TECHNISCHE UNIVERSITÄT WIEN

Variational Analysis, Dynamics and Operations Research

SWM VADOR

An immuno-epidemiological model with waning immunity after infection or vaccination

G. Angelov, R. Kovacevic, N. I. Stilianakis, V. M. Veliov

Research Report 2023-03

November 2023

ISSN 2521-313X

Variational Analysis, Dynamics and Operations Research
Institute of Statistics and Mathematical Methods in Economics
TU Wien

Research Unit VADOR
Wiedner Hauptstraße 8 / E105-04
1040 Vienna, Austria
E-mail: vador@tuwien.ac.at

An immuno-epidemiological model with waning immunity after infection or vaccination*

Georgi Angelov†, Raimund Kovacevic‡, Nikolaos I. Stilianakis§, and Vladimir M. Veliov¶

Abstract

In epidemics, waning immunity is common after infection or vaccination of individuals. Immunity levels are highly heterogeneous and dynamic. This work presents an immuno-epidemiological model that captures the fundamental dynamic features of immunity acquisition and wane after infection or vaccination and analyzes mathematically its dynamical properties. The model consists of first order partial differential equations with different transfer velocities. This untypical feature makes the proved existence of a solution novel and non-trivial. The asymptotic behaviour of the model is analysed using the obtained qualitative properties of the solution. An optimal control problem with objective function including the total number of deaths and the costs of vaccination is explored. Numerical results describe the dynamic relationship between contact rates and optimal solutions. The approach can contribute to the understanding of the dynamics of immune responses at population level and may guide public health policies.

1 Introduction

In many infectious diseases immunity acquired from infection is waning over time. The same accounts for immune responses elicited by vaccines. Typical examples are influenza and COVID-19 where immunity wanes within a few months (Rambhia & Rambhia [2019]; Goldberg et al. [2022]). The durability of natural immunity and immune responses triggered by vaccines are crucial for public health and the associated decision making for interventions. Antibodies seem to be the protective mechanism for these infections but often more specific immune responses such as specific T cell groups are needed to build up and maintain immune memory. Immunity waning is highly heterogeneous in the population between individuals and changes over time (Lavine et al. [2021]).

Several mathematical models have been developed to assess effectiveness and the possibility of waning immunity after infection or vaccination (Montalbán et al. [2022]; Iyaniwura et al. [2023];

*This research is supported by he Austrian Science Foundation (FWF) under grant No I 4571-N.

†Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, georgi.angelov@tuwien.ac.at.

‡Department for Economy and Health, University for Continuing Education Krems, Austria, raimund.kovacevic@donau-uni.ac.at.

§European Commission, Joint Research Centre (JRC), Ispra, Italy, and Department of Biometry and Epidemiology, University of Erlangen-Nuremberg, Erlangen, Germany, nikolaos.stilianakis@ec.europa.eu. Disclaimer: The views expressed are purely those of the writer (NIS) and may not in any circumstance be regarded as stating an official position of the European Commission.

¶Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, vladimir.veliov@tuwien.ac.at.

Pell et al. [2022]; Gosh et al. [2022]; Domenech de Celles et al. [2022]; Veliov and Widder [2016]). To a lesser extent models investigated the optimal timing of vaccine administration, accounting for interseasonal waning immunity for infectious diseases such as influenza (Costantino et al. [2019]). A population with heterogeneous immunity is considered in Montalbán et al. (2022). However, individual immunity is modeled as constant over time. In addition, Montalbán et al. (2022) consider no change in immunity levels due to previous infection or vaccination and do not study decision (control) aspects. Iyaniwura et al. (2023) used a distributed delay equations framework to describe the dynamics of waning immunity in a population with vaccine or natural infection induced immunity at an endemic stage. They performed a bifurcation analysis showing that waning immunity from natural infection influences the bifurcation type more than vaccine associated waning immunity. In addition, they derived a control reproduction number and showed the interplay between waning immunity rate and transmission rate of the pathogen. Similar approaches were used by Pell et al. (2022) and Gosh et al. (2022). Domenech de Cellès et al. (2022) showed in a simulation study how immunological heterogeneity plays a role in determining durability of vaccine protection. A model with heterogeneous dynamic immunity where sub-populations were structured with respect to the host immunity was developed and analysed by Veliov and Widder (2016). In all cases the investigation of control aspects were either not investigated or played a rather limited role.

In the present work, we develop a deterministic model in order to study the dynamics of an infectious disease with waning immunity after natural infection or vaccination. The approach models the following characteristic effects: A model with dynamic immunity, formulated as a time dependent function $\omega(t)$. After infection, immunity at the individual level mounts overtime until recovery starts. With the initiation of recovery immunity declines over time, denoting the waning immunity after natural infection or vaccination.

Model features such as existence of solution and asymptotic behaviour are explored. The model is then extended to capture effects of vaccination. Numerical results are presented for several scenarios, including the disease behaviour with and without vaccination and optimal vaccination policies. The latter could be used in the assessment of vaccine administration.

From the mathematical point of view, the model we propose is challenging due to the following reasons:

(i) It consists of a system of two or three first order Partial Differential Equations (PDE) (each of which is of size-structured type, see, e.g., Martcheva & Pilyugin [2006]) with different transfer velocities, hence, with different characteristic lines. This creates a substantial problem in the analysis of the system, because the usual reformulation of the PDE system as an Ordinary Differential Equation (ODE) system in a closed form is not possible.

(ii) The Lipschitz constant of the equations may tend to infinity along the solution, which makes problematic the proof of global existence.

Several authors investigate the disease dynamics taking into account waning immunity, some also regarding gaining immunity during infection or after vaccination. Closer to our model is the work of White and Medley [1998], where equations with different transfer velocities are involved. However, the authors focus on the formal steady state equations and do not study the overall PDE system. Also Rouderfer and Becker [1994], Barbarosa and Röst [2015], Ehrhardt et al. [2019], among others, consider first order PDEs, but either the velocity flows are identical or a single PDE (together with ODEs) is involved.

In the Appendix of this work we prove global existence of a solution, even for more general system than our particular model requires. The proof is not straightforward and can be of independent

mathematical interest (see Subsection 3.2 for more explanations).

The optimal vaccination problem presented in Subsection 4.2 is analyzed only numerically, focusing on meaningful observations about the optimal vaccination policy and the respective evolution of the epidemic. A more profound investigation of the optimization problem, in more general terms, is a subject of future work.

2 The basic model with dynamic immunity

In order to model effects of immunity in the population, we use a function $\omega(t)$, indicating the immunity level at time t. It takes values in the interval $[0, 1]$ and may be chosen, e.g., proportional to the individual immunity status, measured for example by the amount of antibodies per ml blood. The larger this number ω, the higher is the individuals immunity, which implies lower susceptibility and lower infectiousness.

Throughout the paper we assume that after an individual is infected, its immunity level increases until the time or recovery (Yaugel-Novoa et al. [2022]). Therefore, we describe the evolution of the immunity level after infection at time τ by the equation

$$\dot{\omega}(t) = g(\omega(t)), \quad \omega(\tau) = \xi, \quad t \geq \tau, \tag{2.1}$$

where $\xi \in [0, 1]$ is the initial immunity level at the time of infection τ. The function $g : [0, 1] \to [0, +\infty)$ is assumed to be differentiable and to satisfy $g(0) > 0$, and $g(1) = 0$.

In the long run, the immunity level decreases (Yaugel-Novoa et al. [2022]). Therefore, beginning with recovery or shortly after vaccination, immunity wanes over time and we describe its decrease by the equation

$$\dot{\omega}(t) = f(\omega(t)), \quad \omega(\tau) = \xi, \quad t \geq \tau, \tag{2.2}$$

where $\xi \in [0, 1]$ is the immunity level at the time of recovery τ. It is assumed that $f : [0, 1] \to (-\infty, 0]$, is continuously differentiable with $f(0) = 0$, $f(1) < 0$.

While in a population the individual immunity levels change as described above, at any point in time the individuals in a population may have differing immunity levels, depending on their individual history with the disease. Therefore, the immunity status of a whole population can be modeled as a frequency distribution over the possible values of the immunity level, $\omega \in [0, 1]$. Note that ω here denotes just one possible value of the function $\omega(\cdot)$.

Based on these considerations, we denote by $S(t, \omega)$ and $I(t, \omega)$ the size of the susceptible, respectively infected, population of immunity level ω at time t. Thus the total population at time t is

$$N(t) = \int_0^1 [S(t, \omega) + I(t, \omega)] \, \mathrm{d}\omega,$$

Throughout this paper it is assumed that, given a particular disease, the susceptibility of an individual depends only on the immunity level ω. This immune memory may have been acquired through previous exposure to the relevant pathogen and infection or previous vaccination. We denote the susceptibility by $\sigma(\omega) \geq 0$, where the function $\sigma : [0, 1] \to [0, \infty)$ is continuous and presumably decreasing in ω. Similarly, the infectiousness of infected individuals is modeled as solely dependent on the immunity level ω. It is denoted by $i(\omega)$, where i is a continuous non-negative function.

We denote by $c > 0$ the contact rate of susceptible individuals, while the contact rate of infected individuals is denoted by $c_I \in (0, c]$. In principle, the contact parameters may be extended to

depend on ω, because people who know that they are well protected by immunity may have more contacts. Moreover, dependence on time may be used for the description of seasonal or other time dependent behaviour of the individuals. However, in this paper we assume for simplicity that c and c_I are constant.

Under the assumption of weighted random mixing, the infectiousness of the environment in which susceptible individuals contact infected individuals is represented as

$$D(t) = \frac{c_I \int_0^1 i(\omega) I(t,\omega)\,\mathrm{d}\omega}{c_I \int_0^1 I(t,\omega)\,\mathrm{d}\omega + c \int_0^1 S(t,\omega)\,\mathrm{d}\omega}. \tag{2.3}$$

Finally, the mortality rate of infected individuals is denoted by $\mu(\omega)$, and the recovery rate from infection is denoted by $\rho(\omega)$. Both parameters are modeled as functions, depending on the immunity level. In the present paper we will not take into account demographic influences such as births and deaths (all-cause mortality), which play a role in the long run.

Based on these assumptions and the related notations, it is now possible to describe the time dependent dynamics of the classes of susceptible and infected individuals for different immunity levels in terms of a system of PDEs for the population sizes S and I of the susceptible and infected individuals.

$$\frac{\partial}{\partial t} S(t,\omega) + \frac{\partial}{\partial \omega}(f(\omega) S(t,\omega)) = -cD(t)\sigma(\omega) S(t,\omega) + \rho(\omega) I(t,\omega), \tag{2.4}$$

$$\frac{\partial}{\partial t} I(t,\omega) + \frac{\partial}{\partial \omega}(g(\omega) I(t,\omega)) = cD(t)\sigma(\omega) S(t,\omega) - (\rho(\omega) + \mu(\omega)) I(t,\omega), \tag{2.5}$$

with the initial conditions

$$S(0,\omega) = S^0(\omega), \quad I(0,\omega) = I^0(\omega), \quad \omega \in [0,1], \tag{2.6}$$

(S^0 and I^0 are initial data) and the zero flux boundary conditions

$$f(\omega) S(t,\omega) = 0, \quad g(\omega) I(t,\omega) = 0, \quad \omega \in \{0,1\}, \quad t \geq 0.$$

In this system, the left hand side is obtained as the total derivative of S, respectively I with respect to t, taking into account the dynamics of $\omega(\cdot)$ for susceptible - see (2.2) - and infected - see (2.1) - individuals. The right hand side models the change in the compartment at immunity level ω due to new infections, recovery and death. This derivation of the next equations follows the same (conservation of mass) argument as that for the advection-reaction equation for a compressible gas.

We note that due to the assumptions $f(0) = g(1) = 0$ and $f(1) < 0$, $g(0) > 0$, the initial conditions and the zero-flux condition are equivalent to

$$S(t,1) = 0, \quad I(t,0) = 0, \quad t > 0. \tag{2.7}$$

Moreover, due to the meaning of ω, and for consistency of the initial and boundary conditions it is natural to assume that $S^0(1) = I^0(0) = 0$.

3 Existence of solution and asymptotic behaviour

3.1 Notion of solution

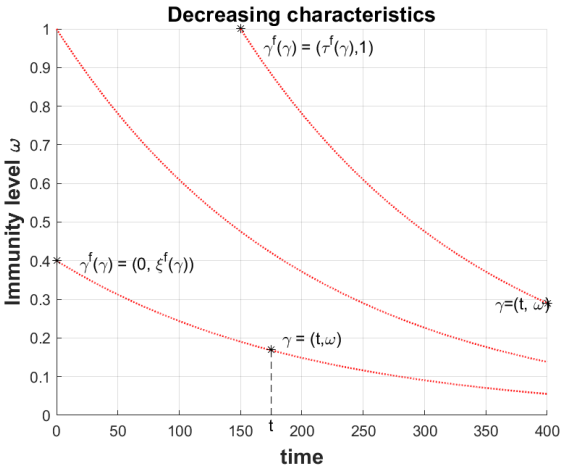
The following assumptions hold throughout the paper.

Standing Assumptions. The functions f and g are differentiable with Lipschitz derivatives, defined in a neighborhood of $[0,1]$, with $f(0) = g(1) = 0$, and the derivatives $f'(\omega) < 0$ on $(0,1]$, $g'(\omega) < 0$ in $\omega \in [0,1)$. The function $i : [0,1] \to [0,\infty)$ is measurable and bounded, the functions $\sigma, \rho, \mu, S^0, I^0 : [0,1] \to [0,\infty)$ are continuous, differentiable on $(0,1)$ except a finite number of points[1], and the derivatives are Lipschitz continuous in each interval of existence. Moreover, $S^0(1) = I^0(0) = 0$ and $\int_0^1 (S^0(\omega) + I^0(\omega))\,\mathrm{d}\omega = 1$.

