, with region $\mathcal{R}_3$ highlighted. The center plot shows cases of convergence to the maximum possible equilibrium point within this region. This happens when the initially infected proportion is high enough. The plot on the right shows cases of convergence towards the smallest equilibrium point in this region, obtained for sufficiently small values of $i(0)$.

Then we have that

$$\frac{g_1(i,\tilde{s})}{i\tilde{s}} = g(1-i-\tilde{s},i,\tilde{s})\beta\left(\frac{1-i-\tilde{s}}{\tilde{s}}\right) + \varphi - (\gamma+\mu)\frac{1}{\tilde{s}},$$

$$\frac{g_2(i,\tilde{s})}{i\tilde{s}} = \frac{\gamma}{\tilde{s}} - \varphi - \frac{\mu}{i}.$$

This implies that

$$\frac{\partial}{\partial i}\left(\frac{g_1(i,\tilde{s})}{i\tilde{s}}\right) + \frac{\partial}{\partial\tilde{s}}\left(\frac{g_2(i,\tilde{s})}{i\tilde{s}}\right) = \left[\left(\frac{\partial g}{\partial i} - \frac{\partial g}{\partial s}\right)\beta\left(\frac{1-i-\tilde{s}}{\tilde{s}}\right)\right] - \frac{\beta g(1-i-\tilde{s},i,\tilde{s})}{\tilde{s}} - \frac{\gamma}{\tilde{s}^2}.$$

This is negative when $\frac{\partial g}{\partial i} - \frac{\partial g}{\partial s} < 0$, and using the definition of $g(\cdot)$ given by (1.2), this is equivalent to $C^s < C^i$. Applying the Dulac criterion, we obtain the non-existence of limit cycles in the region $\{(i,\tilde{s}) \in \mathbb{R}^2, i > 0, \tilde{s} > 0, i+\tilde{s} < 1\}$ for this system of differential equations.

4. Numerical results

4.1. Stable equilibrium points

We explore the equilibrium results in a simulation of disease scenarios. Let us consider Figure (1) with the case $\kappa = 0.8$ and $\theta = 1.7$ (and all the other model parameters as in that example). Using $C^i = 3$ we obtain $R_0 \approx 1.01057$. This R_0 is found in the $\mathcal{R}_3$ region in the bifurcation plot. Solving the corresponding cubic equation, we obtain three possible theoretical equilibrium points: $i^* \in \{i_1^* = 0.004914, i_2^* = 0.010455, i_3^* = 0.238099\}$.

Of these possibilities, i_3^* and i_1^* are asymptotically stable equilibrium points. The system could converge to each point depending on its initial conditions. The middle point, which is unstable, actually works as a threshold value as solutions drift away from it. If we take initial conditions $a_0 = (S(0), I(0), \tilde{S}(0))$ with $N = S(0) + I(0) + \tilde{S}(0)$, and let $i(0) = \frac{I(0)}{N}$, then the highest equilibrium will attract all solutions when $i(0) > i_2^*$, otherwise it is the lowest equilibrium to which the system converges.

The following graphs show cases for convergence to each equilibrium point. Figure (3) displays $i(t)$ through time using initial conditions $a_0 = (N - \rho N - 10, \rho N, 10)$, where $\rho \in [0, 1]$ and $N = 10000$. On the left are some simulations using $\rho > i_2^*$, in which the system converges to i_3^*, the highest equilibrium possible. On the right, a system is solved with $\rho < i_2^*$, where the final point obtained is i_1^*, although with a much slower convergence rate. We included the bifurcation plot on the left, highlighting the region of interest.

4.2. Effect of (κ, θ)

Now we explore how the values of κ and θ affect the size of the final equilibrium points.[1] To compare within a given R_0, we fix $C^i = 3$ (thus obtaining $R_0(C^i)$ as the examples above) and vary the values of C^s and $C^{\tilde{s}}$ using κ and θ respectively, we set $\kappa \in [0, 1]$ and $\theta \in [0, 2]$. Figure (4) shows the effect of increasing both values on the stable steady states attained in the model for two different initial conditions. We observe that increasing κ or θ yields a reduction in the final equilibrium of the system.

