Mechanistic intracellular and within-host models of bacterial and viral infections

Bevelynn Fay Williams

Department of Applied Mathematics

University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

December, 2022

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated.

The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Joint publications

Almost all of the work in Chapter 3 has been refereed and published, as follows:

- **Williams, B.**, López-García, M., Gillard, J.J., Laws, T.R., Lythe, G., Carruthers, J., Finnie, T. and Molina-París, C., (2021). A stochastic intracellular model of anthrax infection with spore germination heterogeneity. *Frontiers in immunology*, 12, p.688257.

In addition, the work in Chapter 4 and Chapter 5 is in preparation:

- **Williams, B.**, López-García, M., Gillard, J.J., Laws, T.R., Lythe, G., Carruthers, J. and Molina-París, C. A mathematical within-host model of inhalational anthrax infection.
- **Williams, B.**, Carruthers, J., Gillard, J.J., Laws, T.R., Lythe, G., Molina-París, C. and López-García, M. On the reproduction number probability distribution in stochastic models of viral dynamics with non-exponential infectious period.

Acknowledgements

Firstly, I would like to thank my supervisors at the University of Leeds, Martín López-García, Grant Lythe, and Carmen Molina-París, for their constant support and guidance throughout this project, especially through the difficult and isolated times of working from home. I am greatly appreciative of the encouragement and advice that they have offered, and for sharing with me their knowledge and enthusiasm for mathematical biology.

I would like to express my gratitude for the funding of this research, without which I would not have been able to complete this work. The research has been funded by an Engineering and Physical Sciences Research Council CASE Studentship, jointly provided by the University of Leeds and the Defence Science and Technology Laboratory (Dstl).

From Dstl, I would like to thank Joseph Gillard for his invaluable help with the development of the mathematical models and with directing the research. I would also like to thank Thomas Laws for letting me pick his brain about biology during many useful conversations. I am hugely grateful to them for investing their time in the project and for all the miles travelled between Salisbury and Leeds (pandemic permitting) to meet and discuss ideas together. Thank you also to all of those who have shared experimental data to help this research.

Next, I would like to thank my peers from the Mathematical Biology and Medicine research group, past and present, and others in the school. All the help, encouragement, and laughs they have provided has made working on this project thoroughly enjoyable.

Finally, I would like to thank my friends and family for all their support. A special mention goes to Luke, for always believing in me, even when I didn't, and for constantly suggesting to "just add them up and see what you get", which has been surprisingly helpful at times.

Abstract

This thesis explores mathematical models for the infection dynamics of two types of pathogens, namely the specific bacterium *Bacillus anthracis*, which is the causative agent of anthrax, and viruses more generally. I present a stochastic multi-scale model for describing *B. anthracis* infection at the intracellular and within-host levels. The intracellular model uses a Markov chain to describe the key interaction between *B. anthracis* spores and host cells, and to predict the distribution of outcomes from this interaction. Key outputs are then used to connect this intracellular model to a within-host model of inhalational anthrax, which aims to offer a realistic mechanistic description of the infection dynamics, as well the probability of infection for a given inhaled dose. The multi-scale model is calibrated via a Bayesian approach, using various experimental *in vitro* and *in vivo* data. Stochastic models of virus kinetics are also studied, with a focus on finding probability distributions for the burst size and cellular reproduction number. It is shown how these distributions are affected by modelling choices, and how these distributions in turn affect the probability that the virus and infected cells will be eliminated before an established infection can occur.

Contents

List of Figures

LIST OF FIGURES

List of Tables

Chapter 1

Introduction

1.1 Biological introduction

The immune system needs a diverse array of mechanisms to deal with the diversity of pathogens that it may encounter, which can have vastly different sizes and lifecycles. For example, viruses are usually around 0.1 μm (Burrell *et al.* (2017)) and must be intracellular in order to replicate. On the other hand, bacteria are usually around 1-2 μm and may replicate in the intracellular or extracellular environment, depending on the particular bacterium. The primary defences against such pathogens are physical blockades such as the skin and mucosa. If a pathogen manages to evade these barriers and enter the body, the second line of defence is the innate immune response. The innate response involves generalised recruitment and activation of a wide range of cells for the removal of the infectious agent. If needed, this is followed by an adaptive immune response, which involves specific pathogen-directed responses by a select group of cells. This thesis will explore mathematical models for the infection dynamics of two types of pathogens - the specific bacterium *Bacillus anthracis*, and viruses more generally.