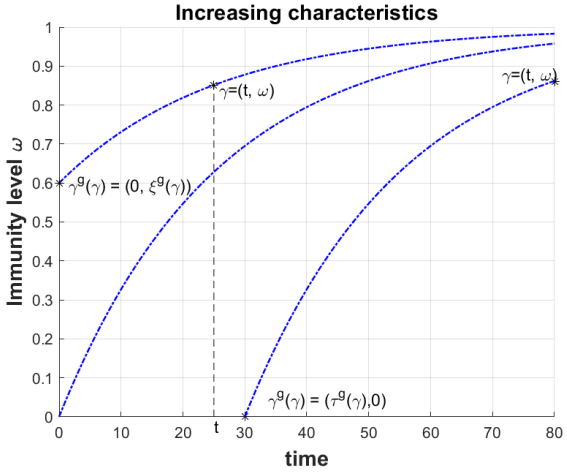
The solution of the system (2.3)–(2.7) can be defined in several ways (cf. Kato & Torikata [1997]). Here, we define the notion of solution by the method of characteristics. For reasons of further analysis, we restrict the definition to the case of Lipschitz continuous solutions (although the solutions may be discontinuous for general initial/boundary data).

Denote $\Gamma := [0,T] \times [0,1]$, and let $\tilde{\Gamma} \subset \mathbb{R}^2$ be an open neighborhood of Γ. For $\gamma := (\tau, \xi) \in \tilde{\Gamma}$ we denote by $\omega^f[\gamma](\cdot)$ and $\omega^g[\gamma](\cdot)$ the solutions of (2.2) and (2.1), respectively. Due to the assumptions for f and g, the set $[0,1]$ is an invariant domain for both equations, hence, considering a sufficiently small neighborhood $\tilde{\Gamma}$ of Γ, the solutions are defined on $[0,T]$ for every $\gamma \in \tilde{\Gamma}$.

Further, denote $\Gamma_f := (\{0\} \times [0,1]) \cup ([0,T] \times \{1\})$ (the left-upper boundary of Γ), $\Gamma_g := (\{0\} \times [0,1]) \cup ([0,T] \times \{0\})$ (the left-lower boundary of Γ). Due to the assumptions for f and g, we have that $\cup_{\gamma \in \Gamma_f} \omega^f[\gamma](t) = [0,1]$. Similarly, $\cup_{\gamma \in \Gamma_g} \omega^g[\gamma](t) = [0,1]$. Again due to the properties of f and g, there are unique functions $\gamma^f : \tilde{\Gamma} \to \Gamma_f$ and $\gamma^g : \tilde{\Gamma} \to \Gamma_g$ such that $\omega^f[\gamma^f(t,\omega)](t) = \omega$ and $\omega^g[\gamma^g(t,\omega)](t) = \omega$ for all $\tilde{\Gamma}$. Moreover, due to the (Lipschitz) continuous dependence of the solutions of (2.2) and (2.1) on the initial data, the functions $\omega^f, \omega^g, \gamma^f, \gamma^g$ have Lipschitz continuous derivatives with respect to γ and t.



(a) Characteristic lines for $\omega^f[\gamma_0](\cdot), \gamma_0 \in \Gamma_f$



(b) Characteristic lines for $\omega^g[\gamma_0](\cdot), \gamma_0 \in \Gamma_g$

Figure 3.1: Characteristic lines and illustration of the notations.

[1]The assumption can be relaxed by replacing "finite number" with "countable number".

For the (dummy) real numbers t, ω, d, s, i, denote (in relation with (2.4)–(2.5))

$$F^S(t,\omega,d,s,i) := -cd\sigma(\omega)s + \rho(\omega)i - f'(\omega)s, \tag{3.1}$$
$$F^I(t,\omega,d,s,i) := cd\sigma(\omega)s - (\rho(\omega)+\mu(\omega))i - g'(\omega)i, \tag{3.2}$$

where the argument t is included for further use. For shortness we introduce the notations

$$\gamma^f(\gamma) := (\tau^f(\gamma), \xi^f(\gamma)), \quad \gamma^g(\gamma) := (\tau^g(\gamma), \xi^g(\gamma)),$$

for $\gamma \in \Gamma$, and and for $\gamma \in \Gamma_f$ and $\gamma \in \Gamma_g$, respectively:

$$\bar{S}^0(\gamma) := \begin{cases} S^0(\xi) & \text{if } \gamma = (0,\xi), \\ 0 & \text{if } \gamma = (\tau, 1), \end{cases} \qquad \bar{I}^0(\gamma) := \begin{cases} I^0(\xi) & \text{if } \gamma = (0,\xi), \\ 0 & \text{if } \gamma = (\tau, 0). \end{cases}$$

for $\gamma \in \Gamma_f$ and $\gamma \in \Gamma_g$, respectively.

Definition 3.1. The pair of continuous functions $S, I : \Omega \to \mathbb{R}$ is called a solution of system (2.3)–(2.7) if the following equations are satisfied for all $\gamma = (t,\omega) \in \Gamma$:

$$S(\gamma) = \int_{\tau^f(\gamma)}^{t} F^S\big(s, \omega^f[\gamma](s), D(s), S(s,\omega^f[\gamma](s)), I(s,\omega^f[\gamma](s))\big)\,\mathrm{d}s + \bar{S}^0(\gamma^f(\gamma)), \tag{3.3}$$
$$I(\gamma) = \int_{\tau^g(\gamma)}^{t} F^I\big(s, \omega^g[\gamma](s), D(s), S(s,\omega^g[\gamma](s)), I(s,\omega^g[\gamma](s))\big)\,\mathrm{d}s + \bar{I}^0(\gamma^g(\gamma)), \tag{3.4}$$

together with the relation (2.3).

For readers who are not familiar with this kind of definition we mention that if S and I are differentiable and satisfy the equations (2.3)–(2.7) in classical sense, then they also solve equations (3.3), (3.4). This fact is not straightforward, however it can be directly checked using the identities $\frac{\partial}{\partial t}\omega^f[\gamma](s) + f(\omega)\frac{\partial}{\partial \omega}\omega^f[\gamma](s) = 0$ for all $\gamma = (t,\omega) \in \Gamma_f$, and a few more similar identities that appear when plugging the expressions of S and I in (3.3)–(3.4) into (2.4)–(2.5). An alternative way of verifying (2.3)–(2.7) is to use the representation of S and I along the characteristic lines, as presented in Part 5 of the proof of Theorem 6.1 in Appendix.

3.2 Existence of a "smooth" solution

In the next subsection we prove existence of a solution of system (2.3)–(2.7) which is regular enough to enable performing the subsequent analysis. Although the proof is based on the Banach contraction mapping theorem, it is not straightforward due to two reasons: (i) the Lipschitz constants of F^S and F^I may tend to infinity with the time due to the expression (2.3) for D, which makes the existence on $[0,\infty)$ problematic; (ii) due to the involvement of different flow velocity fields f and g, the system (2.4)–(2.5) cannot be reduced to a closed form ODE system along the characteristics; (iii) proving the non-negativity of the solution is not straightforward at all. Therefore, we present in Appendix a detailed proof of the existence. In fact, we prove a more general theorem assuming a few properties of the functions F^S and F^I in (3.3)–(3.4) and not necessarily the specific form in (3.1)–(3.2).

Theorem 3.1. *Under the standing assumptions, system* (2.3)–(2.7) *has a solution* (S, I) *on* $[0,\infty) \times [0,1]$ *which is Lipschitz continuous on every set* $[0,T] \times [0,1]$, $T > 0$. *Moreover, for each* $\omega \in [0,1]$ *and* $T \in (0,\infty)$ *the derivatives* $\frac{\partial}{\partial t} S(t,\omega)$, $\frac{\partial}{\partial t} I(t,\omega)$ *exist on* $(0,T]$ *except of finite number of points, and for each* $t \in (0,\infty)$ *the derivatives* $\frac{\partial}{\partial \omega} S(t,\omega)$, $\frac{\partial}{\partial \omega} I(t,\omega)$ *exist on* $(0,1)$ *except of finite number of points.*

Remark 3.1. The differentiability property in the claim of the theorem is important, because it allows to integrate equations (2.4)–(2.5) in ω, and changing the order of integration and differentiation. This results, having in mind the zero flux conditions (given after (2.6)), in the relations

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \int_0^1 S(t,\omega)\,\mathrm{d}\omega &= \int_0^1 [-cD(t)\sigma(\omega)S(t,\omega) + \rho(\omega)I(t,\omega)]\,\mathrm{d}\omega, \\ \frac{\mathrm{d}}{\mathrm{d}t} \int_0^1 I(t,\omega)\,\mathrm{d}\omega &= \int_0^1 [cD(t)\sigma(\omega)S(t,\omega) - (\rho(\omega)+\mu(\omega))I(t,\omega)]\,\mathrm{d}\omega, \end{aligned}$$

where the derivatives exist for all t except a finite number of points on every bounded set $]0,T]$. Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^1 (S(t,\omega) + I(t,\omega))\,\mathrm{d}\omega = -\int_0^1 \mu(\omega))I(t,\omega)\,\mathrm{d}\omega \tag{3.5}$$

We mention that local existence of a solution is proved in the Appendix for a more general system than (2.3)–(2.7). Moreover, including more equations with different characteristic curves (such as the system with vaccination in the next section) does not change the proof. Since we allow dependence of the functions F^S and F^I on the time in the proof, also the presence of control function in the equations is covered by the existence theorem as proved in the Appendix.

We also mention that the above theorem does not claim uniqueness. The proof given in Appendix employ two times contraction mapping theorem, therefore uniqueness in a restricted set of functions is implied.

3.3 Long run behaviour

The goal of this subsection is to obtain conditions under which the epidemic converges to a disease-free state. For that we assume the differentiability of the solution in Theorem 3.1. Let (S, I, D) be the solution of (2.3)–(2.7) on $[0,\infty) \times [0,1]$. Denote

$$\hat{S}(t) := \int_0^1 S(t,\omega)\,\mathrm{d}\omega, \qquad \hat{I}(t) := \int_0^1 I(t,\omega)\,\mathrm{d}\omega.$$

According to Remark 3.1, the next equations are fulfilled:

$$\begin{aligned} \dot{\hat{S}}(t) &= -cD(t)\int_0^1 \sigma(\omega)S(t,\omega)\,\mathrm{d}\omega + \int_0^1 \rho(\omega)I(t,\omega)\,\mathrm{d}\omega, \\ \dot{\hat{I}}(t) &= cD(t)\int_0^1 \sigma(\omega)S(t,\omega)\,\mathrm{d}\omega - \int_0^1 (\rho(\omega)+\mu(\omega))I(t,\omega)\,\mathrm{d}\omega. \end{aligned}$$

Using the obvious estimation

$$cD(t) \le \frac{c_I \int_0^1 i(\omega))I(t,\omega)\,\mathrm{d}\omega}{\hat{S}(t)}, \tag{3.6}$$

we obtain from the equation for $\hat{I}$ that

$$\dot{\hat{I}}(t) \le \int_0^1 [c_I \bar{\sigma} i(\omega) - (\rho(\omega) + \mu(\omega))] I(t,\omega)\,\mathrm{d}\omega, \qquad \bar{\sigma} := \sup\{\sigma(\omega) : \omega \in [0,1]\}. \tag{3.7}$$

Thus we obtain the following proposition.

Proposition 3.2. *At any time, if the current density of infected individuals, $I(\omega)$, satisfies the inequality*

$$\int_0^1 [c_I \bar{\sigma} i(\omega) - (\rho(\omega) + \mu(\omega))] I(\omega)\,\mathrm{d}\omega < 0,$$

then the number of infected individuals strictly decreases at this time.

Corollary 3.3. *Let the following inequality be fulfilled:*

$$\lambda := \min_{\omega\in[0,1]} \{\rho(\omega) + \mu(\omega)\} - c_I \bar{i} \bar{\sigma} > 0, \qquad \bar{i} := \sup\{i(\omega) : \omega \in [0,1]\}. \tag{3.8}$$

Then for any initial state $(S^0(\cdot), I^0(\cdot))$ the disease dies out exponentially, namely, $\hat{I}(t) \le e^{-\lambda t}\hat{I}(0)$. Moreover, if the initial amount of infected, $\hat{I}(0) = \int_0^1 I^0(\omega)\,\mathrm{d}\omega$, is sufficiently small, then the solution converges to a non-void disease free population. More precisely, using the notation $b := \max\{0, c_I\bar{\sigma}\bar{i} - \min_{\omega\in[0,1]} \rho(\omega)\}$, we have that for all $t \ge 0$

$$\hat{S}(t) \ge \hat{S}(0) - \frac{b}{\lambda + b} > 0,$$

provided that $\hat{I}(0) < \frac{\lambda}{\lambda+b}$.