---

[1] For each point on the grid, the convergence speed of the system varies. Each simulation was performed at a point in time when the difference between successive time states fell below a machine precision threshold.
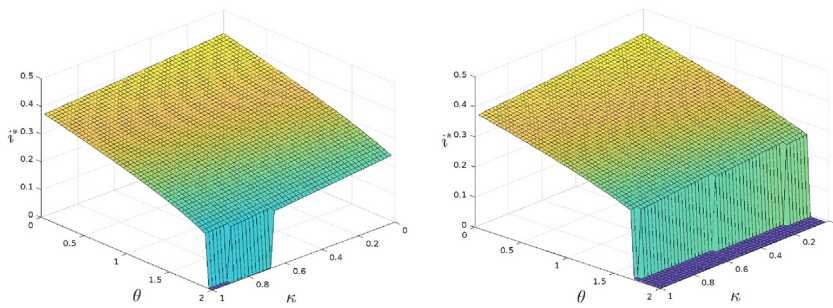
Fig. 4. Effect of (κ, θ) for two different initial condition scenarios. On the left using $i(0) = 0.1$, on the right $i(0) = 0.02$. For low initial infected populations, a high value of θ yields disease eradication, independently of κ.

However, we note that the value of this state is more sensible to θ than κ, indicating that, in the relapse case, contacts made by recovered individuals have a stronger impact on the disease outcome. Furthermore, we can see that high values of (κ, θ) may induce the system to attain a semi-disease-free steady state. This state, naturally, is more likely to be obtained when $i(0)$ is small, as seen by comparing both cases in Figure (4). In other words, a considerable infected population makes this population more likely to become established.

Region $\mathcal{R}_3$ offers the most interesting behavior. In other regions, the disease is either maintained at a high steady prevalence or eradicated. In region $\mathcal{R}_3$, there is the possibility of a low equilibrium state, without achieving the disease disappearance from the population. However, the window for this behavior is small. For $\theta > \theta_1$ (as in Theorem 3.4), we find a window $[1, R_{0,\max}(\kappa, \theta)]$ which defines region $\mathcal{R}_3$, the next figure shows the upper limit of this interval, depending on (κ, θ). We can again infer a similar situation. This window becomes larger when $\|(\theta, \kappa)\|$ increases, however, the effect of θ is more prominent. In this case, $\theta_1 \approx 1.4$.

Remark 4.1. Note that in these simulations, we decided to place less focus on the disease parameter β. The reason for this lies in the scale of the incidence rate function $g(\cdot)$. By using contact rates $C^i = 3$ and $C^s = C^i/\kappa$, $C^r = C^i/\theta$, the scale forces us to reduce the value of β to work with incidence rates that produce valuable epidemic scenarios. For example, if we use $\beta = 0.0096$, as in (Sanchez et al., 2007, 2023), this scenario would give us a basic reproductive number of $R_0 \simeq 10.105$ using these contacts, an exceedingly high and unrealistic number in many applications, which brings the model to a biological scenario with a single stable equilibrium point, resulting of an overestimation of the incidence term $\beta g(\cdot) SI/N$ of the system. Therefore, we note the importance of keeping in mind the scales of incidence rate functions when using the contact information in incidence rate functions for these models.

4.3. Other examples

In this subsection, we consider some numerical results and discussions on possible extensions of the contact rate disaggregation approach taken in the present study.

Example 4.2. (Disaggregated contacts for Influenza, a non-relapse case). We study the effect of disaggregated contact rates in epidemics models for individual-based transmitted diseases such as influenza. In this example, we use parameters for influenza transmission based on estimations performed in (Chowell et al., 2008). We then consider model 2.1 with the following epidemic parameters.

- $\beta = 0.07943065$,

- $\gamma = 0.243902 = 1/4.1$ (equivalent to 4.1 recovery period as mentioned in [24]),

- $\varphi = 0$, and $\mu = 0.0005$.