1.1.1 Anthrax

Anthrax is an infectious disease, caused by the gram-positive bacterial pathogen *Bacillus anthracis*. Although anthrax has been associated with human history for a long time, dating back to pre-biblical times, it was in the 1870s that Robert

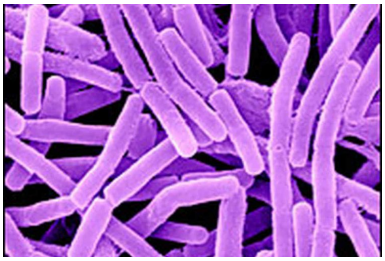

Figure 1.1: *Bacillus anthracis* bacteria. Image taken from European Pharmaceutical Review (2013).

Koch discovered that *B. anthracis* is the causative agent of the disease (Moayeri *et al.* (2015)). *B. anthracis* bacteria are rod-shaped, measuring 1-1.2 μm in width and 3-5 μm in length (Goel (2015)), and usually form chains of bacterial cells as they multiply, shown in Figure 1.1. Under adverse conditions, *B. anthracis* forms a dormant spore, which is highly resistant and unable to replicate. These spores can remain viable for centuries in soil, air, and water, with a half-life of around 100 years (Goel (2015)). The spores monitor their environment, and when favourable conditions are detected, such as the nutrient content of a host, the spores begin to germinate into vegetative bacteria (Setlow (2003)). These bacteria can then replicate inside the host.

Anthrax is a zoonotic disease, which means that it can be transmitted from animals to humans. It commonly affects domestic and wild animals in agricultural regions in some areas of the world, and can be transferred to humans through contact with infected animals or contaminated animal products. There are three main routes of exposure to *B. anthracis* spores: the cutaneous, gastrointestinal, and inhalation routes. If a human comes into contact with contaminated animal products, then *B. anthracis* spores can enter the body through a cut or scratch in the skin, which can lead to cutaneous anthrax. This is the most common type of anthrax, but can usually be resolved with timely treatment, and only about 20% of untreated cases are fatal. Consuming infected meat can lead to gastrointestinal anthrax, which is the second most common form of anthrax infection and has fatality rates of 25-60%. Thirdly, *B. anthracis* spores can enter the air from industrial processing of contaminated materials such as wool, hide, or hair. Spores in the surrounding air inhaled into the lungs can lead to inhalational anthrax,

which is usually fatal if not rapidly detected and treated (Cote *et al.* (2011)). Even with treatment, fatality rates for inhalational anthrax can be rather high if the treatment is not started early enough after exposure and symptoms onset. One of the reasons that inhalational anthrax is a such a dangerous form of the disease is that the initial symptoms resemble those of flu, so it is often difficult to diagnose early (Goel (2015)). Once it is correctly diagnosed, it may be too late to treat the infection. *B. anthracis* is viewed as a potential biological terrorism threat because the spores can be produced and preserved and then deliberately spread by release in the air (Goel (2015)). In 1979, an outbreak of 96 human cases of inhalational anthrax occurred in Sverdlovsk in the Soviet Union, due to an accidental release of *B. anthracis* spores from a Soviet facility (Moayeri *et al.* (2015)). The first confirmed use of the bacterium as a weapon occurred in the United States in 2001, when envelopes containing *B. anthracis* spores were posted to several news media companies and government officials. This led to 22 cases of the disease, including 5 fatalities (Jernigan *et al.* (2002)).