Proof. Indeed, from (3.7) we obtain that

$$\dot{\hat{I}}(t) \le \int_0^1 [c_I \bar{\sigma} i(\omega) - (\rho(\omega) + \mu(\omega))]\, I(t,\omega)\,\mathrm{d}\omega \le -\lambda \hat{I}(t),$$

hence $\hat{I}(t) \le e^{-\lambda t}\hat{I}(0)$. Moreover, from the equation for $\hat{S}$ and (3.6) we have

$$\dot{\hat{S}}(t) \ge \int_0^1 [-c_I \bar{\sigma}\bar{i} + \rho(\omega)] I(t,\omega)\,\mathrm{d}\omega \ge -b\hat{I}(t).$$

Integrating, we obtain

$$\hat{S}(t) \ge \hat{S}(0) - b\int_0^t \hat{I}(s)\,\mathrm{d}s \ge \hat{S}(0) - \frac{b}{\lambda}\hat{I}(0) \ge \hat{S}(0) - \frac{b}{\lambda}\frac{\lambda}{\lambda+b} = \hat{S}(0) - \frac{b}{\lambda+b} > 0.$$

The last inequality follows from $1 = \hat{S}(0) + \hat{I}(0) < \hat{S}(0) + \frac{\lambda}{\lambda+b}$. □

Inequality (3.8) is only a sufficient condition for extinction of the disease. In the benchmark case considered in Section 5, the condition in Proposition 3.2 is not fulfilled and the disease converges to an endemic state (see Figure 5.4a). Another example is given in which the disease extincts (Figure 5.4b). The condition (3.8) fails in this case, but the condition in Proposition 3.2 is fulfilled.

For comparison, we mention that in the case of data independent of ω, condition (3.8) can be represented as

$$R_0 := \frac{c_I i \sigma}{\rho + \mu} < 1,$$

which the usual expression for the basic reproduction number (in our notations).

A way of formal derivation of the steady states of the system (2.3)–(2.7) is given for a similar model in Veliov and Widder [2016]. It can easily be obtained that the disease free steady state is $S^*(\omega) = \alpha\delta_0(\omega)$, where δ_0 is Dirac's delta function concentrated at zero, and $\alpha \in [0, 1]$ (and equals one in the case of zero mortality).

4 Modeling and optimization of vaccination

In this section we introduce a control variable that represents the vaccination efforts and consider a class of optimization problems for the vaccination policy.

4.1 Modelling vaccination

Including vaccination requires an extension of the basic model (2.3)–(2.7). We assume that only susceptible individuals are vaccinated. It is necessary then, to model the act of vaccination together with the immunological behavior of vaccinated individuals.

We denote by $v(t, \omega)$ the vaccination effort applied to susceptible individuals of immunity level ω at time t. This means, that $v(t, \omega)S(t, \omega)$ individuals of immunity level ω become vaccinated at time t. As far as the immunity level of people is not measured in reality, the dependence of v on ω is an idealization. However, the time from last vaccination or from the last infection could be considered (as practiced in reality) as a proxy for the individual immunity level.

The effect of vaccination on immunity is not immediate. Like newly infected individuals, vaccinated individuals gain immunity over time, until their immunity level reaches a maximum, possibly depending on the immunity level before vaccination. After that, the immunity level slowly decreases in the same way as that of all susceptible individuals with the same immunity level (Goel et al. [2021]).

Therefore, we augment the model by an additional compartment, representing the newly vaccinated individuals acquiring immunity after vaccination. Similarly as for susceptible and infected individuals, the size of this compartment is counted separately for each immunity level over time, and is denoted by $V(t, \omega)$. The process of acquiring immunity from vaccination (in a relatively short period after the vaccination) is modeled in a similar way as the increase of immunity during illness, namely by the equation

$$\dot{\omega}(t) = h(\omega(t)), \quad \omega(\tau) = \xi, \quad t \geq \theta, \tag{4.1}$$

where $\xi \in [0, 1]$ is the initial immunity level at the time of vaccination τ. The function $h(\omega) : [0, 1] \to [0, \infty)$, represents how fast immunity is built up at the current immunity level ω. Presumably, it is a decreasing function, with $h(0) > 0$ and $h(1) = 0$, that is, with similar properties as the function g.

When reaching their individual maximum immunity level, newly vaccinated individuals leave the compartment of the vaccinated and are counted as susceptible individuals with the attained

new immunity level. This means that their decrease of immunity will change in the same manner as for susceptible individuals of the same immune level. The transition process from vaccinated to susceptible occurs with a rate $r(\omega)$, so that $1/r(\omega)$ is the average duration of increase of immunity depending on the current immunity level.

Since the vaccinated individuals behave in their activities as susceptible, the infectiousness of the environment, $D(t)$, takes the form

$$D(t) = \frac{c_I \int_0^1 i(\omega) I(t,\omega)\,\mathrm{d}\omega}{c_I \int_0^1 I(t,\omega)\,\mathrm{d}\omega + c\int_0^1 (S(t,\omega)+V(t,\omega))\,\mathrm{d}\omega}. \tag{4.2}$$

The overall model with vaccination takes the form

$$\frac{\partial}{\partial t}S(t,\omega) + \frac{\partial}{\partial \omega}(f(\omega)S(t,\omega)) = -\left(cD(t)\sigma(\omega) + v(t,\omega)\right) S(t,\omega) + \rho(\omega)I(t,\omega) + r(\omega)V(t,\omega), \tag{4.3}$$

$$\frac{\partial}{\partial t}I(t,\omega) + \frac{\partial}{\partial \omega}(g(\omega)I(t,\omega)) = cD(t)\sigma(\omega)(S(t,\omega)+V(t,\omega)) - (\rho(\omega)+\mu(\omega))I(t,\omega), \tag{4.4}$$

$$\frac{\partial}{\partial t}V(t,\omega) + \frac{\partial}{\partial \omega}(h(\omega)V(t,\omega))) = -\left(cD(t)\sigma(\omega) + r(\omega)\right) V(t,\omega) + v(t,\omega)S(t,\omega), \tag{4.5}$$

with initial conditions

$$S(0,\omega) = S^0(\omega), \quad I(0,\omega) = I^0(\omega), \quad V(0,\omega) = 0, \quad \omega \geq 0. \tag{4.6}$$

and boundary conditions

$$S(t,1) = 0, \quad I(t,0) = 0, \quad V(t,0) = 0, \quad t \geq 0. \tag{4.7}$$

4.2 Optimal vaccination policies

In this subsection we use the model involving the vaccination rate $v(t,\omega)$ to formulate an optimal control problem reflecting the desire of acting in an rational way. There is a number of reasonable objectives considered in the literature, involving burden on hospitals, number of sick individuals, cost of policy measures, direct or indirect economic losses due to the disease, e.g., work absenteeism, etc. (see e.g. Bloom et al. [2020], Caulkins et al. [2021]). We focus on the following three objectives posed on a fixed time-horizon $[0,T]$:

(i) the total number of deaths; on average this number represents also the total number of sick people, hence it also reflects the economic cost of the epidemics in absence of additional restrictive measures such as partial lock-down (not employed anymore in most of the countries after 2022);

(ii) the social tension created by the pressure that people experience when the vaccination effort is high; the decision makers, i.e. governments have to take into account this social tension, especially in countries in which vaccination is obligatory or semi-obligatory due to the need of vaccination certificates for many activities;

(iii) the cost of vaccination, which is perhaps less significant than the first two especially in public health emergency situations.

The first objective is clearly contradictory to the other two. Therefore, in the spirit of Pareto's approach, we define the weighted aggregated objective to be minimized as

$$J(v) := \int_0^t \int_0^1 \mu(\omega) I(t,\omega)\,\mathrm{d}\omega\,\mathrm{d}t + \alpha \int_0^t \int_0^1 v(t,\omega)^2\,\mathrm{d}\omega\,\mathrm{d}t + \beta \int_0^t \int_0^1 v(t,\omega)S(t,\omega)\,\mathrm{d}\omega\,\mathrm{d}t \tag{4.8}$$

Here, $\alpha \geq 0$ and $\beta \geq 0$ are weighting parameters. The optimization is subjected to the constraints (4.2)–(4.7) and the control constraint $v(t,\omega) \geq 0$.

The optimal control problem (4.2)–(4.8) is non-standard, due to the presence of first order PDEs with different characteristic equations. We do not deeply investigate, in this paper, issues as existence of a solution, necessary optimality conditions, convergence of numerical methods, etc. As seen in Appendix, even the proof of existence of a solution of system (4.2)–(4.7) for a fixed control v is non-straightforward. However, due to the linear-convex form of the objective functional and the structure of the equations (4.3)–(4.5), one may expect that an optimal solution exists and the optimal control is Lipschitz continuous. Although this is far not enough to claim convergence of our numerical approach, the results of the numerical experiments we have made (see the next section) and the pertaining sensitivity analysis support such a claim.

The numerical approach we employ is the so-called *direct method* in optimal control, which consists of direct discretization of the equations and the objective functional in time and space (for ω), as briefly described in Subsection 5.1.

5 Numerical experiments

In the following section we provide several purely illustrative numerical experiments for the evolution of the model dynamics with and without vaccination. Moreover, we also analyze the impact of optimal vaccination policies among sub-populations with differing immunity level.

5.1 Numerical approximation

In order to obtain numerical solution to (4.2)–(4.7) we use the so called upwind scheme which is of first order accuracy, see LeVeque [2002]. We can write equation (4.3) in the following way:

$$\frac{\partial}{\partial t}S(t,\omega) + \frac{\partial}{\partial \omega}\big(f(\omega)S(t,\omega)\big) = F^S\big(t,\omega,D(t),Z(t,\omega)\big), \tag{5.1}$$

with the initial/boundary conditions (4.6) and (4.7). Here, $Z(t,\omega) := \big(S(t,\omega), I(t,\omega), V(t,\omega)\big)$ and F^S is the right-hand side of (4.3).

In order to describe the numerical scheme, we define a mesh $\omega_j, j = 1, ..., M+1$, in the ω-dimension with step size $\Delta\omega = \omega_{i+1} - \omega_i$. Respectively, in t-dimension we construct also a mesh $t_i, i = 1, ..., N+1$, with step size $\Delta t = t_{i+1} - t_i$. The upwind scheme is the following:

$$\frac{S(t_{i+1},\omega_j) - S(t_i,\omega_j)}{\Delta t} = -f(\omega_j)\Big(\frac{S(t_i,\omega_{j+1}) - S(t_i,\omega_j)}{\Delta\omega}\Big) + F^S(t_i,\omega_j,D(t_i),Z(t_i,\omega_j)), \tag{5.2}$$

for $i = 1, ..., N$ and $j = 1, ..., M$. From the boundary condition (4.7) we have that $S(t_i,\omega_{M+1}) = 0$, for every grid point t_i.

The numerical scheme has to take also the sign of the functions f, g and h into consideration. For the equations (4.4)–(4.6) we have to change f with g or h, the numerator on the right-hand side to $I(t_i,\omega_j) - I(t_i,\omega_{j-1})$ or $V(t_i,\omega_j) - V(t_i,\omega_{j-1})$, and account for the zero boundary condition (4.7).

A necessary condition for the convergence of the numerical upwind scheme is the Courant–Friedrichs–Lewy condition (CFL), see Courant [1967]. In our case this conditions is resolved to

$\frac{u\Delta t}{\Delta\omega} \leq C < 1$, where $u = \max_\omega\{f(\omega), g(\omega), h(\omega)\}$ and C is the Courant number for the problem. We tested the numerical scheme with various step sizes that satisfy the CFL condition in order to check consistency in the results.

5.2 Model parameters

The parameters used for numerical experiments are described in form of a baseline scenario, which varies later on in order to analyze the sensitivity with respect to some parameters. All parameter values are chosen for illustrative purposes only and do not refer to a specific disease. While the present work describes a general model and analyzes some of its properties, substantial empirical work remains to be done in order to apply it to observed data.

We choose the initial distribution of the population compartments to be consistent with the boundary conditions provided in (2.7), respectively (4.7). For the distribution of the susceptible population at time zero, we take a linear function in the ω variable. It has a maximum at $\omega = 0$ and it is decreasing to zero at $\omega = 1$. The susceptible population at the first moment is chosen to be 95% of the whole initial population. In a similar manner, for distribution of the initial infected population a parabola is chosen with a maximum at $\omega = 0.5$ and zero at the boundary immunity levels $\omega = 0$ and $\omega = 1$. The infected population at time zero is 5% of the total initial population. All numerical simulations have zero vaccinated people at the start.

The parameters and functions for modeling contact rates, infectiousness, recovery and mortality are summarized in Table 1.

Model parameters and functions		
contact rate of susceptibles	c	8
contact rate of infected individuals	c_I	3
susceptibility	$\sigma(\omega)$	$(1.5-\omega)^3$
recovery	$\rho(\omega)$, $r(\omega)$	$\frac{120^{\omega^{1.3}}}{40}$
infectiousness	$i(\omega)$	$0.3(1-0.9\omega)$
mortality	$\mu(\omega)$	$0.01(1-\omega)$
immunity decrease (susceptibles)	$f(\omega)$	-0.005ω
immunity increase (infected individuals)	$g(\omega)$	$0.04(1-\omega)$
immunity increase (vaccinated individuals)	$h(\omega)$	$0.02(1-\omega)$
convex cost parameter	α	0.00001
administration cost parameter	β	0.0005

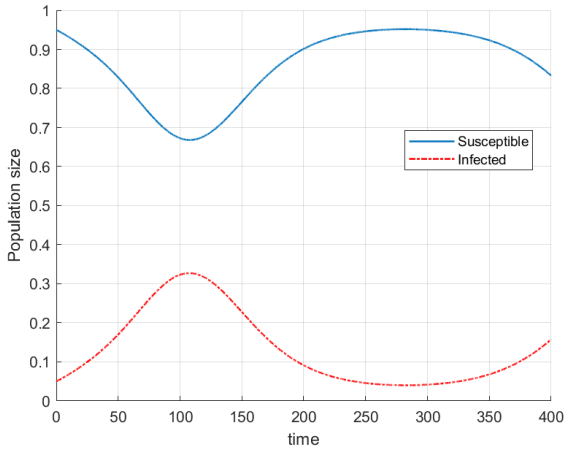
Table 1: Parameters and functions used in the numerical examples, $\omega \in [0, 1]$.

