



Technische Universität München

TUM School of Computation, Information and Technology

Multi-scale Modeling of Cholera: Linking Within-host and Between-host Dynamics

Beryl Onyuma Musundi

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr. Daniel Matthes

Prüfer*innen der Dissertation: 1. Prof. Dr. Johannes Müller
2. Prof. Zhilan Feng, Ph.D.

Die Dissertation wurde am 20.02.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 26.04.2023 angenommen.

With my signature below, I assert that the work in this thesis has been composed by myself independently and no source materials or aids other than those mentioned in the thesis have been used.

München, July 7, 2023

Place, Date

Signature

This work is licensed under the Creative Commons Attribution 3.0 Germany License. To view a copy of the license, visit <http://creativecommons.org/licenses/by/3.0/de>

Or

Send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

München, July 7, 2023

Place, Date

Signature

Kurzfassung

Cholera, eine Infektionskrankheit welche durch das pathogene Bakterium *Vibrio cholerae* verursacht wird, ist nach wie vor ein bedeutendes Problem für die Öffentliche Gesundheit. Die Übertragungsdynamik der Krankheit erstreckt sich über mehrere miteinander verknüpfte Skalen, vom einzelnen Individuum über die Bevölkerung bis hin zur Umwelt. Daher gestaltet sich die Erfassung der Krankheitsdynamik durch herkömmliche immunologische oder epidemiologische Modelle basierend auf einer einzigen Skala als schwierig. In dieser Arbeit entwickeln wir Multiskalenmodelle zur Untersuchung der Übertragungsdynamik von Cholera. Zunächst stellen wir ein Modell vor, das die Krankheitsdynamiken durch die Strukturierung eines epidemischen Modells mithilfe einer intraindividuellen Dynamik verbindet. Hier stellt das System innerhalb des Wirts die Interaktion zwischen dem Erreger und der Immunantwort dar. Annahmen bezüglich der Zeitskala werden getroffen, um die Dynamik innerhalb des Wirts zu analysieren, und die Ergebnisse werden hochskaliert, um das epidemische Modell zu strukturieren. Wir analysieren die Existenz und die Stabilität der Gleichgewichtslösungen des epidemischen Modells mithilfe von Linearisierungstechniken und Halbgruppen- und Lyapunov-Argumenten. Im Anschluss entwickeln wir ein verfeinertes Mehrskalenmodell, bei dem der Übergang von suszeptibel zu infiziert von der Erregerbelastung im Wirt abhängig ist. Jeder suszeptiblen Person wird eine Erregerlast zugewiesen, welche sich durch die Aufnahme von kontaminiertem Wasser (Booster-Ereignis), die zu zufälligen Zeitpunkten erfolgt, erhöht und durch den Einsatz einer Immunreaktion abnimmt. Durch die Skalierung dieser Dynamiken wird dann die suszeptible Population strukturiert. Sobald der Erreger eine kritische Schwelle innerhalb eines Individuums überschreitet, wird der Übergang von suszeptibel zu infiziert ausgelöst. Für die Analyse wird ein Zeitskalenansatz gewählt, da angenommen wird, dass die Prozesse innerhalb des Wirts schneller ablaufen, als die epidemische Dynamik. Auf der schnellen Zeitskala wird die Existenz einer invarianten Verteilung der Erregerlast mithilfe von Halbgruppenmethoden und Spektralanalyse gezeigt. Die Ergebnisse der Spektralanalyse werden dann verwendet um das Mehrskalenmodell auf ein SIR-Modell in der langsamen Zeitskala zu reduzieren. Außerdem werden numerische Simulationen zur Ermittlung des Langzeitverhaltens durchgeführt. Schließlich entwickeln wir ein Multiskalenmodell, das die individuelle Heterogenität der Wirte berücksichtigt. Dazu weisen wir jedem Individuum eine Erregerlast zu, deren intrinsische Wachstumsrate vom Zustand des Darmmikrobioms abhängig ist. Die Individuen nehmen den Erreger aus der Umwelt auf, und der Erreger wächst in der Umwelt, nachdem sie ihn ausscheiden. Wir verwenden Entkopplungsannahmen, um das Modell auf eine Erneuerungsgleichung zu reduzieren und analysieren das asymptotische Verhalten mit Hilfe von Laplace-Transformationen.

Abstract

Cholera, an infectious disease caused by the pathogenic *Vibrio cholerae* bacteria, remains a significant public health concern. The transmission dynamics for the disease span multiple interconnected scales, from within-individual to the population and the environment, making it challenging for traditional single-scale immunological or epidemiological models to capture the disease dynamics. In this thesis, we develop multi-scale models to study the transmission dynamics of cholera. First, we present a model that links disease dynamics by structuring the epidemic model using within-host immune dynamics. Here, the within-host system depicts the interaction of the pathogen and the immune response. Time-scale assumptions are used to analyze the within-host dynamics with the results scaled up to structure the epidemic model. For the epidemic model, we analyze the existence and stability of the equilibrium solutions through linearization techniques and the use of semigroup and Lyapunov arguments. Next, we develop a refined multi-scale model where the transition from susceptible to infected is dependent on the within-host pathogen load. Here, each susceptible person is assigned a pathogen load. The pathogen load increases through the ingestion of contaminated water (booster event) that takes place at random times and declines through the actions of the immune response. These dynamics are then scaled to structure the susceptible population. The transition from susceptible to infected is triggered when the pathogen exceeds the critical threshold. The within-host dynamics are considered to occur faster than the epidemic dynamics and thus the analysis follows the fast-slow approach. On the fast scale, the existence of pathogen load's invariant distribution is established through the use of semigroup methods and spectral analysis. The results of the spectral analysis are then used to reduce the multi-scale model to a SIR model on the slow scale and numerical simulations are conducted on the SIR model to establish its long-term behavior. Finally, we develop a multi-scale model that accounts for individual host heterogeneity. We do this by assigning each individual a pathogen load whose intrinsic growth rate is dependent on the state of the gut microbiome. Individuals ingest the pathogen from the environment, and the pathogen grows in the environment after they shed it. We use decoupling properties to reduce the model to a renewal equation and analyze the asymptotic behavior with the help of Laplace transforms.

Acknowledgement

First, I would like to express my sincere gratitude to my supervisor Prof. Dr. Johannes Müller, for his continuous support, guidance, and encouragement at each stage of this research. His positive attitude and constant check-up on work progress eased the process. I'm also thankful to Prof. Zhilan Feng for hosting me at Purdue University for my research stay. I would also like to extend my gratitude to my mentor Dr. Nada Sissouno for her invaluable advice, especially at the formative stages of the project.

My doctoral project was funded by the German Academic Exchange Service (DAAD) which I would like to acknowledge. I'm also grateful to the TUM IGSSE graduate school for funding my research stay and conferences under the project GENOMIE_QADOP.

It was a pleasure working with Prof. Tellier, Dr. Sona, and my doctoral colleagues Augustine, Usman, Franz, and Legrace. I would also like to thank Prof. Donna Ankerst for the encouragement and the many enjoyable lunches we had together.

I thank my parents Chrispinus and Jean and siblings Christine, Sebastian, and Reagan for always supporting me throughout my academic journey. Their love, patience, and understanding were significant sources of strength. To my friends Hazel, Badru, Sarita, Sara, Renata, Walter, Elvis, Deria, Ikenna, Oscar, and Alex, thank you for keeping me balanced and creating a home far away from home.

Contents

Contents	4
1 Introduction	6
1.1 Biological Background of Cholera	8
1.1.1 History of Cholera	8
1.1.2 <i>Vibrio cholerae</i> and the Aquatic Environment	8
1.1.3 Transmission, Susceptibility, and Immunity	9
1.1.4 Pathogenesis, Symptoms, and Treatment	9
1.2 Mathematical Background	10
1.2.1 Fast-Slow Systems	11
1.2.2 Semigroups	12
1.2.3 Spectral Theory	13
1.2.4 Laplace Transforms	13
2 Multi-scale Epidemic Modeling and Modeling of Cholera	15
2.1 Multi-scale Epidemic Modeling	15
2.1.1 Epidemiological models	15
2.1.2 Immunological models	20
2.1.3 Linking Mechanisms	21
2.2 Modeling of Cholera	23
2.2.1 Between-host Cholera Models	23
2.2.2 Within-host Cholera Models	24
2.2.3 Multi-scale Cholera Models	24
3 Linking the Dynamics through Immune Responses	26
3.1 The Within-host Model	26
3.2 Time-scale Analysis	27
3.2.1 Fast System	28
3.2.2 Slow System	29
3.3 Between-Host Model	32
3.3.1 Existence and Uniqueness of Solutions	33
3.3.2 Basic Reproduction Number and Stability of the DFE	35

3.3.3	Existence of the Endemic Equilibrium	39
3.3.4	Local Stability of the Endemic Equilibrium	41
3.4	Modified Between-host Model	43
3.4.1	Global Stability of Equilibrium Points	43
3.5	Summary	47
4	Multi-scale to SIR model using First Principles	49
4.1	Model Description	50
4.2	Pathogen Distribution for Constant Environmental Pathogen Load	55
4.2.1	Existence of the Semigroup	56
4.2.2	Stationary Solution and Spectral Gap	61
4.3	Reduced Model	73
4.3.1	Fast-Slow Analysis	73
4.3.2	Behavior of the Reduced Model: A Simulation Study	75
4.4	Summary	79
5	Heterogeneity in the Within-Host Model	80
5.1	Model Structure	81
5.2	Analysis of the Model: Linear Case	81
5.2.1	Renewal Equation	82
5.2.2	Existence of Solutions	83
5.2.3	Asymptotic behaviour	84
5.3	Summary	86
6	Discussion and Conclusion	87
6.1	Discussion	87
6.2	Conclusion	90
A	Appendix	91
A.1	Parameter values for Figure (4.5)	91
	List of Tables	92
	Bibliography	93

Chapter 1

Introduction

Infectious diseases are still a major cause of mortality and morbidity despite the advances that have been made in medicine. The emergence of novel diseases such as Covid-19 and the reemergence of old disease epidemics has further intensified the threat posed by these diseases and highlighted the need for the development of proper strategies for management. Mathematical modeling is one of the effective tools that has been used to understand the dynamics of infectious diseases. The models describe the transmission principles, forecast the future course of an epidemic, and assess the effectiveness of intervention strategies [70, 58, 12]. Consequently, information derived from the models can be used to inform policy decisions about health interventions.

Disease dynamics have largely been studied at two scales: immunological and epidemiological. The epidemiological scale (between-host) is centered on disease transmission among members of the host population [11, 99, 85], whereas the immunological scale (within-host) considers infection dynamics within a single host [104, 8]. Traditionally, the two scales have been modeled independently. However, the infection process spans multiple scales: From the within-host scale to the between-host scale, and the environment among others. Therefore, it is necessary to bridge the dynamics at multiple scales to comprehensively describe the entire infection process. Additionally, models that link multiple scales (multi-scale models) provide novel insights into disease mechanisms [19, 38, 74]. This applies particularly to cholera, an environmentally transmitted disease, whose transmission dynamics evolve from complex within-host processes to between-host transmission with interactions from the cholera pathogen in the environment. As such, in this thesis, we develop multi-scale models to study the transmission dynamics of cholera.

The thesis is organized as follows: In this chapter, we give a biological introduction to cholera and define some mathematical concepts that will be used later in the work.

In Chapter 2, we introduce the concept of multi-level epidemic modeling. We discuss the underlying framework of each of the components of a multi-scale model. In Section 2.2, we give a brief review of the notable literature on cholera modeling starting from the between-

and within-host levels to the multi-scale level.

In Chapter 3, we formulate and analyze an immuno-epidemiological model that links the within-host dynamics to the between-host dynamics of cholera through the within-host immune response. To that end, we first derive a within-host model depicting the interaction of the pathogen and the adaptive immune response. We take pathogen dynamics to be faster than the immune response. Analysis through the separation of time scales allows us to characterize a single infected individual by the state of the immune response. We then scale up the dynamics of an individual infected person to construct an epidemic model in which the infected population is structured by individual immunological dynamics. For this physiologically structured epidemic model, we establish the existence of stationary solutions and study their local asymptotic properties by utilizing linearization techniques and semigroup arguments. Finally, in Section 3.4, we modify the epidemic model to include a maximum age for the bacteria and analyze the global asymptotic properties of the resulting model. The underlying assumption of the work in this chapter is that one infectious contact is enough to push contacts over the threshold.

In Chapter 4, we refine the underlying assumption in Chapter 3 to formulate a multi-scale model where the transition from susceptible to infected occurs at a pathogen load-dependent rate, that is, infection occurs when the pathogen load in an individual exceeds the critical threshold. To do this, we assign a single susceptible individual with a pathogen load that increases through the consumption of contaminated water (booster event) and declines through immune responses. We also assign maximal and minimal times for booster events and define the time since the last booster event. We then scale the dynamics of a single susceptible person to build an epidemic model. In this model, the susceptible population is structured by the pathogen load and the time since the last booster event. By considering population dynamics to be slower than within-host dynamics, we analyze the pathogen load's invariant distribution on a fast time scale. We utilize semigroup methods and spectral analysis for the analysis. From the results of the spectral analysis, we reduce the multi-scale model to a refined SIR model on a slow time scale. Finally, we perform numerical simulations on the SIR model to study its long-term behavior.

In Chapter 5, we formulate a multi-scale model that captures the heterogeneity of individual hosts. We do this by assigning an individual with a pathogen load whose intrinsic growth is dependent on the nature of the gut microbiome. We also include an environmental bacterial compartment that contributes to the within-host pathogen load through ingestion and grows through shedding by individuals. The contact structure of the model is only through a single compartment (the environmental bacterial compartment), and as such, we use decoupling properties to reduce the model into a Volterra renewal equation in the case where intrinsic pathogen growth is linear. Lastly, we analyze the asymptotic behavior of the model with the help of Laplace transforms.

Finally, in Chapter 6 we give a discussion of the work and conclude the thesis.

1.1 Biological Background of Cholera

Cholera, an acute diarrheal disease, affects millions of people worldwide with an estimated 1.4 – 3 million cholera cases and 21,000 – 143,000 related deaths reported yearly [109]. Sub-Saharan Africa and South-East Asia account for the majority of the cholera cases [2] with lack of access to safe drinking water and improper sanitation being the drivers of the disease. In this section, we give a brief overview of the disease.

1.1.1 History of Cholera

Cholera is an ancient disease. The earliest description of a disease resembling cholera on the Indian subcontinent can be found in the Sanskrit text *Sushruta Samhita* written around 500-400 B.C [25]. Modern accounts of the disease gained prominence in 1817 after an outbreak in the Ganges river delta spread beyond the subcontinent of India. This was considered to be the first cholera pandemic. Six cholera pandemics were recorded between 1817-1923 [50]. The ongoing (seventh) pandemic began in 1961 in Indonesia and has spread to most parts of the world. The disease is now endemic predominantly in Africa and South Asia [93].

The etiological agent responsible for cholera was identified in the 19th century [92]. The first description of the agent was by Pacini in 1854. He found curved bacteria in the intestinal contents of cholera victims that he named *Vibrio cholera* [53]. However, his initial discovery was not recognized until later. In 1883, Robert Koch, who was studying cholera in Egypt, demonstrated that the disease was caused by this comma-shaped organism that he referred to as *Kommabazillen*. The subsequent name *Vibrio comma* was used for several years and was later renamed *Vibrio cholerae* after the recognition of Pacini's work [53].

1.1.2 *Vibrio cholerae* and the Aquatic Environment

Vibrio cholerae are curved, gram-negative bacteria that belong to the Vibrionaceae family commonly found in estuaries and marine waters [50, 92]. The organism is classified into serogroups based on the O antigen. There are over 200 serogroups. Out of these, only the O1 and O139 cause cholera epidemics as they carry the genes encoding the cholera toxin (CT) and the toxin co-regulated pilus (TCP) that are considered to be the virulence factors for the disease [87, 66]. Infection with non-O1 and non-O139 serogroups is seen to be less clinically significant [92]. The O1 serogroup is further subdivided into two biotypes: Classical and El Tor. The Classical biotype is responsible for earlier pandemics while the ongoing seventh pandemic is due to the El Tor biotype [87, 50]. It's worth noting that not all strains of the O1 and O139 serogroups cause cholera epidemics [53].

In marine ecosystems, *V. cholerae* attach themselves to various surfaces including plants, zooplanktons, crustaceans, green algae, and insects. A co-relation between cholera outbreaks and the seasonal occurrence of algae has been observed, even though this has not been linked to the increase in toxigenic strains associated with cholera epidemics [92]. *V.*

cholerae can survive environmental stressors including changes in water salinity, temperature, and nutrient deprivation through adaptive mechanisms that include transformation into viable but non-culturable forms and attachment to biotic and abiotic forms as biofilms [66]. This makes it possible for the bacteria to persist in the aquatic environment. The ability of the bacteria to exist in a viable but non-cultural form contributes to seasonal outbreaks of the disease. In such cases, the contribution from infected humans is not necessary [25].

1.1.3 Transmission, Susceptibility, and Immunity

Pathogenic bacteria are transmitted to human hosts through the ingestion of contaminated food and water. Direct human-to-human transmission through the fecal-oral route is also possible, though uncommon [94].

Susceptibility to the disease depends on several host factors including genetics, nutrition and the chemical state of the gut [87, 22]. Genetically, individuals with the O phenotype that corresponds to the H antigen show a decreased risk of infection but an increased risk of severe disease in case of infection [87, 63]. Nutritional factors such as deficiency of vitamin A and Zinc, which are vital in mucosal immunity, increase the susceptibility to cholera since they influence innate and adaptive immunity [44]. The chemical state of the gut, defined by the gut microbiome, determines an individual's susceptibility or resistance to infection with a healthy microbiome community shown to induce resistance to the infection [1].

Cholera infection induces immunity: Studies with North American volunteers have shown that infection by *V. cholerae* induces protective immunity against subsequent infections. Infection with the classical biotype offers 100% immunity while the El Tor biotype offers 90% immunity [87]. Immunity to infection can also result from vaccination. Three oral vaccines: Dukoral, Shanchol, and Euvichol-Plus are available. However, they only offer short-term immunity.[109].

1.1.4 Pathogenesis, Symptoms, and Treatment

The cholera pathogen has human and environmental phases in its life cycle (see Figure (1.1) below). After ingestion, the bacteria must first survive the stomach acids before penetrating the mucosal lining of the intestinal epithelial cells [92]. The majority of the bacteria are killed by gastric acid in the stomach [50]. A high infectious dose of $10^6 - 10^{11}$ colony-forming units is required for infection because of the bacteria's acid sensitivity [92]. The bacteria that survive adhere to the epithelial cells and colonize them leading to the production of the cholera toxin [92]. The cholera toxin then sets off the onset of cholera symptoms that include vomiting and acute watery diarrhea. These symptoms often occur abruptly after the incubation period (18 hours-5 days). In cases of severe cholera, fluid loss can lead to severe dehydration and death can occur within a few hours of the onset of the symptoms [93]. Symptomatic and asymptomatic infected individuals shed the organisms back into the environment. Passage of the bacteria through the gut transforms it into a

hyperinfectious state. The hyperinfectious state is short-lived as the bacteria decay to a less infectious state after some time [45]. A low infectious dose of freshly shed bacteria is required to cause infection: The dose is 10-100 times lower than that of non-human shed *Vibrio* [50].

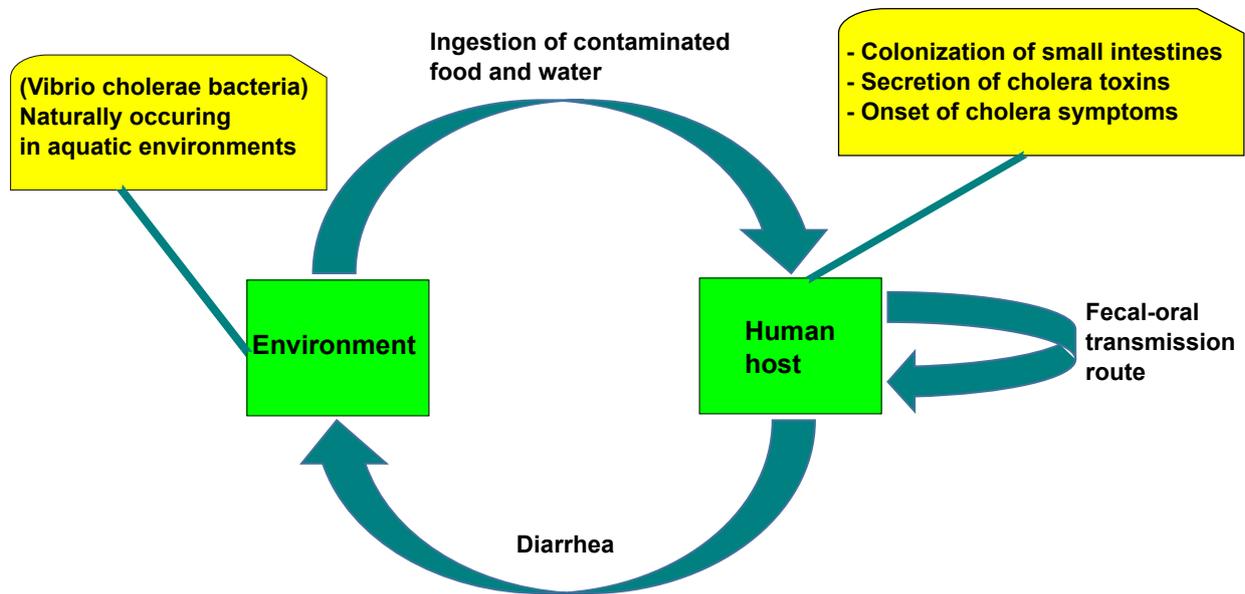


Figure 1.1: The infection cycle of *Vibrio cholerae*.

The disease is managed with rehydration therapies. Milder cases are treated with Oral Rehydration Solutions (ORS) that replace the fluid deficits and maintain hydration while severe cases first require emergency intravenous polyelectrolyte solution for rehydration followed by ORS to maintain hydration [93]. Symptomatic patients are also treated with antibiotics to lessen diarrhea and shorten the infection period. The recovery period without antibiotics (only rehydration) is 4-5 days, while the recovery period with the use of antibiotics is 2-3 days. That said, widespread resistance to antibiotic treatment has also been reported [93].

1.2 Mathematical Background

In this section, we give a brief overview of mathematical concepts that are used later in the work. The definitions and theorems that are stated are derived from established work found in the associated references.

1.2.1 Fast-Slow Systems

The structure of a system of ordinary differential equations with fast and slow time-scale variables is written as

$$\begin{aligned}\frac{dx}{dt} &= f(x, y) \\ \frac{dy}{dt} &= \epsilon g(x, y)\end{aligned}\tag{1.2.1}$$

where ϵ is small and $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ [82]. Moreover, the x variable is referred to as the fast variable while the y variable is referred to as the slow variable. Rescaling the time such that $\tau = \epsilon t$, gives the system

$$\begin{aligned}\epsilon \frac{dx}{d\tau} &= f(x, y) \\ \frac{dy}{d\tau} &= g(x, y).\end{aligned}\tag{1.2.2}$$

We refer to t as the slow time scale and τ as the fast time scale.

The analysis of the fast-slow systems follows the singular perturbation theory. The general idea behind it is that one can understand the behavior of the full system by studying the two limiting subsystems.

Definition 1.2.1. [62] *The system of ordinary differential equations derived by setting $\epsilon = 0$ on the fast time scale formulation (1.2.1) is called the fast subsystem, that is*

$$\begin{aligned}\frac{dx}{dt} &= f(x, y) \\ \frac{dy}{dt} &= 0.\end{aligned}\tag{1.2.3}$$

Definition 1.2.2. [62] *The differential-algebraic equation derived by setting $\epsilon = 0$ in the slow time scale formulation (1.2.2) is called the slow subsystem, that is*

$$\begin{aligned}0 &= f(x, y) \\ \frac{dy}{d\tau} &= g(x, y).\end{aligned}\tag{1.2.4}$$

This subsystem is also referred to as the reduced problem.

Definition 1.2.3. *The line of stationary points of the fast system $f(x, y) = 0$ is referred to as the slow manifold [82] and the value $\epsilon = 0$ is called the singular limit.*

Generally, the solutions of the slow subsystem, which include the fast subsystem variables, are used to approximate the long-term behavior of the original model.

1.2.2 Semigroups

Definition 1.2.4. (a) A one-parameter family $T(t)$, $t \geq 0$ of bounded linear operators from a Banach X to itself is called a semigroup of bounded linear operators on X if it satisfies the following equations

1. $T(0) = I$
2. $T(t+s) = T(t)T(s) \quad \forall \quad t, s \geq 0.$

(b) A semigroup of bounded linear operators $T(t)$, $t \geq 0$ on a Banach space X is called uniformly continuous if

$$\lim_{t \rightarrow 0^+} \|T(t) - I\| = 0.$$

(c) The infinitesimal generator of a semigroup $T(t)$ is the linear operator A defined by

$$Ax = \lim_{t \rightarrow 0^+} \frac{T(t)x - x}{t} = \left. \frac{d^+ T(t)x}{dt} \right|_{t=0} \quad \text{for } x \in D(A)$$

with

$$D(A) = \{x \in X : \lim_{t \rightarrow 0^+} \frac{T(t)x - x}{t} \text{ exists}\}$$

as the domain of A .

(d) A semigroup $T(t)$, $t \geq 0$ of bounded linear operators on a Banach space X is a strongly continuous semigroup or simply a C^0 -semigroup if

$$\lim_{t \rightarrow 0^+} T(t)x = x \quad \text{for all } x \in X.$$

(e) A strongly continuous semigroup $T(t)$ is called compact for $t > t_0$ if $T(t)$ is a compact operator for every $t > t_0$. $T(t)$ is said to be compact if it is compact for all $t > 0$ [90, Definitions 1.1, 2.1, 3.1].

Definition 1.2.5. A strongly continuous semigroup $T(t)$, $t > 0$ is called eventually compact if there exists $t_0 > 0$ such that $T(t_0)$ is compact [35, Definition 4.23].

Definition 1.2.6. A strongly continuous semigroup $T(t)$ is called quasi-compact if $T(t) = T_1(t) + T_2(t)$ with operator families $T_1(t), T_2(t)$, such that $\|T_1(t)\| \rightarrow 0$ as $t \rightarrow \infty$ and $T_2(t)$ is eventually compact. That is, there exists $t_0 > 0$ such that $T_2(t)$ is compact for all $t > t_0$ [73].

Given A as the infinitesimal generator of a strongly continuous semigroup $T(t)$ on a Banach space X . Then for all $x \in D(A)$, $u(t) = T(t)x$ is the solution to the initial value problem

$$\begin{aligned} \frac{du}{dt} &= Au(t) \quad \text{for } t \geq 0 \\ u(0) &= x \end{aligned}$$

Consequently, the semigroup theory is useful in the study of evolution problems which can be written in terms of equation (1.2.5) [35].

Theorem 1.2.1. Arzelà-Ascoli Theorem

Let K be a compact metric space and H be a bounded subset of the Banach space $C(K)$, continuous functions over K with the supremum norm. Assume that H is uniformly equicontinuous, that is for all $\epsilon > 0$ there exists $\delta > 0$ such that

$$d(x_1, x_2) < \delta \text{ implies that } |f(x_1) - f(x_2)| < \epsilon \quad \forall f \in H.$$

Then the closure of $C(K)$ in H is compact [13, Theorem 4.25].

1.2.3 Spectral Theory

Let X be a complex topological space and T be a closed linear operator.

Definition 1.2.7. The resolvent set of T denoted by $\rho(T)$ is defined by

$$\rho(T) = \{\lambda \in \mathbb{C} : \lambda I - T \text{ is bijective from } X \text{ onto } X\}.$$

The spectrum of T denoted by $\sigma(T)$ is the complement of the resolvent set, that is

$$\sigma(T) = \mathbb{C} \setminus \rho(T).$$

For $\lambda \in \rho(T)$ the operator $R(\lambda, T) = (\lambda I - T)^{-1} : X$ maps to X is called the resolvent. The point spectrum of T denoted by $\sigma_p(T)$ is the set of $\lambda \in \mathbb{C}$ for which $\lambda I - T$ is not one-to-one. For λ in the point spectrum, the equation $T\theta = \lambda\theta$ has a solution $\theta \neq 0$. Consequently, λ is called the eigenvalue of T and θ is called the eigenvector of T . The null space $N(\lambda I - T)$ is called the eigenspace of T and its dimension is called the multiplicity of λ .

The spectral radius of T denoted by $r(T)$ is defined by

$$r(T) = \sup\{|\lambda| : \lambda \in \sigma(T)\}.$$

The spectral radius is finite and satisfies $r(T) \leq \|T\|$. (See e.g. [111, 29, 35]).

1.2.4 Laplace Transforms

Definition 1.2.8. Laplace Transform [48, 27, 30]

Let $f(t)$ be a function defined on some interval $0 \leq t < \infty$. The Laplace transform $\hat{f}(\lambda)$ of the function is defined by

$$\mathcal{L}[f] = \hat{f}(\lambda) = \int_0^{\infty} e^{\lambda t} f(t) dt$$

granted that the integral exists. The Laplace integral is said to be absolutely convergent if

$$\mathcal{L}[f] = \hat{f}(\lambda) = \lim_{\omega \rightarrow \infty} \int_0^{\omega} |e^{\lambda t} f(t)| dt$$

exists. The inverse Laplace transform denoted as $\mathcal{L}^{-1}[\cdot]$ maps \hat{f} to f .

Definition 1.2.9. A function $f(t)$ is said to be of bounded exponential growth if there exist constants $c \in \mathbb{R}$, $N > 0$ and $a > 0$ such that

$$|f(t)| < Ne^{ct}$$

for all $t > a$. If $f(t)$ is a function of bounded exponential growth with

$$\int_0^a |f(t)| dt$$

existing and finite, then $\hat{f}(\lambda)$ exists for $\lambda \in \mathbb{C}$ and $\operatorname{Re}(\lambda) > c$ where $\operatorname{Re}(\lambda)$ is the real part of $\lambda \in \mathbb{C}$ [41].

Functions of bounded exponential growth have Laplace transforms.

Proposition 1.2.1. Convolution of integral equations

Let $f(t)$ and $g(t)$ be Laplace transformable functions. Then, the convolution of the functions

$$H(t) = f * g(t) = \int_0^t f(\tau)g(t - \tau) d\tau \quad \text{for } t \geq 0$$

is also Laplace transformable and

$$\hat{F}(\lambda) = \hat{f}(\lambda)\hat{g}(\lambda)$$

for $\operatorname{Re}(\lambda) > \max\{\sigma_f, \sigma_g\}$ where σ_f, σ_g are the abscissa of convergence [48, Proposition A.6].

Chapter 2

Multi-scale Epidemic Modeling and Modeling of Cholera

2.1 Multi-scale Epidemic Modeling

Mathematical models of infectious diseases have been fundamental in generating a deep understanding of disease dynamics, making predictions, and proposing optimal intervention strategies for the control of diseases [70, 58]. Standard disease models restrict the dynamics to a single scale with the most popular scales being the between-host scale and the within-host scale [38, 5]. On the within-host (immunological) scale, the focus is on the evolution of the infection within a single individual whereas the between-host (epidemiological) scale is interested in the spread of the disease between individuals [37, 36]. Models that link the two scales of infection and give rise to novel insights into disease dynamics are referred to as multi-scale (immuno-epidemiological) models [38]. For some diseases like Dengue, where the severity of the disease depends on the strength of the immune response, multi-scale models become very handy. Other benefits that arise from the use of multi-level models are: the models enable the prediction of epidemiological quantities such as the reproduction number from the immunological processes, they explain the role of within-host dynamics on pathogen evolution and data exist on both scales [74]. Multi-scale models usually consist of three parts, namely, the between-host model, the within-host model, and the mechanism of linkage. In the next sections, we give an overview of each of the components.

2.1.1 Epidemiological models

Models that describe disease dynamics in the population have evolved since the seminal works of Kermack and McKendrick [59, 60]. The models subdivide the population into several compartments that represent the state of individuals at a particular point in time. A simple model, studied in depth by Kermack and McKendrick [59], that describes the dynamics of an epidemic is the *SIR* model. The *S* denotes susceptible individuals (in-

dividuals who are yet to be infected), I denotes infected individuals who are assumed to be infectious and can spread the disease when they come in contact with susceptible individuals and R represents individuals who were previously infected but have recovered and cannot be infected again. $S(t)$, $I(t)$, and $R(t)$ are functions of time since the number of persons in the classes changes with time. The total population $N(t)$ is then taken to be the sum of individuals from the three compartments, that is,

$$N(t) = S(t) + I(t) + R(t).$$

The flow diagram of such a model is shown in Figure 2.1.

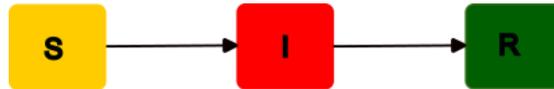


Figure 2.1: A SIR model, the black arrows indicate the movement of individuals within the classes.