Inhalational anthrax is initiated by ungerminated, dormant *B. anthracis* spores, inhaled by a host. The spores travel through the air passages and eventually reach the alveoli of the lungs. There is some evidence that spores may be able to germinate extracellularly in the lungs (Powell *et al.* (2015)). However the generally accepted model of inhalational anthrax infection is the Trojan horse model, which assumes that ungerminated spores must be engulfed by alveolar phagocytes before they begin to germinate (Guidi-Rontani (2002)). Once the spores have been phagocytosed, the infected phagocytes migrate into the nearby lymph nodes in the mediastinum. These initial steps are outlined in Figure 1.2. Since the vegetative bacteria are less resistant than the spore form of *B. anthracis*, it is possible for the bacteria to die or be killed by the host phagocyte. Hence, the phagocytes are sometimes able to recover if they eliminate all intracellular bacteria. The host phagocytes that do not recover will eventually rupture and release the bacteria into the extracellular environment (Day *et al.* (2011); Pantha *et al.* (2018)). It is believed that the host phagocytes migrate to the lymph nodes before rupturing. This is because bacteria are not normally found within the lungs after systemic bacteraemia (Cote *et al.* (2011)), and inhalation anthrax does not usually cause pneumonia (Day *et al.* (2011)). During the first stage of the infection, in which

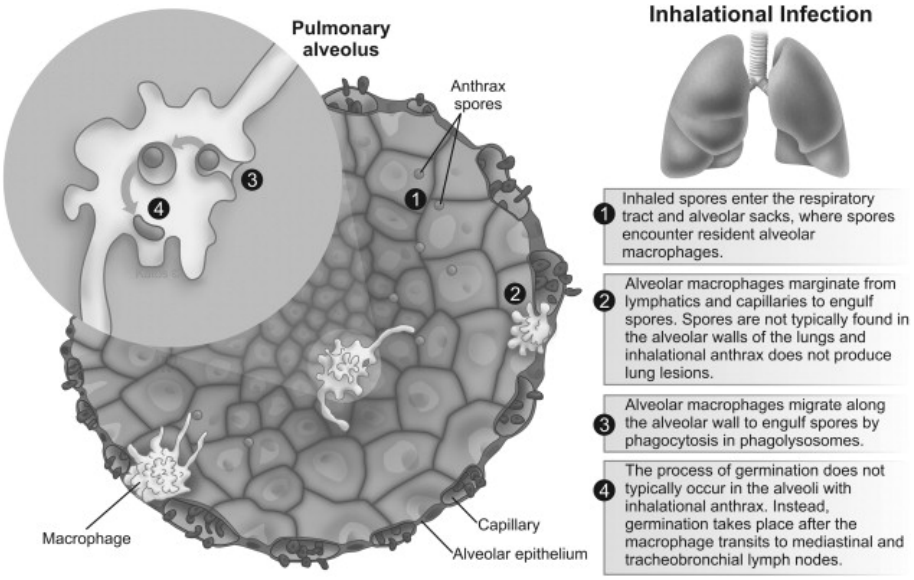

Figure 1.2: The initial steps of the infection process of inhalational anthrax. Figure taken from Gupta (2015).

bacteria germinate and replicate inside host cells, few symptoms present in the host. Once host cells rupture, the extracellular bacteria continue to multiply, leading to oedema and haemorrhage of the mediastinal lymph nodes. This results in large amounts of fluid entering the pleural cavity, which can severely affect breathing (Guarner *et al.* (2003)). The bacteria can also spread into the bloodstream and other organs to establish a systemic infection (Moayeri *et al.* (2015)). In cases of inhalational anthrax, treatment is usually started after the onset of this stage because this is when the flu-like symptoms start to appear (Goel (2015)). Unfortunately, it is very difficult to cure the disease by antibiotic therapy at this stage if the initial dose of spores was high.