Table 1 also shows the specific functions f, g and h, which are chosen as linear functions such that their sign determines the gaining or loosing immunity over time. Figure 3.1 shows the resulting characteristic curves for solving (2.2) and (2.1).

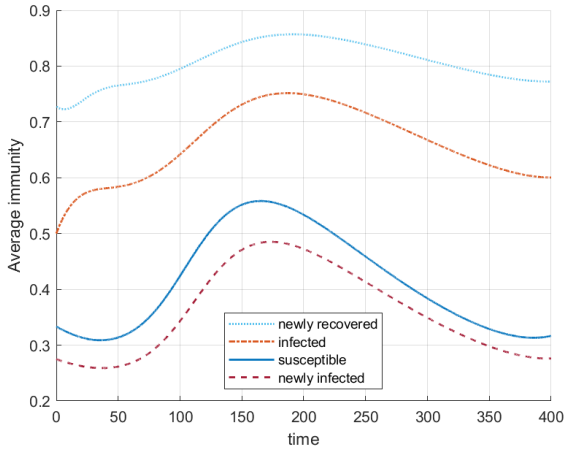
5.3 Numerical results without vaccination

We start with numerical simulations for the model without vaccination in the baseline case.

Figure 5.1 gives an overview of the development of the epidemic over the simulation horizon. Figure 5.1a shows the first wave of the epidemic plus the emergence of a second wave. Here the numbers of individuals in the compartments of susceptible and infected individuals are aggregated over all immunity levels ω. Figure 5.1b shows the development of the average immunity level for susceptible and infected individuals, weighted by the sizes of subgroups with different immunity level. In addition, the average immunity level of newly infected and newly recovered individuals is shown. In particular, one can observe an increase of the average immunity level in all compartments after the peak of the number of infected individuals in the first plot 5.1a. In the compartment of susceptible individuals the average immunity increases due to the inflow of recovered individuals, until the number of infected individuals becomes very low again.

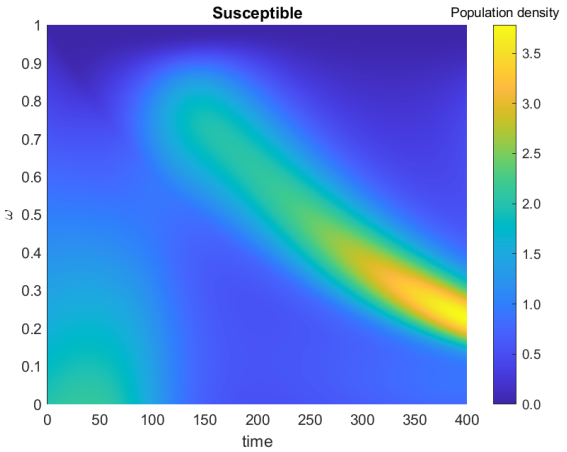


(a) Total number of susceptible and infected individuals

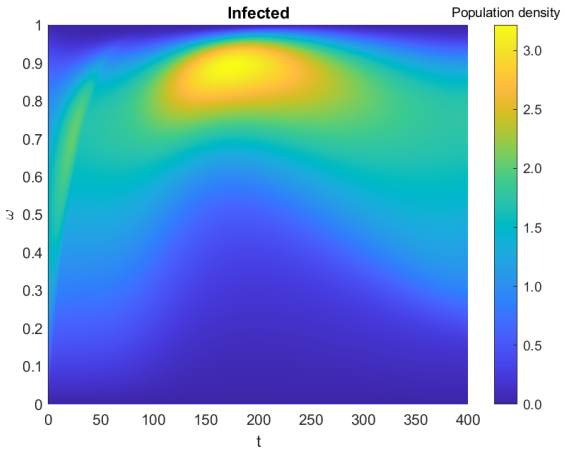


(b) Change of average immunity in time for different groups

Figure 5.1: Evolution of epidemiological population groups without vaccination

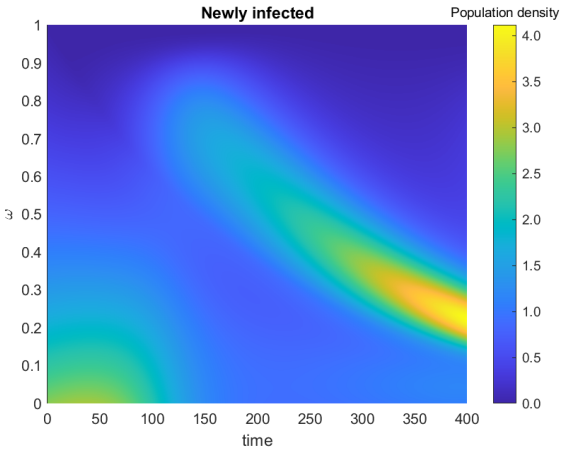


(a) Change of the distribution of susceptible individuals

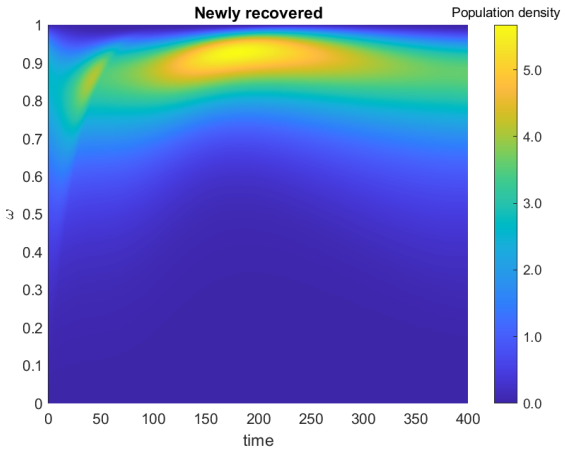


(b) Change of the distribution of infected individuals

Figure 5.2: Evolution of susceptible and infected individuals without vaccination



(a) Change of the distribution of newly infected individuals



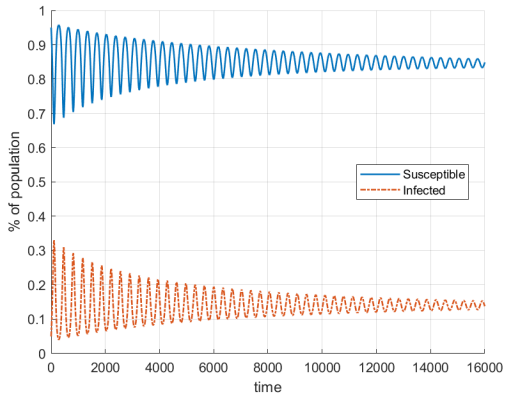
(b) Change of the distribution of newly recovered individuals

Figure 5.3: Evolution of newly infected and newly recovered individuals

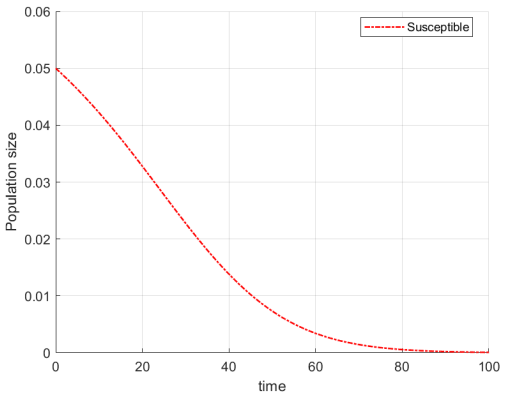
Figures 5.2 shows the time dependence of the distribution of immunity levels within the epidemiological groups: the susceptible and infected individuals and the newly infected and recovered individuals. For each point in time (on the horizontal axis), the related vertical cut pictures the distribution of immunity levels as a density, i.e. the integral over the immunity levels is one. The bright yellow spots on the graphs correspond to higher concentration of individulas with given immunity and the dark blue color corresponds to a lower concentration.

The groups of newly infected/newly recovered individuals at each time are described by the corresponding transition rates between the compartments of susceptible and infected individuals. As these rates have instantaneous magnitudes we also plot the time dependent normalized distribution of the immunity levels for the groups of newly infected and newly recovered in figure 5.3. One expected difference between the newly infected and newly recovered individuals, as shown previously, is the higher immunity level of the latter. Moreover the distribution of immunity levels for the newly infected is a slightly distorted version of the distribution for the susceptibles, where the newly infected individuals come from. The distribution of immunity levels within the newly recovered individuals on the other hand is concentrated at the highest levels of immunity, seen within the compartment of infected individuals, which includes individuals with increasing immunity until recovery.

In order to observe the long term equilibrium behaviour of the model, we simulate the scenario where there is no mortality $\mu(\omega) = 0$, and run the simulations for a longer period. In figure 5.4a we see the obtained behaviour numerically for running the simulation with $\mu(\omega) = 0$ for 16000 days. An oscillatory behaviour is observed with declining amplitude suggesting convergence to an equilibrium. In figure 5.4b is shown the decrease of the group of infected individuals when the contact rate c_I of infected individuals is reduced to 0.1. For this value the expression in proposition 3.2 is satisfied and we observe convergence to an infection-free equilibrium.



(a) Evolution of susceptible and infected individuals, long-term estimation



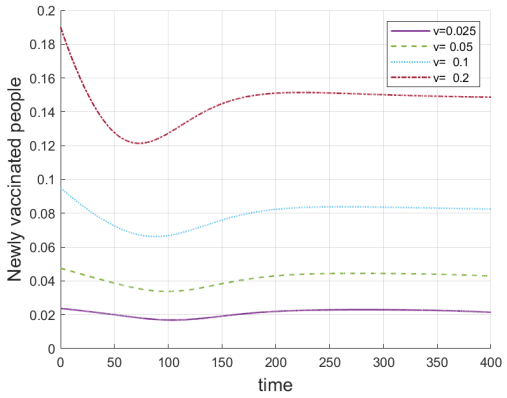
(b) Evolution of infected individuals, low contact rate $c_I = 1.2$

Figure 5.4: Evolution of epidemiological groups with no mortality rate

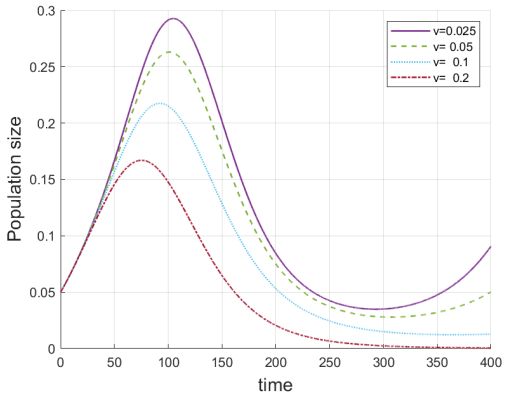
5.4 Simulations with constant vaccination

Although the vaccination rate $v(t, \omega)$ is modeled to depend on time t and immunity level ω, we consider first the special case of constant vaccination: at every point in time and for every immunity level we vaccinate susceptible individuals with the same intensity.

Results for different constant vaccination intensities are shown in Figure 5.5 where the number of vaccinated individuals can be seen in 5.5a. Due to the varying number of susceptibles over time, the number of vaccinated individuals also changes over time. In figure 5.5b one can observe the numbers of infected individuals, depending on the vaccination rate. Choosing $v(t, \omega) = 0.2$ drives the number of infected individuals to almost zero.



(a) Number of newly vaccinated individuals for different vaccination intensities



(b) Number of infected individuals obtained after different vaccination intensities

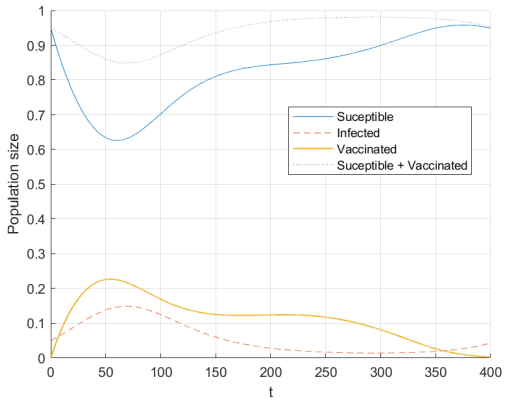
Figure 5.5: Evolution of epidemiological population groups with constant vaccination

5.5 Optimal vaccination

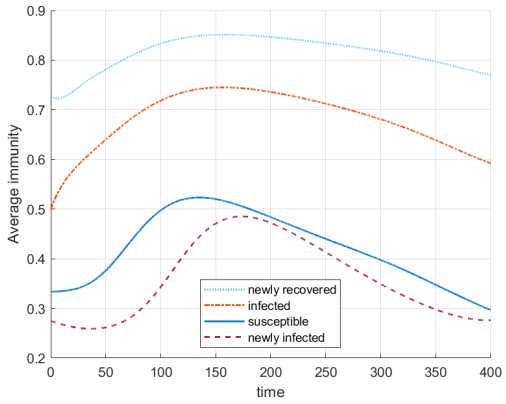
We obtain numerically the optimal vaccination policy by solving the formulated problem in Subsection 4.2. We minimize the number of deaths with an additional vaccination cost. The objective function is directly discretized over the same mesh that was used for equations (4.3) – (4.6). Afterwards we utilize an optimization procedure to obtain numerically optimal solution. For the discrete optimization problem we choose the SQP optimization solver provided by the MATLABs Optimization Toolbox.

For illustration we chose $\alpha = 0.00001$ and $\beta = 0.0005$, as shown in Table 1.