The estimation of the β infection parameter was performed as follows. For influenza (Chowell et al., 2008), obtains a natural reproduction number estimation, $R_p \simeq 1.3$ for influenza seasons in different countries from 1972 to 1997. The number R_p is defined as $R_p := R_0(1 - p)$ where p is assumed to be a proportion of susceptible individuals that have been successfully immunized before an epidemic. We perform simulations with R_0 obtained using $p = 0.2$, and $C^i = 5$, giving $R_0 \approx 1.625$.

Although this model presents a non-relapse scenario, we can still incorporate the contact information and obtain similar numerical results as before. For example, as observed in Fig. 6, the effect of θ over the peak epidemic prevalence of the model seems to be stronger than the effect of κ, thus indicating a similar behavior in the non-relapse case in terms of the contact proportions κ, θ. In this non-relapse scenario, our focus shifts towards the peak prevalence, as the final equilibrium will have a null infected population (see Fig. 7).
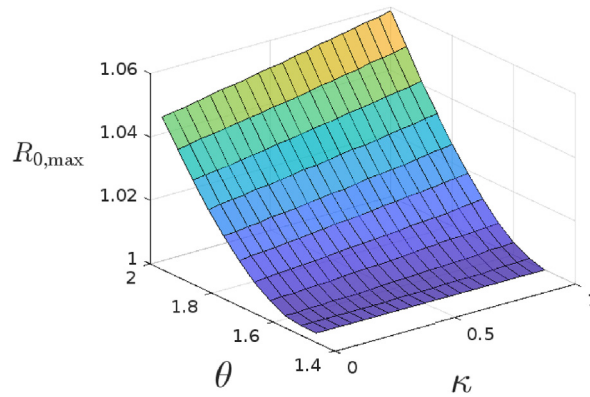
Fig. 5. Effect of (κ, θ) on the $\mathcal{R}_3$ window length. When θ increases, there is more room to observe this region with more unstable behavior.
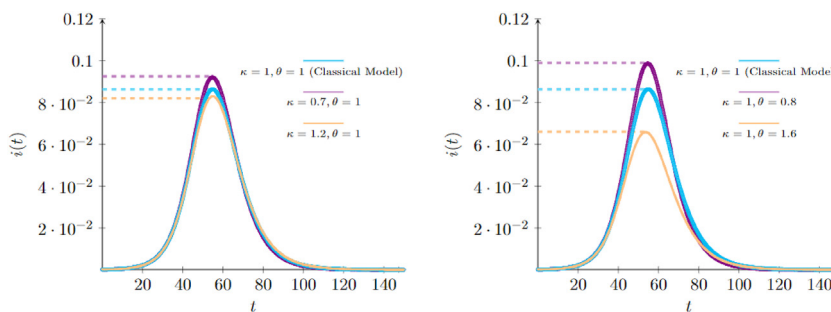

Fig. 6. Different infected results varying both contact proportions κ and θ. We keep $\theta = 1$ on the left and vary the proportion κ. Conversely, we keep $\kappa = 1$ on the right and vary the value of θ. We observe a bigger effect on the peak prevalence obtained in the figure on the right, that is, varying θ.

We expand these results for several combinations of values for (κ, θ) in the surface plot in Figure (7). We obtain a parallel situation as in the relapse simulations: we observe a bigger slope in the θ axis than in the κ axis.

We note that, in general, a reduction in peak prevalence is expected using the disaggregated contact rates in comparison to the classical SIR model. For the non-relapse case, this was pointed out within simulations performed in (Fenichel et al., 2011) and also expanded by the results of this example; for the non-relapse case this observation has been supported by our simulations in the previous sections.

Example 4.3.   (Disaggregated contacts approach for more complex models: Discussion).

Incidence rate functions control the entrance of susceptible individuals into the infected population. In the classical SIR model, this occurs because of contact with other infected individuals. This is reflected in the numerator of the incidence rate function, being $C^s C^i N$, consisting of the susceptible contacts multiplied by the contribution of infected contacts.[2] The denominator constitutes the total population activity, given by $\sum_h C^h h$, where the sum is over all possible health statuses, this yields the incidence rate as a proportion of the contact activity of the infected in terms of the total contact activity.