Numerous *in vitro* and *in vivo* studies have been undertaken to improve understanding about important events during *B. anthracis* infections. Alveolar macrophages are one of the types of phagocytes that play a key role in the early infection stages of inhalational anthrax. These cells can induce microbicidal defences against intracellular pathogens and help to clear the infection (Guidi-Rontani *et al.* (1999)), but are also involved with transportation of *B. anthracis* spores to the mediastinal lymph nodes. Dixon *et al.* (2000) studied an *in vitro*

system and found that once spores are internalized by macrophages, they begin to germinate inside the phagosome. They also observed that 1-2 hours after phagocytosis, newly germinated bacteria were able to escape from macrophage phagosomes and begin to replicate in the cytosol. These bacteria were later released from the macrophage into the extracellular environment when the host cell ruptured and died. Akoachere *et al.* (2007) have examined *in vitro* macrophage-spore interactions by fluorescence microscopy. To do this, they used two different dyes, SYTO 13 and PI, which both bind to DNA. SYTO 13 is a green dye that is cell permeable, whereas PI is a red dye that is impermeable to cell membranes. Therefore, when staining macrophages with these dyes, the nucleus of healthy cells shows green fluorescence and the nucleus of dead cells shows red fluorescence. After infecting murine macrophages with *B. anthracis* spores at a spore to macrophage ratio of 20:1, they noticed a large increase in PI fluorescence intensity within 6 hours after *B. anthracis* inoculation. In particular, they observed 20% PI-positive macrophages at 3.5 hours after exposure, and 90% at 7 hours after exposure. They also observed that vegetative bacteria, which were stained green by the SYTO 13 dye, were present extracellularly 7 hours following *B. anthracis* inoculation. This is consistent with observations from Dixon *et al.* (2000), who found that the release of bacteria from macrophages occurred 4-6 hours after incubation of macrophages with spores. Dendritic cells are also thought to play a role in the trafficking of *B. anthracis* to the lymph nodes during the early stages of infection. These cells have been found to readily engulf *B. anthracis* spores (Brittingham *et al.* (2005); Cleret *et al.* (2006)), and transport them to the lymph nodes in a mouse model of inhalational anthrax (Cleret *et al.* (2007)).

Most pathogenic bacteria form virulence factors. These are structures of the bacteria, or molecules produced by the bacteria, that add to its effectiveness and ability to cause disease. Virulence factors help bacteria to evade or inhibit the host's immune system, enter and exit cells, and obtain nutrients from the host. An important factor for the survival of *B. anthracis* bacteria in the host is the anti-phagocytic capsule, which is a layer that lies outside of the cell wall and is composed of poly-γ-D-glutamic acid. The capsule allows extracellular, vegetative bacteria to avoid eradication by the immune system by preventing the bacteria being phagocytosed and destroyed (Sharma *et al.* (2020)). It does this

by covering molecules on the bacteria cell surface that might otherwise cause an immune response in the host, and therefore allows the bacteria to go unnoticed by the immune system (Cress *et al.* (2014); Jang *et al.* (2011)).

Another of the characteristic virulence factors that contributes to the pathogenic success of *B. anthracis* is the production of three proteins that are collectively termed anthrax toxin: protective antigen (PA), oedema factor (EF), and lethal factor (LF). The anthrax toxin follows the A-B model, which comprises an active component, "A", and a binding component, "B". The active component is an enzyme that has a toxic effect on cells by interfering with their functions, for example by inactivating a cell protein or signalling pathway. The binding component binds the toxin to a receptor molecule on the surface of the host cell membrane. In the case of the anthrax toxin, the B-component is PA and there are two alternative A-components, EF and LF. PA binds to receptors on host cells and is cleaved by a host protease called furin. This creates a binding site for either EF or LF. Once the A-B complex is bound to the host cell receptor, the complex is taken into the cell by receptor-mediated endocytosis. Then, acidification of the vesicle causes the A and B components to separate, which allows the A component to enter the cytosol of the cell and exert its toxic effect (Bhatnagar & Batra (2001)). This process is shown in the diagram in Figure 1.3. EF and LF are individually non-toxic, since they are not able to enter cells without PA molecules present. EF in combination with PA forms the oedema toxin, and LF in combination with PA forms the lethal toxin. These two toxins cause different cellular responses and are essential factors for the survival of *B. anthracis* in an infected host. Lethal toxin disrupts cell signalling pathways of macrophages and some other cells, leading to cell death. On the other hand, oedema toxin inhibits the phagocytosis of bacteria by neutrophils (Banks *et al.* (2005)). In some cell types, oedema toxin also increases the levels of cyclic adenosine monophosphate, which is a chemical messenger that plays a major role in controlling many intracellular processes. During systemic infection, the toxins are distributed via the bloodstream to different organs, where they affect the functioning of organs such as the spleen, lymph nodes, liver, kidney, heart and brain. Together, the anthrax toxins cause suppression of the host's immune system, often leading to death of the host.