Figure 5.6 shows the effect of the vaccine strategy in the baseline case. In addition to the overall numbers of susceptible and infected individuals, the yellow line in 5.6a indicates the total number of individuals in the vaccinated group. These are individuals who are in the process of acquiring immunity due to the vaccination. As expected, the overall level of infected individuals is greatly reduced compared to the baseline case without vaccination in figure 5.1a. In figure 5.6 we see the average immunity of all epidemiological population groups. Comparing the previous plot 5.1b and 5.6b, the average immunity of the different compartments does not change significantly. This suggest that the vaccination efforts do not change the qualitative effect of the infection within the population, but affects the group of the vaccinated individuals and the overall infection numbers.



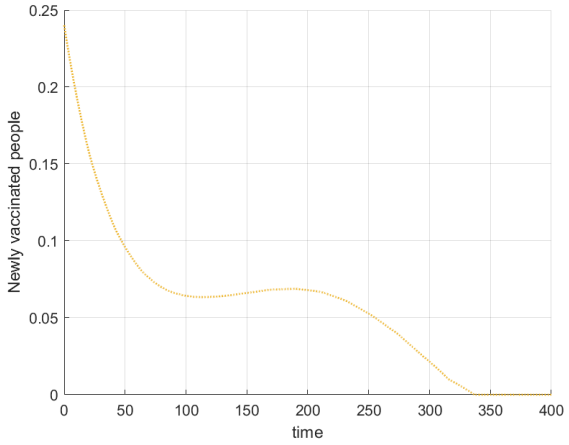
(a) Number of susceptible and infected individuals



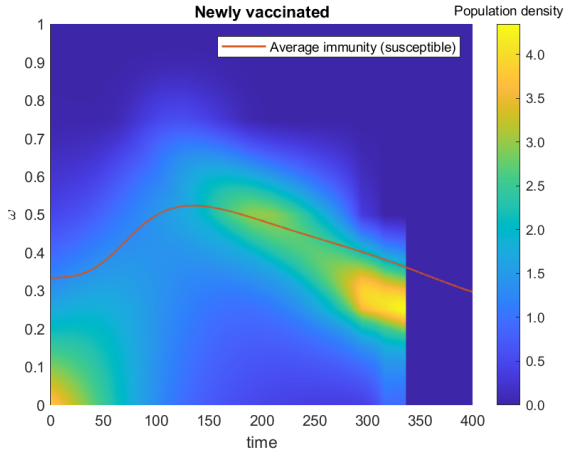
(b) Average immunity within the epidemiological population groups

Figure 5.6: Evolution of epidemiological population groups with optimal vaccination applied

The optimal vaccination policy is analyzed in Figure 5.7. Figure 5.7a shows the optimal total number of newly vaccinated individuals. Due to the finite horizon of the optimal control, the vaccination is terminated before the end of the simulation period. However, the optimal vaccination policy is not only a matter of the overall level and timing of vaccination. Figure 5.7b shows the distribution of the application of vaccines to individuals with differing immunity level over time. The abrupt change at the end of the horizon is due to the stop of vaccination. It can be seen that, at the beginning, vaccination tends to be given to individuals with lower immunity level than the average immunity level in the susceptible group. The levels of immunity of vaccinated individuals catch up with the average immunity level of susceptible individuals after around 150 days, and then follow the general decrease of the average level of immunity. Figure 5.8 shows the increase of the total number of deaths. The optimal vaccination shows a visible reduction of the number of deaths.



(a) Number of newly vaccinated individuals



(b) Normalized distribution of the newly vaccinated individuals and average immunity of the susceptible group

Figure 5.7: Administered vaccines, i.e. newly vaccinated individuals and comparison with the average immunity level of susceptible

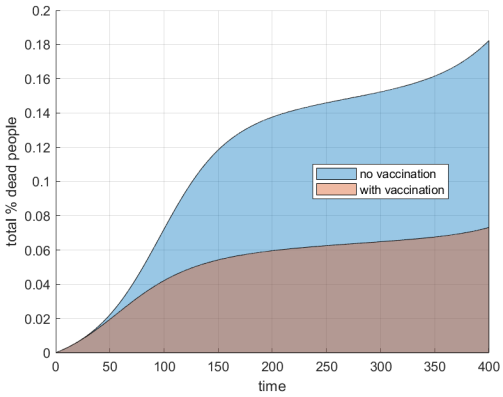


Figure 5.8: Comparison between the proportion of deaths resulting from no vaccination and application of optimal vaccination

The objective (4.8) puts together several aims, the number of deaths and two kinds of vaccination costs. All these aims are relevant for decision making in public health, but also are contradicting. In order to formulate the overall objective, weights are applied, modeling the relative importance of the individual aims.

While the analysis of the optimization problem so far is targeted on the analysis of one ("standard") choice for these weights, it is possible to go deeper by analyzing the efficient (or Pareto) frontier of costs and deaths: Iterating over the possible weight combination one gets for each cost the minimum number of deaths that can be achieved by vaccination.

As the vaccination cost in our case is composed of two parts, we vary only the administration cost in order to plot an efficient frontier curve (holding fixed the other parameters of the baseline case). We range the parameter β in the objective from 0.001 to 0.02, we calculate the optimal vaccination policy, and estimate the corresponding administration vaccination cost. In figure 5.9,

the vaccination administration cost is compared to the percentage of increase or decrease in the number of deaths relative to the baseline case with the standard choice of weights. The more one goes to the left, the more costly becomes additional reduction of the number of deaths. The more one goes to the right, the more additional deaths will occur per monetary unit saved.

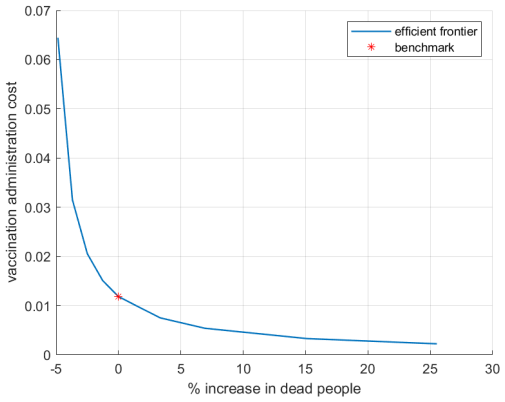


Figure 5.9: Vaccination administration cost compared to the relative increase of the number of deaths for the besaline case

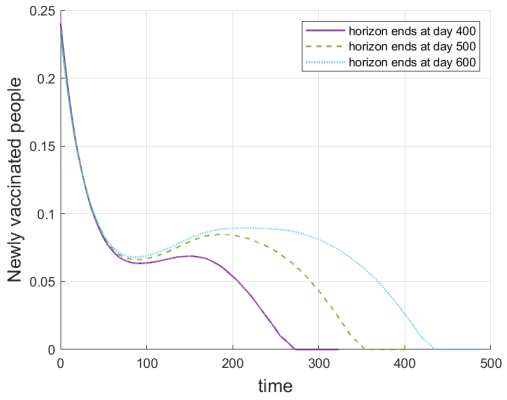
In practice, any vaccination policy has to be revised after some time, in order to catch up with new information. In particular, the improved medical understanding of the disease, changes in the death rates, new variants of the pathogens may emerge and enhanced vaccines may be developed. In terms of control, a new optimization is done after some time with updated information, which is known as Model Predictive Control in the literature.

Following such an approach, we may solve the optimization problem on a relatively short horizon, e.g., $t = 400$ days in the baseline case, apply the obtained solution during an even shorter time horizon, say 70 days, then update the model parameters and the current real state of the epidemic, solve the problem on the next 400 days horizon, and so on.

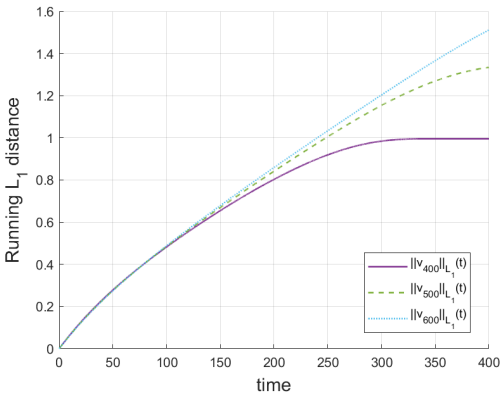
Such an approach only works well, if the results with different planning horizons do not vary too much over the shorter time period (here 70 days). This is tested in Figure 5.10, which shows the dependence of the optimal vaccination policy on the chosen time horizon $[0, t]$ on which the optimization problem is solved. The left plot indicates that the total number of optimally vaccinated people is practically independent of the time horizon over the first 70 to 100 days. More relevant, the same applies (even on a longer horizon) to the aggregated (in ω) vaccination effort (the right Figure 5.10). So it seems to be reasonable to apply Model Predictive Control.

5.6 Comparative analysis

An interesting case for model sensitivity is the effect of parameter changes for the objective function values. As contact rate plays a significant role in pathogen transmission, we estimate the mortality and vaccination cost for different contact rates. Using the baseline values for $c = 8$ and $c_I = 3$, we multiply these parameters by a factor ranging from 0.8 to 1.25. For each parameter value we estimate the optimal vaccination policy and compare the respective objective function values. Comparison is made also between the mortality and vaccination cost parts in the cost objective. In Figure 5.11, the x-axis shows the values for the contact rate c and on the y-axis are the obtained values of the objective function using the respective optimal vaccination policy. In the first plot



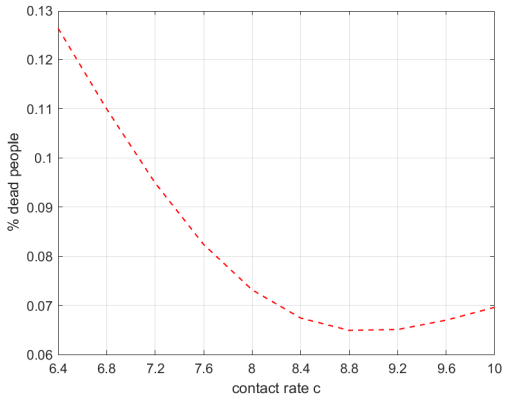
(a) Number of vaccinated individuals for different time horizons



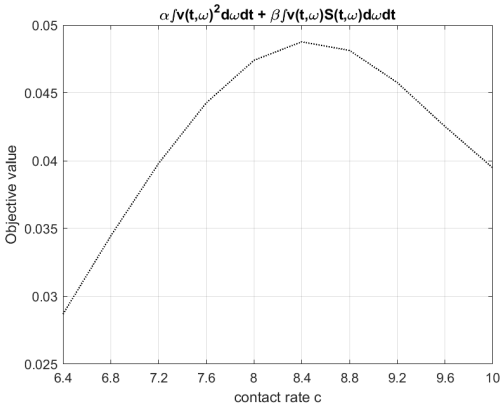
(b) Aggregated optimal control trajectories for different time horizons

Figure 5.10: Dependence of the optimal vaccination policy on the time horizon $[0, t]$, $t = 400, 500, 600$

5.11a the lowest mortality cost is for the contact rate $c = 8.8$ and in the second plot 5.11b the highest vaccination cost occurs for $c = 8.4$. From the two plots one can observe that lower contact rate does not necessary mean lower mortality. As the contact rate increases, the optimal vaccination efforts also increase up to a point. The vaccination leads to lower mortality rates, but we can also observe from figure 5.11b, that the vaccination cost for highest values of the contact rate is reduced. This fact can be explained by the effect of herd immunity. With higher contact rate more people obtain immunity from infection and the effect of vaccination is relatively smaller.



(a) Total number of deaths for different contact rates



(b) Vaccination cost for different contact rates

Figure 5.11: Objective function components for different contact rates

6 Discussion

In this work we develop an epidemiological model that explicitly embeds the effects of waning immunity after infection or vaccination. The basic model with dynamic immunity is formulated as a system of two PDEs of first order. The model is mathematically challenging because the transport flows in the equations are different from each other. Systems of this type seem not to be investigated in the literature. Existence of a global solution is proved in the paper, and its asymptotic behaviour is investigated. In a second step we embed vaccination strategies and formulate an optimal control problem with the following three objectives: the total number of deaths, the social tension created by the pressure that people experience when the vaccination effort is high, and the costs of vaccination. Using plausible scenarios of vaccination, numerical results provide insights into the dynamics of the epidemiological populations involved including the waning immunity without and with vaccination. With respect to the optimal vaccination strategy the model provides insights into the influence of different contact rates between individuals on the number of deaths or the vaccination costs. An interesting fact, for example, is that vaccination efforts and herd immunity act in a certain sense as substitutes: above a threshold value of the contact rate, further increase of the contact rate leads to lower vaccination level. Moreover, one can determine the efficient frontier between vaccine administration (direct and indirect) costs and the number of deaths.

As already mentioned in the introduction the epidemiological model presented here that describes the development of immunity at the epidemiological population level differs substantially from previous approaches (White and Medley [1998]; Rouderfer and Becker [1994]; Barbarosa and Röst [2015]; Ehrhardt et al. [2019]). It has a broader potential for inclusion of heterogeneity aspects especially about immunity and more generally applicable results. Moreover, the model is used to investigate optimal control policies.

Although the model has some striking features such as the description and coupling of the elicitation of immune responses to the epidemiological process it can only be considered as a first step towards a more detailed description of the immune responses and their waning over time. The model captures the immune response in a general manner as a whole and does not distinguish between antibody and cellular immune responses and their differing dynamical characteristics such as the time delay between the two and their strength depending on the infectious disease one may explore. There is still substantial clinical and epidemiological empirical work to be done where relevant immuno-epidemiological data could feed this type of models. Nevertheless, the model shows that a mathematical description of these dynamical processes is possible and this may lead to better understanding of the processes themselves and the evaluation of associated interventions.