The SIR model can be described by a system of ordinary differential equations that show the changes in the different compartments with respect to time. The easiest way to formulate the equations for the model is to assume a closed population with no demography, that is, there are no births or deaths taking place. In addition, recovered individuals become permanently immune to the disease. An additional assumption could be homogeneous mixing, where a person contacts other individuals in the population randomly. The SIR model can then be described by the following equations

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I. \end{aligned} \tag{2.1.1}$$

Here, β is the rate of infection, and subsequently, βSI is the number of individuals who get infected per unit of time (incidence). The parameter γ is the rate of recovery/removal and $\frac{1}{\gamma}$ gives the average infectious period. The infectious period can often be estimated from epidemiological data. The SIR model (2.1.1), although simple in formulation, cannot be solved explicitly. That is, the exact analytical expressions of the dynamics of S or I through time cannot be obtained. Numerical methods and the visualization of phase portraits are employed to gain insight into the dynamics. Even so, the SIR model can be used to describe two qualitative epidemiological characteristics: (a) the threshold phenomenon, which indicates whether an epidemic will spread or whether it will die out, and (b) epidemic burnout, which describes factors that will lead to the breaking down of the transmission chain.

2.1.1.1 Threshold Condition

At the initial stage, we have $S(0)$ number of susceptible. If a small number of infected individuals are introduced to the population, we get from the system (2.1.1)

$$\left. \frac{dI(t)}{dt} \right|_{t=0} = I(0)(\beta S(0) - \gamma) \implies \left. \frac{dI(t)}{dt} \right|_{t=0} > 0 \quad \text{if} \quad \frac{\beta S(0)}{\gamma} > 1. \quad (2.1.2)$$

The value $\frac{\beta S(0)}{\gamma}$ is the threshold condition that determines if the disease will invade the population or die out. The threshold condition is commonly referred to as the basic reproduction number \mathcal{R}_0 . The intuitive definition of the basic reproduction number is the average number of secondary infections that result from the introduction of a single infected individual into a purely susceptible population. The disease invades the population when $\mathcal{R}_0 > 1$ and dies out when $\mathcal{R}_0 < 1$. In the case above, if we take the initial population to be purely susceptible, that is $S(0) = N$ then $\mathcal{R}_0 = \frac{\beta N}{\gamma}$.

2.1.1.2 Final size of the epidemic

The long-term state of the system can be studied to get a view of how the infection breaks down. We can leave out the dynamics of I in eq. (2.1.1) by dividing the equation for S with R ,

$$\frac{dS}{dR} = -\frac{\beta}{\gamma} S.$$

Integration of the equation yields

$$S = S(0)e^{-\frac{\beta}{\gamma}R} \geq S(0)e^{-\frac{\beta}{\gamma}N} > 0. \quad (2.1.3)$$

Note that S always remains positive and as such some susceptible individuals will always manage to escape infection. This implies that an epidemic does not end because of the total lack of susceptible individuals but due to the decline of infected persons. Since the total population, $N = S + I + R$ and the epidemic dies out when $I = 0$, the long-term behavior of eq. (2.1.3) can be written as

$$S_\infty = N - R_\infty = S(0)e^{-\frac{\beta}{\gamma}R_\infty} \implies N - R_\infty - S(0)e^{-\frac{\beta}{\gamma}R_\infty} = 0 \quad (2.1.4)$$

where S_∞ and R_∞ denote the final proportion of susceptible and recovered persons respectively. The quantity R_∞ is equal to the total number of people that got infected during the epidemic and is referred to as the *final size of an epidemic*. We can rewrite eq. (2.1.4) to get the *final size relation*, that is

$$\ln \frac{S_\infty}{S(0)} = \frac{\beta}{\gamma}(N - S_\infty) = \mathcal{R}_0 \left(1 - \frac{S_\infty}{N}\right). \quad (2.1.5)$$

The final size relation is an equation that relates the basic reproduction number to the final size of the epidemic.

2.1.1.3 Solving the Basic SIR Model

To solve the basic SIR model (2.1.1), we only consider the equation for S and I , since the equation for R can be recovered from the relation $N = S + I + R$. That is,

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I.\end{aligned}$$

Therewith, we obtain

$$\frac{dI}{dS} = -1 + \frac{\gamma}{\beta S}.$$

We integrate the equation to get,

$$I(t) = I(0) + S(0) - S(t) + \frac{\gamma}{\beta} \ln\left(\frac{S(t)}{S(0)}\right).$$

The solution of $I(t)$ cannot be evaluated explicitly, thus, numerical methods are used to find an approximate solution. Figure 2.2 shows the behavior of the SIR model.

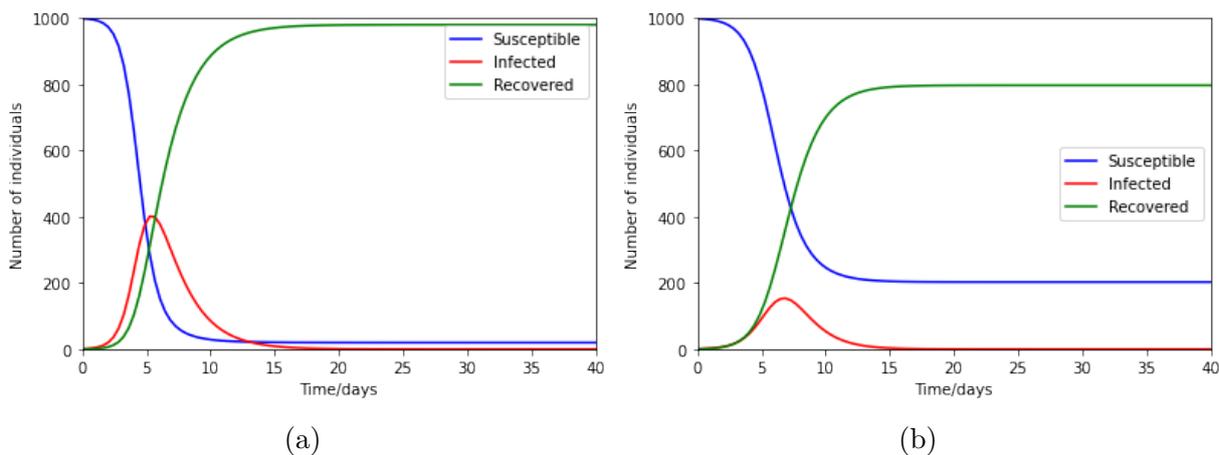


Figure 2.2: Trajectories of the SIR model with different rates of recovery. In (a) $\beta = 0.002$, $\gamma = 0.5$, $S(0) = 999$, $I(0) = 1$. In (b) $\beta = 0.002$, $\gamma = 1$, $S(0) = 999$, $I(0) = 1$.

The basic SIR model also referred to as the Kermack-McKendrick Model, is the prototype for epidemic models. It can be expanded to include vital dynamics (deaths and births). A relaxation of the assumption of permanent immunity gives rise to the Susceptible-Infected-Recovered-Susceptible model ($SIRS$), where immunity wanes and the individual in the recovered class becomes susceptible again. If there's no immunity against the disease a model with Susceptible-Infected-Susceptible classes (SIS model) is used to represent the dynamics. For diseases with no recovery e.g HIV, a model with only susceptible and

infected classes (*SI* model) is suitable for describing the dynamics. Other compartments can be included in the *SIR* model to represent additional dynamics. These include exposed or latent periods, treatment, symptomatic and asymptomatic infections, and quarantine among others.

2.1.1.4 Structured Epidemic Model

In some instances, individual characteristics (physiological variables) such as size, age, and status determine factors like birth, death, and growth rates. Epidemiological models based on such individual characteristics are referred to as structured epidemic models [29, 26]. Partial differential equations are used to model structured populations with a continuous state space of age or size. The earliest example of the inclusion of age structure in an epidemic model is the McKendrick model of 1926 [75]. The Lotka-McKendrick equation is the basic linear model that has been used to describe the evolution of age structure in a population [48]. The model is defined by the following set of equations

$$\begin{aligned} \frac{\partial p(a, t)}{\partial t} + \frac{\partial p(a, t)}{\partial a} &= -\mu(a)p(a, t) \\ p(0, t) &= \int_0^{\bar{a}} \beta(a)p(a, t)da = B(t) \\ p(a, 0) &= \theta(a). \end{aligned} \quad (2.1.6)$$

Here, $p(a, t)$ is the population age density with the age $a \in [0, \bar{a}]$ and time $t > 0$. The population grows to a maximum age of \bar{a} with the fertility rate $\beta(a)$ and mortality rate $\mu(a)$ being age specific. The total number of newborns $B(t)$ is the sum of births in the different age classes and $\theta(a)$ is the initial age distribution. Some basic assumptions are usually prescribed to the fertility and mortality functions to make them biologically relevant and to allow for the mathematical treatment of the model. For the model (2.1.6), the vital functions read as follows:

1. $\mu \in L^1_{loc}[0, \bar{a})$, $\mu(a) \geq 0$ a.e in $[0, \bar{a}]$, and $\int_0^{\bar{a}} \mu(s)ds = +\infty$ such that no individual survives past the maximal age \bar{a} .
2. $\beta \in L^\infty(0, \bar{a})$, $0 \leq \beta(a) \leq \beta_+$ a.e in $[0, \bar{a}]$,
3. $\theta \in L^1(0, \bar{a})$, $\theta(a) \geq 0$ a.e in $[0, \bar{a}]$.

Solving a Structured Epidemic Model

Several approaches can be used to solve the Lotka-McKendrick equation (2.1.6). One of the approaches involves integrating the PDE along characteristic lines (using the method of characteristics) to get the Volterra integral equation of the second kind (renewal equation) [48, 29] defined as

$$B(t) = F(t) + \int_0^t K(t-a)B(a)da \quad (2.1.7)$$

where F , K are known functions. Analytical properties including the existence and uniqueness of solutions can then be investigated using the renewal equation (2.1.7). The other approach is the use of semigroup theory [49, 77, 68, 90].

Ultimately, in multi-scale modeling, the modeling framework at the epidemiological scale depends on the epidemic model used. That is the epidemic model can take the form of an ODE model, age-of-infection (nested) models, size-structured PDE models, or network models [71].

2.1.2 Immunological models

Models at this scale are interested in the infection process within the host and as such, they represent the interaction of the pathogen with the host replication machinery or with the host immune defenses [74]. A basic within-host pathogen model contains certain basic characteristics for the disease in question. However, the format of the basic model for the different types of pathogens has very small variation [23]. Within-host models can be classified into three groups: those depicting the pathogen reproduction process, those showing the interaction of the pathogen with the immune responses, and those showing both the reproduction process and the immune responses. The first two types are commonly used in multi-scale modeling [74].

Models that describe the pathogen replication process show how a viral pathogen reproduces with the help of host cells (target cells). Such models, also referred to as TIV (Target cells, Infected cells, Virus) models, are commonly used to describe the evolution of viral diseases [96, 36, 64] (see Figure 2.3 below).

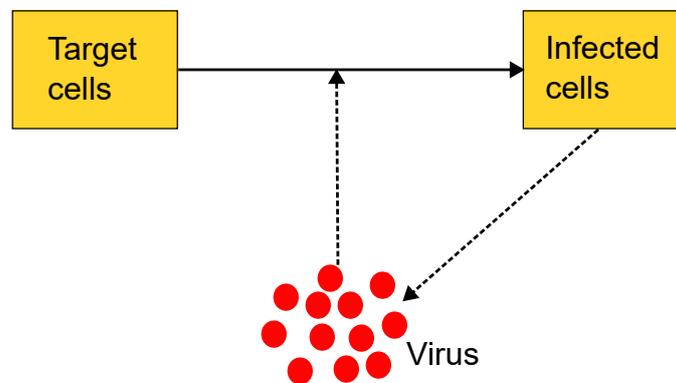


Figure 2.3: A flow diagram for a TIV model depicting within-host dynamics.

Within-host models with immune responses consider either the innate or adaptive immune responses that help eliminate the pathogen. A simple formulation of a within-host model describing the interaction of the pathogen (P) and adaptive immune response (B cells (B))

is given by the following system of ordinary differential equations [74]

$$\begin{aligned}\frac{dP}{dt} &= \gamma P \left(1 - \frac{P}{K}\right) - \delta P B \\ \frac{dB}{dt} &= cP - \mu B.\end{aligned}\tag{2.1.8}$$

Here, γ is the growth rate of the pathogen, K is the carrying capacity, δ is the killing rate of the immune response, c is the activation rate of the immune response, and μ the clearance of the immune response. The model (2.1.8) can be analyzed by finding equilibrium points and their asymptotic behavior. The model has two equilibrium points

$$(P_0, B_0) = (0, 0) \quad \text{and} \quad (P^*, B^*) = \left(\frac{\gamma\mu K}{\gamma\mu + \delta c K}, \frac{c\gamma K}{\gamma\mu + \delta c K} \right)$$

and its dynamics are visualized in Figure 2.4 below. Figure 2.4a shows the chronic case where the pathogen cannot be cleared by the immune response whereas Figure 2.4b shows the case where the pathogen can eventually be cleared out.

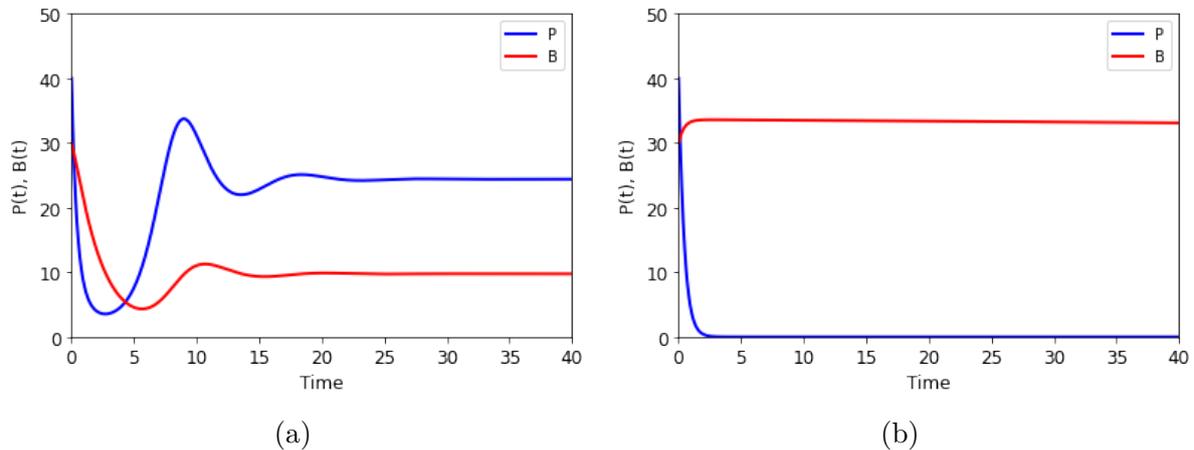


Figure 2.4: The changes in the pathogen load and immune response with respect to time. The parameter values for (a) are $\gamma = 1$, $\delta = 0.2$, $c = 0.1$, $\mu = 0.5$, and $K = 1000$. The parameter values for (b) are $\gamma = 1$, $\delta = 0.2$, $c = 0.1$, $\mu = 0.0004$ and $K = 1000$.

2.1.3 Linking Mechanisms

The linking mechanism establishes the connection of the within-host model to the between-host model and vice versa. The models can be linked in three different ways: between-to-within, within-to-between, or bidirectionally [21]. For deterministic models, the following linkage frameworks have been developed [39].

1. Linking through nesting principles: This is achieved by connecting the dynamics of the immunological to the epidemiological model by the use of a structural variable

or a parameter of the between-host model. In the case of a structural variable, the between-host model must be structured by the time since infection and this variable is then used as an independent variable in the within-host model. In the case of linking via parameters, parameters in the between-host model appear as functions of the dependent variables of the within-host model. For instance, the between-host transmission rate can be assumed to be a function of the within-host pathogen load [15, 101].

2. Linking through developing an immunological inspired between-host model: In this approach, a physiologically structured between-host model is developed, where the physiological aspect is derived from the immunological properties, for example, the immune response [72].
3. Linking through environmental contamination: For infections with free-living environmental pathogens, the dynamics at the environmental scale link the within and between host systems. Typically, the host, pathogen, and environment interact appropriately for the disease to occur, and the within-host, between-host, and environment dynamics are assigned different time scales [37].
4. Linking through network modeling principles: In network modeling, individuals are taken as nodes in a network. In this approach, the pathogen load of an individual is linked with that of an adjacent individual in a randomly-distributed network [103].

Figure 2.5 below depicts an immuno-epidemiological model with linked between-and within-host dynamics.

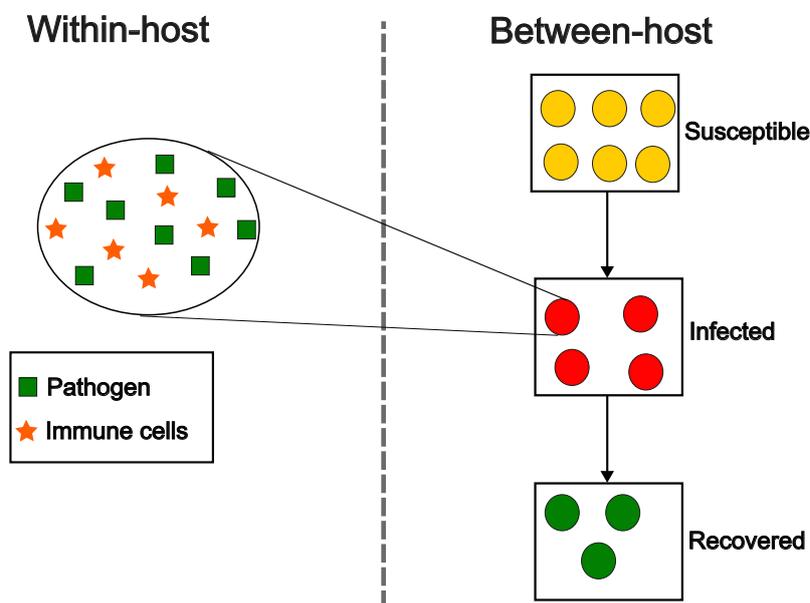


Figure 2.5: An immuno-epidemiological model.

2.2 Modeling of Cholera

In this section, we give a brief review of some of the cholera models that have been developed at the between-host and within-host scales and at the multi-scale level. Cholera is an environmentally transmitted disease. For an infection to occur, there should be adequate interaction between the host, the pathogen, and the environment. Thus, its contact structure not only includes the interactions within the human population (direct) but also the interaction with the environmental bacterial population (indirect). This makes its structure slightly different from the Kermack-McKendrick model where there is direct contact between interacting populations.

2.2.1 Between-host Cholera Models

Mathematical modeling of cholera as an infectious disease has seen advancements in recent years. Among the pioneering epidemic models for the disease was the model developed in 1979 by Capasso to describe the 1973 cholera epidemic in Italy [18]. The model comprised two ordinary differential equations that represented the bacterial population in seawater and the infected human population and formed a basis for the formulation of other cholera models.

To study the role of the aquatic reservoir in the persistence of endemic cholera, Codeço [24] extended the model by Capasso and Paveri-Fontana [18] by including the population of susceptible individuals. The model was the first to apply the *SIR* compartmental modeling approach to the study of the disease. It also elucidated the importance of the environmental reservoir in disease transmission dynamics. Other models were then formulated to explain different aspects of the disease.

To illustrate the importance of hyperinfectious cholera bacteria, Hartley et al. [45] extended the model by Codeço [24] to include a transient hyperinfectious state of the bacteria in the environment. The hyperinfectivity resulted from the passage of the *V. cholerae* bacteria through the human gut. Bacteria in this hyperinfectious state, usually freshly shed, were short-lived and degraded to a less infectious state after a few hours. They however had a competitive advantage in that a lower infectious dose was required to cause infection. Additionally, freshly shed vibrios were more likely to come into contact with humans than bacteria in the environment. Hyperinfectivity of bacteria was seen as one of the causes of the explosive nature of the disease.

Mukandavire et al. [81] simplified the model by Hartley et al. [45] by classifying the pathway for transmission based on the infectiousness of the bacteria. The bacteria in the hyperinfectious state were responsible for the human-human (fast) transmission of the disease, while those in the less infectious state were responsible for the environment-human (slow) transmission of the disease. The model was then used to estimate the prevalence of cholera in Zimbabwe. The dual transmission pathways for the disease were also considered by Tien and Earn [99].

Other aspects such as the role of human mobility in the spread of the disease and the effect

of intervention strategies have been studied by [104, 69, 78, 85, 6]. Although most of the models mentioned above are non-structured (ODE) models, structured epidemic models for the disease have been advanced in [11, 65].

Between-host cholera models provide effective approximations of the epidemic process. Non-structured models are also comparatively easier to analyze. However, the underlying biochemical mechanisms that shape the infectiousness of individuals are ignored thus providing a drawback to the consideration of models at only this scale.

2.2.2 Within-host Cholera Models

Fewer within-host cholera models have been developed in comparison to epidemic models. Wang and Wang [106] formulated deterministic and stochastic versions of a within-host cholera model to study the interactions between cholera bacteria and viruses (phages). The bacteria ingested from the environment (environmental vibrios) were transformed into highly infectious human vibrios through interaction with cholera phages (virus). They derived conditions in which human vibrios would grow and lead to cholera infection. In addition, they derived conditions in which human vibrios would not grow and ingested environmental vibrios would not lead to cholera infection. The model was later extended by [8] to include immune responses.

A limitation of single-scale within-host or between-host cholera models is that they cannot comprehensively describe the infection process. This is because the spread and transmission of the disease are linked to the infectiousness of an individual, which consequently depends on the pathogen load. Thus, host-pathogen interactions within an individual greatly influence the transmission of the disease at the population level.

2.2.3 Multi-scale Cholera Models

Recent advances have been made in the multi-scale modeling of cholera. To our knowledge, the first attempt to develop a multi-scale model was made by Wang and Wang [105]. The within-host model depicts the evolution of highly infectious human vibrios in the body whereas the between-host model is a *SIR* system that includes a compartment for environmental bacteria. The highly infectious human vibrios contribute to the environmental growth of bacteria as well as disease transmission among humans, providing a linkage between the two scales. The within-host dynamics are taken to occur at a faster time scale (hours) in comparison to the between-host dynamics (months) allowing the use of fast-slow analysis for a detailed study of the dynamics at each scale. The advantage of the model is that it allows for the linkage of the two systems. A drawback of the model is that the within-host system takes a single ordinary differential equation form and thus fails to account for other biological processes in the body which are significant in the infection process. Subsequently, the physiological characteristics of individual hosts are assumed to be homogeneously distributed thus not accounting for heterogeneity that occurs from person to person caused by factors such as differences in susceptibility.

The work in [105] is extended by Ratchford and Wang [91]. This is done through the

inclusion of a compartment for viruses and immune responses in the within-host model. This allows for a more detailed examination of the dynamics of disease within a host. The complete model is decoupled into three subsystems, with the between-host system being on the intermediate time scale, the environmental pathogen on the slow time scale, and the within-host dynamics on the fast time scale. The separation of time scales allows for a detailed study of the asymptotic behavior of the models at each scale. The results indicate that the basic reproduction number is dependent on both the direct and indirect transmission pathways of the disease. However, similar to [105] the model assumes homogeneity in the individual hosts.

The nesting approach is used in [15] to couple the within-host and between-host dynamics of cholera via the infection age of individuals. Subsequently, the between-host parameters: direct transmission rate, death, recovery, and shedding rates, explicitly depend on the within-individual pathogen load and immune strength. Consequently, within-host immunopathology is seen to contribute to disease transmission risk at the population level. An advantage of the nested approach is that it helps to account for different infection states of the individual and their contribution to disease transmission. A limitation of the model is that the between-host dynamics have no impact on the within-host dynamics which is not realistic for environmentally driven diseases. For instance, bacterial concentrations in the environment can influence the evolution of pathogens within a human host. Subsequently, the severity of the disease depends on the inoculum [40], which may depend on the disease transmission route.

Based on the existing models, in the next chapters, we aim to develop multi-scale models that allow for some level of heterogeneity in individuals by structuring the epidemic models using within-host immunological properties. Subsequently, we will also introduce heterogeneity caused by differences in susceptibility in a single individual host.

Chapter 3

Linking the Dynamics through Immune Responses

In this chapter, we make the first attempt at formulating an immuno-epidemiological model to link the dynamics of cholera. We begin with a single individual and formulate an immunological (within-host) model that depicts the interaction of the cholera pathogen with the immune response. We distinguish pathogen dynamics from immune response dynamics by assigning different time scales. This turns out to also be mathematically convenient as it allows for the use of the fast-slow approach in the analysis. From the results of the fast-slow analysis, we can follow the fate of an infected individual such that we can characterize the individual by the state of their immune response. We then scale up the dynamics of a single infected person to the population. It is logical that we structure the infected population by the state of the immune response. The epidemiological model (between-host) is a size-structured model that considers human population dynamics as well as the dynamics of bacteria in the environment. Using the standard theory of structured models [48, 29, 26, 108], we analyze the epidemiological model to determine the long-term behavior of solutions. We then modify the epidemiological model to include a maximum age for the bacteria. We establish the global stability of the modified model and finally show the equivalence of the adjusted model to the original model. Parts of this section are available in the preprint [83].

3.1 The Within-host Model

We aim to construct a model that outlines the interaction of cholera pathogens with immune responses within an individual. We follow a similar approach described in [74]. In our case, we consider the growth of the pathogen to be influenced by Allee effects. Allee effects depict the co-relation between population size and fitness of a species [32, 3]. Populations exhibiting this effect show reduced growth at low pathogen densities. The effects are classified as strong if they result in a critical population density and weak if they don't

result in a critical density [32]. We consider the case of strong Allee effects and assume that the *Vibrio* pathogens grow above a critical density for infection to occur. This turns out to be biologically relevant since microbial populations with quorum sensing mechanisms such as *Vibrio cholerae* bacteria have been shown to exhibit Allee effects [56, 51]. Subsequently, cholera bacteria, ingested from the environment, must first penetrate the mucus lining of the intestinal epithelial cells and escape the innate immunity defenses to be able to proliferate [92, 4]. At low densities, the innate immune responses can fight off the bacteria, and therefore a high infectious dose of $10^8 - 10^{11}$ cells is required for the bacteria to colonize the small intestines and cause infection [87]. With that in mind, we formulate the within-host model. We consider that after ingestion of bacteria, an individual has a pathogen load P that interacts with the adaptive immune response W , that is,

$$\begin{aligned}\frac{dP}{dt} &= \alpha P \left(1 - \frac{P}{K}\right) (P - \beta) - \gamma P - \delta PW \\ \frac{dW}{dt} &= \varepsilon (\kappa P - cW).\end{aligned}\tag{3.1.1}$$

To generate the Allee effect, we use a cubic growth term for the pathogen where α is the intrinsic growth rate, β is the critical density (Allee threshold) below which the pathogen has reduced growth rates and K is the carrying capacity. The pathogen is removed through natural death at rate γ and through clearance by the immune response at rate δ . The immune response is activated in the presence of the pathogen at rate κ and self-deactivated at rate c . Adaptive immune responses are known to be slower to respond to pathogens than innate responses [20]. As such, we consider that the pathogen grows at a much faster rate than the immune response. We therefore explicitly distinguish pathogen dynamics from immune responses by prescribing different time scales. The parameter $\varepsilon \ll 1$ is the time scale parameter that represents the slow time scale of the immune response. This turns out to be mathematically convenient for analysis. Time-scale separation methods have become common in the study of biological systems after their advancement through the work of FitzHugh and Nagumo [82, 62]. The analysis of the resulting systems (fast and slow systems) involves splitting the full system into smaller subsystems, the fast system and the slow system, and analyzing the dynamics separately (see Section 1.2.1 for an introduction). The approach has been used to analyze biological systems in [82, 62, 36, 10]. Next, we follow the same procedure for the analysis of the within-host model.

3.2 Time-scale Analysis

Pathogen dynamics and the dynamics of the immune response occur on separate time scales. The pathogen load is considered to grow on a fast time scale while the immune response grows on a slow time scale. We split the system (3.1.1) into fast and slow systems and analyze the dynamics of each system.

3.2.1 Fast System

We consider the pathogen load P to be the fast variable. On the fast time scale, the immune response W does not change. Therefore, we let W be constant and only consider the change in the pathogen load. For $\varepsilon \rightarrow 0$, the fast system is given by

$$\begin{aligned} \frac{dW}{dt} &= 0 \\ \frac{dP}{dt} &= \alpha P \left(1 - \frac{P}{K}\right) (P - \beta) - \gamma P - \delta PW. \end{aligned} \quad (3.2.1)$$

The trajectories of the fast system tend to stationary points as seen in the next proposition.

Proposition 3.2.1. *System (3.2.1) exhibits three branches of stationary states: One trivial uninfected branch $(W, 0)$ and two infected branches $(W, P_{1,2})$ with*

$$P_{1,2} = \frac{\alpha(\beta + K) \pm \sqrt{\alpha^2(\beta + K)^2 - 4\alpha K(\alpha\beta + \gamma + \delta W)}}{2\alpha}$$

which exist for $\alpha \geq \alpha_0 = \frac{4K(\gamma + \delta W)}{(\beta - K)^2}$.

The trivial and the upper infected branches are locally asymptotically stable, while the lower infected branch is unstable. At $\alpha = \alpha_0$ a saddle-node bifurcation takes place.

Note that we can represent the infected branches by $(W, P) = (\phi(P), P)$ with

$$\phi(P) = \frac{\alpha}{\delta} \left(1 - \frac{P}{K}\right) (P - \beta) - \frac{\gamma}{\delta}$$

as can be seen in the proof below.

Proof. As $\frac{dW}{dt} = 0$, W is fixed. We only consider $\frac{dP}{dt}$ for a given $W \in \mathbb{R}_+$. To find the equilibrium points, we set the right-hand side of system (3.2.1) to zero

$$0 = \left(\alpha \left(1 - \frac{P}{K}\right) (P - \beta) - \gamma - \delta W\right) P. \quad (3.2.2)$$

Equation (3.2.2) defines the slow manifold (see Definition 1.2.3) of the system which can also be expressed as

$$\{P = 0\} \cup \{\phi(P) = W\} \quad (3.2.3)$$

where $\phi(P) = \frac{\alpha}{\delta} \left(1 - \frac{P}{K}\right) (P - \beta) - \frac{\gamma}{\delta}$. We let

$$f(P, W) = \alpha P \left(1 - \frac{P}{K}\right) (P - \beta) - \gamma P - \delta PW$$

such that $\frac{dP}{dt} = f(P, W)$. Then, the line $P = 0$ is the infection-free equilibrium and for any point $P_0 = (0, W)$ on the line,

$$\frac{\partial}{\partial P} f(P_0) = -(\alpha\beta + \gamma + \delta W) < 0, \quad (3.2.4)$$

thus, the infection-free equilibrium point is locally asymptotically stable. The curve $\phi(P)$ yields the non-trivial equilibrium. We can rewrite it as

$$\begin{aligned} 0 &= \alpha P^2 - \alpha(\beta + K)P + (\gamma + \delta W + \alpha\beta)K \\ P_{1,2} &= \frac{\alpha(\beta + K) \pm \sqrt{\alpha^2(\beta + K)^2 - 4\alpha K(\alpha\beta + \gamma + \delta W)}}{2\alpha}. \end{aligned} \quad (3.2.5)$$

It then follows that two positive non-trivial stationary points $P_{1,2}$ exist whenever

$$\alpha \geq \alpha_0 = \frac{4K(\gamma + \delta W)}{(\beta - K)^2}.$$

A saddle-node bifurcation (fixed point created and destroyed) occurs at the point $\alpha = \alpha_0$. Moreover, the stability of the non-trivial equilibrium point is a direct consequence of $f(P)$ being a polynomial of the third order. The upper branch of $\phi(P)$ is therefore stable while the lower branch is unstable. \square

Next, we study the slow system. The slow system gives us a reduced problem from which generalizations about the complete system can be made.

3.2.2 Slow System

We consider the immune response W to be the slow variable. Denoting the slow time scale as τ , i.e. $\tau = \varepsilon t$, we express the slow system dynamics as

$$\begin{aligned} \varepsilon \frac{dP}{d\tau} &= \alpha P \left(1 - \frac{P}{K}\right) (P - \beta) - \gamma P - \delta P W \\ \frac{dW}{d\tau} &= \kappa P - cW. \end{aligned}$$

On the singular limit, the system reduces to

$$0 = \alpha P \left(1 - \frac{P}{K}\right) (P - \beta) - \gamma P - \delta P W \quad (3.2.6)$$

$$\frac{dW}{dt} = \kappa P - cW. \quad (3.2.7)$$

Notice that eq. (3.2.6) is the slow manifold of the fast system (3.2.1). We sketch the slow manifold in Figure 3.1a below. We take the upper branch of the slow manifold ($\phi(P)$) to represent the fate of an infected individual (infected branch) and the lower branch ($P = 0$) to represent the recovery process (recovered branch). On the fast scale (Figure 3.1b), the fast dynamics drive the trajectories towards the recovered branch and the upper segment of the infected branch since both are stable. We then use eq. (3.2.7), where P is now given in terms of W , to investigate what happens to the stable branches of the slow manifold. The line $\dot{W} = \kappa P - cW = 0$ (from eq. (3.2.7)) intimates the direction of movement on the

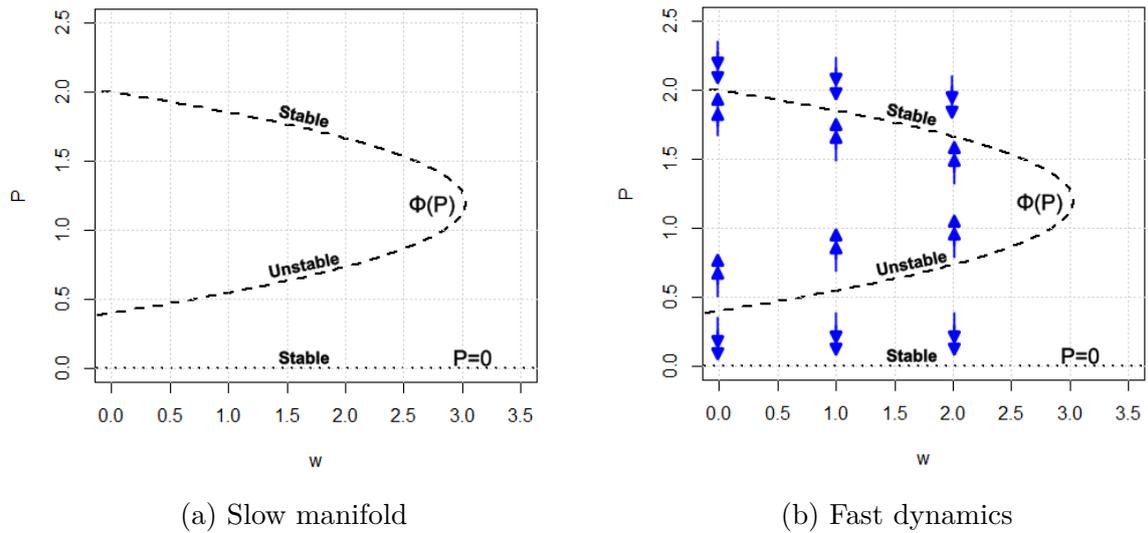


Figure 3.1: The slow manifold and the trajectories on the fast scale. The blue arrows represent the direction of movement of the trajectories on the fast scale.

manifold, that is if movement is to the left or right of the manifold.