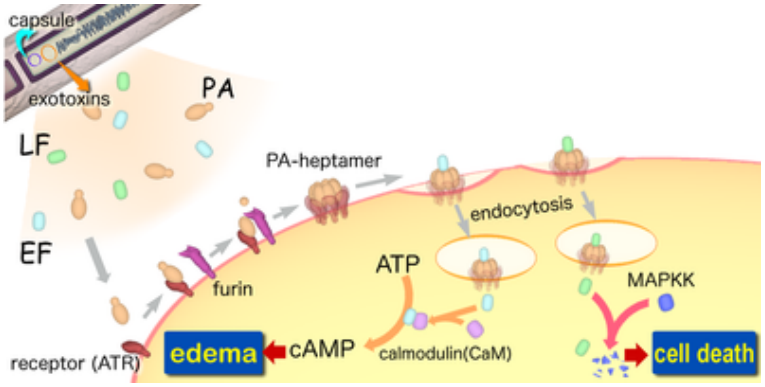

Figure 1.3: The process of intoxication by anthrax toxin follows the A-B model. In this case there are two alternative A-components, lethal factor (LF) and oedema factor (EF). The B-component is protective antigen (PA), which binds to receptors on host cells and is cleaved by a protease, creating a binding site for either LF or EF. Once the A-B complex is bound to the host cell receptor, the complex is taken into the cell by receptor-mediated endocytosis. Then, acidification of the vesicle causes the A and B components to separate, which allows the A component to enter the cytosol of the cell and exert its toxic effect. Figure taken from Tambe (2005).

The genes that encode the anthrax toxin and capsule are found on two plasmids in the cytosol of the bacteria, pXO1 and pXO2. The pXO1 plasmid carries genes that encode the production of the components of the anthrax toxin, and the pXO2 plasmid carries genes that encode the production of the capsule. The Sterne strain of anthrax, discovered in the 1930s by Max Sterne, contains the anthrax toxin plasmid, pXO1, but lacks the capsule forming plasmid, pXO2. This means that the Sterne strain has a reduced ability to cause illness, in comparison to strains that carry both plasmids. The Sterne strain has been used to vaccinate livestock since 1939 because immunisation with this strain has been found to provide a protective immune response (Centers for Disease Control and Prevention (2009)). The Sterne strain is also often used in laboratory settings (for example in the *in vitro* experiment by Akoachere *et al.* (2007) described above) since it poses a reduced infection risk to laboratory workers. Research with virulent strains of *B. anthracis*, such as the Ames strain, requires enhanced biosafety laboratories (Twenhafel (2010)). Another strain of *B. anthracis*, called A16R, which also produces anthrax toxin but not the capsule, is currently used as a

human anthrax vaccine in China (Zai *et al.* (2016)). The toxigenic properties of *B. anthracis* were discovered in 1954 and this new experimental knowledge led to the development of more effective, acellular vaccines for humans in the UK and USA. These acellular vaccines are composed of culture filtrates containing toxin proteins expressed by avirulent vaccine strains of *B. anthracis*. These vaccines aim to counter toxin action by initiating the body to generate antibodies against PA (Splino *et al.* (2005)). Anthrax vaccine adsorbed (AVA) was approved for use in the United States in 1965. Although it contains all three toxin components, PA is the most abundant protein and the main immunogen in this vaccine (Pomerantsev *et al.* (2011)). Anthrax vaccine precipitated (AVP) is licensed in the UK and contains significant amounts of PA and LF.