Appendix: Proof of Theorem 3.1

Since the horizon T may change in the subsequent considerations, at some places we use the notation $\Gamma^T := [0,T]\times[0,1]$. The space of all continuous functions from a set $X \subset \mathbb{R}^n$ to $\mathbb{R}$ or $\mathbb{R}^2$ (the l^1-norm will be used in the latter) is denoted by $C(X)$, with the usual norm denoted by $\|\cdot\|_{C(X)}$. The spaces $L^1(0,T)$ and $L^\infty(0,T)$ are defined as usual. Further, $\mathcal{L}(X) \subset C(X)$ is the subspace of all Lipschitz continuous functions with $\mathrm{Lip}(x)$ denoting the Lipschitz constant of $x \in \mathcal{L}(X)$. We abbreviate

$$F = \begin{pmatrix} F^S \\ F^I \end{pmatrix} \qquad Z = \begin{pmatrix} S \\ I \end{pmatrix} \quad \text{or} \quad Z = (S,I), \quad \text{etc.}$$

The existence theorem presented below will be formulated in the terms of a general function F in the equations (3.3), (3.4), and with a general relation between the functions D and (S,I). Namely, instead of equation (2.3) we consider a general relation between Z and D:

$$D = \mathcal{D}(Z), \quad \mathcal{D} : \mathrm{dom}\,\mathcal{D} := \{Z \in \mathcal{L}(\Gamma^T) : \, Z \geq 0\} \to L^1(0,T), \tag{6.1}$$

where $\mathcal{D}$ has the following properties: there exists constants $a > 1$ and $L_{\mathcal{D}}$ such that

$$\mathcal{D}(Z)(t) \geq 0, \tag{6.2}$$

$$|\mathcal{D}(Z_1)(t) - \mathcal{D}(Z_2)(t)| \leq L_{\mathcal{D}} \max_{\omega\in[0,1]} |Z_1(t,\omega) - Z_2(t,\omega)|, \tag{6.3}$$

$$|\mathcal{D}(Z)(t)| \leq a - 1, \quad \text{for a.e. } t \in [0,T], \quad Z, Z_1, Z_2 \in \operatorname{dom}\mathcal{D}. \tag{6.4}$$

Keeping in mind the specific form of the functions F^S and F^I in (3.1), (3.2), we assume in the general case that there exist constants M and L such that

$$|F(t,\omega,d,s,i)| \leq Ma(|s| + |i|), \tag{6.5}$$

$$|F(t,\omega,d,s,i)| - F(t_1,\omega_1,d_1,s_1,i_1)| \leq La\big((|s| + |i|)(|t - t_1| + |\omega - \omega_1| + |d - d_1|) + |s - s_1| + |i - i_1|\big) \tag{6.6}$$

for all $t, t_1 \in [0,1]$, $\omega, \omega_1 \in [0,1]$, $d, d_1, s, s_1, i, i_1 \in \mathbb{R}$ with $|d|$, $|d_1| \leq a - 1$. Moreover, the following property is fulfilled: for any $d \geq 0$, $(t,\omega) \in \Gamma$

$$F^S(t,\omega,d,0,i) \geq 0 \quad \forall i \geq 0, \qquad F^I(t,\omega,d,s,0) \geq 0 \quad \forall s \geq 0. \tag{6.7}$$

Theorem 6.1. *Let the functions f, g, S^0, I^0 satisfy the Standing Assumptions (at the beginning of Subsection 3.1). Let, in addition, the conditions , (6.2)–(6.4) and (6.5)–(6.7) be fulfilled. Then there exists $T > 0$, independent of the particular initial data (S^0, I^0), such that the system (3.3), (3.4), (6.1) has a nonnegative Lipschitz continuous continuous solution (S, I) on Γ^T.*

Proof. 1. We begin with some preliminary facts and notations. Due to the properties of f and g, the functions $(\gamma, s) \to \omega^f[\gamma](s)$ and $(\gamma, s) \to \omega^g[\gamma](s)$ are continuously differentiable on a neighborhood of $\Gamma^T \times [0,T]$ ($T > 0$ is arbitrary). Denote by λ_ω a common Lipschitz constant of these functions on $\Gamma^1 \times [0,1]$. Moreover, the functions $\Gamma^1 \ni \gamma \to \gamma^f(\gamma)$ and $\Gamma^1 \ni \gamma \to \gamma^g(\gamma)$ are Lipschitz continuous and we denote by λ_γ a common Lipschitz constant. We remind that $\omega^f[\gamma](t)$, $\omega^g[\gamma](t) \in [0,1]$ for every $\gamma \in \Gamma^1$ and $t \in [0,1]$.

Let us fix the number $T > 0$ such that

$$T \leq \min\left\{1, \frac{1}{4L\lambda_\omega\, a}, \frac{1}{2La}\right\}. \tag{6.8}$$

Notice that T does not depend on the initial data (S^0, I^0).

Let us fix a $D \in L^\infty(0,1)$ with $\|D\|_{L^\infty} + 1 \leq a$. Set

$$\lambda := \max\left\{4\lambda_\gamma,\, 8ae^{2Ma}(2M(1 + \lambda_\gamma) + L\lambda_\omega)\right\} \tag{6.9}$$

and define the set

$$\begin{aligned} \mathcal{K}_{T,a} \quad := \quad & \Big\{Z = (S,I) \in \mathcal{L}(\Gamma^T) :\, \operatorname{Lip}(Z) \leq (\|Z^0\|_{C(0,1)} + \operatorname{Lip}(Z^0))\lambda, \\ & \|Z\|_{C(\Gamma^t)} \leq 2e^{2Mat}\|Z^0\|_{C(0,1)} \ \forall t \in [0,T], \ (2.6) \text{ and } (2.7) \text{ are satisfied}\Big\}. \end{aligned} \tag{6.10}$$

On $\mathcal{K}_{T,a}$ we define the mapping $\mathcal{F}_{[D]}$ as

$$\mathcal{F}^S_{[D]}(Z)(\gamma) \quad := \quad \int_{\tau^f(\gamma)}^t F^S(s, \omega^f[\gamma](s), D(s), Z(s, \omega^f[\gamma](s)))\,\mathrm{d}s + \bar{S}^0(\gamma^f(\gamma)), \tag{6.11}$$

$$\mathcal{F}^I_{[D]}(Z)(\gamma) \quad := \quad \int_{\tau^g(\gamma)}^t F^I(s, \omega^g[\gamma](s), D(s), Z(s, \omega^g[\gamma](s)))\,\mathrm{d}s + \bar{I}^0(\gamma^g(\gamma)), \tag{6.12}$$

where $\gamma = (t,\omega) \in \Gamma^T$. In the next three parts of the proof we shall prove that $\mathcal{K}_{T,a}$ is a nonempty complete metric space, $\mathcal{F}_{[D]}$ maps $\mathcal{K}_{T,a}$ to $\mathcal{K}_{T,a}$, and $\mathcal{F}_{[D]}$ is contractive.

2. Let us prove that $\mathcal{K}_{T,a}$ is not empty. For $\gamma \in \Gamma^T$ we set $S^\#(\gamma) := \bar{S}^0(\gamma^f(\gamma))$ and $I^\#(\gamma) := \bar{I}^0(\gamma^g(\gamma))$ (representing the evolution of the initial/boundary data if $F \equiv 0$). Since $Z^0 \in \mathcal{L}(0,1)$ and $S^0(1) = I^0(0) = 0$, the function $\bar{S}^0$ is Lipschitz continuous with $\operatorname{Lip}(\bar{S}^0) = \operatorname{Lip}(S^0)$. Thus $\operatorname{Lip}(S^\#) \leq \operatorname{Lip}(S^0)\operatorname{Lip}(\gamma^f) = \operatorname{Lip}(Z^0)\lambda_\gamma$. The same applies to $I^\#$, thus the first inequality in the definition of $\mathcal{K}_{T,a}$ is fulfilled by $Z^\#$. The second inequality is also fulfilled since $\|\mathcal{Z}^\#\|_C \leq \|Z^0\|_C$. The conditions (2.6) and (2.7) are apparently also fulfilled, thus $\mathcal{K}_{T,a} \neq \emptyset$.

Due to the uniform Lipschitz property in the definition of the set $\mathcal{K}_{T,a}$, it is a complete metric space in the metric induced by the norm in $C(\Gamma^T)$.

3. (Invariance of $\mathcal{K}_{T,a}$.) Obviously for $\gamma = (0,\xi)$ we have $\mathcal{F}^S_{[D]}(Z)(\gamma) = \bar{S}^0(\gamma^f(0,\xi)) = S^0(\xi)$, and for $\gamma = (\tau,1)$ we have $\mathcal{F}^S_{[D]}(Z)(\gamma) = \bar{S}^0(\gamma^g(\tau,1)) = \bar{S}^0(\tau,1) = 0$, thus $\mathcal{F}^S_{[D]}(Z)$ satisfies the side conditions in (2.6) and (2.7). The same applies to $\mathcal{F}^I_{[D]}(Z)$.

Fix an arbitrary $Z \in \mathcal{K}_{T,a}$. Using (6.5), we have for any $\gamma = (t, \omega) \in \Gamma^T$

$$\begin{aligned} |\mathcal{F}^S_{[D]}(Z)(\gamma)| &\leq |\bar{S}^0(\gamma^f(\gamma))| + \int_{\tau^f(\gamma)}^t Ma(|S(\theta, \omega^f[\gamma](\theta))| + I(\theta, \omega^f[\gamma](\theta))|)\,\mathrm{d}\theta \\ &\leq \|Z^0\|_C + \int_{\tau^f(\gamma)}^t 2Mae^{2Ma\,\theta}\|Z^0\|_C\,\mathrm{d}\theta. \end{aligned}$$

Then

$$|\mathcal{F}_{[D]}(Z)(\gamma)| \leq 2\|Z^0\|_C\Big(1 + 2Ma\int_0^t e^{2Ma\,\theta}\,\mathrm{d}\theta\Big) = 2\|Z^0\|_C e^{2Ma\,t}.$$

Thus $\mathcal{F}_{[D]}(Z)$ fulfills the growth condition in the definition of $\mathcal{K}_{T,a}$.

For any $Z \in \mathcal{K}_{T,a}$, $\gamma_1 = (t_1, \omega_1), \gamma_2 = (t_2, \omega_2) \in \Gamma^T$ we have

$$\begin{aligned} &|\mathcal{F}^S_{[D]}(Z)(\gamma_1) - \mathcal{F}^S_{[D]}(Z)(\gamma_2)| \leq |\bar{S}^0(\gamma^f(\gamma_1)) - \bar{S}^0(\gamma^f(\gamma_2))| \\ &\quad + \Big|\int_{\tau^f(\gamma_1)}^{t_1} F^S(\theta_1, \omega^f[\gamma_1](\theta), D(\theta), z_1(\theta))\,\mathrm{d}\theta - \int_{\tau^f(\gamma_2)}^{t_2} F^S(\theta, \omega^f[\gamma_2](\theta), D(\theta), z_2(\theta))\,\mathrm{d}\theta\Big|, \end{aligned}$$

where $z_i(\theta) := S(\theta, \omega^f[\gamma_i](\theta))$. Denote $[\tau', \tau''] := [\tau^f(\gamma_1), t_1] \cap [\tau^f(\gamma_2), t_2]$. Then we split the above integrals into three parts (in each of the integrals only two parts may be non-degenerate). The integration in the first part is on an interval of length $|\tau^f(\gamma_1) - \tau^f(\gamma_2)| \leq \lambda_\gamma|\gamma_1 - \gamma_2|$, and in the third part – of length $|t_1 - t_2| \leq |\gamma_1 - \gamma_2|$. In view of (6.5) and the growth condition in the definition of $\mathcal{K}_{T,a}$, the integrands can be mojorated by $Ma(2e^{2Ma}\|Z^0\|_{C(0,1)})$. Then the sum of these integrals is at most $4Ma(1 + \lambda_\gamma)e^{2Ma}\|Z^0\|_{C(0,1)})|\gamma_1 - \gamma_2|$. The integral on $[\tau', \tau'']$ can be estimated by

$$\begin{aligned} &\int_{\tau'}^{\tau''} |F^S(\theta_1, \omega^f[\gamma_1](\theta), D(\theta), z_1(\theta)) - F^S(\theta, \omega^f[\gamma_2](\theta), D(\theta), z_2(\theta))|\,\mathrm{d}\theta \\ &\leq \int_{\tau'}^{\tau''} L\Big(a\left(2e^{2Ma}\|Z^0\|_{C(0,1)}\right)|\omega^f[\gamma_1](\theta) - \omega^f[\gamma_2](\theta)| + a\,|z_1(\theta) - z_2(\theta)|\Big)\,\mathrm{d}\theta \\ &\leq TL\lambda_\omega a\Big(2e^{2Ma}\|Z^0\|_{C(0,1)} + (\|Z^0\|_{C(0,1)} + \mathrm{Lip}(Z^0))\lambda\Big)|\gamma_1 - \gamma_2|, \end{aligned}$$

where we make use of (6.5), the growth condition and the Lipschitz property in the definition of $\mathcal{K}_{T,a}$. Combining the obtained estimations and using the same estimations for $\mathcal{F}^I_{[D]}(Z)$, we obtain that

$$\begin{aligned} &|\mathcal{F}_{[D]}(Z)(\gamma_1) - \mathcal{F}_{[D]}(Z)(\gamma_2)| \leq \Big(2\lambda_\gamma \mathrm{Lip}(Z^0) + 8Ma(1 + \lambda_\gamma)e^{2Ma}\|Z^0\|_{C(0,1)} \\ &+ 2TL\lambda_\omega a\big(2e^{2Ma}\|Z^0\|_{C(0,1)} + (\|Z^0\|_{C(0,1)} + \mathrm{Lip}(Z^0))\lambda\Big)|\gamma_1 - \gamma_2| \\ &= \Big((2\lambda_\gamma + 2TL\lambda_\omega a\,\lambda)\mathrm{Lip}(Z^0) + \big(4ae^{2Ma}(2M(1 + \lambda_\gamma) + L\lambda_\omega) + 2TL\lambda_\omega a\lambda\big)\|Z^0\|_{C(0,1)}\Big)|\gamma_1 - \gamma_2| \\ &\leq \Big[\Big(\frac{1}{2} + \frac{1}{2}\Big)\lambda\,\mathrm{Lip}(Z^0) + \Big(\frac{1}{2} + \frac{1}{2}\Big)\lambda\,\|Z^0\|_{C(0,1)}\Big]|\gamma_1 - \gamma_2| \leq \lambda\,(\mathrm{Lip}(Z^0) + \|Z^0\|_{C(0,1)})|\gamma_1 - \gamma_2|, \end{aligned}$$

where in the last inequality we have used (6.8) and (6.9). This completes the proof of the invariance of $\mathcal{K}_{T,a}$.