When adding new compartments or modifying the formulation of the model and trying to use the disaggregated contact approach discussed in the present study, this incidence rate formula should follow the same pattern: for the numerator to consider the contact information of compartments that might cause infection to susceptible individuals, and the denominator to reflect the total activity. We discuss some cases of how to apply this method to introduce such functions into more general epidemiological models.

First, consider the model proposed in (Espinoza et al., 1127). Here, the authors provide a modified SIR model of risk-taker (S_1) and risk-evader (S_2) susceptible to study COVID-19 epidemic scenarios. This non-relapse epidemic model was fitted using the adaptive approach, which implies the use of contact-disaggregated incidence rate functions. For each group of susceptible, the infection might come as a result of contact either with infectious exposed -both risk-takers and risk-evaders (E_1, E_2)-, infectious asymptomatic -both risk-takers and risk-evaders (A_1, A_2)- and infected symptomatic (I). Therefore, the incidence rate for both susceptible compartments is given by

---

[2] We could also argue that a better formulation for this denominator would be $C^s C^i I$, by reducing the effect only to consider the volume of the infected population. Looking at equation (4.1), we see that (Espinoza et al., 1127) follows this idea. We decided instead to use the first formulation as it was the approach taken in the original adaptive setting reference: (Fenichel et al., 2011).
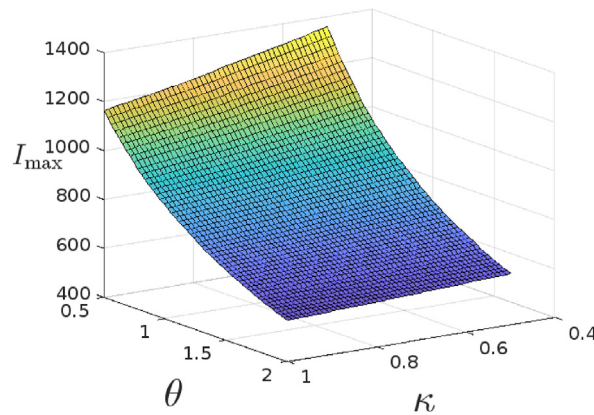
Fig. 7. Disease peak infected prevalence varying both contact proportions κ and θ for the influenza simulation example.

$$g_1(\cdot) := C^{S_1}\frac{\rho(C^{E_1}E_1 + C^{E_2}E_2) + \alpha(C^{A_1}A_1 + C^{A_2}A_2) + C^I I}{\sum_h C^h h}$$

$$g_2(\cdot) := C^{S_2}\frac{\rho(C^{E_1}E_1 + C^{E_2}E_2) + \alpha(C^{A_1}A_1 + C^{A_2}A_2) + C^I I}{\sum_h C^h h},$$

(4.1)

where $h \in \{S_1, S_2, E_1, E_2, A_1, A_2, I, R\}$ (the model also has a recovered compartment, naturally), $\varepsilon \in (0, 1)$ is a reduction in infectious chances by taking a risk-evader approach, and ρ, α are reduction constants for non-symptomatic infectious populations. This example gives us an application of the abovementioned principle in constructing contact-based incidence rate functions.

Another interesting example of the application of this principle consists of vector-borne diseases. In these cases, susceptible humans become infected not by contact with other infected human individuals, but rather by contact with infected vectors. Furthermore, susceptible vectors become infected by contact with infected humans. This dynamic requests a change of form in the incidence rate functions. Let us take, for example, the dengue-chikungunya vector-borne epidemic model proposed in (Sanchez et al., 2018). This model considers two populations: hosts (h) and vectors (v), and it is based on the following system of differential equations.