Experimental studies have been conducted to determine the biological activities of the toxin proteins and evaluate the quantities expressed by different strains in culture conditions. For example, Zai *et al.* (2016) explored the expression of PA and LF components by the A16R strain and the Sterne strain. The growth kinetics of the bacteria were observed to follow a sigmoidal growth curve, reaching a stationary phase of around 10^7 viable cells per ml. They found that the amount of LF and PA increased as the bacteria grew, peaked after around 12-16 hours of bacterial growth, and then declined rapidly. This rapid decline may be caused by downregulation of the toxin genes due to the depletion of glucose in the culture (Puziss & Wright (1959)). Zai *et al.* (2016) suggested that an accumulation of proteases in the culture could be an alternative reason for the decrease in the levels of toxin proteins. This is because it has been shown that certain proteases secreted by *B. anthracis*, such as immune inhibitor A1 (InhA1), can cleave the anthrax toxin proteins (Pflughoeft *et al.* (2014)). When protease inhibitors were added to the culture, Zai *et al.* (2016) observed that the amount of PA and LF increased and then maintained a high level, rather than decreasing. Charlton *et al.* (2007) have simulated the AVP vaccine manufacturing process, and measured the bacterial growth as well as PA and LF concentrations for up to 32 hours. They observed much higher levels of PA and LF in the culture supernatants compared to those of Zai *et al.* (2016), which could be due to different culture conditions. Charlton *et al.* (2007) also did not observe increased breakdown of the toxin proteins after the peak. In their experiment, once the glucose was exhausted,

the bacteria appeared to use amino acids as an alternative carbon source. One notable difference between the experimental setups of the two studies is that Charlton *et al.* (2007) incubated the *B. anthracis* cultures statically, whereas Zai *et al.* (2016) incubated the bottles with vigorous agitation. It is possible that if the cultures were observed over a longer timescale in the experiment by Charlton *et al.* (2007), a decline of the toxin proteins may have been observed. This would be consistent with a study by Puziss & Wright (1959) in which the bacteria were grown statically, and the levels of PA peaked at 38 hours before declining, which is much later than the timing of the decline observed by Zai *et al.* (2016). As well as these studies of the *in vitro* expression of toxin by *B. anthracis*, other studies have aimed to quantify the toxin proteins in the serum of infected animals (Boyer *et al.* (2007, 2009, 2015); Mabry *et al.* (2006); Tang *et al.* (2009)).

Animal species can have very different susceptibilities to anthrax. This depends on their susceptibility to the effects of the toxins (Moayeri & Leppla (2009)), and their susceptibility to the infection itself (i.e. their ability to control the bacteria). Therefore different animal models can be useful for answering specific questions. Mice are the most widely used animal model, and are highly susceptible to *B. anthracis* infection. Elimination of toxin production does not alter virulence in a mouse model of *B. anthracis* infection, and both toxigenic and non-toxigenic strains lead to uncontrolled infection. Although this may be quite different to the human case, the use of the mouse model with an encapsulated non-toxigenic strain allows exploration of the host-bacterial interactions and immune control mechanisms, without the confounding effects of the toxins (Goossens (2009)). Guinea pigs have also been used to gain understanding of the first steps in inhalational anthrax (Ross *et al.* (1957)). This species seems to be sensitive to both the infection and the toxin, and non-toxigenic encapsulated strains are attenuated. On the other hand, rabbits are quite resistant to the infection facet of anthrax, and they manage to eliminate encapsulated non-toxigenic strains of *B. anthracis* (Lovchik *et al.* (2012)).

1.1.2 Viruses

Viruses are very diverse. Some viruses replicate rapidly and cause acute infections, for example influenza and Ebola virus. These viruses must be eliminated by the immune system in order for host survival. On the other hand, some viruses such as Human Immunodeficiency viruses (HIV) cause chronic infections in which the virus will remain at a low level in the host for a long time. In the case of HIV, the virus itself is not lethal, but the depletion of important immune system cells can result in the infected individual becoming extremely susceptible to other infections that an uncompromised immune system would easily be able to deal with. Some viruses can result in either acute or chronic infection, depending on genetic or environmental factors that determine the susceptibility of the host to acute or chronic infection. An example of such a virus is hepatitis C virus (HCV).