On the slow scale, solutions below the line $\dot{W} = 0$, that is on the recovered branch, move to the left towards the origin $(P, W) = (0, 0)$, which is a locally stable fixed point while solutions above the line move to the right (Figure 3.2a).

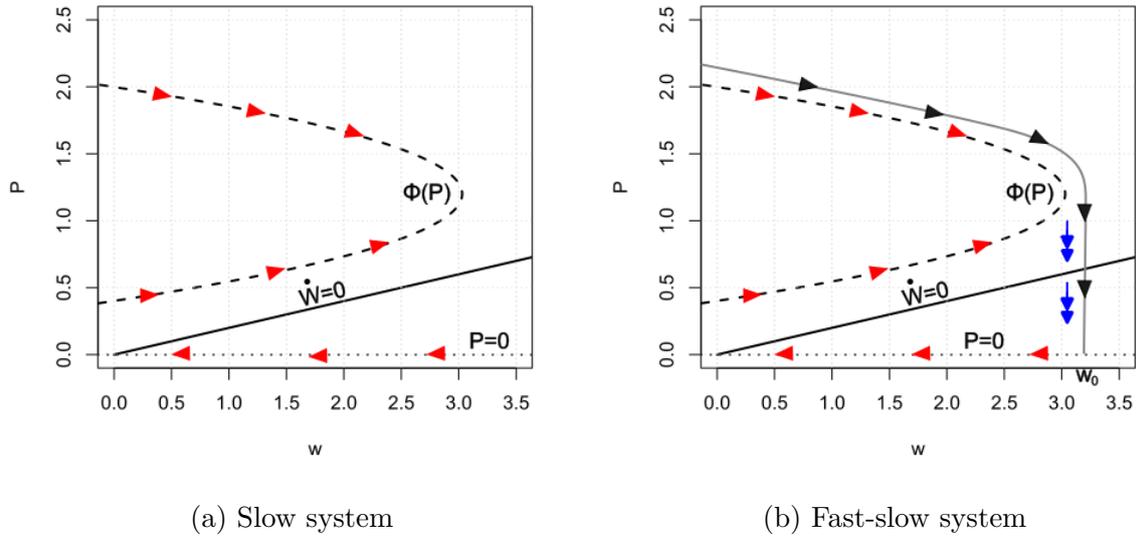


Figure 3.2: Trajectories on the slow scale and the full system. The blue arrows represent movement on the fast scale, the red arrows represent movement on the slow scale and the black arrows represent the movement of the trajectories in the full system.

We observe some minimum pathogen threshold dynamics in Figure 3.2a. Note that we only consider the case where $\frac{\epsilon}{\kappa}$ is small enough that $\{\kappa P = cW\} \cap \{W = \phi(P)\} = \emptyset$, that is, there is no chronic infection stationary state.

Since the slow system can be used to make generalizations about the full within-host system (3.1.1), we conclude the infection process by shifting our attention to the infected branch of the slow manifold ($\phi(P)$). We discuss this below in more detail. The infection starts once the pathogen threshold is surpassed. As shown in Figure 3.2b solutions on the upper stable branch of the infected manifold are driven to the right until the tip of the manifold is reached. At this point, due to the lower branch of $\phi(P)$ being unstable, movement along the manifold is inhibited and the fast dynamics force a jump into the recovered branch ($P = 0$) of the manifold which is also stable. The immune response is heightened during the infection process. Consequently, the state of the immune response at the start of infection is different from that at the point of recovery. We note the state of the immune response at different points. We let the immune response at the start of infection be $W = 0 = W_0^*$ and at the point of recovery (where the jump to the recovered branch takes place) be $W = W_0$. It then follows that we can describe the dynamics of the immune response by

$$\frac{dW}{dt} = \kappa\phi^{-1}(W) - cW = g(W) \quad (3.2.8)$$

where the function $\phi^{-1}(W)$ comes from $W = \phi(P)$ being a stationary solution (see eq.

(3.2.3). We take note that the dynamics of P can be recovered from the equation and describe the function $g(W)$ to be the growth rate of the immune response. That is

$$g(W) = \kappa\phi^{-1}(W) - cW, \quad \dot{W} = g(W), \quad W \in [0, W_0]. \quad (3.2.9)$$

Additionally, we have $g(W) > 0$ for all $W \in [0, W_0]$ such that the immune response grows with time. We can now describe the dynamics within a single infected person in terms of the immune response. We note that at the beginning of an infection, the state of the immune response is denoted by W_0^* ($W_0^* = 0$), and at the point of recovery it is denoted by W_0 . Notice that we only focus on immunity changes along the infected branch. For the recovery branch $P = 0$, we simply note that the response declines exponentially. For simplicity, we refer to the state of the immune response as immune status. In the next section, we scale up the dynamics within a single infected individual to structure the infected population.

3.3 Between-Host Model

We derive the between-host model based on immunological properties. It turns out that a structured epidemiological model naturally emerges from within-host dynamics. The use of physiologically structured models to study populations has been advanced by the work of Metz and Diekmann [77], Diekmann et al. [29], Cushing [26], Magal and Ruan [68]. The structuring variables include age, size, immunity, and many others. In Angulo et al. [7], Martcheva and Pilyugin [72] immunological variables are used to structure the population. Similarly, using the immune status, we formulate the between-host model. We consider a population made up of Susceptible S , Infected I , and Immune V individuals and include an additional compartment B to represent the bacterial concentration in the environment. Susceptible individuals are recruited into the population at rate r and removed by natural death at rate μ_1 . Individuals in the infected class are structured by their within-host immune status such that the total infected population

$$I(t) = \int_0^{W_0} i(t, W) dW,$$

where $i(t, W)$ is the density of infected individuals with immune status W . Transmission of the disease occurs through the consumption of contaminated food and water (indirect) and human-human contact (direct) [45]. As such, we take both transmission pathways into account by assigning a rate β_h as the direct transmission rate and β_e as the indirect transmission rate. We consider the infectivity of an infectious person to be dependent on the within-host pathogen load P . Given that the pathogen load is influenced by the state of immune response along the way addressed in Section 3.2, we take $P = P(W)$. Therefore, the force of infection in directly transmitted cases is proportional to the pathogen load. Natural death removes infected individuals at a rate of $\mu_2(W)$, and recovery occurs when immunity builds to the point where the pathogens are eliminated from the body. The recovery rate at that point is given as $g(W_0)$. Immunity wanes at a rate of ρ such that

an immune individual becomes susceptible once again. Immune individuals are further removed through natural death at rate μ_3 . The bacteria in the environment grow through shedding by infected individuals. Shedding, which occurs at rate $\xi(W)$, is proportional to the pathogen load in an infected individual. The bacteria decay at rate σ . Figure 3.3 illustrates the flow in the between-host model.

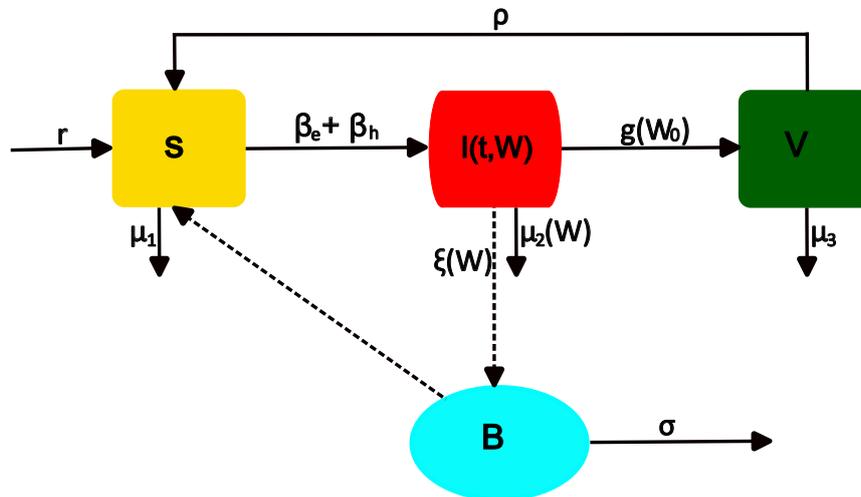


Figure 3.3: Model flow diagram

All in all, the between-host model reads,

$$\begin{aligned}
 \frac{dS(t)}{dt} &= r - \mu_1 S(t) - S(t) \int_0^{W_0} \beta_h P(W) i(t, W) dW - \beta_e S(t) B(t) \\
 &\quad + \rho V(t) \\
 \partial_t i(t, W) + \partial_W (g(W) i(t, W)) &= -\mu_2(W) i(t, W) \\
 g(0) i(t, 0) &= S(t) \int_0^{W_0} \beta_h P(W) i(t, W) dW + \beta_e S(t) B(t) \quad (3.3.1) \\
 \frac{dV(t)}{dt} &= g(W_0) i(t, W_0) - \rho V(t) - \mu_3 V(t) \\
 \frac{dB(t)}{dt} &= \int_0^{W_0} \xi(W) P(W) i(t, W) dW - \sigma B(t),
 \end{aligned}$$

with initial conditions $S(0) = S_0$, $V(0) = V_0$, $B(0) = B_0$, $i(0, W) = \Phi(W)$, and $g(W)$ is defined in equation (3.2.9).

3.3.1 Existence and Uniqueness of Solutions

The standard approach for showing the existence and uniqueness of solutions for structured models such as system (3.3.1) is to first transform the PDE problem into a renewal equation by integration along the characteristic curves. Thereafter, the ODEs are rewritten in terms of the renewal equation and finally fixed point theorems are applied to the resulting system

to show the existence and uniqueness of solutions [16, 61]. Similarly, we integrate the initial and boundary value problem for i along the characteristics.

Proposition 3.3.1. *Assume $i(t, 0)$ is given, the solution of the PDE in the system (3.3.1) with the initial and boundary conditions is given by*

$$i(t, W) = \begin{cases} \Phi(G^{-1}(G(W) - t)) \frac{g(G^{-1}(G(W) - t))}{g(\tau)} e^{-\int_{G^{-1}(G(W) - t)}^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} & W \geq 0 \\ H(t - G(W)) \frac{1}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} & W < 0, \end{cases}$$

with $G(W) = \int \frac{dW}{g(W)}$ and $H(t) = g(0)i(t, 0)$.

Proof. We rewrite the PDE in system (3.3.1) as

$$\partial_t i(t, W) + g(W) \partial_W i(t, W) = -\mu_2(W) i(t, W) - g'(W) i(t, W).$$

The characteristic curves of the PDE are given by

$$\frac{dt}{ds} = 1, \quad \frac{dW}{ds} = g(W), \quad \frac{dz}{ds} = -(\mu_2(W) + g'(W))z.$$

We solve the ODE's to get,

$$\begin{aligned} \frac{dt}{ds} &= 1 \implies t = s + c_0, \\ \frac{dW}{ds} &= G(W) \implies \frac{dW}{g(W)} = ds \implies G(W) = s + G(c_1) \implies W = G^{-1}(s + G(c_1)), \\ \frac{dz}{z} &= -(\mu_2(W) + g'(W)) ds \implies z = z(0) e^{-\int_0^s \mu_2(W) + g'(W) ds'} \\ &\implies z = z(0) e^{-\int_0^s \mu_2(G^{-1}(s' + G(c_1))) + g'(G^{-1}(s' + G(c_1))) ds'} \end{aligned}$$

where c_0 and c_1 are arbitrary constants, $z(0) = i(c_0, c_1)$ and the function $G(W) = \int \frac{dW}{g(W)}$. Along the initial conditions, that is $i(0, W) = \Phi(W)$, we obtain

$$\begin{aligned} t &= s, \\ W &= G^{-1}(t + G(c_1)) \implies c_1 = G^{-1}(G(W) - t) \end{aligned}$$

and

$$i(t, W) = \Phi(G^{-1}(G(W) - t)) e^{-\int_0^t \mu_2(G^{-1}(s' + G(W) - t)) + g'(G^{-1}(s' + G(W) - t)) ds'}. \quad (3.3.2)$$

We simplify the above equation by letting $\tau = G^{-1}(s' + G(W) - t)$, that is

$$G(\tau) = s' + G(W) - t \implies \int \frac{d\tau}{g(\tau)} = s' + G(W) - t \quad \text{thus} \quad ds' = \frac{d\tau}{g(\tau)}$$

such that

$$i(t, W) = \Phi(G^{-1}(G(W) - t)) e^{-\int_{G^{-1}(G(W)-t)}^W \frac{\mu_2(\tau) + g'(\tau)}{g(\tau)} d\tau}.$$

Thus for $0 \leq t \leq G(W)$, $W \geq 0$, we obtain the solution along the initial conditions as

$$i(t, W) = \Phi(G^{-1}(G(W) - t)) \frac{g(G^{-1}(G(W) - t))}{g(W)} e^{-\int_{G^{-1}(G(W)-t)}^W \frac{\mu_2(\tau)}{g(\tau)} d\tau}.$$

Similarly, we look for solutions along the boundary. For the sake of convenience, we denote the boundary condition as $H(t) = g(0)i(t, 0) = S(t) \int_0^{W_0} \beta_h P(W) i(t, W) dW + \beta_e S(t) B(t)$. For $t > G(W)$, $W < 0$ we obtain the appropriate integral along the boundary. That is

$$t = s + c_0, \quad G(W) = s, \quad z = z(0) e^{-\int_0^{G(W)} \mu_2(G^{-1}(s')) + g'(G^{-1}(s')) ds'}$$

such that

$$\begin{aligned} i(t, W) &= H(t - G(W)) e^{-\int_0^{G(W)} \mu_2(G^{-1}(s')) + g'(G^{-1}(s')) ds'} \\ &= H(t - G(W)) e^{-\int_0^W \frac{\mu_2(\tau) + g'(\tau)}{g(\tau)} d\tau} = H(t - G(W)) \frac{1}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau}. \end{aligned}$$

Thus, the complete solution of the PDE in system (3.3.1) with the initial and boundary conditions is given as

$$i(t, W) = \begin{cases} \Phi(G^{-1}(G(W) - t)) \frac{g(G^{-1}(G(W)-t))}{g(W)} e^{-\int_{G^{-1}(G(W)-t)}^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} & W \geq 0 \\ H(t - G(W)) \frac{1}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} & W < 0. \end{cases} \quad (3.3.3)$$

□

Now that we have the solution of $i(t, W)$ along the characteristics, the next steps for showing the existence and uniqueness of solutions, which we don't go into detail, can be evaluated similarly to [16, 108, 61, 55].

3.3.2 Basic Reproduction Number and Stability of the DFE

The basic reproduction number \mathcal{R}_0 is defined as the expected number of secondary infections produced when a single infected person is introduced into a purely susceptible population [28]. It is clear that a trivial disease-free equilibrium (DFE) for system (3.3.1) given by $\mathcal{E}_0 = (S_0^*, 0, 0, 0)$ where $S_0^* = \frac{r}{\mu_1}$ always exists. In this section, we investigate the local stability of the DFE by linearizing the system (3.3.1) around the disease-free equilibrium. In so doing, we derive the threshold condition for the spread of the disease that we consider to be the basic reproduction number. We note that we consider stability analysis with respect to the point spectrum. This is generally a sufficient condition for

linear stability analysis, that is, we presume that the point spectrum is contained in the left half plane $\text{Re}(z) < 0$ (see [49, Section 3.1.2]). To complete the proof of local stability we will need to show the connection between the roots of the characteristic equation and the stability of the equilibrium. For the PDE model, we only have information on the eigenvalues of the generator of the semigroup associated with the linear perturbed system and not the semigroup itself. Compact, eventually compact, and quasi-compact semigroups (see Section 1.2.2) relate the eigenvalues of the generator to the long-term behavior of the semigroup [73]. The following theorem suffices to complete the proof of local stability for stationary solutions.

Theorem 3.3.1. [73, Theorem B.1] *Let $T(t)$ be a quasi-compact C^0 -semigroup and A be its infinitesimal generator. Then $e^{\delta t} \|T(t)\| \rightarrow 0$ as $t \rightarrow \infty$ for some $\delta > 0$ if and only if all eigenvalues of A have strictly negative real parts.*

Later, we use the Lyapunov approach to prove the global stability of the appropriate stationary solutions.

Theorem 3.3.2. *The disease-free equilibrium is locally asymptotically stable when $\mathcal{R}_0 < 1$ and unstable if $\mathcal{R}_0 > 1$, where*

$$\mathcal{R}_0 = \frac{r\beta_h}{\mu_1} \int_0^{W_0} \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} dW + \frac{r\beta_e}{\mu_1} \int_0^{W_0} \frac{\xi(W)}{\sigma} \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} dW. \quad (3.3.4)$$

Proof. We let $S(t) = S_0^* + x(t)$, $i(t, W) = i_1(t, W)$, $V = y(t)$ and $B = z(t)$ where $x(t)$, $i_1(t, W)$, $y(t)$ and $z(t)$ are the perturbation variables and S_0^* is the trivial equilibrium point. We substitute these expressions in the system (3.3.1) to get

$$\begin{aligned} \frac{dx(t)}{dt} &= r - \mu_1(S_0^* + x(t)) - (S_0^* + x(t)) \int_0^\infty \beta_h P(W) i_1(t, W) dW \\ &\quad - \beta_e(S_0^* + x(t)) z(t) + \rho(V^* + y(t)) \\ \partial_t i_1(t, W) + \partial_w(g(W) i_1(t, W)) &= -\mu_2(W) i_1(t, W) \\ g(0) i_1(t, 0) &= (S_0^* + x(t)) \int_0^\infty \beta_h P(W) i_1(t, W) dW + \beta_e(S_0^* + x(t)) z(t) \\ \frac{dy(t)}{dt} &= g(W_0) i_1(t, W_0) - \rho(V^* + y(t)) - \mu_3(V^* + y(t)) \\ \frac{dz(t)}{dt} &= \int_0^\infty \xi(W) P(W) i_1(t, W) dW - \sigma z(t). \end{aligned} \quad (3.3.5)$$

Multiplying through by the perturbation terms yields

$$\begin{aligned}
 \frac{dx(t)}{dt} &= r - \mu_1 S_0^* - \mu_1 x(t) - S_0^* \int_0^\infty \beta_h P(W) i_1(t, W) dW \\
 &\quad - x(t) \int_0^\infty \beta_h P(W) i_1(t, W) dW - \beta_e S_0^* z(t) - \beta_e x(t) z(t) \\
 &\quad + \rho V^* + \rho y(t) \\
 \partial_t i_1(t, W) + \partial_w(g(W) i_1(t, W)) &= -\mu_2(W) i_1(t, W) \\
 g(0) i_1(t, 0) &= S_0^* \int_0^\infty \beta_h P(W) i_1(t, W) dW + x(t) \int_0^\infty \beta_h P(W) i_1(t, W) dW \\
 &\quad + \beta_e S_0^* z(t) + \beta_e x(t) z(t) \\
 \frac{dy(t)}{dt} &= g(W_0) i_1(t, W_0) - \rho V^* - \rho y(t) - \mu_3 V^* - \mu_3 y(t) \\
 \frac{dz(t)}{dt} &= \int_0^\infty \xi(W) P(W) i_1(t, W) dW - \sigma z(t).
 \end{aligned}$$

We simplify the model by neglecting quadratic perturbation terms that we assume to be much smaller than the perturbations. We, therefore, get the linearized system

$$\begin{aligned}
 \frac{dx(t)}{dt} &= -\mu_1 x(t) - S_0^* \int_0^{W_0} \beta_h P(W) i_1(t, W) dW - \beta_e S_0^* z(t) \\
 &\quad + \rho y(t) \\
 \partial_t i_1(t, W) + \partial_W(g(W) i_1(t, W)) &= -\mu_2(W) i_1(t, W) \\
 g(0) i_1(t, 0) &= S_0^* \int_0^{W_0} \beta_h P(W) i_1(t, W) dW + \beta_e S_0^* z(t) \quad (3.3.6) \\
 \frac{dy(t)}{dt} &= g(W_0) i_1(t, W_0) - \rho y(t) - \mu_3 y(t) \\
 \frac{dz(t)}{dt} &= \int_0^{W_0} \xi(W) P(W) i_1(t, W) dW - \sigma z(t).
 \end{aligned}$$

Next, we analyze the stability of the linearized system to determine the long-term behavior of the solutions. Since the system is linear, we expect the solutions to be exponential. Thus, we look for solutions of the form $x(t) = \bar{x}e^{\lambda t}$, $i_1(t, W) = \bar{i}_1(W)e^{\lambda t}$, $y(t) = \bar{y}e^{\lambda t}$, $z(t) = \bar{z}e^{\lambda t}$ where \bar{x} , $\bar{i}_1(W)$, \bar{y} , \bar{z} and λ are to be determined. The λ represents an eigenvalue. Substituting the appropriate formulation in system (3.3.6) gives us the eigenvalue problem

$$\begin{aligned}
 \lambda \bar{x} &= -\mu_1 \bar{x} - S_0^* \int_0^{W_0} \beta_h P(W) \bar{i}_1(W) dW - \beta_e S_0^* \bar{z} + \rho \bar{y} \\
 \partial_W(g(W) \bar{i}_1(W)) &= -(\mu_2(W) + \lambda) \bar{i}_1(W) \\
 g(0) \bar{i}_1(0) &= S_0^* \int_0^{W_0} \beta_h P(W) \bar{i}_1(W) dW + \beta_e S_0^* \bar{z} \quad (3.3.7) \\
 \lambda \bar{y} &= g(W_0) \bar{i}_1(W_0) - \rho \bar{y} - \mu_3 \bar{y} \\
 \lambda \bar{z} &= \int_0^{W_0} \xi(W) P(W) \bar{i}_1(W) dW - \sigma \bar{z}.
 \end{aligned}$$

We are now interested in obtaining an equation in terms of λ (characteristic equation), We need to eliminate \bar{x} , $\bar{i}_1(W)$, \bar{y} , \bar{z} to achieve this. We start by solving the second equation in system (3.3.7)

$$\begin{aligned} g'(W)\bar{i}_1(W) + g(W)\frac{d\bar{i}_1(W)}{dW} &= -(\mu_2(W) + \lambda)\bar{i}_1(W) \\ \frac{d\bar{i}_1(W)}{dW} &= -\frac{(\mu_2(W) + \lambda + g'(W))}{g(W)}\bar{i}_1(W) \\ \bar{i}_1(W) &= \bar{i}_1(0)e^{-\int_0^W \frac{\mu_2(\tau)+\lambda}{g(\tau)}d\tau} e^{-\int_0^W \frac{g'(\tau)}{g(\tau)}d\tau} \\ \bar{i}_1(W) &= \frac{\bar{i}_1(0)g(0)}{g(W)}e^{-\int_0^W \frac{\mu_2(\tau)+\lambda}{g(\tau)}d\tau}. \end{aligned}$$

Substituting $\bar{i}_1(W)$ in the the fifth equation of system (3.3.7) enables us to express \bar{z} as

$$\bar{z} = \frac{\bar{i}_1(0)g(0)}{\lambda + \sigma} \int_0^{W_0} \xi(W) \frac{P(W)}{g(W)} e^{-\int_{0th e}^W \frac{\mu_2(\tau)+\lambda}{g(\tau)}d\tau} dW.$$

We take the expression of \bar{z} and $\bar{i}_1(W)$ and substitute it into the third equation of system (3.3.7) to get

$$\begin{aligned} g(0)\bar{i}_1(0) &= S_0^*\bar{i}_1(0)g(0) \left[\beta_h \int_0^{W_0} \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)+\lambda}{g(\tau)}d\tau} dW \right. \\ &\quad \left. + \frac{\beta_e}{\lambda + \sigma} \int_0^{W_0} \xi(W) \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)+\lambda}{g(\tau)}d\tau} dW \right]. \end{aligned} \quad (3.3.8)$$

Respectively, we obtain the characteristic equation for λ : $G(\lambda) = 1$ with

$$G(\lambda) = S_0^* \left[\int_0^{W_0} \beta_h \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)+\lambda}{g(\tau)}d\tau} dW + \frac{\beta_e}{\lambda + \sigma} \int_0^{W_0} \xi(W) \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)+\lambda}{g(\tau)}d\tau} dW \right]. \quad (3.3.9)$$

We then check the roots of the characteristic equation to deduce stability. The DFE is stable if the roots of the characteristic equation have negative real parts and it's unstable otherwise.

A non-zero solution to eq. (3.3.8) exists only if there is a number $\lambda \in \mathbb{R}$ such that $G(\lambda) = 1$. Differentiating equation (3.3.9) with respect to λ yields $G'(\lambda) < 0$ and thus $G(\lambda)$ is a strictly decreasing function, additionally, $\lim_{\lambda \rightarrow \infty} G(\lambda) = 0$. If $\hat{\lambda}$ is a unique real solution of eq. (3.3.9) then $\hat{\lambda} > 0$ provided $G(0) > 1$ and $\hat{\lambda} < 0$ provided $G(0) < 1$.

We can let $H = S_0^*\beta_h \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)}d\tau}$ and $J = S_0^*\beta_e \sigma(W) \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)}d\tau}$ such that

$$G(\lambda) = \int_0^{W_0} H e^{-\int_0^W \frac{\lambda}{g(\tau)}d\tau} dW + \frac{1}{\lambda + \sigma} \int_0^{W_0} J e^{-\int_0^W \frac{\lambda}{g(\tau)}d\tau} dW.$$

Suppose $G(0) < 1$ and $\lambda = a \pm bi$ is a complex solution to equation (3.3.9) with $a \geq 0$. Then,

$$\begin{aligned}
 |G(\lambda)| &= \left| \int_0^\infty H e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW + \frac{1}{\lambda + \sigma} \int_0^\infty J e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW \right| \\
 &= \left| \int_0^\infty H e^{-\int_0^W \frac{a \pm bi}{g(\tau)} d\tau} dW + \frac{1}{a \pm bi + \sigma} \int_0^\infty J e^{-\int_0^W \frac{a \pm bi}{g(\tau)} d\tau} dW \right| \\
 &\leq \int_0^\infty H \left| e^{-\int_0^W \frac{a \pm bi}{g(\tau)} d\tau} \right| dW + \frac{1}{|a \pm bi + \sigma|} \int_0^\infty J \left| e^{-\int_0^W \frac{a \pm bi}{g(\tau)} d\tau} \right| dW \\
 &= \int_0^\infty H \left| e^{-\int_0^W \frac{a \pm bi}{g(\tau)} d\tau} \right| dW + \frac{1}{\sqrt{(a + \sigma)^2 + b^2}} \int_0^\infty J \left| e^{-\int_0^W \frac{a \pm bi}{g(\tau)} d\tau} \right| dW \\
 &\leq \int_0^{W_0} H e^{-\int_0^W \frac{a}{g(\tau)} d\tau} dW + \frac{1}{(a + \sigma)} \int_0^{W_0} J e^{-\int_0^W \frac{a}{g(\tau)} d\tau} dW = G(a) \leq G(0) < 1.
 \end{aligned}$$

It follows then that equation (3.3.9) has a complex solution $\lambda = a \pm ib$ if $a < 0$, that is, every solution of eq. (3.3.9) must have a negative real part. We consider $G(0) = 1$ to be the threshold condition for the stability of the disease-free equilibrium. According to [28] this can be defined as the basic reproduction number, that is $G(0) = \mathcal{R}_0$ where

$$\mathcal{R}_0 = \frac{r\beta_h}{\mu_1} \int_0^{W_0} \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} dW + \frac{r\beta_e}{\mu_1} \int_0^{W_0} \frac{\xi(W)}{\sigma} \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} dW.$$

By Theorem 3.3.1 the disease-free equilibrium is locally asymptotically stable. \square

We can interpret the basic reproduction number to be the total infectivity given by

$$\mathcal{R}_0 = \mathcal{R}_d + \mathcal{R}_i$$

$$\mathcal{R}_d = \frac{r\beta_h}{\mu_1} \int_0^{W_0} \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} dW, \quad \mathcal{R}_i = \frac{r\beta_e}{\mu_1} \int_0^{W_0} \frac{\xi(W)}{\sigma} \frac{P(W)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau} dW.$$

\mathcal{R}_d represents the new infections occurring due to direct contact with an infected individual while \mathcal{R}_i are the infections resulting from consumption of contaminated water containing bacteria shed to the environment by infected individuals.

3.3.3 Existence of the Endemic Equilibrium

Proposition 3.3.2. *A unique positive endemic equilibrium of system (3.3.1) given by $\mathcal{E}^* = (S^*, i^*(W), V^*, B^*)$ exists if $\mathcal{R}_0 > 1$.*

Proof. The endemic equilibrium satisfies the following equations

$$\begin{aligned}
 0 &= r - \mu_1 S^* - S^* \int_0^{W_0} \beta_h P(W) i^*(W) dW - \beta_e S^* B^* + \rho V^* \\
 \partial_W(g(W) i^*(W)) &= -\mu_2(W) i^*(W) \\
 g(0) i^*(0) &= S^* \int_0^{W_0} \beta_h P(W) i^*(W) dW + \beta_e S^* B^* \\
 0 &= g(W_0) i^*(W_0) - \rho V^* - \mu_3 V^* \\
 0 &= \int_0^{W_0} \xi(W) P(W) i^*(W) dW - \sigma B^*.
 \end{aligned} \tag{3.3.10}$$

Solving the second equation of system (3.3.10) yields $i^*(W) = \frac{i^*(0)g(0)}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau}$.