Hosts	Vectors
$\dfrac{dS_h}{dt} = \mu_h N_h - \beta g_h(\cdot)S_h\dfrac{I_v}{N_v} - \mu_h S_h$	$\dfrac{dS_v}{dt} = \mu_v N_v - \beta_v g_v(\cdot)S_v\dfrac{I_h}{N_h} - \mu_v S_v$
$\dfrac{dE_h}{dt} = \beta g_h(\cdot)S_h\dfrac{I_v}{N_v} - (\mu_h + \alpha_h)E_h$	$\dfrac{dE_v}{dt} = \beta g_v(\cdot)S_v\dfrac{I_h}{N_h} - (\mu_v + \alpha_v)E_v$
$\dfrac{dI_h}{dt} = \alpha_h E_h - (\mu_h + \gamma)I_h$	$\dfrac{dI_v}{dt} = \alpha_v E_v - \mu_v I_v$
$\dfrac{dR_h}{dt} = \gamma I_h - \mu_h R_h$	

Here, the host population has five health classes: S_h, susceptible hosts, E_h, exposed hosts, I_h, infected hosts, and R_h, recovered hosts, and the vector population has three: S_v, susceptible vectors, E_v, exposed/latent vectors, and I_v, infected vectors. Total populations are N_h for hosts and N_v for vectors. There are no recovered vectors, as they die with the disease. For each population, we inserted incidence rate functions $g_h(\cdot), g_v(\cdot)$, which are constant and equal to 1 in (Sanchez et al., 2018).

Following the abovementioned principle, we can propose the following formulas for both functions.

$$g_h(\cdot) := \frac{C^{S_h}C^{I_v}I_v}{\sum_j C^j j}$$

$$g_v(\cdot) := \frac{C^{S_v}C^{I_h}I_h}{\sum_j C^j h},$$

(4.2)

where $j \in \{S_h, E_h, I_h, R_h, S_v, E_v, I_v\}$ goes through all possible health statuses. This formulation considers the infection dynamics of vector-borne diseases: hosts become infected after contact with infected vectors, and vectors become infected after contact with infected hosts. Equations 4.2 offer an alternative for researching the possible impact of non-linear incidence rate functions in more complex scenarios, such as vector-borne diseases. Introducing contact rates between susceptible hosts and infected vectors could offer a mathematical approach to model the interactions between these two populations and understand further indirect contact-impacting measures, such as protection against vectors. We believe this offers an opportunity to further understand the dynamics of this biological scenario, especially in light of real data-based analysis, such as performed in (Sanchez et al., 2018).

5. Discussion

Motivated by the recent advances in the adaptive setting framework, we proposed a model that incorporates non-linear relapse and contact behavior among individuals of different health classes. Our study aimed to explore the analytical properties of this model and investigate the effects of disaggregating contact rates on disease dynamics. We found that the model exhibits a high sensitivity to initial conditions and the relationships between contact rates, with significant implications for disease control strategies.

To gain insights into the behavior of our model, we performed numerical simulations that revealed several important features. First, we observed that the model's dynamics are highly dependent on the values of the basic reproductive number (R_0), which reflects the behavior of infected individuals. We established explicit conditions for the existence of multiple stable infected populations, which are highly sensitive to the initial conditions of the model. Furthermore, we found that the impact of the contact rates for recovered individuals with relapse (θ) is more substantial than that of infected individuals (κ), with larger differences required to achieve complete disease control for higher initial epidemic volumes.

Models that incorporate relapse phenomena highlight the significant impact of recovered individuals on the progress of diseases. Such models exhibit different dynamics compared to those without relapse, with distinct results regarding recovery and relapse. Our study supports this view, with our conclusions showing that changes in the contact behavior of recovered individuals (θ) have a more substantial effect on epidemic equilibria than corresponding changes in contact rates for infected individuals (κ), after normalizing with respect to the susceptible contact rate. The differences in behavior when recovering from the disease (or addiction) play a crucial role in determining the prevalence of the disease. Our findings suggest that a low θ value, indicating a lack of meaningful contact engagement by recovered individuals after infection, can result in the establishment of a considerable epidemic burden. This highlights the importance of the successful reintegration of recovered individuals into society, which can reduce the likelihood of significant epidemics. Similar conclusions regarding the impact of recovered individuals on bifurcation plots have been observed in other relapse models (Tasman et al., 1002).

Our results are closely tied to the behavior of infected individuals, as captured by the basic reproductive number R_0. We established explicit conditions for the existence of a region $1 < R_0 < R_{0,\max}$, characterized by multiple stable infected populations, which are highly sensitive to the initial conditions of the model. We also found that this region becomes wider as the contacts of recovered individuals with relapse increase. In our simulations, we observed that the impact of (κ, θ) is intertwined with the initial infected population size, with larger initial epidemic volumes requiring more significant differences in contact rates to achieve disease control.