A virion consists of a genome (either DNA or RNA) surrounded by a protein coating, called a capsid. Together, the capsid and the nucleic acid core are called a nucleocapsid. For some viruses, the nucleocapsid is surrounded by an outer envelope made of protein and phospholipid membranes derived from the host cell (Burrell *et al.* (2017)). In a general virus lifecycle, the virus will attach to a host cell and be internalised into an endosome. Then the viral nucleocapsid is released into the cytoplasm of the cell. For some viruses, viral genome replication occurs in the cytoplasm, whereas for others it can occur in the nucleus. The genome consists of genes coding for different structural proteins needed to form new virus particles, and non-structural proteins needed for genome replication. If a virus has a positive-sense genome, this means that the genome can be directly translated into viral proteins. On the other hand, negative-sense genomes do not directly code for proteins and have to first be transcribed into positive-sense mRNA. Newly synthesised viral genomes and proteins form nucleocapsids, which are then delivered to sites of viral assembly where they are packaged into new virus particles. For some viruses, such as poliovirus, progeny virions accumulate inside the host cell and eventually numerous virus particles exit at once in a burst, killing the cell. On the other hand, for most enveloped viruses, new virus particles are released at the cell membrane throughout the lifetime of the infected

cell via a process called budding (Pornillos *et al.* (2002)). These released virus particles can then go on to infect new cells and start the lifecycle again.

Chapter 5 of this thesis focusses on using stochastic viral dynamics models to calculate a probability distribution for the number of new cells that become infected due to an initial infected cell. The mathematical viral dynamics models considered in Chapter 5 are very simple and hence can be applied to a wide variety of viruses. To present numerical results of the models, a few specific examples are considered: Ebola virus, influenza, HIV, and HCV. Parameter values applicable to these viruses have been obtained from previous modelling studies in the literature (Guedj *et al.* (2013); Liao *et al.* (2020); Yan *et al.* (2020); Yuan & Allen (2011)). All of these examples of viruses are enveloped and so are released from host cells by budding. Ebola and influenza both have negative-sense RNA, whereas HIV and HCV have positive-sense RNA. The simplest model considered in Chapter 5 consists of uninfected cells, eclipse phase cells that are infected with virus but have not yet started to release virus, actively infected cells that are releasing virus, and free (i.e., extracellular) virus particles. In this model, the actively infected cells (also referred to as infectious phase cells) are assumed to release virus particles at a constant rate until the cell dies. This model is analysed using the parameter estimates for Ebola virus, influenza, and HIV. However, the model considered in Chapter 5 for HCV is slightly different. In the HCV model, the rate at which the virus particles are released from the host cell is assumed to depend on the length of time since the cell became infectious, and the amount of intracellular viral RNA. This assumption may also be realistic in the case of many other viruses, but is only considered in the HCV case in Chapter 5. This is because the model was previously applied to HCV infection in order to study the potential effects of direct-acting antiviral agents (DAAs) on specific intracellular processes (Guedj *et al.* (2013)).

1.2 Thesis objectives

Inhalational anthrax is generally a very rare disease, but it is also very dangerous, and is a potential biological terrorism threat. If *B. anthracis* spores were made into an aerosol and released, the spores could infect a large number of people in a

short space of time. Although some cases of inhalational anthrax can be resolved with antibiotic treatment, the fatality rates with treatment are still rather high because often treatment is not started early enough. Due to the significant risk posed if the bacterium was to be intentionally used as a bioweapon, it is important to be able to defend against such an attack and to make optimal decisions about treatment strategies. The aim of Chapters 3 and 4 of this thesis is therefore to start to develop a mechanistic multi-scale mathematical model for inhalational anthrax, by linking an intracellular model of anthrax infection with a within-host model. This builds on the novel multi-scale methodologies recently developed for *Francisella tularensis* infection (Carruthers *et al.* (2018, 2020)). The aim of these multi-scale models is to help to improve understanding and quantify key mechanisms of the infection. Ultimately, mathematical models like this would aim to accurately predict how many individuals would become infected in different exposure scenarios and also on what timescale this would occur. This could then aid in the identification of a time period during which treatment would be most effective.