We let $\pi(W) = \frac{1}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau}$ and rewrite

$$i^*(W) = i^*(0)g(0)\pi(W).$$

Substituting $i^*(W)$ to the fourth and fifth equations of system (3.3.10) gives

$$B^* = i^*(0)g(0) \int_0^{W_0} \frac{\xi(W)}{\sigma} P(W) \pi(W) dW, \quad V^* = \frac{g(W_0) i^*(0) g(0) \pi(W_0)}{(\rho + \mu_3)}.$$

Plugging $i^*(W)$ and B^* into the boundary equation of system (3.3.10) yields

$$S^* = \frac{1}{\int_0^{W_0} \beta_h P(W) \pi(W) dW + \int_0^{W_0} \beta_e \frac{\xi(W)}{\sigma} P(W) \pi(W) dW}.$$

To find the value of $i^*(0)$, we use the first equation of system (3.3.10), that can be rewritten as

$$r - \mu_1 S^* - g(0) i^*(0) + \rho V^* = 0. \tag{3.3.11}$$

Notice that S^* can be expressed in terms of \mathcal{R}_0 , that is $S^* = \frac{r}{\mu_1 \mathcal{R}_0}$. Substituting S^* and V^* in eq. (3.3.11) yields

$$r - \frac{r}{\mathcal{R}_0} - g(0) i^*(0) + i^*(0) g(0) \frac{\rho g(W_0) \pi(W_0)}{(\rho + \mu_3)} = 0.$$

Thus, $i^*(0) = \frac{r(1 - \frac{1}{\mathcal{R}_0})}{g(0)(1 - \frac{\rho g(W_0) \pi(W_0)}{(\rho + \mu_3)})}$. We can then express

$$i^*(W) = \frac{r(1 - \frac{1}{\mathcal{R}_0})}{(1 - \frac{\rho g(W_0) \pi(W_0)}{(\rho + \mu_3)})} \pi(W). \tag{3.3.12}$$

Since $\frac{\rho g(W_0) \pi(W_0)}{(\rho + \mu_3)} = \frac{\rho e^{-\int_0^{W_0} \frac{\mu_2(\tau)}{g(\tau)} d\tau}}{(\rho + \mu_3)} < 1$, the denominator in eq. (3.3.12) is positive. $i^*(W)$ is only positive if $\mathcal{R}_0 > 1$, thus, the endemic equilibrium $\mathcal{E}^* = (S^*, i^*(W), V^*, B^*)$ exists only if $\mathcal{R}_0 > 1$. \square

3.3.4 Local Stability of the Endemic Equilibrium

We assume that $\mathcal{R}_0 > 1$ and linearize system (3.3.1) around the endemic equilibrium. We start by introducing perturbation terms, that is let $S(t) = S^* + x(t)$, $i(t, W) = i^*(W) + i_1(t, W)$, $V = V^* + y(t)$ and $B = B^* + z(t)$. We substitute these expressions in system (3.3.1), simplify using the set of equations in system (3.3.10) and neglect quadratic perturbation terms to get the linearized system

$$\begin{aligned}
 \frac{dx(t)}{dt} &= -\mu_1 x(t) - S^* \int_0^{W_0} \beta_h P(W) i_1(t, W) dW - x(t) \int_0^{W_0} \beta_h P(W) i^*(W) d(W) \\
 &\quad - \beta_e S^* z(t) - \beta_e B^* x(t) + \rho y(t) \\
 \partial_t i_1(t, W) + \partial_W (g(W) i_1(t, W)) &= -\mu_2(W) i_1(t, W) \\
 g(0) i_1(t, 0) &= S^* \int_0^{W_0} \beta_h P(W) i_1(t, W) dW + x(t) \int_0^{W_0} \beta_h P(W) i^*(W) d(W) \\
 &\quad + \beta_e S^* z(t) + \beta_e B^* x(t) \\
 \frac{dy(t)}{dt} &= g(W_0) i_1(t, W_0) - \rho y(t) - \mu_3 y(t) \\
 \frac{dz(t)}{dt} &= \int_0^{W_0} \xi(W) P(W) i_1(t, W) dW - \sigma z(t).
 \end{aligned} \tag{3.3.13}$$

We then look for solutions of the form $x(t) = x e^{\lambda t}$, $i_1(t, w) = i_1(W) e^{\lambda t}$, $y(t) = y e^{\lambda t}$, $z(t) = z e^{\lambda t}$. Substituting the appropriate formulation in system (3.3.13) yields the eigenvalue problem

$$\begin{aligned}
 \lambda x &= -\mu_1 x - S^* \int_0^{W_0} \beta_h P(W) i_1(W) dW - x \int_0^{W_0} \beta_h P(W) i^*(W) d(W) \\
 &\quad - \beta_e S^* z - \beta_e B^* x + \rho y \\
 \partial_W (g(W) i_1(W)) &= -(\mu_2(W) + \lambda) i_1(W) \\
 g(0) i_1(0) &= S^* \int_0^{W_0} \beta_h P(W) i_1(W) dW + x \int_0^{W_0} \beta_h P(W) i^*(W) d(W) \\
 &\quad + \beta_e S^* z + \beta_e B^* x \\
 \lambda y &= g(W_0) i_1(W_0) - \rho y - \mu_3 y \\
 \lambda z &= \int_0^{W_0} \xi(W) P(W) i_1(W) dW - \sigma z.
 \end{aligned} \tag{3.3.14}$$

We solve for x, y and z in system (3.3.14) to get

$$\begin{aligned}
 x &= \frac{\rho g(W_0) i_1(W_0) - g(0) i_1(0) (\lambda + \rho + \mu_3)}{(\lambda + \rho + \mu_3) (\lambda + \mu_1)}, & y &= \frac{g(W_0) i_1(W_0)}{\lambda + \rho + \mu_3}, \\
 z &= \frac{1}{\lambda + \sigma} \int_0^{W_0} \xi(W) P(W) i_1(W) dW.
 \end{aligned}$$

The solution to the PDE in system (3.3.14) yields

$$i_1(W) = i_1(0) g(0) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau}, \quad \pi_1(W) = \frac{1}{g(W)} e^{-\int_0^W \frac{\mu_2(\tau)}{g(\tau)} d\tau}.$$

Denoting $K = \int_0^{W_0} \beta_h P(W) i^*(W) d(W)$ and substituting x, y, z and $i_1(W)$ in the third equation of system (3.3.14) gives the characteristic equation

$$1 = S^* \int_0^{W_0} \beta_h P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW + \frac{\beta_e S^*}{\lambda + \sigma} \int_0^{W_0} \xi(W) P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW \\ + \frac{(K + \beta_e B^*) (\rho g(W_0) \pi_1(W_0) e^{-\int_0^{W_0} \frac{\lambda}{g(\tau)} d\tau})}{(\lambda + \rho + \mu_3)(\lambda + \mu_1)} - \frac{K + \beta_e B^*}{\lambda + \mu_1}.$$

We can rewrite this equation as

$$\frac{\lambda + \mu_1 + K + \beta_e B^*}{\lambda + \mu_1} = S^* \int_0^{W_0} \beta_h P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW \\ + \frac{\beta_e S^*}{\lambda + \sigma} \int_0^{W_0} \xi(W) P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW \\ + \frac{(K + \beta_e B^*) (\rho g(W_0) \pi_1(W_0) e^{-\int_0^{W_0} \frac{\lambda}{g(\tau)} d\tau})}{(\lambda + \rho + \mu_3)(\lambda + \mu_1)}. \quad (3.3.15)$$

Studies have shown that a cholera infection confers immunity against subsequent infection [87]. We consider the case of a single cholera epidemic, where most individuals become immune such that the loss of immunity is negligible. We show that the endemic equilibrium is asymptotically stable.

Theorem 3.3.3. *Given no loss of immunity ($\rho = 0$), the endemic equilibrium is locally asymptotically stable if $\mathcal{R}_0 > 1$.*

Proof. Considering no loss of immunity, the characteristic equation (3.3.15) reduces to

$$\frac{\lambda + \mu_1 + K + \beta_e B^*}{\lambda + \mu_1} = S^* \int_0^{W_0} \beta_h P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW \\ + \frac{\beta_e S^*}{\lambda + \sigma} \int_0^{W_0} \xi(W) P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW. \quad (3.3.16)$$

If we let $\lambda = a + ib$ and assume that $a \geq 0$, then for $\Re(\lambda) \geq 0$ the left hand side of equation (3.3.16) gives

$$\left| \frac{\lambda + \mu_1 + K + \beta_e B^*}{\lambda + \mu_1} \right| = \frac{\sqrt{(a + \mu_1 + K + \beta_e B^*)^2 + b^2}}{\sqrt{(a + \mu_1)^2 + b^2}} > 1,$$

while for $a \geq 0$ the right hand side yields

$$\left| S^* \int_0^{W_0} \beta_h P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW + \frac{\beta_e S^*}{\lambda + \sigma} \int_0^{W_0} \xi(W) P(W) \pi_1(W) e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} dW \right| \\ \leq S^* \int_0^{W_0} \beta_h P(W) \pi_1(W) \left| e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} \right| dW + \frac{\beta_e S^*}{|\lambda + \sigma|} \int_0^{W_0} \xi(W) P(W) \pi_1(W) \left| e^{-\int_0^W \frac{\lambda}{g(\tau)} d\tau} \right| dW \\ \leq S^* \int_0^{W_0} \beta_h P(W) \pi_1(W) e^{-\int_0^W \frac{a}{g(\tau)} d\tau} dW + \frac{\beta_e S^*}{\sigma} \int_0^{W_0} \xi(W) P(W) \pi_1(W) e^{-\int_0^W \frac{a}{g(\tau)} d\tau} dW \\ \leq S^* \int_0^{W_0} \beta_h P(W) \pi_1(W) dW + \frac{\beta_e S^*}{\sigma} \int_0^{W_0} \xi(W) P(W) \pi_1(W) dW = \frac{S^* \mathcal{R}_0 \mu_1}{r} = 1.$$

Thus, given λ with $\Re(\lambda) \geq 0$, the left side of equation (3.3.16) is strictly greater than one while the right side of equation (3.3.16) is strictly less than one, which is a contradiction. Therefore, any λ with non-negative real parts can not satisfy the characteristic equation. By Theorem 3.3.1 the endemic equilibrium is locally asymptotically stable. \square

3.4 Modified Between-host Model

In this section, we modify the between-host model (3.3.1) by assuming permanent immunity and assigning an age structure to the bacteria in the environment, that is, $B(t) = \int_0^{\bar{a}} b(t, a) da$. We consider this to be more realistic and biologically meaningful. See Remark 3.4.1 below. We use the methods from Meehan et al. [76] to check for the global stability of the equilibrium points of the modified model and give conditions in which the modified model is equivalent to the original model. The modified model is described by

$$\begin{aligned}
 \frac{dS(t)}{dt} &= r - \mu_1 S(t) - S(t)F(t) \\
 \partial_t i(t, W) + \partial_W(g(W)i(t, W)) &= -\mu_2(W)i(t, W) \\
 g(0)i(t, 0) &= S(t)F(t) \\
 (\partial_t + \partial_a)b(t, a) &= -\sigma b(t, a) \\
 b(t, 0) &= \int_0^{W_0} \xi(W)P(W)i(t, W)dW \\
 S(0) = S_0 \quad V(0) = V_0, \quad b(0, a) &= \beta(a), \quad i(0, W) = \Phi(W),
 \end{aligned} \tag{3.4.1}$$

where the force of infection $F(t) = \int_0^{W_0} \beta_h P(W)i(t, W)dW + \int_0^{\bar{a}} \beta_e(a)b(t, a)da$.

Assumption 3.4.1. *The rate of environmental transmission of bacteria $\beta_e(a)$ has compact support for $a \in [0, \bar{a}]$ such that $\beta_e(a) \geq 0$ for $a \in [0, \bar{a}]$ and zero elsewhere.*

Remark 3.4.1. *Assumption 3.4.1 is based on the premise that bacteria shed into the environment are in a hyperinfectious state and only decay into a less infectious state over time [45]. Therefore, bacteria that contribute to the infection process are considered to be young. If we assign a maximal age \bar{a} to the bacteria, we expect that bacteria with age $\bar{a} > 0$ will have no contribution to the infection process.*

3.4.1 Global Stability of Equilibrium Points

With Assumption 3.4.1 in mind, we turn to check the global stability of the equilibrium points of the model (3.4.1). We follow the approach used by Meehan et al. [76] for systems of renewal equations. First, we give a summary of their work.

Meehan et al. [76] describe a susceptible class experiencing a force of infection $F(t)$ as

$$\begin{aligned}\frac{dS(t)}{dt} &= \lambda - \mu_1 S(t) - S(t)F(t) \\ F(t) &= \int_0^{\bar{\tau}} A(\tau)S(t-\tau)F(t-\tau)d\tau,\end{aligned}\tag{3.4.2}$$

where $A(\tau)$ is the contribution of individuals infected for time τ to the force of infection and the infectivity kernel $A \geq 0$. Two assumptions are prescribed to the system, that is, the maximal age of infection $\bar{\tau} < \infty$, and the infection confers permanent immunity. Using Lyapunov functionals, they conclude from the integral form, with compact support of the integral kernel, the global stability of the DFE when $\mathcal{R}_0 \leq 1$ and the global stability of the endemic equilibrium when $\mathcal{R}_0 > 1$. We state the associated theorems below.

Theorem 3.4.1. [76, Theorem 1] *The infection-free equilibrium point P_0 of the system (3.4.2) is globally asymptotically stable in an infinite dimensional phase space Ω for $\mathcal{R}_0 \leq 1$. However, if $\mathcal{R}_0 > 1$, solutions of (3.4.2) starting sufficiently close to P_0 in Ω leave a neighborhood of P_0 , except those starting within the boundary region $\partial\Omega$ which approach P_0 .*

Theorem 3.4.2. [76, Theorem 2] *If $\mathcal{R}_0 > 1$, there is a unique endemic equilibrium \bar{P} which is globally asymptotically stable in Ω (for initial conditions away from the boundary region $\partial\Omega$).*

We aim to reformulate system (3.4.1) in terms of these results to establish global stability.

Theorem 3.4.3. *Given Assumption 3.4.1, the infection-free equilibrium and endemic equilibrium of the system (3.4.1) are globally asymptotically stable when $\mathcal{R}_0 \leq 1$ and $\mathcal{R}_0 > 1$ respectively.*

Proof. We aim to rewrite $F(t)$ as $F(t) = \int_0^{\bar{\tau}} A(\tau)S(t-\tau)F(t-\tau)d\tau$. Similar to what is done in Proposition 3.3.1, we can define

$$i(t, W) = \begin{cases} S(t - G(W))F(t - G(W))\frac{1}{g(W)}e^{-\int_0^W \frac{\mu_2(W')}{g(W')}dW'} & W \leq W_0 \\ 0 & W > W_0. \end{cases}\tag{3.4.3}$$

Using the method of characteristics we solve the boundary value problem for $b(t, a)$. That is, the characteristic curves are given by

$$\frac{dt}{ds} = 1, \quad \frac{da}{ds} = 1, \quad \frac{dz}{ds} = -\sigma z.$$

Solving the curves with the boundary conditions gives,

$$\begin{aligned}t &= s + t_0, & a &= s, & z &= b(t_0, 0)e^{-\sigma s} \\ b(t, a) &= b(t - a, 0)e^{-\sigma a} \\ &= e^{-\sigma a} \int_0^{W_0} \xi(W)P(W)i(t - a, W)dW.\end{aligned}\tag{3.4.4}$$

We substitute the value of i from equation (3.4.3) in $b(t, a)$ to obtain

$$b(t, a) = e^{-\sigma a} \int_0^{W_0} \xi(W)P(W)S(t-a-G(W))F(t-a-G(W))\frac{1}{g(W)}e^{-\int_0^W \frac{\mu_2(W')}{g(W')}dW'}dW.$$

To find the total bacterial concentration, we integrate $b(t, a)$ over a and obtain the integral equation

$$B(t) = \int_0^{\bar{a}} b(t, a)da = \int_0^{\bar{a}} e^{-\sigma a} \int_0^{W_0} \frac{\xi(W)P(W)}{g(W)}S(t-a-G(W))F(t-a-G(W)) \cdot e^{-\int_0^W \frac{\mu_2(W')}{g(W')}dW'}dWda.$$

We then find the bacterial contribution to the force of infection. Recall that $G(W) = \int \frac{dW}{g(W)}$. If we let $a + G(W) = \theta$ then $d\theta = \frac{1}{g(W)}dW$ and

$$\begin{aligned} \int_0^{\bar{a}} \beta_e(a)b(t, a)da &= \int_0^{\bar{a}} \beta_e(a)e^{-\sigma a} \int_{a+G(0)}^{a+G(W_0)} \xi(G^{-1}(\theta-a))P(G^{-1}(\theta-a))S(t-\theta) \\ &\quad \cdot F(t-\theta)e^{-\int_0^{G^{-1}(\theta-a)} \frac{\mu_2(W')}{g(W')}dW'}d\theta da, \\ &= \int_{G(0)}^{\bar{a}+G(W_0)} K_e(\theta)S(t-\theta)F(t-\theta)d\theta. \end{aligned}$$

with (see Figure 3.4)

$$K_e(\theta) = \begin{cases} \int_0^{\theta-G(0)} \beta_e(a)e^{-\sigma a}\xi(G^{-1}(\theta-a))P(G^{-1}(\theta-a))e^{-\int_0^{G^{-1}(\theta-a)} \frac{\mu_2(W')}{g(W')}dW'}da & \text{for } \theta \in (G(0), G(W_0)) \\ \int_{\theta-G(W_0)}^{\theta-G(0)} \beta_e(a)e^{-\sigma a}\xi(G^{-1}(\theta-a))P(G^{-1}(\theta-a))e^{-\int_0^{G^{-1}(\theta-a)} \frac{\mu_2(W')}{g(W')}dW'}da & \text{for } \theta \in (G(W_0), \bar{a}+G(0)) \\ \int_{\theta-G(W_0)}^{\bar{a}} \beta_e(a)e^{-\sigma a}\xi(G^{-1}(\theta-a))P(G^{-1}(\theta-a))e^{-\int_0^{G^{-1}(\theta-a)} \frac{\mu_2(W')}{g(W')}dW'}da & \text{for } \theta \in (\bar{a}+G(0), \bar{a}+G(W_0)). \end{cases}$$

Now that we have the force of infection from indirect transmission, we evaluate the same for the case of direct transmission. This yields

$$\int_0^{W_0} \beta_h P(W)i(t, W)dW = \int_0^{W_0} \beta_h P(W)S(t-G(W))F(t-G(W))\frac{1}{g(W)}e^{-\int_0^W \frac{\mu_2(W')}{g(W')}dW'}dW.$$

If we let $\tau = G(W)$

$$\int_0^{W_0} \beta_h P(W)i(t, W)dW = \int_{G(0)}^{G(W_0)} K_h(\tau)S(t-\tau)F(t-\tau)d\tau$$

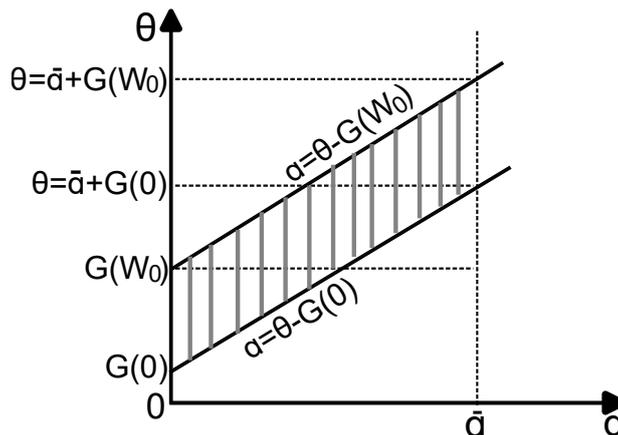


Figure 3.4: Change of the order of integration.

with

$$K_h(\tau) = \beta_h P(G^{-1}(\tau)) e^{-\int_0^{G^{-1}(\tau)} \frac{\mu_2(W')}{g(W')} dW'}.$$

We can then write the total force of infection as

$$\begin{aligned} F(t) &= \int_0^{W_0} \beta_h P(W) i(t, W) dW + \int_0^\infty \beta_e(a) b(t, a) da \\ &= \int_{G(0)}^{G(W_0)} K_h(\tau) S(t - \tau) F(t - \tau) d\tau + \int_{G(0)}^{\bar{a} + G(W_0)} K_e(\theta) S(t - \theta) F(t - \theta) d\theta \\ &= \int_{G(0)}^{\bar{a} + G(W_0)} A(W) S(t - W) F(t - W) dW, \end{aligned}$$

where

$$A(W) = \begin{cases} K_h(W) + K_e(W) & W \leq G(W_0) \\ K_e(W) & G(W_0) < W \leq \bar{a} + G(W_0). \end{cases}$$

The susceptible class of system (3.4.1) is now described by the system

$$\begin{aligned} \frac{dS(t)}{dt} &= r - \mu_1 S(t) - S(t) F(t) \\ F(t) &= \int_0^{W_0} A(W) S(t - W) F(t - W) dW. \end{aligned}$$

From the results of Meehan et al. [76], the disease-free equilibrium and the endemic equilibrium of the above system are globally asymptotically stable when $\mathcal{R}_0 \leq 1$ and $\mathcal{R}_0 > 1$ respectively. \square

We next check the equivalence of the two models.

Proposition 3.4.1. *The original model (3.3.1) is equivalent to the modified model (3.4.1) whenever $\bar{a} \rightarrow \infty$.*

Proof. Recall that, $B(t) = \int_0^{\bar{a}} b(t, a) da$. From equation (3.4.4) we find $b(t, a) = b(t - a, 0)e^{-\sigma a}$. We define

$$\varphi(a) = \begin{cases} e^{-\sigma a} & \text{for } a \in [0, \bar{a}] \\ 0 & \text{for } a > \bar{a} \end{cases}$$

such that $b(t, a) = b(t - a, 0)\varphi(a)$. For simplicity, we write $b(t - a, 0) = b(t - a)$ then,

$$\frac{dB(t)}{dt} = \frac{d}{dt} \int_0^{\bar{a}} b(t - a)\varphi(a) da = - \int_0^{\bar{a}} \left(\frac{d}{da} b(t - a) \right) \varphi(a) da.$$

Integrating by parts yields

$$\begin{aligned} \frac{dB(t)}{dt} &= -\varphi(a)b(t - a) \Big|_0^{\bar{a}} + \int_0^{\bar{a}} \varphi'(a)b(t - a) da \\ &= -\varphi(\bar{a})b(t - \bar{a}) + b(t) - \sigma \int_0^{\bar{a}} \varphi(a)b(t - a) da = b(t) - \sigma B(t) - \varphi(\bar{a})b(t - \bar{a}) \\ &= \int_0^{W_0} \xi(W)P(W)i(t, W) dW - \sigma B(t) - \varphi(\bar{a})b(t - \bar{a}). \end{aligned} \quad (3.4.5)$$

For $\bar{a} \rightarrow \infty$

$$\frac{dB(t)}{dt} = \int_0^{W_0} \xi(W)P(W)i(t, W) dW - \sigma B(t). \quad (3.4.6)$$

□

Brauer et al. [11] has examined the global stability of systems similar to (3.4.1) for the case $g(W) = 1$. Although the case of waning immunity has not been considered, analysis of models with waning immunity can be viewed in Barbarossa and Röst [9], Nakata et al. [86].

3.5 Summary

We have presented a within-host model in which pathogen growth is denoted by a cubic term. We have been able to characterize a single infected individual by the state of his immune response. We find that, unlike other within-host cholera models, our modeling approach allows for the possibility of recovery as the pathogen is cleared from the body after a finite time. Subsequently, we have also found that a minimum pathogen load is required to activate an immune response reinforcing the ideas behind the critical infectious dose needed for infection.

The results from the within-host analysis have allowed us to structure the epidemic model on immune properties. The resulting epidemic model has an advantage over standard SIR-type models in that immunological variables can determine the ability of the disease to

spread. For instance, a sustained immune response within the host could alter the time course of an infection in the population. Therefore, disease spread is not just a consequence of epidemiological factors but also of immunological factors. Using the basic reproduction number of the model, we can estimate the contribution of bacteria from the environment and human-human contribution to the infection process. We have shown that for the DFE, the disease will be eradicated if $\mathcal{R}_0 < 1$ and persist otherwise. We have considered the case where there is no loss of immunity and outlined the conditions for the local stability of the endemic equilibrium. Further, for the modified model we have shown global stability for both the disease-free and endemic equilibrium.

In the study, one infectious contact is assumed to bring the contacts over the threshold. This is not the case when an individual has a sub-critical pathogen load. We will consider the sub-critical case in the next chapter of the thesis.

Chapter 4

Multi-scale to SIR model using First Principles

In this chapter, we take up recent ideas on multi-scale modeling and return to first principles to derive a refined model of the within and between-host dynamics of cholera. The aim is to introduce a model structure that is closer to the immunological and environmental processes associated with the disease. We also aim to provide a rigorous way for the analysis of such a model structure. We formulate the model by explicitly addressing the pathogen level in susceptible individuals. This sounds contradictory at first glance since susceptible individuals are usually thought to be pathogen-free. However, in some instances, the consumption of contaminated food and water does not immediately trigger infection. This is because cholera infection only occurs when the pathogen load of an individual exceeds a critical threshold [87, 95] (which varies from individual to individual, e.g., caused by differences in the gut microbiome [1, 22]). That is, individuals take up *V. cholerae*, but the innate immune system fights off and eliminates the bacteria as long as the critical pathogen threshold is not exceeded. Furthermore, laboratory experiments on mice show that a moderate increase in the infectious dose leads to an increase in the pathogen burden over time (with a time scale of approximately 12 h) [40], and not to infection. We are thus led to a model, where the susceptible are structured by their pathogen load. The transition into the infected class happens at a rate dependent on the pathogen load. This step allows for a better connection between the within- and between-host dynamics, as the somewhat arbitrary distinction between a susceptible person and an infected person is softened.

Mathematically, the model structure resembles that of fragmentation–aggregation equations: we have a time-continuous process of pathogen clearance (we assume that pathogens are degraded at a constant rate within a host), and at randomly distributed times, by food uptake, we have booster events that instantaneously increase the pathogen load. Accordingly, the analysis of the invariant distribution for the pathogen load follows the theory developed for aggregation–fragmentation equations, particularly for cell division, developed in the 1980s by Heijmans, Gyllenberg, and others [46, 47], which was later extended, e.g.,

by Doumic and others [31, 17, 97, 98]. The understanding of the pathogen load's invariant distribution and the assumption of a time scale separation then allows the reduction of the rather complex multi-scale model to a population-level only model, where the incidence term assumes a form derived from the multi-scale model. We then analyze the resultant model numerically. This chapter is based on the work published in [84].

4.1 Model Description

We subdivide the human population into three compartments that represent susceptible (S), infected (I), and recovered (R) individuals. Additionally, we have a compartment for the bacterial concentration in the environment B . The pathogen (*Vibrio cholerae*) can survive and reproduce in the environment without interaction with human hosts. We simply assume an environmental pathogen production rate a and a pathogen death rate σ . As a detailed model of the pathogen's life cycle is out of the scope of the present work, we focus on the human population.

We are interested in a model connecting the within-host level with the population level. Therefore, we assign each susceptible person a pathogen load P . Although transmission can take place directly through person-to-person contact, we only consider the case of indirect transmission since direct transmission has been shown to be rare [94]. That is, we consider that contact with contaminated food (booster event) increases the pathogen load of an individual at the rate ψB . We assume that food uptake does not take place completely at random, that is, there is some minimal and maximal time between uptake events, and as such, we also take the time since the last booster event τ into account. We can now describe the susceptible class $S = S(t, \tau, P)$ such that the class is structured according to the pathogen load and the time since the last uptake event. It becomes apparent that this structure is not only more realistic but also mathematically convenient. We model the timing of booster events by the rate $\rho(\tau)$, which we will discuss below in more detail.

It is well known that a critical pathogen load is required for a cholera infection to occur [80, 87]. Under sub-critical pathogen levels, the innate immune system is able to control the pathogens. We take that into account by defining a clearance rate γ for the within-host pathogens. To be precise, within a single susceptible individual, we have an interplay between the booster event and the degradation of pathogens, which establishes a stochastic process as shown in Figure 4.1. Experiments done on mice with a moderate pathogen level show the exact kind of decline in the pathogen load addressed by this sub-model for the susceptible [40]. Furthermore, the infection rate β is a function depending on the pathogen load, $\beta = \beta(P)$. Here, $\beta(P) = 0$ if the pathogen load is sub-critical. β only becomes positive if the pathogen load P exceeds some value and persons with such a high pathogen load are transferred to the infected class. An infected individual sheds pathogens into the environment at rate ξ and recovers at the rate α . We now define some assumptions for the model.

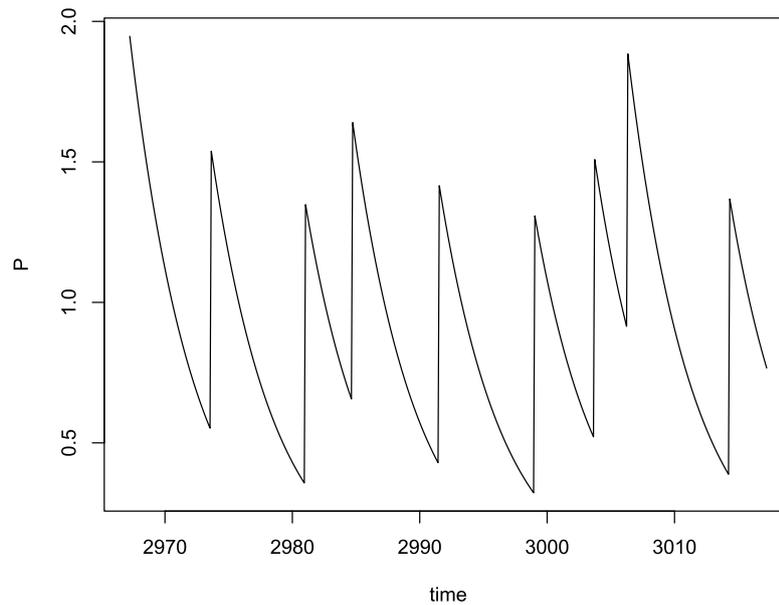


Figure 4.1: Time course of the pathogen load within a single susceptible individual: booster events (uptake of pathogens) instantaneously increase the load, in-between two booster events the immune system leads to an exponential decrease in pathogens. The time between two booster events is given by i.i.d. random variable T . $\psi B = 1$ is taken constant here, the initial time span is dismissed as a burn-in phase s.t. time runs here from 1970 to 3010.

- *Recovered individuals become immune and stay immune.* The time scale of our model covers months, which corresponds to the time scale of one cholera epidemic, and not years, where recovered individuals become susceptible once again.
- *The dynamics at the population level are slow while the within-individual pathogen dynamics are fast.* A typical cholera epidemic has a time scale of weeks or months, while the time between two booster events is rather hours, and accordingly, the time scale of the native immune event to handle the ingested pathogens is the same as the time scale for the uptake events (otherwise the pathogens would accumulate and eventually an infection would be inevitable).

From the second assumption, we can now distinguish the dynamics within the host and at the population level by considering time scales. The time scale parameter ϵ is used to express slow dynamics. Figure 4.2 depicts the movement between the different compartments.

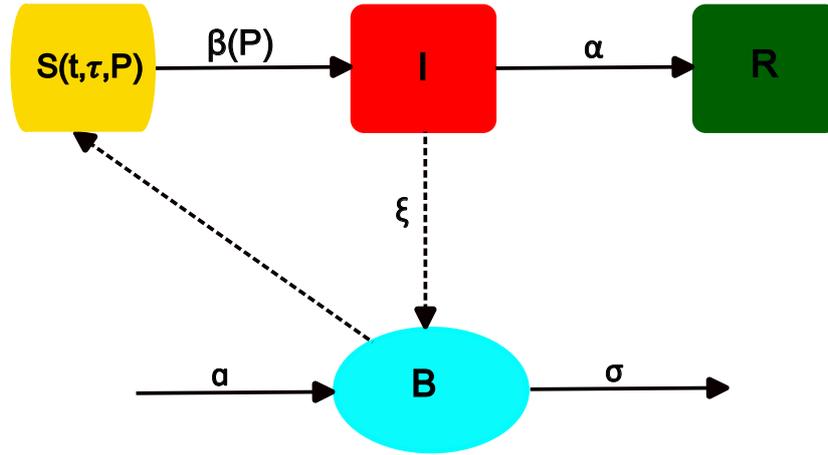


Figure 4.2: Model flow diagram at population level

All in all, the model reads

$$\begin{aligned}
 \partial_t S(t, \tau, P) + \partial_\tau S(t, \tau, P) + \partial_P(-\gamma P S(t, \tau, P)) &= -\rho(\tau) S(t, \tau, P) - \epsilon \beta(P) S(t, \tau, P) \\
 S(t, 0, P) &= \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau) S(t, \tau, P - \psi B) d\tau \\
 S(t, \tau, P) &= 0 \quad \text{for } P < 0 \\
 \dot{I} &= \int_0^\infty \int_{\check{\tau}}^{\hat{\tau}} \epsilon \beta(P) S(t, \tau, P) d\tau dP - \epsilon \alpha I \\
 \dot{R} &= \epsilon \alpha I \\
 \dot{B} &= \epsilon (a + \xi I - \sigma B) \\
 S(0, \tau, P) &= S_0(\tau, P), \quad I(0) = I_0, \quad R(0) = R_0, \quad B(0) = B_0.
 \end{aligned} \tag{4.1.1}$$

The boundary condition $S(t, 0, P)$ defines the pathogen density in the susceptible class right after a booster event. The time since the last booster event τ is zero since the booster event has just occurred, and the delay term $P - \psi B$ implies that the boosted individuals had a lower pathogen load before the jump.

Now that we have defined the model equations, we turn to discuss a central aspect, the timing of booster events coded by $\rho(\tau)$. The idea is that beneath our deterministic model, there is an underlying stochastic process describing the uptake of pathogens. We will later use this idea to formulate a numerical method to obtain a stationary solution to our model. The time between two booster events is i.i.d. as a random variable T with probability density $\varphi(t) \in C^1(\mathbb{R}_+)$. That is, $P(T < \tau) = \int_0^\tau \varphi(t) dt$. We assume that there is a minimal time $\check{\tau} > 0$ between two booster events, $P(T < \check{\tau}) = 0$, as well as a maximal time $\hat{\tau} > \check{\tau}$, $P(T < \hat{\tau}) = 1$. Therefore,

$$\varphi(\tau) = 0 \quad \text{for } \tau \in [0, \check{\tau}) \cup (\hat{\tau}, \infty), \quad \int_{\check{\tau}}^{\hat{\tau}} \varphi(\tau) d\tau = 1. \tag{4.1.2}$$

These assumptions are sensible given the application, and, as we shall see, are also convenient for the analysis of the model. We next show that $\rho(\tau)$ can be defined in terms of $\varphi(\tau)$.