Building on the mathematical analysis presented in (Sanchez et al., 2023), we were able to confirm the conclusions and discussions regarding the influence of recovered individuals on the prevalence of diseases with relapse.

The incorporation of non-linear relapse significantly alters the dynamics of the SIR model, leading to more complex equilibria and bifurcation considerations. In our analysis, we adopted a non-linear relapse formulation (1.2) that assumes fixed contact rates among health compartments. Our aim was to obtain analytical results that can be compared to future studies using a complete adaptive formulation. Such comparisons will be made against the non-linear non-adaptive model proposed in this article.

Our study underscores the importance of incorporating adaptive behavior and contact heterogeneity into epidemiological models, particularly in the presence of relapse phenomena. The results obtained can inform public health policy decisions and provide a foundation for future research into the behavior of complex disease systems.

Declaration of competing interest

All authors declare no conflicts of interest in this paper.

Acknowledgments

Appendix

For the proof of theorem (3.4), we use the general theory of Sturm chains. Here we set some basic definitions and properties. We base this treatment on the theory detailed in (Eisermann).

Definition 6.1. A sequence $S = \{p_0(x), p_1(x), p_2(x), ..., p_n(x)\}$ of polynomials in $\mathbb{R}[x]$ is called a **Sturm chain** with respect to an interval I if it satisfies the Sturm property:
If $\alpha \in I$ is a real root of $p_i(x)$, for some i with $0 < i < n$. Then $p_{i-1}(\alpha)p_{i+1}(\alpha) < 0$.

Definition 6.2. Let f be a real rational function. The Cauchy index of f at x is defined by

$$\text{Ind}_x(f) := \text{Ind}_x^+(f) - \text{Ind}_x^-(f), \quad \text{where for } \sigma \in \{+, -\} \text{ we have}$$

$$
\text{Ind}_x^\sigma(f) :=
\begin{cases}
+\dfrac{1}{2} & \text{if } \lim\limits_{y \to x^\sigma} f(y) = +\infty, \\[2mm]
-\dfrac{1}{2} & \text{if } \lim\limits_{y \to x^\sigma} f(y) = -\infty, \\[2mm]
0 & \text{otherwise}.
\end{cases}
$$

The Cauchy index of f at $[a, b]$ is then given by

$$\text{Ind}_a^b(f) := \text{Ind}_a^+(f) - \text{Ind}_b^-(f) + \sum_{x \in a,b} \text{Ind}_x(f). \tag{6.1}$$

Remark 6.3. We can assume that f is in its reduced form, that is, the numerator and denominator have no common factors. In this case, $\text{Ind}_x(f)$ is non zero (with values 1 or -1) only for odd-multiplicity roots of the denominator, and since there are only finitely many such points, the sum in 6.1 is well defined.

We state the following generalization of the classical *Sturm Theorem*.

Theorem 6.4. ((Eisermann), Theorem 3.11). If $S = \{p_0(x), p_1(x), ..., p_{n-1}(x), p_n(x)\}$ is a Sturm chain in $\mathbb{R}[x]$ with respect to $[a, b]$, then

$$\text{Ind}_a^b\left(\frac{p_1}{p_0}\right) + \text{Ind}_a^b\left(\frac{p_{n-1}}{p_n}\right) = V_S(a) - V_S(b). \tag{6.2}$$

Where $V_S(c)$ is the number of sign changes in the values of consecutive polynomials of the chain S at a point $x = c$; that is, the number of those $j \in \{1, ..., n\}$ for which $p_{j-1}(c)p_j(c) < 0$.

This theorem can be applied to a special case of non-repeated roots, but first, we need the following auxiliary lemma, which can be proved by standard calculus arguments.

Lemma 6.5. Let $p_0(x)$ be a polynomial of degree n with n distinct real roots. Suppose that $c \in \mathbb{R}$ such that $p'(c) = 0$, then $p(c) p''(c) < 0$.