4. (Contractivity of $\mathcal{F}_{[D]}$.) For $Z, Z_1 \in \mathcal{K}_{T,a}$ we have, using (6.6) and (6.8),

$$\begin{aligned} |\mathcal{F}^S_{[D]}(Z)(\gamma) - \mathcal{F}^S_{[D]}(Z_1)(\gamma)| &\leq \int_{\tau^f(\gamma)}^t La|Z(s, \omega^f[\gamma](s)) - Z_1(s, \omega^f[\gamma](s))|\,\mathrm{d}s \\ &\leq TLa\|Z - Z_1\|_{C(\Gamma^T)} \leq \frac{1}{2}\|Z - Z_1\|_{C(\Gamma^T)}. \end{aligned}$$

According to the Banach contraction mapping theorem, for any function $D \in L^\infty(0, T)$ with $\|D\|_{L^\infty} + 1 \leq a$, there exists a unique $(S, I) = (S[D], I[D]) \in K_{T,D}$ that solves the system (3.3)–(3.4).

5. (Properties of $(S[D], I[D])$.) So far we have proved that there exists $T > 0$ such that for any $D \in L^\infty(0, 1)$ with $\|D\|_{L^\infty} + 1 \leq a$, the system (3.3)–(3.4) has a unique continuous solution $Z[D] = (S[D], I[D])$ on Γ^T. Here T is independent of Z^0. Moreover, $Z[D]$ is a Lipschitz continuous function with a constant $\lambda^* := (\|Z^0\|_{C(0,1)} + \mathrm{Lip}(Z^0))\lambda \leq 2\mathrm{Lip}(Z^0)\lambda$ (see (6.9) and (6.10)).

For a fixed D as above, we shorten the notation $Z[D] = (S[D], I[D])$ to $Z = (S, I)$. For any $\gamma = (\tau, \xi) \in \Gamma_f$ the function $z^S[\gamma](t) := S(t, \omega^f[\gamma](t))$ satisfies (due to the identities $\omega^f[t, \omega^f[\gamma](t)](s) = \omega^f[\gamma](s)$ and $\gamma^f(\omega^f[\gamma](t)) = \gamma$) the relation

$$z^S[\gamma](t) = \int_\tau^t F^S(s, \omega^f[\gamma](s), D(s), z^S[\gamma](s), I(s, \omega^f[\gamma](s)))\,\mathrm{d}s + \bar{S}^0(\gamma),$$

and an analogical equation is satisfied by $z^I[\gamma'](t) := I(t, \omega^g[\gamma'](t))$, $\gamma' \in \Gamma_g$. Differentiating these relations, we obtain the following ODEs satisfied by the Lipschitz functions $z^S[\gamma]$ and $z^I[\gamma']$ on $[0, T]$:

$$\dot{z}^S[\gamma](t) = F^S(t, \omega^f[\gamma](t), D(t), z^S[\gamma](t), I(t, \omega^f[\gamma](t))), \quad \gamma \in \Gamma_f, \tag{6.13}$$

$$\dot{z}^I[\gamma'](t) = F^I(t, \omega^g[\gamma'](t), D(t), S(t, \omega^g[\gamma'](t)), z^I[\gamma'](t)), \quad \gamma' \in \Gamma_g \tag{6.14}$$

(the so-called equations representing the solution along the characteristic lines). One can inversely express $S(t, \omega) = z^S[\gamma^f(t, \omega)](t)$ and similarly for I.

Now we shall prove that if D is non-negative, then (S, I) is also non-negative, making use of the property (6.7). Denote

$$p(t) := \min\big\{0, \min_{\omega \in [0,1]} S(t, \omega)\big\}, \qquad q(t) := \min\big\{0, \min_{\omega \in [0,1]} I(t, \omega)\big\}.$$

We have $p(0) = q(0) = 0$ and both functions are Lipschitz continuous. Let $[0, t_p]$ be a maximal sub-interval of $[0, T]$ such that $p(t) = 0$ for all $t \in [0, t_p]$. Similarly, let $[0, t_q]$ be the maximal interval on which $q(t) = 0$. If $t_p < t_q \le T$, then for every $\gamma \in \Gamma_f$ and $t \in [0, t_q]$ we have $I(t, \omega^f[\gamma](t)) \ge 0$, hence $F^S(t, \omega^f[\gamma], 0, I(t, \omega^f[\gamma](t))) \ge 0$. Then, by a standard argument, the set $s \ge 0$ is invariant with respect to (6.13) on $[0, t_q]$ for any $\gamma \in \Gamma_f$, thus $z^S[\gamma](t) \ge 0$ on $[0, t_q]$. This contradicts the definition of t_p and implies $t_q \le t_p$. Similarly we can prove that $t_p \le t_q$, thus $t_p = t_q =: \bar{t}$.

Assume that $\bar{t} < T$ and take an arbitrary $\gamma \in \Gamma_f$ and $t \in (\bar{t}, T]$. Consider two cases:
(i) $z^S[\gamma](t) \ge 0$;
(ii) $z^S[\gamma](t) < 0$.

In the second case there is a minimal number $t' \in [0, t)$ such that $z^S[\gamma](s) < 0$ on $(t', t]$. Since $Z^S[\gamma](\bar{t}) \ge p(\hat{t}) = 0$, we have that $t' \in [\bar{t}, t)$ and $z^S[\gamma](t') = 0$. In the expressions below we skip the first three arguments of F^S, namely, $s, \omega^f[\gamma](s), D(s)$, since they stay the same in all formulas. We have

$$\begin{aligned} z^S[\gamma](t) &= \int_{t'}^t F^S(z^S[\gamma](s), I(s, \omega^f[\gamma](s)))\,\mathrm{d}s \\ &= \int_{t'}^t \big(F^S(z^S[\gamma](s), I(s, \omega^f[\gamma](s))) - F^S(0, I(s, \omega^f[\gamma](s)))\big)\,\mathrm{d}s \\ &\quad + \int_{t'}^t F^S(0, I(s, \omega^f[\gamma](s)))\,\mathrm{d}s \\ &\ge -\int_{t'}^t La|z^S[\gamma](s)|\,\mathrm{d}s - \int_Q Ma|I(s, \omega^f[\gamma](s))|\,\mathrm{d}s, \end{aligned}$$

where $Q := \{s \in [t', t] : F^S(s, 0, I(s, \omega^f[\gamma](s))) < 0\}$. Notice that according to (6.7), $I(s, \omega^f[\gamma](s)) < 0$ on this set, hence $-|I(s, \omega^f[\gamma](s))| = I(s, \omega^f[\gamma](s)) \ge q(t)$. Also $-|z^S[\gamma](s)| = z^S[\gamma](s) \ge p(s)$. Then

$$z^S[\gamma](t) \ge La\int_{t'}^t p(s)\,\mathrm{d}s + Ma\int_Q q(s)\,\mathrm{d}s \ge C\int_{t'}^t (p(s) + q(s))\,\mathrm{d}s \ge C\int_{\bar{t}}^t (p(s) + q(s))\,\mathrm{d}s,$$

where $C := a\max\{L, M\}$. Combining the two cases we obtain that

$$z^S[\gamma](t) \ge \min\Big\{0, C\int_{\bar{t}}^t (p(s) + q(s))\,\mathrm{d}s\Big\} = C\int_{\bar{t}}^t (p(s) + q(s))\,\mathrm{d}s,$$

Since $\gamma \in \Gamma_f$ is arbitrary, this inequality implies

$$p(t) \ge C\int_{\bar{t}}^\theta (p(s) + q(s))\,\mathrm{d}s. \tag{6.15}$$

By the same argument, a similar inequality is fulfilled by q. Summing the two inequalities we obtain that

$$p(t) + q(t) \ge 2C\int_{\bar{t}}^t (p(s) + q(s))\,\mathrm{d}s, \quad t \in [\bar{t}, T].$$

Since p and q are Lipschitz continuous non-positive functions, we conclude that $p(t)+q(t)=0$ on $[\bar{t},T]$, hence also on $[0,T]$. Then (6.15) implies that $p(t)=0$ and similarly $q(t)=0$. This proves the nonnegativity of S and I.

The next step is to proof that the solution $(S[D],I[D])$ of (3.3)–(3.4) on $[0,T]$ depends in a Lipschitz way on D in a sense that will become clear in the next lines. For any two functions $D_1,D_2\in L^\infty(0,1)$ with $\|D_1\|_{L^\infty}$, $\|D_2\|_{L^\infty}\le a-1$, denote $\Delta(t):=\sup_{\omega\in[0,1]}|Z[D_1](t,\omega)-Z[D_2](t,\omega)|$. For any $\gamma=(t,\omega)\in\Gamma^T$ we have from (6.6) that

$$|S[D_1](\gamma)-S[D_2](\gamma)|\le L\int_{\tau^f(\gamma)}^t\big(A(s)\,|D_1(s)-D_2(s)|+a\big|Z[D_1](s,\omega^f[\gamma](s))-Z[D_2](s,\omega^f[\gamma](s))\big|\big)\,\mathrm{d}s,$$

where $A(s):=|Z[D_1](s,\omega^f[\gamma](s))|\le 2e^{2Ma}\|Z^0\|_C$, according to the growth condition in the definition of $\mathcal{K}_{T,a}$. A similar inequality holds for I. Summing the two, and taking the supremum in $\omega\in[0,1]$ on the left-hand side, we obtain that

$$\Delta(t)\le 2L\int_{\tau^f(\gamma)}^t\big(2e^{2Ma}\|Z^0\|_C\,|D_1(s)-D_2(s)|+a\Delta(s)\big)\,\mathrm{d}s.$$

Using the Grünwal inequality we obtain that

$$\Delta(t)\le\int_{\tau^f(\gamma)}^t e^{2aL(t-s)}4Le^{2Ma}\|Z^0\|_C\,|D_1(s)-D_2(s)|\,\mathrm{d}s\le L_Z\|D_1-D_2\|_{L^1(0,t)},$$

where $L_Z:=4Le^{2aLT+2aM}\|Z^0\|_C$. This inequality gives the meaning of the Lipschitz property of $Z[D]$.

6. (End of the proof of Theorem 6.1.) Define the set

$$\mathcal{N}_T:=\{D\in L^1(0,T):\,0\le D(t)\le a-1,\,t\in[0,T]\}.$$

Then the solution $Z[D]$ of (3.3)–(3.4) exists or every $D\in\mathcal{N}_T$ and we may define

$$\mathcal{G}(D)=\mathcal{D}(Z[D]),\quad D\in\mathcal{N}_T.$$

Due to the properties of $\mathcal{D}$ and $\mathcal{N}_T$ the latter is invariant with respect $\mathcal{G}$. Apparently, it is a complete metric space. We shall show that the mapping $\mathcal{G}$ is contractive with respect to the norm $\|D\|_N:=\int_0^T e^{-tN}|D(t)|\,\mathrm{d}t$, where $N>2L_DL_Z$. Indeed,

$$\begin{aligned}
\|\mathcal{G}(D_1)-\mathcal{G}(D_2)\|_N &= \int_0^T e^{-tN}|\mathcal{D}(Z[D_1])(t)-\mathcal{D}(Z[D_2])(t)|\,\mathrm{d}t\\
&\le L_D\int_0^T e^{-tN}\max_{\omega\in[0,1]}|Z[D_1](t,\omega)-Z[D_2](t,\omega)|\,\mathrm{d}t\\
&\le L_{\mathcal{D}}L_Z\int_0^T e^{-tN}\|D_1-D_2\|_{L^1(0,t)}\,\mathrm{d}t\\
&\le L_{\mathcal{D}}L_Z\int_0^T|D_1(s)-D_2(s)|\int_s^T e^{-tN}\,\mathrm{d}t\,\mathrm{d}s\\
&\le \frac{L_DL_Z}{N}\int_0^T|D_1(s)-D_2(s)|e^{-sN}=\frac{L_DL_Z}{N}\|D_1-D_2\|_N\\
&\le \frac{1}{2}\|D_1-D_2\|_N.
\end{aligned}$$

Thus $\mathcal{G}$ is a contraction on $\mathcal{N}_T$ hence it has a fix point D. Obviously the pair $(Z[D],D)$ satisfies the system (3.3), (3.4), (6.1). Moreover, $Z[D]\in\mathcal{K}_{T,a}$ is a Lipschitz function, which completes the proof of Theorem 6.1. □

Now, we continue with the proof of Theorem 3.1 concerning the specific system (2.3)–(2.7). The conditions (6.5)–(6.7) are apparently fulfilled in this case. Inspecting the proof of Theorem 6.1 we see that the conditions (6.2)–(6.4) are not used in parts 1–5 (they are only used in part 6). We have proved (in parts 1–5) that for any $D\in L^\infty(0,T)$ with $0\le D(t)\le a-1$ for a.e. $t\in[0,T]$, there exists a non-negative Lipschitz continuous solution $Z[D]\in\mathcal{K}_{T,a}$ of (3.3)–(3.4), hence of (2.4)–(2.7). The existence of $Z[D]$ was obtained due to the contractivity of the operator $\mathcal{F}_{[D]}$ on $\mathcal{K}_{T,a}$. Hence, $Z[D]$ can be considered as the uniform limit of the sequence of functions $\{Z_k\}$ generated by

$$Z_{k+1}=\mathcal{F}_{[D]}(Z_k),\quad Z_0=\mathcal{Z}^\#.$$

Notice that due to (6.5) and the definition of $\mathcal{F}_{[D]}$ we have

$$\|Z_{k+1}\|_C \le TMa\|Z_k\|_C + \|Z^0\|_C,$$

which implies the estimate

$$\|Z_k\|_C \le 2\|Z^0\|_C, \quad \text{provided that } T \le \frac{1}{2Ma}. \tag{6.16}$$

Further, we shall choose T satisfying the last inequality, in addition to (6.8).