Proposition 4.1.1. *The parameter $\rho(\tau)$ in the model (4.1.1) is the hazard rate defined by $\varphi(\tau)$, that is*

$$\rho(\tau) = \frac{\varphi(\tau)}{1 - \int_0^\tau \varphi(s) ds}.$$

Proof. Given $\rho(\tau)$ as the timing of a booster event, the time distribution of the booster event $\varphi(\tau)$ can be taken as the arrival time of the process, that is,

$$\varphi(\tau) = \frac{\rho(\tau)e^{-\int_0^\tau \rho(s) ds}}{\int_0^\infty \rho(\tau')e^{-\int_0^{\tau'} \rho(s) ds} d\tau'}.$$

We aim at an expression of $\rho(\tau)$ in terms of $\varphi(\tau)$. We solve the denominator to get

$$\int_0^\infty \rho(\tau')e^{-\int_0^{\tau'} \rho(s) ds} d\tau' = - \int_0^\infty \frac{d}{d\tau'} e^{-\int_0^{\tau'} \rho(s) ds} d\tau' = 1 - e^{-\int_0^\infty \rho(s) ds} = c$$

such that $\varphi(\tau) = \frac{\rho(\tau)e^{-\int_0^\tau \rho(s) ds}}{c}$. Taking $c = 1$

$$\varphi(\tau) = \rho(\tau)e^{-\int_0^\tau \rho(s) ds}.$$

Integrating both sides

$$\begin{aligned} \int_0^\tau \varphi(s) ds &= \int_0^\tau -\frac{d}{ds} e^{-\int_0^s \rho(\eta) d\eta} ds = 1 - e^{-\int_0^\tau \rho(\eta) d\eta} \\ 1 - \int_0^\tau \varphi(s) ds &= e^{-\int_0^\tau \rho(\eta) d\eta}. \end{aligned}$$

We can rewrite the equation as

$$\int_0^\tau \rho(\eta) d\eta = -\ln\left(1 - \int_0^\tau \varphi(s) ds\right).$$

Differentiating both sides gives

$$\rho(\tau) = \frac{\varphi(\tau)}{1 - \int_0^\tau \varphi(s) ds}.$$

Thus the parameter $\rho(\tau)$ in system (4.1.1) is the hazard rate defined by $\varphi(\tau)$. \square

We superimpose several rather technical assumptions on the hazard rate.

Assumption 4.1.1. The hazard rate $\rho(\tau) = \frac{\varphi(\tau)}{1 - \int_0^\tau \varphi(s) ds}$ stems from a distribution $\varphi(\tau) \in C^1(\mathbb{R}_+)$, where φ satisfies eq. (4.1.2). We also assume that

$$(a) \lim_{\tau \rightarrow \hat{\tau}} \int_0^\tau \rho(\tau) d\tau = \infty, \quad (b) \sup_{\tau \in [0, \hat{\tau}]} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} < \infty, \quad \lim_{\tau \rightarrow \hat{\tau}} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} = 0.$$

$$(c) \int_{\check{\tau}}^{\hat{\tau}} \rho^2(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau < \infty, \quad (d) \int_{\check{\tau}}^{\hat{\tau}} |\rho'(\tau)| e^{-\int_0^\tau \rho(s) ds} d\tau < \infty.$$

We show that these assumptions are satisfied by a wide range of random variables T (resp. their distribution φ).

Proposition 4.1.2. Let $\varphi \in C^1([0, \hat{\tau}))$, $\text{supp}(\varphi) \subset [\check{\tau}, \hat{\tau}]$, $\int_0^{\hat{\tau}} \varphi(s) ds = 1$, and assume that there is $z \in C^1(\mathbb{R}_+)$ and $m > 0$ such that $\varphi(\tau) = z(\tau)(\hat{\tau} - \tau)^m$. Then, the hazard rate $\rho(\tau)$ satisfies Assumption 4.1.1.

Proof. The conditions in eq. (4.1.2) are obviously satisfied. We now show (a)-(d) from Assumption 4.1.1.

(a)

$$\int_0^\tau \rho(\tau) d\tau = \int_0^\tau \frac{\varphi(\tau)}{1 - \int_0^\tau \varphi(s) ds} d\tau = -\ln \left(1 - \int_0^\tau \varphi(s) ds \right)$$

implies that $\lim_{\tau \rightarrow \hat{\tau}} \int_0^\tau \rho(\tau) d\tau = \infty$.

(b) Note that the function $\rho(\tau) e^{-\int_0^\tau \rho(s) ds}$ is continuous for $\tau < \hat{\tau}$. If we can show that this function converges to 0 for $\tau \rightarrow \hat{\tau}$, it is already bounded ($\text{sup}(\dots) < \infty$). We use the formula derived in (a) and obtain

$$\begin{aligned} \lim_{\tau \rightarrow \hat{\tau}} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} &= \lim_{\tau \rightarrow \hat{\tau}} \frac{\varphi(\tau)}{1 - \int_0^\tau \varphi(s) ds} e^{-\int_0^\tau \frac{d}{ds} \left(-\ln \left(1 - \int_0^s \varphi(\tau') d\tau' \right) \right) ds} \\ &= \lim_{\tau \rightarrow \hat{\tau}} \frac{\varphi(\tau)}{1 - \int_0^\tau \varphi(s) ds} \left(1 - \int_0^\tau \varphi(s) ds \right) = \lim_{\tau \rightarrow \hat{\tau}} \varphi(\tau) = 0. \end{aligned}$$

$$\int_{\check{\tau}}^{\hat{\tau}} \rho^2(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau = \int_{\check{\tau}}^{\hat{\tau}} \left(\frac{\varphi(\tau)}{1 - \int_0^\tau \varphi(s) ds} \right)^2 \left(1 - \int_0^\tau \varphi(s) ds \right) d\tau = \int_{\check{\tau}}^{\hat{\tau}} \frac{\varphi^2(\tau)}{1 - \int_0^\tau \varphi(s) ds} d\tau.$$

We discuss the asymptotic of the integrand for $\tau \rightarrow \hat{\tau}$ for the case $\varphi(\tau) = z(\tau)(\hat{\tau} - \tau)^m$ for some smooth $z(\tau)$. For the denominator, we obtain

$$1 - \int_0^\tau \varphi(s) ds = \int_{\check{\tau}}^{\hat{\tau}} \varphi(s) ds = \int_{\check{\tau}}^{\hat{\tau}} z(\tau) (\hat{\tau} - s)^m ds = \mathcal{O}((\hat{\tau} - s)^{m+1}).$$

Therewith,

$$\frac{\varphi^2(\tau)}{1 - \int_0^\tau \varphi(s) ds} = \frac{\mathcal{O}((\hat{\tau} - s)^{2m})}{\mathcal{O}((\hat{\tau} - s)^{m+1})} = \mathcal{O}((\hat{\tau} - s)^{m-1})$$

which is integrable, as we assume $m > 0$.

(d) Note that

$$\begin{aligned} \int_{\hat{\tau}}^{\hat{\tau}} |\rho'(\tau)| e^{-\int_0^{\tau} \rho(s) ds} d\tau &= \int_{\hat{\tau}}^{\hat{\tau}} \left| \frac{d}{d\tau} \left(\frac{\varphi(\tau)}{1 - \int_0^{\tau} \varphi(s) ds} \right) \right| \left(1 - \int_0^{\tau} \varphi(s) ds \right) d\tau \\ &\leq \int_{\hat{\tau}}^{\hat{\tau}} |\varphi'(\tau)| d\tau + \int_{\hat{\tau}}^{\hat{\tau}} \frac{\varphi^2(\tau)}{1 - \int_0^{\tau} \varphi(s) ds} d\tau \end{aligned}$$

which is integrable, as seen previously. \square

4.2 Pathogen Distribution for Constant Environmental Pathogen Load

What follows next is an analysis of the pathogen distribution $S(t, \tau, P)$. We do this by studying the reduced model of the system (4.1.1). To reduce the model, we take ϵ to zero such that no new infections are occurring and assume that the environmental pathogen load B is constant. That is, we consider

$$\begin{aligned} \partial_t S(t, \tau, P) + \partial_{\tau} S(t, \tau, P) + \partial_P(-\gamma P S(t, \tau, P)) &= -\rho(\tau) S(t, \tau, P) \\ S(t, 0, P) = g(t, P) &= \int_{\hat{\tau}}^{\hat{\tau}} \rho(\tau) S(t, \tau, P - \psi B) d\tau \\ S(0, \tau, P) &= S_0(\tau, P) \end{aligned} \tag{4.2.1}$$

for $B > 0$ given, fixed. This section aims to show that there is a stationary solution of eq. (4.2.1) (the invariant distribution of the underlying stochastic model) and that any non-negative initial condition eventually tends to this solution. Thereto, we first show that eq. (4.2.1) defines a strongly continuous and eventually compact semigroup, and then inspect the spectrum of the infinitesimal generator of this semigroup (see Section 1.2.2 for definitions). As we will find out, the generator has an eigenvalue 0 (caused by mass conservation), which is dominant. The non-local term in the boundary condition (a jump of B by a booster event) is not completely straightforward to handle.

Mathematically, this semigroup is close to systems describing cell division, or more generally, aggregation-fragmentation equations. In aggregation-fragmentation equations, entities grow and are reduced by sudden nonlocal fragmentation events, but in our case, the continuous degradation of the immune system decreases P , but the sudden disruptive nonlocal events (booster events) increase P . This is the difference between aggregation-fragmentation equations and our model. However, for the analysis, we mostly follow the strategy used for those equations developed by Heijmans, Gyllenberg and others [43, 47, 46], and advanced in recent years [31, 17].

We adapt the methods above to address a technical problem that arises in the analysis: the lack of strong positivity. That is, we find a compact interval $[P_*, P^*]$, with P_* and P^*

as the lower and upper bounds of the pathogen respectively, for the pathogen load right after a booster event happens. However, the probability mass is zero at the boundary of this interval. To solve this, we need to first regularize the problem before using the standard arguments. Additional effort is necessary if we pass to the limit and de-regularize the operators again, to ensure that the desired results remain valid in the limit.

4.2.1 Existence of the Semigroup

Below, using the methods of characteristics, we show that there is a bounded region Ω , such that the semigroup defined by eq. (4.2.1) has the space of continuous functions $C^0(\Omega)$ as an invariant function space. We thus work with the state space $E = C^0(\Omega)$. For now, we take Ω for granted, and define the operator $A : D(A) \subset E \rightarrow E$ as

$$\begin{aligned} A\phi(\tau, P) &= -\partial_\tau\phi(\tau, P) - \partial_P(-\gamma P\phi(\tau, P)) - \rho(\tau)\phi(\tau, P) \quad \forall\phi \in D(A) \\ D(A) &= \{\phi(\tau, P) | \phi, A\phi \in E, \phi(0, \tau) = \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau)\phi(\tau, P - \psi B)d\tau\}. \end{aligned} \quad (4.2.2)$$

We rewrite eq. (4.2.1) as an abstract Cauchy equation (with the understanding that $S(t) \in E$).

$$\frac{d}{dt} S(t) = AS(t), \quad S(0) = S_0.$$

We next show that the operator A is the infinitesimal generator of a strongly continuous semigroup $\{T(t) | t \geq 0\}$ on E . As is usual for this kind of equation [68], the construction of the existence of the semigroup is based on the method of characteristics. The characteristic curves of eq. (4.2.1) are given by

$$\frac{dt}{ds} = 1, \quad \frac{d\tau}{ds} = 1, \quad \frac{dP}{ds} = -\gamma P, \quad \frac{dz}{ds} = -(\rho(\tau) - \gamma)z.$$

For $t < \tau < \check{\tau}$, we obtain the solution by a pure transport of the initial conditions along the characteristics,

$$\begin{aligned} t = \tau, \quad \tau = s + \tau_0, \quad P = P_0 e^{-\gamma s}, \quad z = S_0(\tau_0, P_0) e^{-\int_0^s \rho(s+\tau_0) - \gamma ds}. \\ S(t, \tau, P) = S_0(\tau - t, P e^{\gamma t}) e^{-\int_{\tau-t}^{\tau} \rho(\tau') - \gamma d\tau'}. \end{aligned}$$

Therewith, for $t < \check{\tau}$, we obtain the boundary values $S(t, 0, P)$ by an appropriate integral over S_0 ,

$$\begin{aligned} g(t, P) &= \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau) S(t, \tau, P - \psi B) d\tau \\ &= \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau') S_0(\tau' - t, (P - \psi B) e^{\gamma \tau'}) e^{-\int_{\tau'-t}^{\tau'} \rho(\tau'') - \gamma d\tau''} d\tau' \end{aligned} \quad (4.2.3)$$

such that we are also able to define $S(t, \tau, P)$ for $\tau < t < \check{\tau}$. That is,

$$\begin{aligned} t &= \tau + t_0, & \tau &= s, & P &= P_0 e^{-\gamma s}, & z &= S(t_0, 0, P_0) e^{-\int_0^s \rho(\tau') - \gamma d\tau'}, \\ S(t, \tau, P) &= S(t - \tau, 0, P e^{\gamma\tau}) e^{-\int_0^\tau \rho(\tau') - \gamma d\tau'} = g(t - \tau, P e^{\gamma\tau}) e^{-\int_0^\tau \rho(\tau') - \gamma d\tau'} \\ &= \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau') S_0(\tau + \tau' - t, (P e^{\gamma\tau} - \psi B) e^{\gamma\tau'}) e^{-\int_{\tau+\tau'-t}^{\tau'} \rho(\tau'') - \gamma d\tau''} d\tau' e^{-\int_0^\tau \rho(\tau') - \gamma d\tau'}. \end{aligned}$$

Combining the two observations, we are able to explicitly define the semigroup for $0 \leq t \leq \check{\tau}$,

$$\begin{aligned} T(t)S_0(\tau, P) &= \begin{cases} S_0(\tau - t, P e^{\gamma\tau}) e^{-\int_{\tau-t}^\tau \rho(\tau') - \gamma d\tau'} & t \geq \tau \\ \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau') S_0(\tau + \tau' - t, (P e^{\gamma\tau} - \psi B) e^{\gamma\tau'}) e^{-\int_{\tau+\tau'-t}^{\tau'} \rho(\tau'') - \gamma d\tau''} d\tau' e^{-\int_0^\tau \rho(\tau') - \gamma d\tau'} & t < \tau. \end{cases} \end{aligned} \quad (4.2.4)$$

We have an explicit representation of the semigroup for $t \in [0, \check{\tau}]$. The condition for strong continuity immediately follows from this representation. We extend that definition to $t \geq 0$. For $t = n\check{\tau} + \delta$ (where $0 \leq \delta < \check{\tau}$ and $n \in \mathbb{N}_0$), we define $T(t)S_0 = T(\check{\tau})^n T(\delta)S_0$. The following theorem is a consequence of the results above.

Theorem 4.2.1. *Equation (4.2.1) defines a strongly continuous semigroup $T(t)$ on $C^0([0, \hat{\tau}] \times \mathbb{R}_+)$.*

As indicated above, the state space $(\tau, P) \in [0, \hat{\tau}] \times \mathbb{R}_+$ is possible but too large, as for $P \gg 1$, the degradation is faster than the boosting and thus the mass of the solution will eventually collect in the bounded region Ω constructed below, which turns out to be invariant under the within-host pathogen dynamics.

We obtain an upper bound P^* for the pathogen load right after a booster event ($\tau = 0$): starting at $(\tau, P) = (0, P^*)$, the characteristic originating in this point is given by $P(\tau) = P^* e^{-\gamma\tau}$. The minimal time interval to the next booster event might take place is $\check{\tau}$, such that $P^* = P^* e^{-\gamma\check{\tau}} + \psi B$ and

$$P^* = \frac{\psi B}{1 - e^{-\gamma\check{\tau}}}. \quad (4.2.5)$$

We also introduce a minimal pathogen load P_* right after a booster event in the same way: now we consider booster events happening at the maximal time span $\hat{\tau}$, such that

$$P_* = \frac{\psi B}{1 - e^{-\gamma\hat{\tau}}}. \quad (4.2.6)$$

Therewith, we can define the invariant region Ω , bounded above and below by the characteristic starting in $(\tau, P) = (0, P^*)$, respectively, in $(\tau, P) = (0, P_*)$, and extending to the right up to $\hat{\tau}$ (see Figure 4.3).

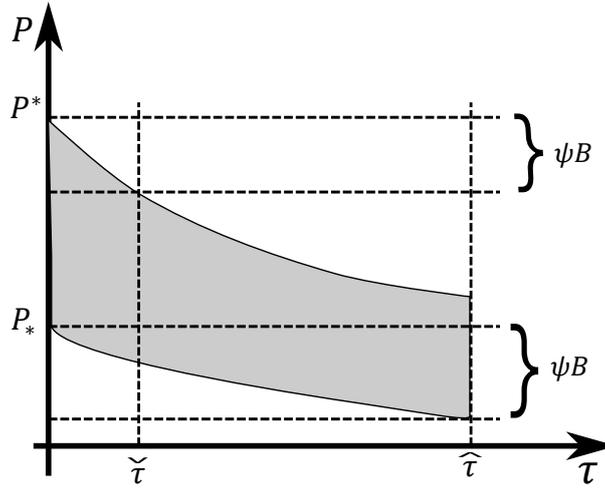


Figure 4.3: Shape of Ω : The upper and lower bound of Ω are given by characteristic curves, originating in $(0, P^*)$ and $(0, P_*)$. The points P^* resp. P_* are constructed in such a way that the characteristics hit $(\check{\tau}, P^* - \psi B)$ resp. $(\hat{\tau}, P_* - \psi B)$.

Corollary 4.2.1. *Let*

$$\Omega = \{(\tau, P) \mid 0 \leq \tau \leq \hat{\tau}, P_* e^{-\gamma\tau} \leq P \leq P^* e^{-\gamma\tau}\}. \quad (4.2.7)$$

If $S_0 \in C^0([0, \hat{\tau}] \times \mathbb{R}_+)$, $\text{supp}(S_0) \subset \Omega$, then for all $t \geq 0$ $\text{supp}(T(t)S_0) \subset \Omega$, such that (in slight abuse of notation) $T(t)(C^0(\Omega)) \subset C^0(\Omega)$.

Note that in the representation of $T(t)$ given by eq. (4.2.4) the integral extends over values (τ, P) outside of Ω , where the solution $S(t, \tau, P)$ is consequently zero. From now on, we consider the semigroup to be acting on $E = C^0(\Omega)$.

Theorem 4.2.2. *If $t > 3\hat{\tau}$, the semigroup $T(t)$ is compact.*

Proof. If $t > \hat{\tau}$, the booster event has already occurred, we have

$$S(t, \tau, P) = g(t - \tau, P e^{\gamma\tau}) e^{-\int_0^\tau \rho(\tau') - \gamma d\tau'}$$

such that,

$$S(t, \tau, P - \psi B) = g(t - \tau, (P - \psi B) e^{\gamma\tau}) e^{-\int_0^\tau \rho(\tau') - \gamma d\tau'}$$

and therefore

$$g(t, P) = \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau) S(t, \tau, P - \psi B) d\tau = \int_{\check{\tau}}^{\hat{\tau}} g(t - \tau, (P - \psi B) e^{\gamma\tau}) \rho(\tau) e^{-\int_0^\tau \rho(\tau') - \gamma d\tau'} d\tau.$$

Assuming $t > 2\hat{\tau}$, we iterate twice to express the boundary value $g(t, P)$ as an integral over the boundary value at earlier times.

We already know that $g(t, P) = \int_{\tilde{\tau}}^{\hat{\tau}} g(t - \tau', (P - \psi B) e^{\gamma\tau'}) \rho(\tau') e^{-\int_0^{\tau'} \rho(s) - \gamma ds} d\tau'$ and thus,

$$g(t - \tau, (P - \psi B) e^{\gamma\tau}) = \int_{\tilde{\tau}}^{\hat{\tau}} g(t - \tau - \tau', ((P - \psi B) e^{\gamma\tau} - \psi B) e^{\gamma\tau'}) \rho(\tau') e^{-\int_0^{\tau'} \rho(s) - \gamma ds} d\tau'$$

which yields

$$\begin{aligned} & g(t, P) \\ &= \int_{\tilde{\tau}}^{\hat{\tau}} \left(\int_{\tilde{\tau}}^{\hat{\tau}} g(t - \tau - \tau', ((P - \psi B) e^{\gamma\tau} - \psi B) e^{\gamma\tau'}) \rho(\tau') e^{-\int_0^{\tau'} \rho(s') - \gamma ds'} d\tau' \right) \rho(\tau) e^{-\int_0^{\tau} \rho(s) - \gamma ds} d\tau \\ &= \int_{\tilde{\tau}}^{\hat{\tau}} \int_{\tilde{\tau}}^{\hat{\tau}} g(t - \tau - \tau', ((P - \psi B) e^{\gamma\tau} - \psi B) e^{\gamma\tau'}) \rho(\tau') e^{-\int_0^{\tau'} \rho(s') - \gamma ds'} \rho(\tau) e^{-\int_0^{\tau} \rho(s) - \gamma ds} d\tau' d\tau. \end{aligned}$$

To transform the integral, we let

$$\begin{aligned} u &= t - \tau - \tau' \\ v &= ((P - \psi B) e^{\gamma\tau} - \psi B) e^{\gamma\tau'} = (P - \psi B) e^{r(\tau + \tau')} - \psi B e^{r\tau'} \end{aligned} \quad (4.2.8)$$

thus

$$v = (P - \psi B) e^{r(t-u)} - \psi B e^{r\tau'} \quad (4.2.9)$$

$$e^{r\tau'} = \frac{(P - \psi B) e^{r(t-u)} - v}{\psi B}, \quad \tau' = \tau'(u, v; t, P) = \frac{1}{\gamma} \ln \left(\frac{(P - \psi B) e^{r(t-u)} - v}{\psi B} \right) \quad (4.2.10)$$

and therefore

$$\tau = \tau(u, v; t, P) = t - u - \tau'(u, v; t, P). \quad (4.2.11)$$

Then, we find the derivatives of $\tau(u, v; t, P)$ and $\tau'(u, v; t, P)$ with respect to t and P .

$$\partial_t \tau'(u, v; t, P) = \frac{1}{r} \left(\frac{\gamma \psi B (P - \psi B) e^{r(t-u)}}{\psi B ((P - \psi B) e^{r(t-u)} - v)} \right) = \frac{(P - \psi B) e^{r(t-u)}}{(P - \psi B) e^{r(t-u)} - v}.$$

Since $(P - \psi B) e^{r(t-u)} - v = e^{r\tau'}$, $\partial_t \tau'(u, v; t, P)$ will be bounded for all feasible values of u, v, t, P .

$$\partial_P \tau'(u, v; t, P) = \frac{1}{r} \left(\frac{\psi B e^{r(t-u)}}{\psi B ((P - \psi B) e^{r(t-u)} - v)} \right) = \frac{1}{\gamma} \left(\frac{e^{r(t-u)}}{(P - \psi B) e^{r(t-u)} - v} \right).$$

This derivative is also bounded. Thus, $\tau'(u, v; t, P)$ is bounded in C^1 with regard to t and P .

Since $\tau(u, v; t, P) = t - u - \tau'(u, v; t, P)$, the function $\tau(u, v; t, P) = t - u - \tau'(u, v; t, P)$ is C^1 with regard to t and P .

Transformation of the integral domain: The integral domain reads $(\tau, \tau') \in [\check{\tau}, \hat{\tau}] \times [\check{\tau}, \hat{\tau}]$. Due to eq. (4.2.8)

$$t - 2\check{\tau} \leq u \leq t - 2\hat{\tau}.$$

Furthermore, for a given v , we have according to eq. (4.2.9)

$$v \in [(P - \psi B)e^{\gamma(t-u)} - \psi B e^{\gamma\check{\tau}}, (P - \psi B)e^{\gamma(t-u)} - \psi B e^{\gamma\hat{\tau}}]$$

which implies that the integral boundaries transform to

$$\int_{\check{\tau}}^{\hat{\tau}} \int_{\check{\tau}}^{\hat{\tau}} (\dots) d\tau' d\tau = \int_{t-2\check{\tau}}^{t-2\hat{\tau}} \int_{(P-\psi B)e^{\gamma(t-u)}-\psi B e^{\gamma\check{\tau}}}^{(P-\psi B)e^{\gamma(t-u)}-\psi B e^{\gamma\hat{\tau}}} (\dots) \left| \det \frac{\partial(\tau, \tau')}{\partial(u, v)} \right| dudv.$$

Transformation of the infinitesimal: From eqs. (4.2.10) and (4.2.11),

$$\begin{aligned} \tau(u, v; t, P) &= t - u - \tau'(u, v; t, P) \\ \tau'(u, v; t, P) &= \frac{1}{\gamma} \ln \left(\frac{(P - \psi B)e^{\gamma(t-u)} - v}{\psi B} \right). \end{aligned}$$

We find the Jacobian of $\tau(u, v; t, P)$ and $\tau'(u, v; t, P)$ with regard to u and v . Therefore, we first consider their derivatives.

$$\begin{aligned} \partial_u \tau'(u, v; t, P) &= \frac{-(P - \psi B)e^{\gamma(t-u)}}{(P - \psi B)e^{\gamma(t-u)} - v} \\ \partial_v \tau'(u, v; t, P) &= \frac{-1}{r((P - \psi B)e^{\gamma(t-u)} - v)} \\ \partial_u \tau(u, v; t, P) &= -1 - \partial_u \tau'(u, v; t, P) \\ \partial_v \tau(u, v; t, P) &= -\partial_u \tau'(u, v; t, P). \end{aligned}$$

All four derivatives are C^1 with regard to t and P . Thus, the determinant of the Jacobian also

$$|\det J(u, v; t, P)| = \left| \det \begin{pmatrix} \partial_u \tau(u, v; t, P) & \partial_v \tau(u, v; t, P) \\ \partial_u \tau'(u, v; t, P) & \partial_v \tau'(u, v; t, P) \end{pmatrix} \right|$$

is in C^1 with regard to t and P . Taking all the elements together we obtain

$$\begin{aligned} &g(t, P) \\ &= \int_{\check{\tau}}^{\hat{\tau}} \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau') g(t - \tau - \tau', ((P - \psi B)e^{\gamma\tau} - \psi B)e^{\gamma\tau'}) e^{-\int_0^{\tau'} \rho(s') - \gamma ds'} \rho(\tau) e^{-\int_0^{\tau} \rho(s) - \gamma ds} d\tau' d\tau \\ &= \int_{t-2\check{\tau}}^{t-2\hat{\tau}} \int_{(P-\psi B)e^{\gamma(t-u)}-\psi B e^{\gamma\check{\tau}}}^{(P-\psi B)e^{\gamma(t-u)}-\psi B e^{\gamma\hat{\tau}}} \rho(\tau'(u, v; t, P)) g(u, v) e^{-\int_0^{\tau'(u, v; t, P)} \rho(s') - \gamma ds'} \\ &\quad \rho(\tau(u, v; t, P)) e^{\int_0^{\tau(u, v; t, P)} \rho(s) - \gamma ds} |\det J(u, v; t, P)| dudv. \end{aligned}$$

As the derivative of t and P no longer acts on $g(\cdot)$ and all the expressions are C^1 with regard to t and P , for $\bar{t} > 2\hat{\tau}$, we can obtain an estimate of the form

$$\begin{aligned} \|g(t, P)\|_{C^1([2\hat{\tau}, \bar{t}] \times [0, P^*])} &\leq C(\bar{t}) \|g(t, P)\|_{C^0([0, \bar{t}] \times [0, P^*])} \\ &\leq C(\bar{t}) \|T_t S_0\|_{C^0([0, \bar{t}], C^0(\Omega))} \leq C(\bar{t}) \|S_0\|_{C^0(\Omega)}. \end{aligned}$$

That is, the boundary conditions are smooth for $t > 2\hat{\tau}$, and as the transport of the boundary conditions into the region Ω is also smooth this inequality implies for $t > 3\hat{\tau}$ that

$$\|T_t S_0\|_{C^0([3\hat{\tau}, \bar{t}], C^1(\Omega))} \leq C(\bar{t}) \|S_0\|_{C^0(\Omega)}.$$

Note that for $t > 3\hat{\tau}$, the integral extends to S_0 . Therefore, the semigroup is eventually compact by Arzelà–Ascoli. \square

4.2.2 Stationary Solution and Spectral Gap

After establishing the semigroup, we now turn to the spectrum of the infinitesimal generator. For this, we follow once more the standard approach [43, 47, 46]: We first convert the eigenvalue problem into a fixed point equation, then analyze the fixed point operator (we derive suitable a priori estimates to show compactness), and use the theory of positive operators (Krein–Rutman) to obtain information about eigenvalues and particularly the dominant eigenvalue. The problem herein is that we will not find a strongly positive fixed point operator, and as such, we will have to first regularize our operator before applying the Krein–Rutman theory and related ideas. Afterwards, we will check that the results hold for the de-regularized operator.

However, the existence of a stationary solution is simply a consequence of mass conservation, which implies that $f_0 = 1$ is an adjoint eigenfunction for eigenvalue 0. Compactness properties imply that there is also an eigenfunction for eigenvalue zero [111]. The involving part is to establish a spectral gap leading to a spectral decomposition of the underlying Banach space and the stability of the stationary solution [108].

4.2.2.1 Eigenvalues and Fixed Point Operator

To check for the existence of solutions, we consider the eigenvalue problem associated with eq. (4.2.1),

$$\begin{aligned} \partial_\tau S + \partial_P(-\gamma PS) &= -(\lambda + \rho(\tau))S \\ S(0, P) &= g(P) = \int_0^\infty \rho(\tau) S(\tau, P - \psi B) d\tau. \end{aligned} \tag{4.2.12}$$

In the same way as above (equation (4.2.3)), we use the method of characteristics to transform eq. (4.2.12) into a linear integral operator. We obtain

$$S(0, P) = \int_{\hat{\tau}}^{\bar{\tau}} S(0, (P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau (\lambda + \rho(s)) ds} d\tau.$$

Note that we know that Ω is invariant for the semigroup, such that $\text{supp}(S(0, P)) \subset [P_*, P^*]$. We are led to the definition of an operator K_λ . The kernel of this operator $e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau (\lambda + \rho(s)) ds}$ is integrable due to Assumption 4.1.1(b). We also know that Ω is invariant for the semigroup, such that $\text{supp}(K_\lambda[S(0, P)]) \subset [P_*, P^*]$. The integral itself extends over points outside of Ω , where consequently $S(0, P) = 0$.

Definition 4.2.1. Let $K_\lambda : C^0[P_*, P^*] \rightarrow C^0[P_*, P^*]$ be defined by

$$K_\lambda[g](P) = \int_{\check{\tau}}^{\hat{\tau}} g((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau. \quad (4.2.13)$$

The integral bounds of K are $\check{\tau}$ and $\hat{\tau}$, with the understanding that $g = 0$ if $(P - \psi B)e^{\gamma\tau}$ is outside of $[P_*, P^*]$ (see also Figure 4.3). To be more precise, we can define $\underline{\tau}(P)$ and $\bar{\tau}(P)$ (see also Proposition 4.2.2) such that

$$K_\lambda[g](P) = \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} g((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau.$$

Particularly, $\bar{\tau}(P) = \underline{\tau}(P)$ for $P \in \{P_*, P^*\}$, such that for all $g \in C^0([P_*, P^*])$

$$K_\lambda[g](P_*) = K_\lambda[g](P^*) = 0 \quad (4.2.14)$$

which leads to technical difficulties as this equation shows that K_λ is not strictly positive.

Corollary 4.2.2. An eigenfunction of eq. (4.2.12) for eigenvalue λ is a fixed point of K_λ . Particularly, a stationary solution of eq. (4.2.1) is a fixed point of K_0 .

4.2.2.2 A Priori Estimates

We shall use the following a priori estimates to show compactness of the operator K_λ . The first statement of the next proposition is basically a fact that the semigroup is mass-preserving.

Proposition 4.2.1. (a) For $g \in L^\infty(P_*, P^*)$, $g \geq 0$, we find

$$\|K_0[g]\|_{L^1(P_*, P^*)} = \|g\|_{L^1(P_*, P^*)}.$$

(b) If $g \in C^0[P_*, P^*]$, then $K_\lambda[g] \in C^0[P_*, P^*]$, and there is $c = c(\lambda) > 0$ such that

$$\|K_\lambda[g]\|_{C^0[P_*, P^*]} \leq c \|g\|_{C^0[P_*, P^*]}.$$

Proof. (a) Note that the integrand is non-negative for $g \geq 0$. Thus,

$$\begin{aligned} \int_{P_*}^{P^*} K_0[g](P) dP &= \int_{P_*}^{P^*} \int_{\check{\tau}}^{\hat{\tau}} g((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau dP \\ &= \int_{\check{\tau}}^{\hat{\tau}} \int_{P_*}^{P^*} g((P - \psi B)e^{\gamma\tau}) dP e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau \end{aligned}$$

where we have exchanged the integrals with the understanding (as above) that g becomes zero outside of $[P_*, P^*]$. We now change the integration variable, $x = (P - \psi B)e^{\gamma\tau}$. For any $\tau \in (\check{\tau}, \hat{\tau})$, we have

$$(P_* - \psi B)e^{\gamma\tau} = P_* e^{-\gamma(\hat{\tau}-\tau)} < P_* \quad \text{and} \quad (P^* - \psi B)e^{\gamma\tau} = P^* e^{-\gamma(\check{\tau}-\tau)} > P^*$$

such that $\int_{P_*}^{P^*} g((P - B)e^{\gamma\tau}) e^{\gamma\tau} dP = \int_{P_*}^{P^*} g(x) dx$ and we proceed

$$\begin{aligned} \int_{P_*}^{P^*} K_0[g](P) dP &= \int_{\check{\tau}}^{\hat{\tau}} \int_{P_*}^{P^*} g(x) dx \rho(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau \\ &= \|g\|_{L^1(P_*, P^*)} \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau = \|g\|_{L^1(P_*, P^*)}. \end{aligned}$$

(b) If $g \in C^0[B, P^*]$, then $\hat{K}_\lambda[g]$ is continuous for any $P \in [P_*, P^*]$, as the integral kernel is smooth. The *a priori* estimate follows as before,

$$\begin{aligned} |K_\lambda[g](P)| &\leq \int_{\check{\tau}}^{\hat{\tau}} |g((P - \psi B)e^{\gamma\tau})| e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \\ &\leq C(\lambda) \int_{\check{\tau}}^{\hat{\tau}} e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau \|g\|_{C^0[P_*, P^*]} \leq c(\lambda) \|g\|_{C^0[P_*, P^*]}. \end{aligned}$$

□

Based on this observation, we obtain a proper $C^{0,1}$ estimate from which we deduce compactness by the Arzelà–Ascoli theorem.