Proposition 6.6. Let $p_0(x)$ be a polynomial of degree n that has n distinct real roots, then the sequence $S_0 = \{p_0(x), p_0'(x), p_0^{(2)}(x), p_0^{(3)}(x), ..., p_0^{(n)}(x)\}$ of higher derivatives of $p_0(x)$ is a Sturm chain with respect to any interval $[a, b]$. Moreover, if $p_0(a)p_0(b) \neq 0$, then the number of roots of $p_0(x)$ in $[a, b]$ equals $V_{S_0}(a) - V_{S_0}(b)$.

Proof. *By a repeated application of* Lemma 6.5, *it is easy to check that the sequence S_0 of higher derivatives of $p_0(x)$ is a Sturm chain with respect to any interval I. Given that $p_0^{(n)}$ is a non-null constant, then $\frac{p_0^{(n-1)}}{p_0^{(n)}}$ is a polynomial which means that* $\text{Ind}_a^b\left(\frac{p_0^{(n-1)}}{p_0^{(n)}}\right) = 0$. *For the first term in 6.2, if we write $p_0(x) = c(x - a_1)\cdots(x - a_n)$, where $a_1, ...a_n$ are the roots of p_0, we have*

$$\frac{p_1(x)}{p_0(x)} = \frac{p_0'(x)}{p_0(x)} = \frac{c\sum_{i=1}^n \prod_{j \neq i}(x - a_j)}{c\prod_{j=1}^n(x - a_j)} = \frac{1}{x - a_1} + \frac{1}{x - a_2} + \cdots \frac{1}{x - a_n}.$$

So, for each $j = 1, ..., n$, $\lim_{x \to a_j^\pm} = \pm\infty$, that is $\text{Ind}_{a_j}\left(\frac{p_0'}{p_0}\right) = 1$. Therefore, the Cauchy index $\text{Ind}_a^b\left(\frac{p_0'}{p_0}\right)$ is the number of roots of p_0 that are contained in $[a, b]$, and by Theorem 6.4, *this coincides with $V_{S_0}(a) - V_{S_0}(b)$.*

References

Anderson, R., & May, R. (1979). Population biology of infectious diseases: Part I. *Nature, 280*, 361–367.

S. Blythe, K. Cooke, C. Castillo-Chavez, Autonomous risk-behavior change, and non-linear incidence rate, in models of sexually transmitted diseases, Biometrics Unit Technical Report B-1048-M.

Brauer, F. (2009). Mathematical epidemiology is not an oxymoron. *BMC Public Health, 9*(Suppl 1), S2. https://doi.org/10.1186/1471-2458-9-S1-S2

Chowell, G., Miller, M., & Viboud, C. (2008). Seasonal influenza in the United States, France, and Australia: Transmission and prospects for control. *Epidemiology and Infection, 136*(6), 852–864. https://doi.org/10.1017/S0950268807009144

M. Eisermann, The fundamental theorem of algebra made effective: An elementary real-algebraic proof via Sturm chains, The American Mathematical Monthly 119. doi:10.4169/amer.math.monthly.119.09.715.

B. Espinoza, M. Marathe, S. Swarup, M. Thakur, Asymptomatic individuals can increase the final epidemic size under adaptive human behavior, Scientific Reports 11 (19744). doi:10.1038/s41598-021-98999-2.

B. Espinoza, S. Swarup, C. L. Barrett, M. Marathe, Heterogeneous adaptive behavioral responses may increase epidemic burden, Scientific Reports 12 (11276). doi:10.1038/s41598-022-15444-8.

Di Lauro, F., Berthouze, L., Dorey, M. D., Miller, J. C., & Kiss, I. Z.. The impact of contact structure and mixing on control measures and disease-induced herd immunity in epidemic models: A mean-field model perspective. Bulletin of Mathematical Biology, 83(117). doi:10.1007/s11538-021-00947-8.