Denote by Ω_0 the set of points in $(0,1)$ on which S^0 or I^0 is non-differentiable, together with the points $\omega = 0$ and $\omega = 1$. Denote

$$\Gamma^{\#} := \{\gamma \in \Gamma^T : \xi^f(\gamma) \in \Omega_0 \text{ or } \xi^g(\gamma) \in \Omega_0\} = \{\omega^f[(\xi,0)](t),\, \omega^g[(\xi,0)](t) : \xi \in \Omega_0,\, t \in [0,T]\}.$$

This set consists of finite number of curves in Γ^T. Observe that the assumption $f', g' < 0$ on $(0,1)$ implies that the set $\bar\Gamma^T = \Gamma^T \setminus \Gamma^{\#}$ consists of finite number of open sets, further called facets. We remind that $\gamma^f(\gamma)$ and $\gamma^g(\gamma)$ have Lipschitz derivatives in a neighborhood of Γ^T. Then the function $Z_0(\gamma) = Z^{\#}(\gamma) = (\bar S^0(\gamma^f(\gamma)), \bar I^0(\gamma^g(\gamma)))$ is differentiable with a Lipschitz derivative on each facet of $\bar\Gamma^T \setminus \Gamma^{\#}$. In addition, for every $\gamma \in \Gamma^T$ the functions $s \mapsto \sigma(\omega^f[\gamma](s))$ and $s \mapsto \sigma(\omega^g[\gamma](s))$ are differentiable and have Lipschitz derivatives on every of the finite number of intervals for s in which the argument of σ belongs to one facet. The same applies to the functions ρ and μ. Thanks to the properties mentioned in this paragraph, we can differentiate Z_{k+1} with respect to $\gamma \in \bar\Gamma^T$ using (6.11)–(6.12). Skipping the cumbersome details, we obtain the following relations:

$$\mathrm{Lip}^{\#}\Big(\frac{\partial Z_{k+1}}{\partial\gamma}\Big) \le c_1 + c_2\mathrm{Lip}^{\#}\Big(\frac{\mathrm{d}S^0}{\mathrm{d}\omega}\Big) + Tc\,\mathrm{Lip}^{\#}\Big(\frac{\partial Z_k}{\partial\gamma}\Big),$$

where $\mathrm{Lip}^{\#}(Q)$ is a common Lipschitz constant of a function Q on each facet of $\bar\Gamma^T$ (for $Q : \bar\Gamma^T \to \mathbb{R}^2$ which is Lipschitz on every facet), $\mathrm{Lip}^{\#}\Big(\frac{\mathrm{d}S^0}{\mathrm{d}\omega}\Big)$ is the Lipschitz constant of $\frac{\mathrm{d}S^0}{\mathrm{d}\omega}$ on each of the intervals of its existence, c_1 and c_2 are constants (which may depend on $\mathrm{Lip}(D)$ and $\|Z^0\|_C$), c is a constant which is independent of Z^0 and D with $0 \le D \le a-1$. The derivation of this recurrent inequality also uses the fact that $\mathcal{F}_{[D]}$ is an affine mapping of Z. Since $\mathrm{Lip}^{\#}\Big(\frac{\mathrm{d}S^0}{\mathrm{d}\omega}\Big)$ is finite, we obtain inductively that for every k

$$\mathrm{Lip}^{\#}\Big(\frac{\partial Z_{k+1}}{\partial\gamma}\Big) \le \Big(c_1 + c_2\mathrm{Lip}^{\#}\Big(\frac{\mathrm{d}S^0}{\mathrm{d}\omega}\Big)\Big)\sum_{j=0}^{k}(Tc)^j \le 2\Big(c_1 + c_2\mathrm{Lip}^{\#}\Big(\frac{\mathrm{d}S^0}{\mathrm{d}\omega}\Big)\Big),$$

provided that $T \le c/2$. We add the last condition to (6.16) and (6.8). The choice of T is still independent of the initial distribution Z^0 and the particular D. For the limit $Z[D]$ of Z_k we obtain

$$\mathrm{Lip}^{\#}\Big(\frac{\partial Z[D]}{\partial\gamma}\Big) \le 2\Big(c_1 + c_2\mathrm{Lip}^{\#}\Big(\frac{\mathrm{d}S^0}{\mathrm{d}\omega}\Big)\Big).$$

Since every horizontal and every vertical line intersects $\Gamma^{\#}$ only finite number of times, the partial derivatives of $Z[D]$ in each of the variables (t,ω) exist, except a finite number of points, for every value of the other variable.

Using the obtained differentiability properties of the solution $(S[D], I[D])$, we may employ (3.5) to estimate

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 (S[D](t,\omega) + I(t,\omega))\,\mathrm{d}\omega = -\int_0^1 \mu(\omega)I[D](t,\omega)\,\mathrm{d}\omega \ge -\|\mu\|_{L^\infty(0,1)}\int_0^1 I[D](t,\omega)\,\mathrm{d}\omega,$$

hence

$$\int_0^1 (S[D](t,\omega) + I[D](t,\omega))\,\mathrm{d}\omega \ge 1 - \|\mu\|_{L^\infty(0,1)}\int_0^t\int_0^1 I[D](t,\omega)\,\mathrm{d}\omega\,\mathrm{d}t \ge 1 - T\|\mu\|_{L^\infty(0,1)} \ge \frac{1}{2},$$

where, if necessary, we choose the number T even smaller, so that $T\|\mu\|_{L^\infty(0,1)} \le 1/2$ (still being independent of the distribution of the initial data). The inequality $\int_0^1 I[D](t,\omega)\,\mathrm{d}\omega \le 1$ used above, follows from the obtained decrease of $\int_0^1 (S[D](t,\omega) + I[D](t,\omega))\,\mathrm{d}\omega$ starting from value 1 at $t = 0$. So we obtain that

$$\int_0^1 (c_I I[D]((t,\omega) + cS[D](t,\omega))\,\mathrm{d}\omega \ge \frac{1}{2}\min\{c_I, c\} > 0. \tag{6.17}$$

Now we return to conditions (6.2)–(6.4). The first and the last are apparently fulfilled for the mapping $\mathcal{D}$ defined by (2.3) (with the convention that $\mathcal{D}(0) = 0$). Condition (6.3) is not fully used in the proof of Theorem 6.1 (part 6). What is used, is the inequality

$$|\mathcal{D}(Z[D_1])(t) - \mathcal{D}(Z[D_2])(t)| \le L_D \max_{\omega\in[0,1]} |Z[D_1](t,\omega) - Z[D_2](t,\omega)|, \qquad D_1,\, D_2 \in \mathcal{N}_T.$$

Due to (6.17) (which holds for every $D \in \mathcal{N}_T$) and the specific form of $\mathcal{D}$, a constant L_D does exist such that the last inequality is fulfilled. For the same reason, $\mathcal{D}(Z[D])(\cdot)$ is Lipschitz continuous for every $D \in \mathcal{N}_T$, uniformly in D. Thus the fixed point

D of $\mathcal{G}$ (which defines a solution $Z[D]$ of (2.3)–(2.7)) is also Lipschitz continuous. Hence, $Z[D]$ has the desired differentiability property.

It remains to prove that the solution (S, I, D) can be extended to $[0, \infty)$. We have proved that in exists on $[0, T]$ and that T is independent of particular distribution of S^0 and I^0, given that $\int_0^1 (S^0(\omega) + I^0(\omega))\,\mathrm{d}\omega = 1$. Taking new initial data $\tilde{S}^0(\omega) = S(T, \omega)/\beta$, $\tilde{I}^0(\omega) = I(T, \omega)/\beta$ with $\beta = \int_0^1 (S(T, \omega) + I(T, \omega))\,\mathrm{d}\omega$ (so that the new initial data are normalized), we may apply the obtained existence result: a solution $(\tilde{S}, \tilde{I}, \tilde{D})$ exists on $[0, T]$. Observe that the system (2.3)–(2.5) is homogeneous of first order. Then $(S(T + t, \omega), I(T + t, \omega), D(T + t)) := (\beta\tilde{S}(t, \omega), \beta\tilde{I}(t, \omega), \beta\tilde{D}(t))$ is a continuation of the solution on $[0, 2T]$. The process can be infinitely continued. This concludes the proof of Theorem 3.1.

References

Ackleh A, Deng K, Hu S (2005) A quasilinear hierarchical size-structured model: well-posedness and approximation. Appl Math Optim 51:35–39

Barbarossa MV, Röst G (2015) Immuno-epidemiology of a population structured by immune status: a mathematical study of waning immunity and immune system boosting. J Math Biol 71:1737–1770

Bloom DE, Kuhn M, Prettner K (2022) Modern infectious diseases: macroeconomic impacts and policy responses. J Econ Lit 66:85–131

Caulkins JP, Grass D, Feichtinger G, Hartl RF, Kort PM, Prskawetz A, Seidl A, Wrzaczek S (2021) The optimal lockdown intensity for COVID-19. J Math Econ 93:102489.

Courant R, Friedrichs K, Lewy H (1967) On the Partial Difference Equations of Mathematical Physics. IBM J Res Dev 11:215–234

Constantino V, Trent M, McIntylre CR (2019) Modelling of optimal timing for influenza vaccination as a function of intrasesonal waning of immunity and vaccine coverage. Vaccine 37: 6768-6775

Domenech de Celles M, Wong A, Barrero Guevara, LA, Rohani P (2022) Immunological heterogeneity informs estimation of the durability of vaccine protection J R Soc Interface 19: 20220070

Ehrhardt M, Gasper J, Kilianova S (2019) SIR-based mathematical modeling of infectious diseases with vaccination and waning immunity. J Comput Sci 37:101027

Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman LS, Ash N, Alroy-Preis S, Huppet A, Milo R (2022) Protection and waning of natural and hybrid immunity to SARS-CoV-2. N Engl J Med 386: 2201–2212

Goel RR, Painter MM, Apostolidis SA, Mathew D, Meng W, Rosenfeld AM, et al. (2021) mRNA vaccines induce durable immune memoryto SARS-CoV-2 and variants of concern. Science 374:eabm0829

Ghosh S, Banerjee M, Volpert V. (2022) Immuno-epidemiological model-based prediction of further covid-19 epidemic outbreaks due to immunity waning. Math Modell Nat Phenom, 17:9

Iyaniwura SA, Musa R, Kong JD (2023) A generalized distributed delay model of COVID-19: An endemic model with immunity waning. Math Biosci Eng 20: 5379-5412

Kato N, Torikata H (1997) Local existence for a general model of size-dependent population dynamics. Abstr. Appl. Anal. 2:(3-4): 207-226

Lavine JS, Bjornstad, ON, Antia R. (2021) Immunological characteristics govern the transition of covid-19 to endemicity. Science 371: 741—745

LeVeque R (2002) Finite Volume Methods for Hyperbolic Problems (Cambridge Texts in Applied Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791253

Martcheva M., Pilyugin S.S. (2006). An epidemic model structured by host immunity. J Biol Sys. 14(02):185–203

Montalbán A, Corder RM, Gomes MGM (2022) Herd immunity under individual variation and reinfection. J Math Biol 85:2. doi: 10.1007/s00285-022-01771-x

Pell B, Johnston MD, Nelson P. (2022) A data-validated temporary immunity model of covid-19 spread in Michigan. Math Biosci Eng 19: 10122–10142

Rambhia KJ, Rambhia MT (2019) Early bird gets the flu: What should be done about waning intraseasonal immunity against seasonal influenza? Clin Infect Dis 68: 1235–1240

Rouderfer V, Becker NG (1994) Waning immunity and its effects on vaccination schedules. Math Biosci 124:59–82

Veliov VM, Widder A (2016) Modelling and estimation of infectious diseases in a population with heterogeneous dynamic immunity. J Biol Dyn 10:457–476

White LJ, Medley GF (1998) Microparasite population dynamics and continuous immunity. Proc R Soc B 265:1977-1983

Yaugel-Novoa M, Bourlet T, Paul S (2022) Role of the humoral immune response during COVID-19: guilty or not guilty? Mucosal Immunol 15:1170–1180