Proposition 4.2.2. *There is a $c > 0$ such that $\|K_\lambda[g]\|_{C^{0,1}[P_*, P^*]} \leq c \|g\|_{C^0[P_*, P^*]}$. The operator $K_\lambda : C^0[P_*, P^*] \rightarrow C^0[P_*, P^*]$ is compact.*

Proof. We already know that $\|K_\lambda[g]\|_{C^0[P_*, P^*]} \leq c(\lambda) \|g\|_{C^0[P_*, P^*]}$. We now estimate $|\frac{d}{dP} K_\lambda[g](P)|$ by $\|g\|_{C^0[P_*, P^*]}$.

Here, we need to use the correct boundaries in the integral defining $K_\lambda[g](P)$ (until now, we have simply stated that g becomes zero if the argument is outside $[P_*, P^*]$). For a given P , we compute the τ -value, where the upper (lower) boundary of Ω hits $P - \psi B$. For the upper bound, we have

$$P - \psi B = P^* e^{-\gamma\tau} \quad \Leftrightarrow \quad \tau = \frac{1}{\gamma} \ln \left(\frac{P^*}{P - \psi B} \right).$$

The intersection with the lower boundary has the same expression where we use P_* instead of P^* . Accordingly, we define

$$\begin{aligned} \bar{\tau}(P) &= \begin{cases} \frac{1}{\gamma} \ln(P^*/(P - \psi B)) & \text{for } P > \psi B + P^* e^{-\gamma\hat{\tau}} \\ \hat{\tau} & \text{else} \end{cases} \\ \underline{\tau}(P) &= \begin{cases} \frac{1}{\gamma} \ln(P_*/(P - \psi B)) & \text{for } P < \psi B + P_* e^{-\gamma\check{\tau}} \\ \check{\tau} & \text{else} . \end{cases} \end{aligned}$$

Note that $\bar{\tau}$ and $\underline{\tau}$ are Lipschitz continuous and the derivative is zero for the constant part, respectively,

$$\bar{\tau}'(P) = \underline{\tau}'(P) = \frac{-1}{\gamma(P - \psi B)}$$

in the non-constant part. As $P \geq P_* > \psi B$, the derivative is bounded. With these preliminaries, we can now start to estimate.

$$\begin{aligned} \left| \frac{d}{dP} K_\lambda[g](P) \right| &= \left| \frac{d}{dP} \int_{\hat{\tau}}^{\hat{\tau}} g((P - \psi B) e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \right| \\ &= \left| \frac{d}{dP} \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} g((P - \psi B) e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \right| \\ &\leq \left| g((P - \psi B) e^{\gamma\bar{\tau}(P)}) e^{\gamma\bar{\tau}(P)} \rho(\bar{\tau}(P)) e^{-\int_0^{\bar{\tau}(P)} \lambda + \rho(s) ds} \bar{\tau}'(P) \right| \\ &\quad + \left| g((P - \psi B) e^{\gamma\underline{\tau}(P)}) e^{\gamma\underline{\tau}(P)} \rho(\underline{\tau}(P)) e^{-\int_0^{\underline{\tau}(P)} \lambda + \rho(s) ds} \underline{\tau}'(P) \right| \\ &\quad + \left| \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} g'((P - \psi B) e^{\gamma\tau}) e^{2\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \right|. \end{aligned}$$

We estimate the three terms, one after the other. For the first term,

$$\begin{aligned} &\left| g((P - \psi B) e^{\gamma\bar{\tau}(P)}) e^{\gamma\bar{\tau}(P)} \rho(\bar{\tau}(P)) e^{-\int_0^{\bar{\tau}(P)} \lambda + \rho(s) ds} \bar{\tau}'(P) \right| \\ &\leq \|g\|_{C^0[P_*, P_*]} e^{\gamma\hat{\tau}} \max\{1, e^{\Re(\lambda)\hat{\tau}}\} \rho(\hat{\tau}) e^{-\int_0^{\bar{\tau}(P)} \rho(s) ds} |\bar{\tau}'(P)| \leq C(\lambda) \|g\|_{C^0[P_*, P_*]} \end{aligned}$$

as $\bar{\tau}(P)$ is Lipschitz-continuous and $\rho(\tau) e^{-\int_0^\tau \rho(s) ds}$ is bounded (Assumption 4.1.1). Similarly, for the second term,

$$\left| g((P - \psi B) e^{\gamma\underline{\tau}(P)}) e^{\gamma\underline{\tau}(P)} \rho(\underline{\tau}(P)) e^{-\int_0^{\underline{\tau}(P)} \lambda + \rho(s) ds} \underline{\tau}'(P) \right| \leq C(\lambda) \|g\|_{C^0[P_*, P_*]}.$$

For the third term, we integrate by parts and proceed

$$\begin{aligned} &\int_{\underline{\tau}(P)}^{\bar{\tau}(P)} g'((P - \psi B) e^{\gamma\tau}) e^{2\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \\ &= \frac{1}{\gamma(P - \psi B)} \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} \frac{d}{d\tau} g((P - \psi B) e^{\gamma\tau}) d\tau \\ &= \frac{1}{\gamma(P - \psi B)} e^{\gamma\bar{\tau}(P)} \rho(\bar{\tau}(P)) e^{-\int_0^{\bar{\tau}(P)} \lambda + \rho(s) ds} g((P - \psi B) e^{\gamma\bar{\tau}(P)}) \\ &\quad - \frac{1}{\gamma(P - \psi B)} e^{\gamma\underline{\tau}(P)} \rho(\underline{\tau}(P)) e^{-\int_0^{\underline{\tau}(P)} \lambda + \rho(s) ds} g((P - \psi B) e^{\gamma\underline{\tau}(P)}) \\ &\quad - \frac{1}{\gamma(P - \psi B)} \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} g((P - \psi B) e^{\gamma\tau}) \frac{d}{d\tau} \left(e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} \right) d\tau. \end{aligned}$$

We again estimate the three terms, where the first two terms follow the same scheme as above,

$$\left| \frac{1}{\gamma(P - \psi B)} e^{\gamma \bar{\tau}(P)} \rho(\bar{\tau}(P)) e^{-\int_0^{\bar{\tau}(P)} \lambda + \rho(s) ds} g((P - \psi B) e^{\gamma \bar{\tau}(P)}) \right| \leq C(\lambda) \|g\|_{C^0[P_*, P^*]}$$

(note that $P - \psi B \geq P_* - \psi B > 0$) and similarly for the second term. For the third term, we remark

$$\begin{aligned} & \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} \left| \frac{d}{d\tau} \left(e^{(\gamma - \lambda)\tau} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} \right) \right| d\tau \\ \leq & (\gamma + |\lambda|) e^{(\gamma + |\Re(\lambda)|)\hat{\tau}} \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} \rho(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau + e^{(\gamma + |\Re(\lambda)|)\hat{\tau}} \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} |\rho'(\tau)| e^{-\int_0^\tau \rho(s) ds} d\tau \\ & + e^{(\gamma + |\Re(\lambda)|)\hat{\tau}} \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} \rho^2(\tau) e^{-\int_0^\tau \rho(s) ds} d\tau. \end{aligned}$$

Due to Assumption 4.1.1, all three integrals are finite. Hence,

$$\left| \frac{1}{\gamma(P - \psi B)} \int_{\underline{\tau}(P)}^{\bar{\tau}(P)} g(P - \psi B) \frac{d}{d\tau} \left(e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} \right) d\tau \right| \leq C(\lambda) \|g\|_{C^0[P_*, P^*]}.$$

Therewith, the $C^{0,1}$ estimate for K_λ is established, and the compactness is a consequence of the Arzelà-Ascoli theorem. \square

4.2.2.3 Regularized Operator

The Perron-Frobenius theory of positive operators will be useful in the proof of the existence of a dominant eigenvalue. Below, we give a summary of ideas and results derived from the theorem [47, 49].

Definition 4.2.2. Positive Linear Operators

Let E be a Banach space and E^* be its dual, that is, E^* is the space of all linear functionals on E . We denote the value $h \in E^*$ at $\Psi \in E$ as $\langle h, \Psi \rangle$. A closed subset $F \subset E$ is called a cone if the following conditions are satisfied:

1. $F + F \subset F$
2. $\lambda F \subset F$ if $\lambda \geq 0$
3. $F \cap (-F) = \{0\}$
4. $F \neq \{0\}$.

With regard to the cone F , we say $x \leq y$ if and only if $y - x \in F$ and $x < y$ if $y - x \in F_+ := \setminus F \setminus \{0\}$. The cone F is called a total cone if the set $\{\Psi - \Phi : \Psi, \Phi \in F\}$ is dense in E . The cone F is said to be a solid cone if it has a non-empty interior F° . We

note that $x \ll y$ if $y - x \in F^\circ$. The dual cone F^* is by definition a subset of E^* composed of all positive linear functionals, that is $h \in F^*$ if and only if $h \in E^*$ with $\langle h, \Psi \rangle \geq 0$ for all $\Psi \in F$. A positive linear functional $h \in F^*$ is said to be strictly positive if $\langle h, \Psi \rangle > 0$ for all $\Psi \in F_+$. Let $B(E)$ be a set of bounded linear operators of E onto E , that is $B(E) : E \rightarrow E$, an operator $T \in B(E)$ is said to be positive with respect to the cone F if $T(F) \subset F$ and strictly positive if $T(F_+) \subset F_+$. The spectral radius of $T \in B(E)$ is denoted by $r(T)$.

The Krein-Rutman theorem below (Theorems 4.2.3 and 4.2.4) is an extension of the Perron-Frobenius theorem to infinite dimensional Banach spaces.

Theorem 4.2.3 (Krein–Rutman Theorem, [49]). *Let S be a total cone, $K : S \rightarrow S$ be a compact positive linear operator and $r(K) > 0$. Then, $r(K)$ is an eigenvalue of S that corresponds to a positive eigenvector $\Psi \in S^+$.*

Theorem 4.2.4 ([49]). *Let S be a solid cone and $K : S \rightarrow S$ be a compact strongly positive linear operator. Then,*

1. $r(K) > 0$, $r(K)$ is a simple eigenvalue with an eigenvector in the non-empty interior S° and no other eigenvalue has a positive eigenvector;
2. $|\lambda| < r(K) \forall$ eigenvalues, $\lambda \neq r(K)$.

Furthermore, the following theorem is well known (e.g., [33, Proposition 1.4]).

Theorem 4.2.5. *If K_i are linear positive operators on the same Banach space and $K_1 \leq K_2$, then $\|K_1\| \leq \|K_2\|$ and $r(K_1) \leq r(K_2)$.*

Regularized Operator K_λ^ε We regularize the integral operator by replacing K_λ with a convex combination of K_λ and a strictly positive rank 1 operator, which also preserves the L^1 norm of positive functions.

Definition 4.2.3. *Let $Q : C^0[P_*, P^*] \rightarrow C^0[P_*, P^*]$ be the rank-1 operator*

$$Q[g] = \frac{1}{P^* - P_*} \int_{P_*}^{P^*} g(\tau) d\tau \quad (4.2.15)$$

For $\varepsilon \in [0, 1]$, introduce $K_\lambda^\varepsilon : C^0[P_*, P^*] \rightarrow C^0[P_*, P^*]$ by

$$K_\lambda^\varepsilon[g] = (1 - \varepsilon)K_\lambda[g] + \varepsilon Q[g]. \quad (4.2.16)$$

Furthermore, we introduce

$$\Lambda^\varepsilon = \{\lambda \in \mathbb{C} \mid \exists g \in C([P_*, P^*], \mathbb{C}) : K_\lambda^\varepsilon[g] = g\}.$$

From the definition and our knowledge about K_λ , we immediately obtain the following corollary.

Corollary 4.2.3. (a) For $\varepsilon \in [0, 1]$, $\lambda \in \mathbb{C}$ the operator K_λ^ε is compact, and for $\varepsilon \in (0, 1)$, $\lambda \in \mathbb{R}$ strongly positive;
 (b) The map $(\lambda, \varepsilon) \rightarrow K_\lambda^\varepsilon$ is continuous with regard to the operator norm;
 (c) Furthermore, for $g \geq 0$, we have

$$\|K_0^\varepsilon[g]\|_{L^1(P_*, P^*)} = \|g\|_{L^1(P_*, P^*)}.$$

The adjoint eigenfunction of K_0^ε for eigenvalue 1 is $f_0^\varepsilon = 1$, independently on $\varepsilon \in [0, 1]$.

Note that Λ^0 coincides with the point spectrum of the infinitesimal generator.

Eigenvalues Later, we will show that the semigroup has a non-negative stationary solution, that is, that K_0 has a fixed point. To prepare for that result, we show that our regularized operator K_0^ε already has a positive fixed point.

Theorem 4.2.6. For $\lambda = 0$, K_λ^ε has a fixed point for all $\varepsilon \in [0, 1]$, which is positive for $\varepsilon \in (0, 1]$.

Proof. Corollary 4.2.3 indicates that there is an adjoint eigenfunction of K_0^ε for eigenvalue 1, which implies using the compactness of the operator that there also is an eigenfunction for this eigenvalue. For $\varepsilon > 0$, the Krein-Rutman theorem implies the positivity of the eigenfunction. \square

4.2.2.4 De-Regularization and Spectral Gap

Uniqueness of the Fixed Point for $\lambda = 0$ Note that the next theorem looks like the Krein-Rutman Theorem. However, as K_0 is not strictly positive, we need to prove the theorem via approximations of K_λ by K_λ^ε . Particularly, for $\lambda = 0$, we already know that there is an eigenfunction. The additional information in this special case is the non-negativity of the eigenfunction.

Theorem 4.2.7. For $\lambda \in \mathbb{R}$, the spectral radius of K_λ is an eigenvalue with a positive eigenfunction. Particularly, K_0 has a non-negative fixed point and spectral radius 1.

Proof. There are positive functions $g_\lambda^\varepsilon \in C^0[P_*, P^*]$ with $K_\lambda[g_\lambda^\varepsilon] = g_\lambda^\varepsilon$ (Theorem 4.2.6), that is,

$$g_\lambda^\varepsilon(P) = (1 - \varepsilon)K_\lambda[g_\lambda^\varepsilon](P) + \varepsilon Q[g_\lambda^\varepsilon].$$

Let $\lambda \in \mathbb{R}$ be arbitrary, fixed. We normalize the eigenfunction to $\|g_\lambda^\varepsilon\|_{L^1(P_*, P^*)} = 1$. Due to Assumption 4.1.1(b), the function $\rho(\tau) e^{-\int_0^\tau \rho(s) ds}$ is bounded (supremum finite), such that

$$K_\lambda[g_\lambda^\varepsilon](P) \leq C \int_{\tilde{\tau}}^{\hat{\tau}} g((P - \psi B)e^{\gamma\tau}) e^{(\gamma-\lambda)\tau} d\tau \leq C \|g^\varepsilon\|_{L^1(P_*, P^*)} = C.$$

and

$$\|g_\lambda^\varepsilon\|_{C^0[P_*, P^*]} = \|K_\lambda[g_\lambda^\varepsilon]\|_{C^0[P_*, P^*]} \leq (1 - \varepsilon) + \varepsilon.$$

The C^0 norm of the family $B = \{g_\lambda^\varepsilon \mid \varepsilon \in (0, 1)\}$ is uniformly bounded. Then, the set $K_0[B]$ is also bounded in $C^{0,1}$ and relatively compact in C^0 , such that we find a subsequence $g_\lambda^{\varepsilon_n}$ in C^0 , $\varepsilon_n \rightarrow 0$ for $n \rightarrow \infty$, that converges in C^0 to a function $g_\lambda \in C^0$. As $g_\lambda^\varepsilon \geq 0$, also $g_\lambda \geq 0$. We need to exclude that $g_\lambda = 0$.

As the topology C^0 is stronger than the L^1 topology, the sequence also converges in L^1 , and hence, (the L^1 norm of g_λ^ε is 1) also $\|g_\lambda\|_{L^1(P_*, P^*)} = 1$, such that $g_\lambda \not\equiv 0$. Additionally, from continuity, $K_\lambda[g_\lambda] = g_\lambda$. With that, we establish the existence of a non-negative fixed point.

The function g_λ is also an eigenfunction for the spectral radius of K_λ : K_λ is the limit of the family K_λ^ε of compact operators depending (with regard to the operator norm) continuously on ε ; as $K_\lambda^\varepsilon[g_\lambda] = r(K_\lambda^\varepsilon)g_\lambda^\varepsilon$, this equation carries over to $\varepsilon = 0$. For $\lambda = 0$ we have $r(K_0^\varepsilon) = 1$ and thus also $r(K_0) = 1$. \square

Then, we ensure that the fixed point of the K_0 solution is still unique, as it is for K_0^ε for $\varepsilon > 0$. Since we used the limit $\varepsilon \rightarrow 0$ to construct a fixed point, it is not clear if possibly two eigenvalues merged in the eigenvalue 1, such that we have a higher dimensional eigenspace. Ultimately, we use the knowledge that the underlying stochastic process mixes well enough to prevent a non-unique invariant measure.

Proposition 4.2.3. *If the space of fixed points of K_0 has at least dimension 2, then there is a fixed point that changes sign.*

Proof. Assume there are two different fixed points g_1, g_2 which are independent (no $\alpha, \beta \in \mathbb{R}$, $|\alpha| + |\beta| > 0$ and $\alpha g_1 + \beta g_2 = 0$). Both functions are non-zero; without restriction, both are non-negative (otherwise, we are done). Furthermore, again without restriction, $\|g_i\|_{L^1} = 1$.

Consider $g = g_1 - g_2 \not\equiv 0$ as $g_1 \neq g_2$. g is again a fixed point. If g assumes positive and negative values, we are done. Otherwise, either $g \geq 0$ or $g \leq 0$.

If $g \geq 0$, then

$$g_1 \geq g_2 \geq 0.$$

As $g_i \geq 0$, $g_i \in C^0$, and the L^1 -norm of g_1 and g_2 are equal, we conclude that $g_1 = g_2$, which is a contradiction. The second case $g \leq 0$ gives us a contradiction by the parallel argument. \square

The next proposition is a way to express that the underlying Markov process is well mixing.

Lemma 4.2.1. *Let $g \geq 0$, $p_0 \in \text{supp}(g)$. Then, for $n \in \mathbb{N}$ and $\lambda \in \mathbb{R}$*

$$\text{supp}(K_\lambda^n[g]) \supset [(1 - e^{-k\gamma\hat{\tau}})P_* + e^{-k\gamma\hat{\tau}}p_0, (1 - e^{-k\gamma\check{\tau}})P^* + e^{-k\gamma\check{\tau}}p_0].$$

Proof. Let $p_0 \in \text{supp}(g)$. Then, $\{P \in [P_*, P^*] : \exists \tau \in [\check{\tau}, \hat{\tau}] : (P - \psi B)e^{\gamma\tau} = p_0\} \subset \text{supp}(K_\lambda[g])$. That is,

$$[\psi B + e^{-\gamma\hat{\tau}} p_0, \psi B + e^{-\gamma\check{\tau}} p_0] \cap [P_*, P^*] \subset \text{supp}(K_\lambda[g]).$$

With the same argument, we find

$$[\psi B + \psi B e^{-\gamma\hat{\tau}} + e^{-2\gamma\hat{\tau}} p_0, \psi B + \psi B e^{-\gamma\check{\tau}} + e^{-2\gamma\check{\tau}} p_0] \cap [P_*, P^*] \subset \text{supp}(K_\lambda^2[g])$$

and if we iterate k times with operator K ,

$$[a_k, b_k] \subset \text{supp}(K_\lambda^k[g])$$

where

$$\begin{aligned} a_k &= \sum_{\ell=0}^{k-1} \psi B e^{-\ell\gamma\hat{\tau}} + e^{-k\gamma\hat{\tau}} p_0 = \psi B \frac{1 - e^{-k\gamma\hat{\tau}}}{1 - e^{-\gamma\hat{\tau}}} + e^{-k\gamma\hat{\tau}} p_0 = (1 - e^{-k\gamma\hat{\tau}})P_* + e^{-k\gamma\hat{\tau}} p_0 \\ b_k &= \sum_{\ell=0}^{k-1} \psi B e^{-\ell\gamma\check{\tau}} + e^{-k\gamma\check{\tau}} p_0 = \psi B \frac{1 - e^{-k\gamma\check{\tau}}}{1 - e^{-\gamma\check{\tau}}} + e^{-k\gamma\check{\tau}} p_0 = (1 - e^{-k\gamma\check{\tau}})P^* + e^{-k\gamma\check{\tau}} p_0. \end{aligned}$$

□

The boundaries of the interval we obtained is a convex combination between p_0 and P_* (resp. P^*). We find that the support of any positive function expands under iteration with K , and becomes $[P_*, P^*]$ after an infinite number of iterations. Unfortunately, the operator is not strictly positive, as $K_\lambda[g](P_*) = K_\lambda[g](P^*) = 0$, and thus, for point measures μ with $\text{supp}(\mu) \subset \{P_*, P^*\}$, the pairing $\langle \mu, K_\lambda^n[g] \rangle = 0$ for all $n \in \mathbb{N}$. In contrast, the proposition indicates that K_λ is semi-supporting in the L^2 -setting.

Theorem 4.2.8. *The eigenspace of K_0 for eigenvalue 1 is one-dimensional.*

Proof. If this is not the case, we have an eigenfunction $g \in C^0$ that changes sign (Proposition 4.2.3). That is, we find two non-negative functions $g_\pm \in C^0$, both not identically zero, with

$$g = g_+ - g_-, \quad \text{supp}(g_+) \cap \text{supp}(g_-) = \emptyset, \quad \text{supp}(g_+), \text{supp}(g_-) \neq \emptyset.$$

As K_0 is linear,

$$K_0[g_+] - K_0[g_-] = K[g] = g = g_+ - g_-.$$

Furthermore, as $g_\pm \geq 0$, we know that $\int_{P_*}^{P^*} K_0[g_\pm](P) dP = \int_{P_*}^{P^*} g_\pm(P) dP$.

Let us focus on g_+ . We know that the support of g_+ is strictly smaller than $[P_*, P^*]$, as $\text{supp}(g_-) \neq \emptyset$. We also know that (Lemma 4.2.1) $\text{supp}(K_0[g_+])$ is strictly larger than that

of g_+ . As the integral of g_+ is preserved by K_0 , it is not possible that $g_+ \leq K_0[g_+]$. There is $x_0 \in \text{supp}(g_+)$ with

$$g_+(x_0) > K_0[g_+](x_0).$$

Note that $g(x_0) = g_+(x_0)$ as $x_0 \in \text{supp}(g_+)$. Therewith,

$$g_+(x_0) > K_0[g_+](x_0) \geq K_0[g_+](x_0) - K_0[g_-](x_0) = K_0[g](x_0) = g(x_0) = g_+(x_0)$$

which is a contradiction. \square

Spectral Gap In the last part of this section, we establish the spectral gap. We do that in two steps: First, we establish the dominance of the eigenvalue $\lambda_d = 0$ in Λ^0 , then we exclude that the real parts of a sequence of elements in Λ^0 can approximate $\lambda_d = 0$.

For a general real λ , we first show that the spectral radius of K_λ is always larger zero.

Lemma 4.2.2. *For $\lambda \in \mathbb{R}$ arbitrary fixed, we find that $r(K_\lambda) > 0$.*

Proof. We use Theorem 4.2.5, and construct a positive operator \tilde{K} which yields a lower bound for $r(K_\lambda)$. Thereto, we rewrite K_λ as

$$\begin{aligned} K_\lambda[g](P) &= \int_{\check{\tau}}^{\hat{\tau}} g((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \\ &= \frac{1}{\gamma(P - \psi B)} \int_{\max\{(P - \psi B)e^{\gamma\check{\tau}}, P_*\}}^{\min\{(P - \psi B)e^{\gamma\hat{\tau}}, P^*\}} g(x) \rho(\tau(x; P)) e^{-\int_0^{\tau(x; P)} \lambda + \rho(s) ds} dx \end{aligned}$$

with

$$x = (P - \psi B)e^{\gamma\tau} \quad \Leftrightarrow \quad \tau = \tau(x; P) = \frac{1}{\gamma} \ln \left(\frac{x}{P - \psi B} \right).$$

Now we check for points in $[P_*, P^*]$ that are in the integration region

$$[\max\{(P - \psi B)e^{\gamma\check{\tau}}, P_*\}, \min\{(P - \psi B)e^{\gamma\hat{\tau}}, P^*\}]$$

and where the integral weight is strictly positive, that is, $\tau(x; P) \in (\check{\tau}, \hat{\tau})$. Simple computations show that for $P_* < P < P^*$, both conditions are given. Choose $\bar{P} = (P_* + P^*)/2$ as a reference point. Then, there are $\delta_1, \delta_2 > 0$ such that

$$K_\lambda[g](P) \geq \int_{\bar{P} - \delta_1}^{\bar{P} + \delta_1} g(x) \delta_2 dx =: \tilde{K}[g](P).$$

Hence, $r(K_\lambda) \geq r(\tilde{K})$. As $\tilde{K}[g]$ is a positive rank one operator (compact) with eigenfunction $g(x) = \chi_{[\bar{P} - \delta_1, \bar{P} + \delta_1]}(P)$ (this is the only eigenfunction for an eigenvalue $\neq 0$), the spectral radius (and only positive eigenvalue) is given by $r(\tilde{K}) = 2\delta_1 \delta_2 > 0$. \square

Proposition 4.2.4. *For $\lambda \in \mathbb{R}$, the operator K_λ has a fixed point if and only if $\lambda = 0$. Furthermore, $r(K_\lambda)$ is strictly decreasing in λ .*

Proof. We already know that $r(K_0) = 1$. Furthermore, the eigenfunction g_λ and the adjoint eigenfunction f_λ of K_λ for eigenvalue $r(K_\lambda)$ are non-negative. Due to Lemma 4.2.1, $g_\lambda > 0$ in the open interval (P_*, P^*) . If the support of f_λ is a subset of $\{P_*, P^*\}$, then there are $a, b \in \mathbb{R}$ such that

$$\langle f_\lambda, K_\lambda[f](P) \rangle = aK_\lambda[g](P_*) + bK_\lambda[g](P^*) = 0$$

according to eq. (4.2.14). That is, in this case, f_λ is an adjoint eigenfunction for eigenvalue 0, which contradicts the fact that $\rho(K_\lambda) > 0$ (Lemma 4.2.2). Thus, the intersection of (P_*, P^*) and the support of f_λ is non-void, and hence $\langle f_\lambda, g_\mu \rangle > 0$ for all $\lambda, \mu \in \mathbb{R}$.

Now, we use an argument by Heijmans [46]: Let $\lambda, \mu \in \mathbb{R}$, $\lambda > \mu$ and g non-negative.

$$\begin{aligned} K_\mu[g](P) &= \int_{\tilde{\tau}}^{\hat{\tau}} g((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \mu + \rho(s) ds} d\tau \\ &\geq e^{(\lambda - \mu)\tilde{\tau}} \int_{\tilde{\tau}}^{\hat{\tau}} g((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau = e^{(\lambda - \mu)\tilde{\tau}} K_\lambda[g](P) \end{aligned}$$

Taking $g = g_\mu$, we obtain $r(K_\mu) g_\mu(P) = K_\mu[g_\mu](P) \geq e^{(\lambda - \mu)\tilde{\tau}} K_\lambda[g_\mu](P)$. If we take the duality pairing with the positive eigenfunctional f_λ on both sides, and use that $\langle f_\lambda, g_\mu \rangle > 0$, we obtain

$$r(K_\mu) \geq e^{(\lambda - \mu)\tilde{\tau}} r(K_\lambda)$$

Hence, $\lambda \rightarrow r(K_\lambda)$ is continuous and strictly decreasing on \mathbb{R} . \square

Theorem 4.2.9. *If $\lambda \in \Lambda^0$, $\lambda \neq \lambda_d = 0$, then $\operatorname{Re} \lambda < \lambda_d = 0$.*

Proof. We basically adapt the argument [46, Theorem 6.13]. Suppose that $\lambda \in \Lambda^0$ and there is a corresponding eigenfunction $g_\lambda \in C^0([P_*, P^*], \mathbb{C})$ such that $K_\lambda[g_\lambda] = g_\lambda$. Then,

$$\begin{aligned} |g_\lambda| = |K_\lambda[g_\lambda]| &= \left| \int_{\tilde{\tau}}^{\hat{\tau}} g_\lambda((P - \psi B)e^{\gamma\tau}) \rho(\tau) e^{\gamma\tau} e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \right| \\ &\leq \int_{\tilde{\tau}}^{\hat{\tau}} |g_\lambda((P - \psi B)e^{\gamma\tau})| e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \Re(\lambda) + \rho(s) ds} d\tau = K_{\Re(\lambda)}[|g_\lambda|]. \end{aligned}$$

If we iterate with $K_{\Re(\lambda)}$, we obtain for $n \in \mathbb{N}$

$$K_{\Re(\lambda)}^n[|g_\lambda|] \leq K_{\Re(\lambda)}^{n+1}[|g_\lambda|].$$

Now, we know that there is a non-negative eigenfunctional $f_{\Re(\lambda)}$ corresponding to the eigenvalue $r(K_{\Re(\lambda)})$ of $K_{\Re(\lambda)}$. We know (by the argument in the proof of Proposition 4.2.4) that the support of $f_{\Re(\lambda)}$ has a non-zero intersection with (P_*, P^*) .

$$\langle f_{\Re(\lambda)}, K_{\Re(\lambda)}^n[|g_\lambda|] \rangle \leq \langle f_{\Re(\lambda)}, K_{\Re(\lambda)}^{n+1}[|g_\lambda|] \rangle = r(K_{\Re(\lambda)}^\varepsilon) \langle f_{\Re(\lambda)}, K_{\Re(\lambda)}^n[|g_\lambda|] \rangle.$$

For a sufficiently large n , due to the expansion property of the support under iteration with $K_{\Re(\lambda)}$ (Lemma 4.2.1), we can ensure that $\langle f_{\Re(\lambda)}, K_{\Re(\lambda)}^n[|g_\lambda|] \rangle > 0$, and thus, $r(K_{\Re(\lambda)}^\varepsilon) \geq 1$. Since $r(K_{\lambda}^\varepsilon), \lambda \in \mathbb{R}$ is a non-increasing function (Lemma 4.2.1) and $r(K_{\lambda_d}^\varepsilon) = 1$, it implies that $\Re(\lambda) \leq \lambda_d$.

Suppose $\lambda = \lambda_d + i\eta \in \Lambda^0$, we show that $\eta = 0$. As $\Re(\lambda) = \lambda_d$, we proceed

$$\begin{aligned} |g| = |K_\lambda[g]| &= \left| \int_{\tilde{\tau}}^{\hat{\tau}} g((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda + \rho(s) ds} d\tau \right| \\ &\leq \int_{\tilde{\tau}}^{\hat{\tau}} |g((P - \psi B)e^{\gamma\tau})| e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \Re(\lambda) + \rho(s) ds} d\tau = K_{\lambda_d}^\varepsilon[|g|], \end{aligned}$$

that is, $K_{\lambda_d}[|g|] \geq |g|$. Assume that $K_{\lambda_d}[|g|] > |g|$. As we know that the adjoint eigenvalue of K_{λ_d} (recall $\lambda_d = 0$) for eigenvalue 1 is $f_0 = 1$ (identical 1 on $[P_*, P^*]$, see Corollary 4.2.3), we have

$$\langle 1, |g| \rangle = \langle 1, K_{\lambda_d}[|g|] \rangle > \langle 1, |g| \rangle$$

which is a contradiction. Thus, $K_{\lambda_d}[|g|] = |g|$.

If we let g_d to be the eigenfunction corresponding to the eigenvalue $r(K_{\lambda_d}) = 1$, we can write that $|g| = cg_d$ for some constant c which we may assume to be one (the eigenspace of K_{λ_d} for eigenvalue 1 is one-dimensional, Theorem 4.2.8). Hence $g(P) = g_d(P)e^{i\zeta(P)}$ for some real-valued function $\zeta(P)$. If we substitute this relation into $|K_\lambda[g]| = |g| = g_p = K_{\lambda_d}[g_d]$, we obtain

$$\begin{aligned} &\int_{\tilde{\tau}}^{\hat{\tau}} g_d((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} \rho(\tau) e^{-\int_0^\tau \lambda_d + \rho(s) ds} d\tau \\ &= \left| \int_{\tilde{\tau}}^{\hat{\tau}} g_d((P - \psi B)e^{\gamma\tau}) e^{\gamma\tau} e^{i\zeta(P)} \rho(\tau) e^{-\int_0^\tau \lambda_d + i\eta + \rho(s) ds} d\tau \right|. \end{aligned}$$

From [47, Lemma 6.12], there exists a constant $b \in \mathbb{C}, |b| = 1$ such that $\zeta(P) - \eta\tau = b$. Substituting this relation in $K_\lambda[g] = g$, we obtain

$$e^{ib} \int_{\tilde{\tau}}^{\hat{\tau}} \rho(\tau) g_d[(P - \psi B)e^{\gamma\tau}] e^{\gamma\tau} e^{-\int_0^\tau \lambda_d + \rho(s) ds} d\tau = g_d(p) e^{i\zeta(p)}.$$

Thus, $e^{ib} K_{\lambda_d}[g_d](P) = g_d(P) e^{i\zeta(P)}$, which in turn implies $b = \zeta(P)$, such that $\eta = 0$. Hence, there is no element with real part λ_d apart from λ_d itself. \square

We then show that the spectral gap follows from an argument by Gyllenberg and Heijmans [43].