Fenichel, E. P., Castillo-Chavez, C., Ceddia, M. G., Chowell, G., Parra, P. A. G., Hickling, G. J., Holloway, G., Horan, R., Morin, B., Perrings, C., Springborn, M., Velazquez, L., & Villalobos, C. (2011). Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences of the United States of America, 108*(15), 6306–6311. https://doi.org/10.1073/pnas.101125010

Funk, S., Salathé, M., & Jansen, V. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: A review. *Journal of The Royal Society Interface, 7*, 1247–1256. https://doi.org/10.1098/rsif.2010.0142

H. Hethcote, The mathematics of infectious diseases, SIAM Review 42 (2). doi:10.1007/BF00160539.

H. Hethcote, P. van den Driessche, Some epidemiological models with nonlinear incidence, Journal of Mathematical Biology. 29 (3). doi:10.1007/BF00160539.

Hu, Z., Ma, W., & Ruan, S. (2012). Analysis of sir epidemic models with nonlinear incidence rate and treatment. *Mathematical Biosciences, 238*(1), 12–20. https://doi.org/10.1016/j.mbs.2012.03.010

Liu, W., Hethcote, H., & Levin, S. (1987). Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of Mathematical Biology, 25*(4), 359–380. https://doi.org/10.1007/BF00277162

G. Li, Y. Zhang, Dynamic behaviors of a modified SIR model in epidemic diseases using nonlinear incidence and recovery rates, PLoS One 12 (4). doi:10.1371/journal.pone.0175789.

S. Moghadas, M. Alexander, Bifurcations of an epidemic model with non-linear incidence and infection-dependent removal rate, Mathematical Medicine and Biology 23 (3). doi:10.1093/imammb/dql011.

B. Morin, E. P. Fenichel, C. Castillo-Chavez, SIR dynamics with economically driven contact rates, Natural Resource Modeling 26 (4). doi:10.1111/nrm.12011.

J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, W. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases, PLoS Medicine 25 (5). doi:10.1371/journal.pmed.0050074.

Sanchez, F., Arroyo-Esquivel, J., & Calvo, J. G. (2023). A mathematical model with nonlinear relapse: Conditions for a forward-backward bifurcation. pMID: 36942364 *Journal of Biological Dynamics, 17*(1), Article 2192238.

Sanchez, F., Barboza, L. A., Burton, D., & Cintrón-Arias, A. (2018). Comparative analysis of dengue versus chikungunya outbreaks in Costa Rica. *Ricerche di Matematica, 67*, 163–174.

Sanchez, F., Calvo, J. G., Segura, E., & Feng, Z. (2019). A partial differential equation model with age-structure and nonlinear recidivism: Conditions for a backward bifurcation and a general numerical implementation. *Computers & Mathematics with Applications, 78*(12), 3916–3930. https://doi.org/10.1016/j.camwa.2019.06.021

Sanchez, F., Wang, X., Castillo-Chavez, C., Gorman, D., & Gruenewald, P. (2007). Drinking as an epidemic: A simple mathematical model with recovery and relapse. In *Therapist's guide to evidence based relapse prevention*. Burlington: Academic Press.

M. Springborn, G. Chowell, M. MacLachlan, E. P. Fenichel, Accounting for behavioral responses during a flu epidemic using home television viewing, BMC Infectious Diseases. 15 (21). doi:10.1186/s12879-014-0691-0.

H. Tasman, D. Aldila, P. Dumbela, M. Ndii, F. Fatmawati, H. F.F, C. Chukwu, Assessing the impact of relapse, reinfection and recrudescence on malaria eradication policy: A bifurcation and optimal control analysis, Travel Medicine and Infectious Disease 263 (7). doi:10.3390/tropicalmed7100263.

Xiao, D., & Ruan, S. (2007). Global analysis of an epidemic model with non monotone incidence rate. *Mathematical Biosciences, 208*(2), 419–429. https://doi.org/10.1016/j.mbs.2006.09.025

Xiao, Y., & Tang, S. (2010). Dynamics of infection with nonlinear incidence in a simple vaccination model. *Nonlinear Analysis: Real World Applications, 11*(5), 4154–4163. https://doi.org/10.1016/j.nonrwa.2010.05.002

J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, H. Yu, Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China, Science 368 (6498). doi:10.1126/science.abb8001.