Theorem 4.2.10. *There exists $\delta > 0$, such that for all $\lambda \in \Lambda^0, \lambda \neq \lambda_d = 0$, it is true that*

$$\Re(\lambda) < \lambda_d - \delta = -\delta.$$

Proof. We know that the desired inequality is true for $\delta = 0$. If no such $\delta > 0$ exists, then there is a sequence of elements $\lambda_n \in \Lambda^0$ with $\Re(\lambda_n) \rightarrow \lambda_d = 0$. For each λ_n , there is a fixed point g_n of K_{λ_n} , which in turn defines an eigenfunction $S_n(\tau, P)$ of eq. (4.2.2). This eigenfunction is also an eigenfunction of T_t for eigenvalue $e^{\lambda_n t}$. As the real values of λ_n converge to $\lambda_d = 0$, the spectrum of the operator T_t has a finite accumulation point. Since T_t is compact for $t > 3\hat{\tau}$ (Theorem 4.2.2), that is impossible. \square

From here, as it is standard for the analysis of aggregation-fragmentation equations [79, 107, 43], a result by Webb [108, 107] immediately implies information about the asymptotic behavior of the semigroup: the underlying Banach space can be decomposed by a spectral projection into the eigenspace of the dominant eigenvalue 0, and a remaining part. If Σ is the stationary solution, then (as 1 is the adjoint eigenfunction), the spectral projector is given by $\Pi[f] = \langle 1, f \rangle \Sigma$. The remaining part $(I - \Pi)[T_t S_0]$ will tend to zero for $t \rightarrow \infty$. To sum it up, we find:

Theorem 4.2.11. *Let $\Sigma \in C^0(\Omega)$ denote the non-negative stationary solution of eq. (4.2.1), normalized to $\|\Sigma\|_{L^1(\Omega)} = 1$. Consider $T(t)S_0$ for a non-negative, non-trivial initial condition $S_0 \in C^0(\Omega)$. Then, with $\tilde{R}(t) = (I - \Pi)[T(t)S_0]$,*

$$T(t)S_0 = \Sigma \langle 1, S_0 \rangle + \tilde{R}(t)$$

and $\tilde{R}(t) \rightarrow 0$ for $t \rightarrow \infty$ exponentially fast in C^0 .

4.3 Reduced Model

We use the theory from Section 4.2 above to reduce the dimension of the model, and investigate the behavior of the resulting equations by numerical simulation.

4.3.1 Fast-Slow Analysis

We intend to use the singular perturbation theory (see Section 1.2.1). As above, we rewrite the semigroup as $\frac{d}{dt}S = AS$. Then, there is a stationary solution $\Sigma(\tau, P) \in C^0(\Omega)$ (an eigenfunction of A for eigenvalue 0). Please note that the generator A depends on the environmental pathogen load B (which we take to be fixed for the moment), such that the eigenfunction Σ also depends on B . The adjoint eigenfunction is simply $\Sigma^*(\tau, P) = 1$, as the semigroup is mass-preserving,

$$\frac{d}{dt} \int_{\Omega} \Sigma^*(\tau, P) S(t, \tau, P) d(\tau, P) = \frac{d}{dt} \int_{\Omega} S(t, \tau, P) d(\tau, P) = 0.$$

Then, Σ and Σ^* are the right and left eigenfunctions of A for eigenvalue 0. Note that $\Sigma^*(\tau, P) = 1$ is independent of B .

We define the spectral projector $\Pi f(\tau, P) = \Sigma(\tau, P; B) \int_{\Omega} f(\tau, P) d(\tau, P) = \Sigma \langle 1, f \rangle$. Then, $A\Pi = \Pi A$ and

$$\frac{d}{dt} \Pi S = A\Pi S = A\Sigma \langle 1, S \rangle = 0, \quad \frac{d}{dt} (I - \Pi)S = A(I - \Pi)S.$$

The definition of the projector Π implies that

$$\Pi S(t, \tau, P) = \Sigma(\tau, P; B) s(t) \quad \text{with} \quad s(t) = \int_{\Omega} S(t, \tau, P) d(\tau, P).$$

We use these two projectors to define a new coordinate system that disentangles the slow and fast dynamics, $S = \Pi S + (I - \Pi)S = \Sigma s + (I - \Pi)S$. Until now, the considerations have been made under the assumption that B is fixed. Even if B (slowly) varies with time, we can still use these new coordinates, but now the projectors Π and $(I - \Pi)$ do not commute with the time derivative. In the case of $I - \Pi$, we simply use the chain rule and the fact that B' scales with $\mathcal{O}(\epsilon)$ to obtain

$$(I - \Pi) \frac{d}{dt} S = \frac{d}{dt} ((I - \Pi)S) + \mathcal{O}(\epsilon).$$

Hence, multiplying the first model equation $\frac{d}{dt} S = AS - \epsilon \beta(P)S$ by $(I - \Pi)$ from the left yields

$$\frac{d}{dt} (I - \Pi)S = A(I - \Pi)S - \epsilon(I - \Pi)[\beta(P)(\Sigma s + (I - \Pi)S)] + \mathcal{O}(\epsilon).$$

In the case of Π , we look slightly more closely, using the fact that the left eigenfunction $\Sigma^* = 1$ does not depend on B , such that

$$\Pi \frac{d}{dt} S(t, \tau, P) = \Sigma(\tau, P) \int_{\Omega} \frac{d}{dt} S(t, \tau, P) d(\tau, P) = \Sigma(\tau, P) \frac{d}{dt} s(t).$$

If we integrate this equation over Ω , we have

$$\int_{\Omega} \Pi \frac{d}{dt} S(t, \tau, P) d(\tau, P) = \frac{d}{dt} s(t).$$

That is, multiplying $\frac{d}{dt} S = AS - \epsilon \beta S$ by Π from the left, and integrating over gives us an equation for $s(t)$ (recall $\Pi A = 0$)

$$\frac{d}{dt} s(t) = -\epsilon s(t) \int_{\Omega} \beta(P) \Sigma(\tau, P; B) d(\tau, P) - \epsilon \langle 1, \beta(P) (I - \Pi)S \rangle.$$

All in all, our model becomes

$$\begin{aligned}\frac{d}{dt}s(t) &= -\epsilon s(t) \int_{\Omega} \beta(P) \Sigma(\tau, P; B) d(\tau, P) - \epsilon \langle 1, \beta(P) (I - \Pi)S \rangle \\ \frac{d}{dt}(I - \Pi)S &= A(I - \Pi)S - \epsilon(I - \Pi)[\beta(P) (\Pi S + (I - \Pi)S)] + \mathcal{O}(\epsilon) \\ \frac{d}{dt}I &= \epsilon \int_{\Omega} \beta(P) (\Sigma(\tau, P; B) s(t) + (I - \Pi)S) dP d\tau - \epsilon \alpha I \\ \frac{d}{dt}R &= \epsilon \alpha I \\ \frac{d}{dt}B &= \epsilon (a + \xi I - \sigma B).\end{aligned}$$

That is, only $(I - \Pi)S$ is the fast variable, while all other variables are slow.

Fast system. If we take $\epsilon = 0$, we find for the fast variable,

$$\frac{d}{dt}(I - \Pi)S = A(I - \Pi)S.$$

Due to the spectral gap, we know that $(I - \Pi)S \rightarrow 0$. Therefore, $(I - \Pi)S = 0$ forms the slow manifold.

Slow system. Now we use the slow time $T = \epsilon t$, and the result for the slow manifold, and obtain the reduced model

$$\frac{d}{dT}s = -s \int_{\Omega} \beta(P) \Sigma(\tau, P; B) d(\tau, P) \quad (4.3.1)$$

$$\frac{d}{dT}I = s \int_{\Omega} \beta(P) \Sigma(\tau, P; B) d(\tau, P) - \alpha I \quad (4.3.2)$$

$$\frac{d}{dT}R = \alpha I \quad (4.3.3)$$

$$\frac{d}{dT}B = a + \xi I - \sigma B. \quad (4.3.4)$$

The model suggests that the distribution of the susceptible population is always in its (quasi) steady state, and hence the force of infection becomes $\int_{\Omega} \beta(P) \Sigma(\tau, P; B) d(\tau, P)$.

4.3.2 Behavior of the Reduced Model: A Simulation Study

We show the dependency of Σ on B .

Lemma 4.3.1. *If $\Sigma(\tau, P)$ is a stationary solution of eq. (4.2.1) for $\psi B = 1$ with $\int_{\Omega} \Sigma d(\tau, P) = 1$, then*

$$S(\tau, P; B) = \frac{1}{\psi B} \Sigma(\tau, P/(\psi B))$$

is a stationary solution for a given value of ψB with $\int_{\Omega} S(\tau, P; B) d(\tau, P) = 1$.

Proof. We suppress the multiplicative factor $\frac{1}{\psi B}$ that only ensures that the norm is preserved. Let $\tilde{P} = P/(\psi B)$,

$$\begin{aligned} \partial_\tau S + \partial_P(-\gamma PS) &= \frac{\partial}{\partial \tau} \Sigma(\tau, P/(\psi B)) + \frac{\partial}{\partial P}(-\gamma P \Sigma(\tau, P/(\psi B))) \\ &= \frac{\partial}{\partial \tau} \Sigma(\tau, \tilde{P}) + \frac{\partial}{\partial \tilde{P}}(-\gamma \tilde{P} \Sigma(\tau, \tilde{P})) \\ &= -\rho(\tau) \Sigma(\tau, P/(\psi B)) = -\rho(\tau) S(\tau, P; B), \end{aligned}$$

and for the boundary value we obtain

$$\int_{\check{\tau}}^{\hat{\tau}} \rho(\tau) S(\tau, (P - \psi B); B) d\tau = \int_{\check{\tau}}^{\hat{\tau}} \rho(\tau) \Sigma(\tau, (P - \psi B)/\psi B) d\tau = \Sigma(0, \tilde{P}) = S(0, P; B).$$

□

As we are not interested in $S(\tau, P)$, but only in the marginal distribution $\int S(\tau, P) d\tau$, it is convenient to use the underlying stochastic process to obtain a numerical approximation of the function (kind of Monte-Carlo integration): we determine the realizations of the random variable T that is distributed according to the time between two booster events. To keep things simple, we use a uniform distribution on $[\check{\tau}, \hat{\tau}]$. With this aspect, it is straightforward to compute a realization of the time course of the pathogen load P_t (also see Figure 4.1). After dismissing a burn-in phase, we sample at discrete time points the values of P_t ; the histogram of those values is proportional to $\int S(\tau, P) d\tau$ (see Figure 4.4).

The transition from susceptible to infected corresponds to massive replication of pathogens, which (at a short time scale), cannot be controlled anymore by the immune system. A branching process, extending from the subcritical to the supercritical parameter range, can function as a toy model for that process. Inspired by that idea, we use the probability of taking off for a branching process as $\beta(P)$ and define $\beta(P) = \beta_0 \max\{0, 1 - \pi_0/P\}$ (Figure 4.4). With these two components and equipped with Lemma 4.3.1, we can determine the incidence's dependency on B . As we have a *SIR* model, we have no stationary solutions to address. However, if we assume that the class of susceptible is reduced relatively slowly, we can identify parameter combinations where infections take off: given that a and I , the environmental pathogen load B asymptotically tends to

$$B = (a + \xi I)/\sigma.$$

If we feed this pathogen level into the incidence, we obtain for the r.h.s. of I'

$$F(a, I) = s \int_{\Omega} \beta(P) \Sigma(\tau, P; B) d(\tau, P) \Big|_{(a+\xi I)/\sigma} - \alpha I.$$

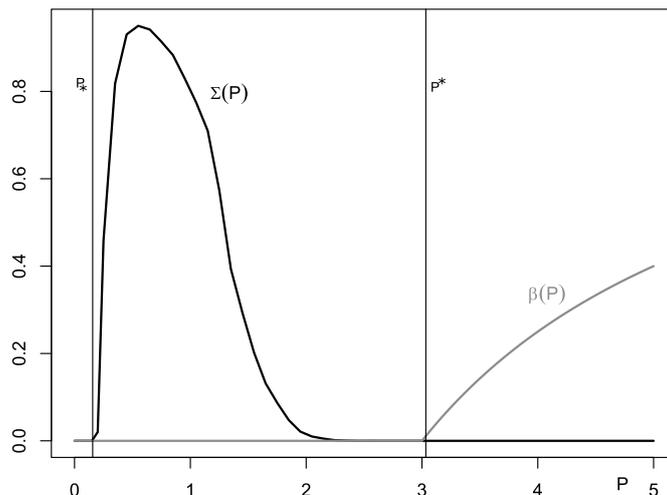


Figure 4.4: Black: equilibrium solution $\Sigma(P) = \int S(\tau, P) d\tau$ (T is assumed to be uniformly distributed between $\check{\tau} = 2$, $\hat{\tau} = 10$, $\psi B = 1$, $\gamma = 0.2$). The support of that function is in the interval $[P_*, P^*]$, as expected. Gray: $\beta(P) = \beta_0 \max\{0, 1 - \pi_0/P\}$, where we use $\pi_0 = 3$ and (for this figure) $\beta_0 = 1$.

The incidence grows if $F(a, I) > 0$, and decreases if $F(a, I) < 0$. We need to emphasize that, here, we use a kind of quasi-steady state for B . In more realistic cases, B and I will change on similar time scales, such that $F(I, a)$ mainly yields a heuristic about the behavior to expect, and not a rigorous threshold argument.

Although cholera is the infection that inspired this model, we are more interested in the model structure than in realistic parameters. Therefore, we do not even try to determine parameters suited for cholera but focus on the discussion of the model for a fairly arbitrary set of parameters (see Appendix A.1), which then yield Figure 4.5. On the left side of the figure, we find a structure resembling typical bistable behavior. This is not a coincidence: If a is small, without additional shedding by infected individuals, the pathogen load in the environment is too small to trigger an epidemic. However, for a certain parameter range of a , positive feedback (infected shed pathogens which in turn additionally infect further individuals) can trigger an epidemic: If I is small, the incidence stays low, if I is large, the number of infections increases further. Only if a is large, such that the pathogen load in the environment becomes supercritical, can an epidemic happen without a considerable number of initial infections.

Accordingly, for the three scenarios shown on the right in Figure 4.5, all parameters are fixed except a , and $I(0) = B(0) = 0$ in all cases. For the chosen parameter values, we observe that there is a threshold value of a between 20.5 and 20.8, which determines

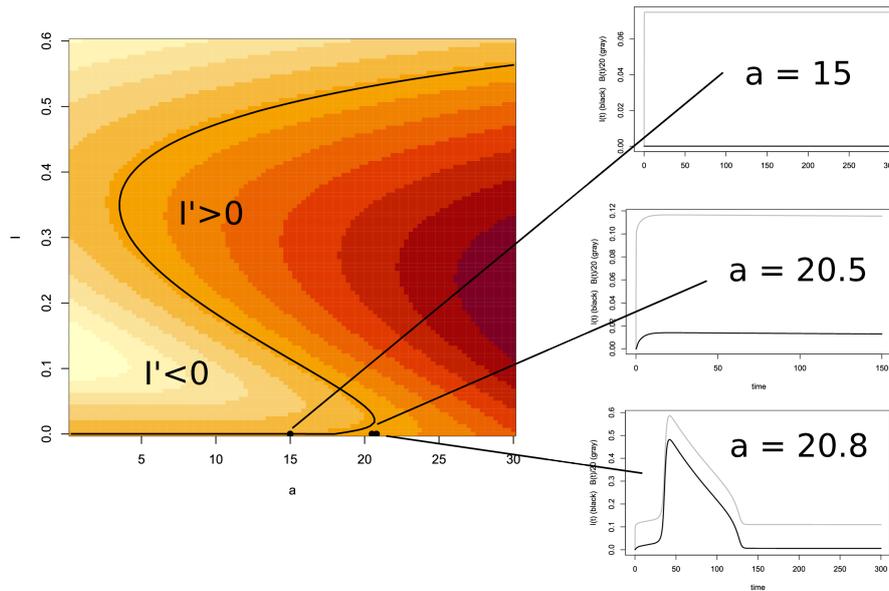


Figure 4.5: Left: $I' = F(a, I)$, right: three simulations differing in the choice of a . Black: $I(t)$, gray: $B(t)/20$ (B is scaled to be comparable with the prevalence). Parameter values are given in the Appendix A.1.

whether an epidemic can take place. For example, for smaller a ($a = 15$), the pathogen load tends to a positive equilibrium, while the incidence is identically zero. The pathogen cannot accumulate sufficiently high in the susceptible host to trigger an infection. If we further increase a ($a = 20.5$), then we again have a positive equilibrium (approximately) for the bacterial compartment, but now there is also a small but positive incidence. The pathogen concentration is large enough to trigger infections occasionally, but we are still in a region, where this small incidence is not able to cause a larger outbreak. Only if we further increase a ($a = 20.8$), then we reach the situation where an epidemic takes place: in the initial time interval, the bacteria reach a plateau, where they are able to increase the prevalence to a certain level. Then, the combined pathogen load due to natural sources (described by a) and shedding is able to trigger the positive feedback, and we obtain a steep increase in the prevalence.

This model behavior can be compared with the incidence of cholera: due to the life cycle of *V. cholerae*, their abundance undergoes seasonal changes, which can be modeled by the different values of a . If they are barely present, the incidence of cholera is zero. However, there are also situations where we have a low, fluctuating incidence and there are distinct outbreaks of larger cholera epidemics. The reason for these different figures is still under discussion, but the main idea is the change of *V. cholerae* abundance, as our model also seems to hint.

It might also be interesting that at least some cholera epidemics share in the onset the

sharp increase in cases, followed by a low decline, e.g., the outbreak in Katsina (Nigeria) in 1982 [102]. An interpretation might be that the pathogen load rapidly increases to a high level, resulting in an increase in prevalence. In contrast to SIR-type models, where the increase is stopped by the depletion of the susceptible class, we rather find here a saturation in the incidence function: also, if B becomes arbitrarily large, the force of infection $\int_{\Omega} \beta(P) \Sigma(\tau, P; B) d(\tau, P)$ stays bounded ($\beta(P)$ is bounded). This property of our model's force of infection is in line with other model approaches [24, 45].

4.4 Summary

We have formulated a multi-scale model where the transition from the susceptible to the infected class is dependent on the within-host pathogen load. This has relaxed the distinction between a susceptible and an infected person. We considered the within-host dynamics to follow a type of velocity-jump process where the jumps are fostered by the ingestion of bacteria from the environment (booster events). We then focused on sub-critical dynamics, which is the case where the pathogen load is below the critical threshold. We took population dynamics to occur on a slower time scale in comparison to within-host dynamics and used semigroup methods and spectral theory to show the existence of the pathogen load's invariant solution on the fast-time scale. We then used the results obtained from the spectral analysis to reduce the dimension of the original multi-scale model on the slow time scale to a SIR model. Numerical simulations done on the SIR model revealed parameter regions where either no cholera cases happened, where cholera was present at a low prevalence, and where a full-blown cholera epidemic took off. Lastly, we also derived an incidence term for the model.

In this chapter, we assume that individual hosts have the same within-host characteristics. We will consider heterogeneity within an individual host in the next chapter.

Chapter 5

Heterogeneity in the Within-Host Model

In this chapter, we aim to study the effects of introducing heterogeneity in the within-host model. Individuals in any given population exhibit distinct characteristics, so it is sensible to include a heterogeneous structure in modeling the dynamics. Structured models are most commonly used to represent population heterogeneity. Simplifications of the somewhat tedious analysis that is associated with structured models can be made by describing the structure of a population by the variation in the population parameters (parametric heterogeneity) such as the birth rate and susceptibility to a disease [54, 57]. That is, parameters are set to be invariant over time in an individual, but they vary from person to person. Models that include parametric heterogeneity help account for the ecological and evolutionary dynamics of a population [89, 34]. For cholera, studies have shown that factors such as host genetics, nutrition, and the chemical state of the gut can affect an individual's susceptibility to the disease [1, 22, 87]. Thus, we propose a model for the disease that accounts for heterogeneity in the susceptibility to the disease within the host. Each individual is assigned a pathogen load whose intrinsic growth depends on the state of the gut microbiome. The individual ingests the pathogen from the environment (consumption of contaminated water) while the pathogen in the environment grows as a result of shedding by individuals. We analyze the asymptotic behavior of the model using Laplace transforms. It is usually difficult to analyze parametric heterogeneous models with direct interactions. In our model, the contact happens through a single compartment (the environmental bacterial compartment B), that is an individual ingests the pathogen from B and excretes it back into B . Thus, individuals decouple given B . This can be used to get an efficient method to analyze the model.

5.1 Model Structure

We consider the bacterial concentration in the environment B and assign each individual a pathogen load p that is obtained through the ingestion of environmental bacteria at rate κ . We take the intrinsic growth of the pathogen to be dependent on the state of the gut microbiome. That is, research has shown that differences in the gut microbiome, which varies from person to person, drive an individual's susceptibility or resistance to a cholera infection [1, 22, 110]. Therefore, we denote the state of the gut microbiome by ω , $\omega \in [0, 1]$ such that the density of the pathogen load for an individual with microbiome state ω at time t is given by $p(t, \omega)$. The intrinsic growth rate of the pathogen is then given as $g(p(t, \omega), \omega)$. The pathogen is removed from the body through bacterial death or shedding at rate γ . In the environment, the bacteria grow through shedding by individuals at rate ξ and decay at rate δ . All in all, the model reads

$$\begin{aligned}\frac{dp(t, \omega)}{dt} &= \kappa B + g(p(t, \omega), \omega) - \gamma p(t, \omega), \\ \frac{dB}{dt} &= \int_0^1 \xi p(t, \omega) \varphi(\omega) d\omega - \delta B,\end{aligned}\tag{5.1.1}$$

where the initial conditions are $p(0, \omega) = p_0(\omega)$, $B(0) = B_0$ and $\varphi(\omega)$ is taken as a density function.

5.2 Analysis of the Model: Linear Case

We aim to first check the fundamental properties of the model and then use the properties as the basis for the evaluation of more complex cases of the model. To that effect, we consider the case where the intrinsic growth rate of the pathogen is represented by a linear term, that is $g(p(t, \omega), \omega) = \alpha(\omega)p(t, \omega)$ such that

$$\begin{aligned}\frac{dp(t, \omega)}{dt} &= \kappa B + \alpha(\omega)p(t, \omega) - \gamma p(t, \omega), \\ \frac{dB}{dt} &= \int_0^1 \xi p(t, \omega) \varphi(\omega) d\omega - \delta B.\end{aligned}\tag{5.2.1}$$

We expect that these results are interesting also in the context of non-linear models as some kind of linearization will most likely lead to a structure that classically resembles the equations we analyze in the present section. However, exploring this idea will not be part of the present thesis. The common strategy for evaluating epidemic models with parametric heterogeneity is to reduce the heterogeneous model to a homogeneous model where the variables are moments of corresponding distributions, and then analyze the resulting system (usually a low dimensional system of ODEs) [89, 88]. Such heterogeneous models assume that the population is closed and the contact process of interacting populations is defined by mass action kinetics (direct interactions). The contact structure in our model is slightly different in that bacteria from the environment are ingested and contribute to pathogen

dynamics within the host. The same is shed back into the environment to contribute to pathogen dynamics outside the host. Thus, the connection is only made through the B compartment. Consequently, in the analysis, we intend to reduce the model (5.2.1) to a linear renewal equation representing the environmental bacterial population with the knowledge that individual pathogen dynamics can be derived from the bacterial population in the environment.

For the analysis, we let $\gamma - \alpha(\omega) = a(\omega)$ and reformulate the model (5.2.1) in terms of a linear Volterra equation (see Section 2.1.1.4 for definition) whose existence and uniqueness can easily be verified. We then derive the expression for the reproduction number, which is useful in determining the long-term behavior of the model, with the help of Laplace transforms. We work with the assumption that $\alpha \in L^\infty(0, 1)$ and $a_- \leq a(\omega) \leq a_+$ for some $a_- \leq 0, a_+ > 0$. We will use the basic linear theory used for structured models in our analysis. Some significant theorems derived from the theory will be summarized in the next sections.

5.2.1 Renewal Equation

Proposition 5.2.1. *Suppose $B(0)$ is given, system (5.1.1) corresponds to the abstract renewal equation*

$$B(t) = G(t) + m \int_0^t K(t-s)B(s)ds$$

with $G(t) = B(0)e^{-\delta t} + \int_0^t \xi F(\tau)e^{-\delta(t-\tau)}d\tau$, $K(t-s) = e^{-\delta(t-s)}H(t-s)$, $m = \xi\kappa$ and $F(t) = \int_0^1 \varphi(\omega)p_0(\omega)e^{-\int_0^t a(\omega)dt}d\omega$, $H(t-s) = \int_0^{t-s} e^{-\delta\theta}T(\theta)d\theta$, $T(t-s) = \int_0^1 \varphi(\omega)e^{-\int_s^t a(\omega)d\tau}d\omega$.

Proof. We assume that $B(0)$ is given, let $\gamma - \alpha(\omega) = a(\omega)$ and solve for $p(t, \omega)$. By variation of constants

$$p(t, \omega) = p_0(\omega)e^{-\int_0^t a(\omega)dt} + \kappa \int_0^t B(s)e^{-\int_s^t a(\omega)d\tau}ds.$$

Plugging $p(t, \omega)$ in the equation of \dot{B} in system (5.1.1) yields

$$\begin{aligned} \frac{dB}{dt} &= \int_0^1 \xi \varphi(\omega) \left(p_0(\omega)e^{-\int_0^t a(\omega)dt} + \kappa \int_0^t B(s)e^{-\int_s^t a(\omega)d\tau}ds \right) d\omega - \delta B \\ &= \xi \int_0^1 \varphi(\omega)p_0(\omega)e^{-\int_0^t a(\omega)dt}d\omega + \xi\kappa \int_0^t \left(\int_0^1 \varphi(\omega)e^{-\int_s^t a(\omega)d\tau}d\omega \right) B(s)ds - \delta B. \end{aligned}$$

Thus

$$\frac{dB}{dt} = \xi F(t) + \xi\kappa \int_0^t T(t-s)B(s)ds - \delta B \quad (5.2.2)$$

with

$$F(t) = \int_0^1 \varphi(\omega)p_0(\omega)e^{-\int_0^t a(\omega)dt}d\omega, \quad T(t-s) = \int_0^1 \varphi(\omega)e^{-\int_s^t a(\omega)d\tau}d\omega.$$

We then integrate eq. (5.2.2) by variation of constants to get the equation for B , that is

$$\begin{aligned} B &= B_0 e^{-\delta t} + \int_0^t e^{-\delta(t-\tau)} \left(\xi F(\tau) + \xi \kappa \int_0^\tau T(\tau-s) B(s) ds \right) d\tau \\ &= B_0 e^{-\delta t} + \int_0^t \xi F(\tau) e^{-\delta(t-\tau)} d\tau + \xi \kappa \int_0^t \int_0^\tau e^{-\delta(t-\tau)} T(\tau-s) B(s) ds d\tau. \end{aligned}$$

We focus on the last term of the expression. Changing the integration variables yields

$$\xi \kappa \int_0^t \int_0^\tau e^{-\delta(t-\tau)} T(\tau-s) B(s) ds d\tau = \xi \kappa \int_0^t \left(\int_s^t e^{-\delta(t-\tau)} T(\tau-s) d\tau \right) B(s) ds.$$

If we let $\theta = \tau - s$, then $d\theta = d\tau$ and

$$\begin{aligned} \xi \kappa \int_0^t \int_0^\tau e^{-\delta(t-\tau)} T(\tau-s) B(s) ds d\tau &= \xi \kappa \int_0^t \left(\int_0^{t-s} e^{-\delta(t-(\theta+s))} T(\theta) d\theta \right) B(s) ds \\ &= \xi \kappa \int_0^t \left(e^{-\delta(t-s)} \int_0^{t-s} e^{-\delta\theta} T(\theta) d\theta \right) B(s) ds \\ &= \int_0^t \xi \kappa e^{-\delta(t-s)} H(t-s) B(s) ds. \end{aligned}$$

with $H(t-s) = \int_0^{t-s} e^{-\delta\theta} T(\theta) d\theta$.

The concentration of bacteria in the environment can now be written as

$$B(t) = B_0 e^{-\delta t} + \int_0^t \xi F(\tau) e^{-\delta(t-\tau)} d\tau + \int_0^t \xi \kappa e^{-\delta(t-s)} H(t-s) B(s) ds. \quad (5.2.3)$$

We rewrite eq. (5.2.3) to obtain the Volterra renewal equation

$$B(t) = G(t) + m \int_0^t K(t-s) B(s) ds \quad (5.2.4)$$

where $G(t) = B_0 e^{-\delta t} + \int_0^t \xi F(\tau) e^{-\delta(t-\tau)} d\tau$, $K(t-s) = e^{-\delta(t-s)} H(t-s)$ and $m = \xi \kappa$. \square

5.2.2 Existence of Solutions

Equation (5.2.4) is the standard Volterra equation addressed in [48, 27, 64]. The usual approach to establishing the existence of solutions to eq. (5.2.4) is to construct Picard iterations and show the convergence of the sequence of Picard iterations, then apply Gronwall's lemma to demonstrate that the solution is unique. We do not show the explicit proof of existence and uniqueness here but refer the reader to [48, Theorem 2.1] for a similar proof.

5.2.3 Asymptotic behaviour

We now investigate the asymptotic behavior of B which is the solution of the renewal equation (5.2.4). We use the theory on Laplace transformations (see Section 1.2.4 for definitions) to establish the asymptotic behavior. To do this, we first show that the functions G and K are Laplace transformable, then rewrite the Laplace transform of B as a convolution and lastly determine the asymptotic behavior of B by investigating the singularities of its Laplace transform. Below, we summarize the theorems derived from the basic linear theory of age-structured models that we use in the analysis.

Theorem 5.2.1. [48, Theorem 2.6] *Let the basic assumptions of the age-structured model (2.1.6) be satisfied. Then, the equation $\hat{K}(\lambda) = 1$ has one and only one real solution α^* , which is a simple root. Moreover*

- $\alpha^* < 0$ if and only if $\int_0^\infty K(t)dt < 1$,
- any other solution α to $\hat{K}(\lambda) = 1$ satisfies $Re(\alpha) < \alpha^*$,
- within any strip $\sigma_1 < Re(\lambda) < \sigma_2$ in the complex plane, there is a finite number of roots of $\hat{K}(\lambda) = 1$.

Theorem 5.2.2. [48, Theorem 2.7] *Let the basic assumptions of the age-structured model (2.1.6) be satisfied and let α^* be the only real root of $\hat{K}(\lambda) = 1$, guaranteed by Theorem 5.2.1. Then, the solution B to the renewal equation $B(t) = F(t) + \int_0^t K(t-s)B(s)ds$ has the following form*

$$B(t) = b_0 e^{\alpha^* t} (1 + \Omega(t))$$

where $b_0 \geq 0$ and $\lim_{t \rightarrow +\infty} \Omega(t) = 0$.

In the next proposition, we show that the functions G and K are Laplace transformable.

Proposition 5.2.2. *There exists c_1 and c_2 in \mathbb{R} such that $|G(t)| \leq M_1 e^{c_1 t}$ and $|K(t)| \leq M_2 e^{c_2 t}$ where M_1, M_2 are constants and $t \geq 0$.*

Proof. The function $G(t) = B_0 e^{-\delta t} + \int_0^t \xi F(\tau) e^{-\delta(t-\tau)} d\tau$, $F(t) = \int_0^1 \varphi(\omega) p_0(\omega) e^{-\int_0^t a(\omega) dt} d\omega$. Then,

$$\begin{aligned} |F(t)| &= \left| \int_0^1 \varphi(\omega) p_0(\omega) e^{-\int_0^t a(\omega) dt} d\omega \right| \\ &\leq \int_0^1 \varphi(\omega) p_0(\omega) |e^{-\int_0^t a(\omega) dt}| d\omega = \int_0^1 \varphi(\omega) p_0(\omega) e^{|\int_0^t a(\omega) dt|} d\omega \leq \bar{c} e^{|\int_0^t a(\omega) dt|} \leq \bar{c}. \end{aligned}$$

and

$$\begin{aligned} G(t) &\leq B_0 e^{-\delta t} + \bar{c} \xi \int_0^t e^{-\delta(t-\tau)} d\tau = B_0 e^{-\delta t} + \bar{c} \xi e^{-\delta t} \int_0^t e^{\delta \tau} d\tau = B_0 e^{-\delta t} + \frac{\bar{c} \xi}{\delta} e^{-\delta t} (e^{\delta t} - 1) \\ G(t) &\leq B_0 e^{-\delta t} - \frac{\bar{c} \xi}{\delta} e^{-\delta t} = (B_0 - \frac{\bar{c} \xi}{\delta}) e^{-\delta t} \end{aligned}$$

Thus $|G(t)| \leq M_1 e^{c_1 t}$ where $M_1 = B_0 + \frac{\bar{c}_1}{\delta}$ and $c_1 = \delta$.

The function $K(t) = e^{-\delta t} H(t)$ with $H(t) = \int_0^t e^{-\delta \theta} T(\theta) d\theta$. Similar to what is done to the function $F(t)$ above, we get $|T(t)| \leq \bar{c}_1$ such that

$$H(t) \leq \bar{c}_1 \int_0^t e^{-\delta \theta} d\theta = \frac{\bar{c}_1}{\delta} (1 - e^{-\delta t}) \leq \frac{\bar{c}_1}{\delta}.$$

Thus $K(t) \leq e^{-\delta t} \frac{\bar{c}_1}{\delta}$ and $|K(t)| \leq M_2 e^{c_2 t}$ with $M_2 = \frac{\bar{c}_1}{\delta}$ and $c_2 = \delta$. \square

By Proposition 5.2.2 the function B is now Laplace transformable and can be written in terms of a Laplace convolution. That is

$$\begin{aligned} B(t) &= G(t) + m \int_0^t K(t-s) B(s) ds \\ \hat{B}(\lambda) &= \hat{G}(\lambda) + m \hat{K}(\lambda) \hat{B}(\lambda) \implies \hat{B}(\lambda) = \frac{\hat{G}(\lambda)}{1 - m \hat{K}(\lambda)} \end{aligned}$$

We rewrite $\hat{B}(\lambda)$ as

$$\hat{B}(\lambda) = \hat{G}(\lambda) + \frac{m \hat{K}(\lambda)}{1 - m \hat{K}(\lambda)} \hat{G}(\lambda). \quad (5.2.5)$$

From eq. (5.2.5) $\hat{B}(\lambda)$ is meromorphic, that is, it has no singularities other than poles, and its poles are among the roots λ of the equation

$$m \hat{K}(\lambda) = 1 \quad (5.2.6)$$

We refer to eq. (5.2.6) as the characteristic equation. From Theorem 5.2.2, the integral equation (5.2.4) has a solution of the form

$$B(t) = B_0 e^{\nu^* t} (1 + \Omega(t)), \quad B_0 \geq 0, \quad \lim_{t \rightarrow +\infty} \Omega(t) = 0, \quad \Omega(t) \in C^0(\mathbb{R}_+) \quad (5.2.7)$$

where ν^* is the intrinsic Malthusian parameter of the bacterial population that determines the growth or reduction of the bacterial population. We relate the Malthusian parameter and the value $m \hat{K}(0)$ using the following lemma.

Lemma 5.2.1. *Given the operator ν^* from eq. (5.2.7), the following relation holds*

$$\begin{aligned} \nu^* < 1 & \text{ if } m \hat{K}(0) < 1 \\ \nu^* = 1 & \text{ if } m \hat{K}(0) = 1 \\ \nu^* > 1 & \text{ if } m \hat{K}(0) > 1. \end{aligned}$$

Thus, we can conjecture the basic reproduction number as $\mathcal{R}_0 = m \hat{K}(0)$.

5.3 Summary

We have formulated a model where individual host heterogeneity is taken into account by the dependence of the intrinsic pathogen growth rate on the state of the gut microbiome. Pathogens within the host are acquired through the ingestion of environmental pathogens while environmental pathogens grow through shedding by individuals. The contact structure of the model is thus a single compartment, unlike other models with parametric heterogeneity. We have considered the simple case where the intrinsic pathogen growth rate is linear and using decoupling assumptions reduced the model to a Volterra equation. Finally, we analyzed the asymptotic properties of the equation and derived an expression for the reproduction number with the help of Laplace transforms.

Chapter 6

Discussion and Conclusion

6.1 Discussion

In this thesis, we have formulated different models to link the within-host and between-host dynamics of cholera. In Chapter 3, we developed a multi-scale model for cholera by structuring the epidemic model using within-host immune dynamics. The immunological model considered the interaction between the pathogen and the adaptive immune response. Through the separation of time scales, we analyzed the dynamics within the host, with pathogen dynamics being considered to change faster than the immune response. From the slow system dynamics, we were able to characterize a single infected individual by the state of their immune response. We then scaled up the dynamics of a single infected individual to structure the infected population based on within-host immune dynamics. We showed the existence of equilibrium points and derived the expression for the basic reproduction number, which was the threshold condition for the spread of the disease. We analyzed the local stability of the equilibrium points using linearization techniques. We then modified the between-host model to include a maximum age for the bacteria and used Lyapunov results from Meehan et al. [76] to establish the global stability of the equilibrium points for the modified model. Finally, we showed the conditions for equivalence of the original to the modified model.

The work has advanced the use of methods in physiologically structured population models [42, 29, 7, 72] in the study of cholera and other diseases. In our case, the growth of the immune response is the velocity vector that describes the changes in the physiological variable. Similar to [105, 91], the model has provided a structure for utilizing the separation of time-scale methods derived from singular perturbation theory [82, 62] to simplify the analysis of disease dynamics. Unlike other within-host cholera models, our modeling approach allows for the possibility of recovery since the pathogen can be cleared after a finite time. Subsequently, the minimum pathogen threshold dynamics for activation of the immune response align with experimental studies that show that a critical infectious dose is

required for the infection [40]. The basic reproduction number \mathcal{R}_0 of the model represents the contribution of the bacteria from the environment (indirect-transmission route) and the human-human (direct transmission) contribution to the infection process, a result that is consistent with single scale models for the disease [81, 99]. This suggests that a reduction in the reproduction number might be enhanced by intervention strategies targeting pathogen eradication at both human and environmental levels as seen in [104, 78]. However, the dependence of \mathcal{R}_0 on immunological variables illustrates the difference between the model and the fore-mentioned models. For the DFE, the disease would be eradicated if $\mathcal{R}_0 < 1$ and persist otherwise. A unique endemic equilibrium existed when $\mathcal{R}_0 > 1$ and in the case of no loss of immunity, the endemic equilibrium was locally asymptotically stable. For the modified epidemic model, both the DFE and the endemic equilibrium were globally asymptotically stable.

We have been able to get an appropriate formulation of the I dynamics taking into account the within-host dynamics. However, an appropriate formulation of the influence of population-level interactions on the within-host pathogen load is lacking. For instance, studies have shown that environmental vibrios shape the infectious dose depending on their nature (hyperinfectious or less infectious) [45, 87, 92] which in turn can affect within-host pathogen evolution. This contribution of environmental vibrios to the within-host process is neglected in the model. Additionally, the underlying assumption of the work is that one infectious contact is enough to push contacts over the threshold. However, a critical pathogen threshold should be exceeded for the infection to occur and for subcritical pathogen loads one infectious contact may not be enough. We address these issues in the next model (Chapter 4).

In Chapter 4 of the thesis, we first developed a multi-scale model (taking into account the shortcomings of the previous model) and then used first principles to reduce the dimension of the model to a SIR model. This was a novel framework for linking the within-host and between-host dynamics of cholera and the rigorous mathematical analysis that we provided could be applied to other multi-scale disease models. We did this by assigning each susceptible individual with a pathogen load that grew through the uptake of contaminated food and water from the environment (booster event) and declined between two booster events through elimination by the immune system. The transition from the susceptible to the infected class took place at a certain pathogen-load-dependent rate. This rate was only positive if a critical pathogen threshold was surpassed. Furthermore, we assumed population dynamics to happen on a slower time scale than within-host dynamics. We analyzed the model on a fast time scale and showed the existence of an invariant solution. We did this by constructing a semigroup of the model and analyzing the spectrum of its infinitesimal generator. We then used the results obtained from the spectral analysis to reduce the dimension of the original multi-scale model on the slow time scale to a SIR model and performed numerical simulations on the resulting model.

Similar to the previous model and other multi-scale models for environmentally driven diseases [105, 91, 37] separation of time scales allowed the simplification of the analysis and more so the reduction of the model to a lower dimension. Further, the work provided

a mathematical framework for utilizing the methods used in aggregation-fragmentation models [47, 97] in the analysis of multi-scale disease models. The interplay between the booster events and the immune responses when the pathogen load is sub-critical can be compared to the mechanisms of integrate and fire neural models, which have been used before to address the spread of infections [100, 67]. The underlying mechanism of our model was a velocity-jump process, where the jumps were fostered by booster events. Velocity-jump processes are close to age-structured models, as both consist of a transport equation with a non-local term. This non-local term, however, is focused at age 0 in age-structured models (the boundary condition—which can be newborns or new infections in other cases—addresses age zero, see, e.g., [14, 52]), but is distributed over all states in velocity-jump processes [42]. With that, velocity-jump processes are often more complex to analyze [31, 98]. The numerical simulations showed three different parameter regions for the infection dynamics: In the first region, the *V. cholerae* bacteria were present in the environment but there were no cholera cases (incidence is zero); in the next region, few cholera cases would arise once in a while (the incidence was low such that larger outbreaks could not occur), and in the third parameter region, a full outbreak of the disease could be observed. Through the results of these simulations, we can qualitatively explain the different outcomes that occur when the disease is introduced into a region. The abundance of bacteria in the environment was seen as a driving force in the occurrence of an epidemic. This was in line with other studies [24] that highlight the role of environmental reservoirs in the infection process. The incidence term derived from our model was close to other saturated incidence terms used for the disease [24, 45, 81]. These terms have been found to be more realistic in modeling the dynamics. However, our incidence can drop to zero if the pathogen levels are low, in contrast with other models that maintain positive values. Moreover, the incidence term still does not adequately capture the transition from susceptible to infected based on the dynamics within the host.

In the two previous models, we introduced heterogeneity in the multi-scale models by structuring the epidemic model by within-host dynamics. In chapter 5 of the thesis, we used a different approach to address heterogeneity. Here, the nature of the gut microbiome of an individual led to differences in susceptibility to the disease, resulting in heterogeneity in the model. We only considered the interaction between the pathogen load in an individual and the pathogen in the environment. We assigned each individual a pathogen load that was ingested from the environment and grew intrinsically at a rate dependent on the state of the gut microbiome. The growth of environmental bacteria was due to the shedding of bacteria by individuals. In the simple case, we analyzed the model with a linear intrinsic pathogen growth rate. Since the contact structure was through a single compartment (the environmental bacteria), we used decoupling properties to reduce the model to a linear renewal equation. Finally, we analyzed the asymptotic properties of the model with the help of Laplace transforms.

The work provides a framework for modeling parametric heterogeneity in within-host cholera models that can be applied to other diseases. Unlike the models in [89, 88], which use moment-generating functions in the analysis, the contact pattern in our model

is through a single compartment and that allows us to use decoupling properties in the mathematical analysis. The renewal equation resulting from the reduction of the model is similar to that of age-structured epidemic models [64, 49, 48], though in our case it is related to the environmental bacterial population rather than the total number of births. Although we have only looked at the case where the intrinsic pathogen growth rate is linear, we intend to explore other cases that can allow us to appropriately define individuals in susceptible or infected states.

6.2 Conclusion

Mathematical models have played an effective role in deepening the understanding of infectious disease dynamics. From explaining transmission phenomena to forecasting future risks and advising on intervention strategies, the benefits of the models cannot be understated. For diseases like cholera, where the infection starts in an individual and spreads to the population, all while interacting with the environment, robust models are needed to gain insights into the dynamics. Multi-scale modeling has been seen as an effective method for obtaining novel insights into transmission dynamics at different scales.

In this thesis, we have presented three different approaches to modeling the dynamics of cholera on multiple scales. In Chapters 3 and 4, we used physiologically structured variables to link the within and between-host dynamics and provided the mathematical theory for the analysis of such models. In Chapter 4, we provided the theory for refining multi-scale models using first principles. Lastly, in Chapter 5, we introduced heterogeneity in an individual which is a close representation of what happens in the population. In all the approaches, the epidemic model utilizes a special structure that includes the bacterial compartment in the environment which is different from standard SIR models.

Although the work has advanced the mathematical theory for multi-scale modeling of cholera models, the principal difficulty is still formulating an appropriate incidence term for the transition from susceptible to infected states based on within-host dynamics. This is an area that can be explored further to create robust models. Other areas of the work that can be extended include: The model in Chapter 3 can be extended by performing an elasticity analysis on the reproduction number to test its dependence on the immunological parameters.

In Chapter 4, we were interested in model structure and mathematical theory, thus the parameters used in numerical simulations were not specific to cholera. Parameters derived from empirical findings or parameter estimation studies could be used to make the model more sophisticated.

A non-linear pathogen growth rate can be considered in Chapter 5, for instance, a cubic growth term, to create a distinction between an infected host and a susceptible host based on the pathogen load.

Further research opportunities to be explored in multi-scale cholera modeling, include the two-way coupling of disease dynamics: within-to-between and between-to-within.

Appendix A

Appendix

A.1 Parameter values for Figure (4.5)

We again emphasize that we did not try to obtain realistic parameter values for cholera, as the aim of the work was to make a proposal for a modeling approach, and the analysis of this approach. For Figure (4.5), we used the following, rather arbitrary parameter values.

Parameter	Value
$\check{\tau}$	2
$\hat{\tau}$	10
γ	0.2
α	12
ξ	2000
σ	10
$\beta(P)$	$\beta_0 \max\{0, 1 - \pi_0/P\}$
β_0	0.01
π_0	3
$s(0)$	100
$I(0), B(0)$	0

Table A.1: Parameter values.

List of Tables

A.1	Parameter values.	91
-----	---------------------------	----

Bibliography

- [1] Salma Alavi, Jonathan D Mitchell, Jennifer Y Cho, Rui Liu, John C Macbeth, and Ansel Hsiao. Interpersonal gut microbiome variation drives susceptibility and resistance to cholera infection. *Cell*, 181(7):1533–1546, 2020.
- [2] Mohammad Ali, Allyson R Nelson, Anna Lena Lopez, and David A Sack. Updated global burden of cholera in endemic countries. *PLoS neglected tropical diseases*, 9(6): e0003832, 2015.
- [3] Warder Clyde Allee. Animal aggregations. *Nature*, 128:940–941, 1931.
- [4] Salvador Almagro-Moreno, Kali Pruss, and Ronald K Taylor. Intestinal colonization dynamics of vibrio cholerae. *PLoS pathogens*, 11(5):e1004787, 2015.
- [5] Alexis Erich S Almcera, Van Kinh Nguyen, and Esteban A Hernandez-Vargas. Multiscale model within-host and between-host for viral infectious diseases. *Journal of Mathematical Biology*, 77(4):1035–1057, 2018.
- [6] Jason R Andrews and Sanjay Basu. Transmission dynamics and control of cholera in haiti: an epidemic model. *The Lancet*, 377(9773):1248–1255, 2011.
- [7] Oscar Angulo, Fabio Milner, and Laurentiu Sega. A sir epidemic model structured by immunological variables. *Journal of Biological Systems*, 21(04):1340013, 2013.
- [8] Jie Bai, Chayu Yang, Xueying Wang, and Jin Wang. Modeling the within-host dynamics of cholera: Bacterial-viral-immune interaction. *Journal of Applied Analysis & Computation*, 11(2):690–710, 2021.
- [9] Maria Vittoria Barbarossa and Gergely Röst. Immuno-epidemiology of a population structured by immune status: a mathematical study of waning immunity and immune system boosting. *Journal of Mathematical Biology*, 71(6-7):1737–1770, 2015.
- [10] Richard Bertram and Jonathan E Rubin. Multi-timescale systems and fast-slow analysis. *Mathematical biosciences*, 287:105–121, 2017.

- [11] Fred Brauer, Zhisheng Shuai, and P Van Den Driessche. Dynamics of an age-of-infection cholera model. *Mathematical Biosciences & Engineering*, 10(5-6):1335–49, 2013.
- [12] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical models in epidemiology*, volume 32. Springer, 2019.
- [13] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.
- [14] Li-Ming Cai, Chairat Modnak, and Jin Wang. An age-structured model for cholera control with vaccination. *Applied Mathematics and Computation*, 299:127–140, 2017.
- [15] Liming Cai, Zhaoqing Li, Chayu Yang, and Jin Wang. Global analysis of an environmental disease transmission model linking within-host and between-host dynamics. *Applied mathematical modelling*, 86:404–423, 2020.
- [16] Àngel Calsina and Joan Saldaña. A model of physiologically structured population dynamics with a nonlinear individual growth rate. *Journal of Mathematical Biology*, 33:335–364, 1995.
- [17] Fabien Campillo, Nicolas Champagnat, and Coralie Fritsch. Links between deterministic and stochastic approaches for invasion in growth-fragmentation-death models. *Journal of mathematical biology*, 73(6):1781–1821, 2016.
- [18] V Capasso and SL Paveri-Fontana. A mathematical model for the 1973 cholera epidemic in the european mediterranean region. *Revue d'épidémiologie et de sante publique*, 27(2):121–132, 1979.
- [19] Xiuli Cen, Zhilan Feng, and Yulin Zhao. Emerging disease dynamics in a model coupling within-host and between-host systems. *Journal of theoretical biology*, 361:141–151, 2014.
- [20] Nigel Chaffey. Alberts, b., johnson, a., lewis, j., raff, m., roberts, k. and walter, p. molecular biology of the cell. 4th edn., 2003.
- [21] Lauren M Childs, Fadoua El Moustaid, Zachary Gajewski, Sarah Kadelka, Ryan Nikin-Beers, John W Smith Jr, Melody Walker, and Leah R Johnson. Linked within-host and between-host models and data for infectious diseases: a systematic review. *PeerJ*, 7:e7057, 2019.
- [22] Jennifer Y Cho, Rui Liu, John C Macbeth, and Ansel Hsiao. The interface of vibrio cholerae and the gut microbiome. *Gut microbes*, 13(1):1937015, 2021.
- [23] Stanca M Ciupe and Jane M Heffernan. In-host modeling. *Infectious Disease Modelling*, 2(2):188–202, 2017.

- [24] Cláudia Torres Codeço. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infectious diseases*, 1(1):1–14, 2001.
- [25] Rita R Colwell. Global climate and infectious disease: the cholera paradigm. *Science*, 274(5295):2025–2031, 1996.
- [26] Jim Michael Cushing. *An introduction to structured population dynamics*. SIAM, 1998.
- [27] Brian Davies. *Integral transforms and their applications*, volume 25. Springer Science & Business Media, 1978.
- [28] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 1990.
- [29] Odo Diekmann, Mats Gyllenberg, and Johan Metz. Physiologically structured population models: towards a general mathematical theory. In *Mathematics for ecology and environmental sciences*, pages 5–20. Springer, 2007.
- [30] Gustav Doetsch. *Introduction to the Theory and Application of the Laplace Transformation*. Springer Science & Business Media, 2012.
- [31] Marie Doumic Jauffret and Pierre Gabriel. Eigenelements of a general aggregation-fragmentation model. *Mathematical Models and Methods in Applied Sciences*, 20(05):757–783, 2010.
- [32] John M Drake and Andrew M Kramer. Allee effects. *Nature Education Knowledge*, 3(10):2, 2011.
- [33] Roman Drnovšek. Spectral inequalities for compact integral operators on banach function spaces. *Mathematical Proceedings of the Cambridge Philosophical Society*, 112(3):589–598, 1992.
- [34] Jonathan Dushoff. Host heterogeneity and disease endemicity: a moment-based approach. *Theoretical population biology*, 56(3):325–335, 1999.
- [35] Klaus-Jochen Engel, Rainer Nagel, and Simon Brendle. *One-parameter semigroups for linear evolution equations*, volume 194. Springer, 2000.
- [36] Zhilan Feng, Jorge Velasco-Hernandez, Brenda Tapia-Santos, and Maria Conceição A Leite. A model for coupling within-host and between-host dynamics in an infectious disease. *Nonlinear Dynamics*, 68(3):401–411, 2012.
- [37] Zhilan Feng, Jorge Velasco-Hernandez, and Brenda Tapia-Santos. A mathematical model for coupling within-host and between-host dynamics in an environmentally-driven infectious disease. *Mathematical Biosciences*, 241(1):49–55, 2013.

- [38] Winston Garira. A complete categorization of multiscale models of infectious disease systems. *Journal of Biological Dynamics*, 11(1):378–435, 2017.
- [39] Winston Garira, Dephney Mathebula, and Rendani Netshikweta. A mathematical modelling framework for linked within-host and between-host dynamics for infections with free-living pathogens in the environment. *Mathematical biosciences*, 256:58–78, 2014.
- [40] Aaron Nicholas Gillman, Anel Mahmutovic, Pia Abel zur Wiesch, and Sören Abel. The infectious dose shapes vibrio cholerae within-host dynamics. *Msystems*, 6(6): e00659–21, 2021.
- [41] Urs Graf. *Applied Laplace transforms and z-transforms for scientists and engineers: a computational approach using a Mathematica package*. Birkhäuser, 2012.
- [42] Mats Gyllenberg. Mathematical aspects of physiologically structured populations: the contributions of jaj metz. *Journal of biological dynamics*, 1(1):3–44, 2007.
- [43] Mats Gyllenberg and Henk JAM Heijmans. An abstract delay-differential equation modelling size dependent cell growth and division. *SIAM journal on mathematical analysis*, 18(1):74–88, 1987.
- [44] Jason B Harris, Regina C LaRocque, Fahima Chowdhury, Ashraful I Khan, Tanya Logvinenko, Abu SG Faruque, Edward T Ryan, Firdausi Qadri, and Stephen B Calderwood. Susceptibility to vibrio cholerae infection in a cohort of household contacts of patients with cholera in bangladesh. *PLoS neglected tropical diseases*, 2(4):e221, 2008.
- [45] David M Hartley, J Glenn Morris Jr, and David L Smith. Hyperinfectivity: a critical element in the ability of v. cholerae to cause epidemics? *PLoS Med*, 3(1):e7, 2005.
- [46] Hendricus Johannes Adrianus Maria Heijmans. An eigenvalue problem related to cell growth. *Journal of mathematical analysis and applications*, 111(1):253–280, 1985.
- [47] Henk JAM Heijmans. The dynamical behaviour of the age-size-distribution of a cell population. In *The dynamics of physiologically structured populations*, pages 185–202. Springer, 1986.
- [48] Mimmo Iannelli and Fabio Milner. The basic approach to age-structured population dynamics. *Models Methods and Numerics*, 10:978–94, 2017.
- [49] Hisashi Inaba. *Age-structured population dynamics in demography and epidemiology*. Springer, 2017.
- [50] Harris JB, LaRocque RC, Qadri F, Ryan ET, and Calderwood SB. Cholera. *Lancet*, 379:2466–2476, 2012.

- [51] Matthew Jemielita, Ned S Wingreen, and Bonnie L Bassler. Quorum sensing controls vibrio cholerae multicellular aggregate formation. *Elife*, 7:e42057, 2018.
- [52] Xin Jiang. Global dynamics for an age-structured cholera infection model with general infection rates. *Mathematics*, 9(23):2993, 2021.
- [53] JB Kaper, JG Morris, and MM Levine. Cholera. *Clinical Microbiology Reviews*, 8(1):48–86, 1995.
- [54] Georgy P Karev and Artem S Novozhilov. How trait distributions evolve in populations with parametric heterogeneity. *Mathematical Biosciences*, 315:108235, 2019.
- [55] Nobuyuki Kato and Hiroyuki Torikata. Local existence for a general model of size-dependent population dynamics. *Abstract and Applied Analysis*, 2(3-4):207 – 226, 1997.
- [56] RajReni B Kaul, Andrew M Kramer, Fred C Dobbs, and John M Drake. Experimental demonstration of an allee effect in microbial populations. *Biology Letters*, 12(4):20160070, 2016.
- [57] Lindsay T Keegan and Jonathan Dushoff. Estimating finite-population reproductive numbers in heterogeneous populations. *Journal of theoretical biology*, 397:1–12, 2016.
- [58] MJ Keeling and L Danon. Mathematical modelling of infectious diseases. *British medical bulletin*, 92(1), 2009.
- [59] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.
- [60] William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics.ii. the problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 138(834):55–83, 1932.
- [61] MY Kim and FA Milner. A mathematical model of epidemics with screening and variable infectivity. *Mathematical and Computer Modelling*, 21(7):29–42, 1995.
- [62] Christian Kuehn. *Multiple time scale dynamics*, volume 191. Springer, 2015.
- [63] MM Levine, DR Nalin, MB Rennels, RB Hornick, S Sotman, G Van Blerk, TP Hughes, S O’Donnell, and D Barua. Genetic susceptibility to cholera. *Annals of human biology*, 6(4):369–374, 1979.
- [64] Xue-Zhi Li, Junyuan Yang, and Maia Martcheva. *Age Structured Epidemic Modeling*, volume 52. Springer Nature, 2020.

- [65] Jiazhe Lin, Rui Xu, and Xiaohong Tian. Global dynamics of an age-structured cholera model with both human-to-human and environment-to-human transmissions and saturation incidence. *Applied Mathematical Modelling*, 63:688–708, 2018.
- [66] Carla Lutz, Martina Erken, Parisa Noorian, Shuyang Sun, and Diane McDougald. Environmental reservoirs and mechanisms of persistence of vibrio cholerae. *Frontiers in microbiology*, 4:375, 2013.
- [67] Ilias N Lympieropoulos. # stayhome to contain covid-19: Neuro-sir–neurodynamical epidemic modeling of infection patterns in social networks. *Expert Systems with Applications*, 165:113970, 2021.
- [68] Pierre Magal and Shigui Ruan. *Structured population models in biology and epidemiology*, volume 1936. Springer, 2008.
- [69] Lorenzo Mari, Enrico Bertuzzo, Lorenzo Righetto, Renato Casagrandi, Marino Gatto, Ignacio Rodriguez-Iturbe, and Andrea Rinaldo. Modelling cholera epidemics: the role of waterways, human mobility and sanitation. *Journal of the royal society interface*, 9(67):376–388, 2012.
- [70] Maia Martcheva. *An introduction to mathematical epidemiology*, volume 61. Springer, 2015.
- [71] Maia Martcheva and Xue-Zhi Li. Linking immunological and epidemiological dynamics of hiv: the case of super-infection. *Journal of biological dynamics*, 7(1):161–182, 2013.
- [72] Maia Martcheva and Sergei S Pilyugin. An epidemic model structured by host immunity. *Journal of Biological Systems*, 14(02):185–203, 2006.
- [73] Maia Martcheva and Horst R Thieme. Progression age enhanced backward bifurcation in an epidemic model with super-infection. *Journal of Mathematical Biology*, 46(5):385–424, 2003.
- [74] Maia Martcheva, Necibe Tuncer, and Colette St Mary. Coupling within-host and between-host infectious diseases models. *Biomath*, 4(2):1510091, 2015.
- [75] AG McKendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1926.
- [76] Michael T Meehan, Daniel G Cocks, Johannes Müller, and Emma S McBryde. Global stability properties of a class of renewal epidemic models. *Journal of Mathematical Biology*, 78(6):1713–1725, 2019.
- [77] Johan A Metz and Odo Diekmann. *The dynamics of physiologically structured populations*, volume 68. Springer, 1986.

- [78] Rachael L Miller Neilan, Elsa Schaefer, Holly Gaff, K Renee Fister, and Suzanne Lenhart. Modeling optimal intervention strategies for cholera. *Bulletin of mathematical biology*, 72(8):2004–2018, 2010.
- [79] Mustapha Mokhtar-Kharroubi. On spectral gaps of growth-fragmentation semigroups with mass loss or death. *Communications on Pure & Applied Analysis*, 2022.
- [80] J Glenn Morris Jr. Cholera - modern pandemic disease of ancient lineage. *Emerging infectious diseases*, 17(11):2099, 2011.
- [81] Zindoga Mukandavire, Shu Liao, Jin Wang, Holly Gaff, David L Smith, and J Glenn Morris. Estimating the reproductive numbers for the 2008–2009 cholera outbreaks in zimbabwe. *Proceedings of the National Academy of Sciences*, 108(21):8767–8772, 2011.
- [82] Johannes Müller and Christina Kuttler. Methods and models in mathematical biology. *Lecture Notes on Mathematical Modelling in Life Sciences*, Springer, Berlin, 2015.
- [83] Beryl Musundi. An immuno-epidemiological model linking between-host and within-host dynamics of cholera. *arXiv preprint arXiv:2105.12675*, 2021.
- [84] Beryl Musundi, Johannes Müller, and Zhilan Feng. A multi-scale model for cholera outbreaks. *Mathematics*, 10(17):3114, 2022.
- [85] Beryl O Musundi, George O Lawi, and Fredrick O Nyamwala. Mathematical analysis of a cholera transmission model incorporating media coverage. *Int. J. Pure Appl. Math*, 111(2):219–231, 2016.
- [86] Yukihiko Nakata, Yoichi Enatsu, Hisashi Inaba, Toshikazu Kuniya, Yoshiaki Muroya, Yasuhiro Takeuchi, et al. Stability of epidemic models with waning immunity. *SUT J. Math*, 50(2):205–245, 2014.
- [87] Eric J Nelson, Jason B Harris, J Glenn Morris, Stephen B Calderwood, and Andrew Camilli. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nature Reviews Microbiology*, 7(10):693–702, 2009.
- [88] Artem S Novozhilov. On the spread of epidemics in a closed heterogeneous population. *Mathematical biosciences*, 215(2):177–185, 2008.
- [89] Artem S Novozhilov. Epidemiological models with parametric heterogeneity: Deterministic theory for closed populations. *Mathematical Modelling of Natural Phenomena*, 7(3):147–167, 2012.
- [90] Amnon Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44. Springer Science & Business Media, 2012.

- [91] Conrad Ratchford and Jin Wang. Modeling cholera dynamics at multiple scales: environmental evolution, between-host transmission, and within-host interaction. *Mathematical Biosciences and Engineering*, 16(2):782–812, 2019.
- [92] Joachim Reidl and Karl E Klose. *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiology Reviews*, 26(2):125–139, 2002.
- [93] David A Sack, R Bradley Sack, G Balakrish Nair, and AK Siddique. Cholera. *The Lancet*, 363(9404):223–233, 2004. ISSN 0140-6736.
- [94] Katharine A Schilling, Emily J Cartwright, John Stamper, Michael Locke, Douglas H Esposito, Victor Balaban, and Eric Mintz. Diarrheal illness among us residents providing medical services in haiti during the cholera epidemic, 2010 to 2011. *Journal of travel medicine*, 21(1):55–57, 2014.
- [95] Paul Schmid-Hempel and Steven A Frank. Pathogenesis, virulence, and infective dose. *PLoS pathogens*, 3(10):e147, 2007.
- [96] Hal L Smith and Patrick De Leenheer. Virus dynamics: a global analysis. *SIAM Journal on Applied Mathematics*, 63(4):1313–1327, 2003.
- [97] Eva Stadler. Eigensolutions and spectral analysis of a model for vertical gene transfer of plasmids. *Journal of Mathematical Biology*, 78(5):1299–1330, 2019.
- [98] Eva Stadler and Johannes Müller. Analyzing plasmid segregation: Existence and stability of the eigensolution in a non-compact case. *Discrete & Continuous Dynamical Systems-B*, 25(11):4127, 2020.
- [99] Joseph H Tien and David JD Earn. Multiple transmission pathways and disease dynamics in a waterborne pathogen model. *Bulletin of mathematical biology*, 72(6):1506–1533, 2010.
- [100] Henry C Tuckwell. *Stochastic processes in the neurosciences*. SIAM, 1989.
- [101] Julius Tumwiine, JYT Mugisha, and Livingstone S Luboobi. A mathematical model for the dynamics of malaria in a human host and mosquito vector with temporary immunity. *Applied mathematics and computation*, 189(2):1953–1965, 2007.
- [102] JU Umoh, AA Adesiyun, JO Adekeye, and M Nadarajah. Epidemiological features of an outbreak of gastroenteritis/cholera in katsina, northern nigeria. *Epidemiology & Infection*, 91(1):101–111, 1983.
- [103] David M Vickers and Nathaniel D Osgood. A unified framework of immunological and epidemiological dynamics for the spread of viral infections in a simple network-based population. *Theoretical Biology and Medical Modelling*, 4(1):1–13, 2007.

- [104] Jin Wang and Chairat Modnak. Modeling cholera dynamics with controls. *Canadian applied mathematics quarterly*, 19(3):255–273, 2011.
- [105] Xueying Wang and Jin Wang. Disease dynamics in a coupled cholera model linking within-host and between-host interactions. *Journal of Biological Dynamics*, 11(sup1): 238–262, 2017.
- [106] Xueying Wang and Jin Wang. Modeling the within-host dynamics of cholera: bacterial–viral interaction. *Journal of Biological Dynamics*, 11(sup2):484–501, 2017.
- [107] Glenn F Webb. An operator-theoretic formulation of asynchronous exponential growth. *Transactions of the American Mathematical Society*, 303(2):751–763, 1987.
- [108] Glenn F Webb et al. *Theory of nonlinear age-dependent population dynamics*. CRC Press, 1985.
- [109] World Health Organization. Cholera, 2022. <https://www.who.int/news-room/fact-sheets/detail/cholera>.
- [110] Tanya Yatsunenko, Federico E Rey, Mark J Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N Baldassano, Andrey P Anokhin, et al. Human gut microbiome viewed across age and geography. *nature*, 486(7402):222–227, 2012.
- [111] Kōsaku Yosida. *Functional analysis*. Springer Science & Business Media, 2012.