In Chapter 3, a stochastic model is introduced to study the key interaction between *B. anthracis* spores and phagocytes, which are cells of the innate immune system that engulf the spores when they first enter the lungs. This Markov chain model describes the dynamics of *B. anthracis* infection inside a single phagocyte, including the germination of spores into vegetative bacteria and replication/death of bacteria inside the cell. The ultimate outcome of the Markov chain is either recovery of the host cell or death of the phagocyte and the release of bacteria. The dynamics of the model are analytically studied through the calculation of a number of summary statistics or "stochastic descriptors" of the process. In particular, an important stochastic descriptor of the intracellular model is the probability distribution for the rupture size, which is the amount of bacteria that are ultimately released from an infected phagocyte. *In vitro* experimental data are used to calibrate the intracellular model in a Bayesian setting, allowing for the uncertainty in parameter estimates to be quantified.

The intracellular model is then connected to a within-host model of inhalational anthrax, presented in Chapter 4. The within-host model aims to provide an overall understanding of the early progression of the infection, and is parametrised

with *in vivo* infection data from studies of rabbits and guinea pigs. Building a model that offers a realistic mechanistic description of these different animal responses to the inhalation of *B. anthracis* spores is an important stepping-stone towards the ultimate aim of obtaining a model that can predict the dynamics of human infection. The model is stochastic so that it can capture the randomness in the biological processes and describe the probabilities of different outcomes, for example how likely it is that the infection will be cleared by the immune system. The model includes variability in the rupture size between cells by incorporating the rupture size distribution from the intracellular model, which is something that has not been considered in previous models of inhalational anthrax. This variability is particularly important for modelling the outcome of exposures that result in a relatively small number of spores entering the lungs. The model also includes the role of anthrax toxin in the infection, because this is a key part of the disease and will also be important for modelling certain treatments in future work. Some of the parameters used for this part of the model are informed by *in vitro* experimental data on toxin production and degradation.

Chapter 5 focusses on using mathematical models to study virus kinetics. Unlike inhalational anthrax, where the bacteria multiply extracellularly after the initial stages of infection, during a viral infection viruses must reproduce intracellularly. The number of progeny virus particles that each infected cell produces and the number of new cells infected by these virus particles determine whether the infection can be sustained. Traditional deterministic models of viral kinetics focus on the average number of new infected cells produced by an initial infected cell, as an indicator of whether or not an infection will become established. This is called the basic reproduction number, and is denoted by R_0. The objective of Chapter 5 in this thesis is to build on this widely used descriptor by considering the stochastic dynamics of infected cells in order to derive a probability distribution of the reproduction number. Various previously published deterministic models are considered in order to analyse the stochasticity of the reproduction number in the Markov chain version of these models.

Chapter 2

Mathematical background

This chapter provides an introduction to probability theory and stochastic processes. Bayesian inference is also briefly introduced, since this approach will be used in the following chapters to calibrate mathematical models with experimental data.

2.1 Probability theory

In this section, some basic principles from probability theory are introduced, which will be necessary for the analysis of the stochastic processes that are the basis of the models in this thesis. The definitions given here can be found in Chapter 1 of Allen (2010).

A random variable, X, is a variable whose value is determined by the outcome of a random phenomenon. It is a function that maps from possible outcomes in a sample space to some measurable space, such as the real numbers. The set of values that X can take is called the *support* of X. If X has a finite or countably infinite support, then it is said to be a *discrete* random variable. If the support is a continuous interval, then X is a *continuous* random variable.

Definition 2.1.1 *Suppose X is a real-valued random variable. The* ***cumulative distribution function (cdf)*** *of X is the function $F_X : \mathbb{R} \to [0, 1]$, defined by*

$$F_X(x) = \mathbb{P}(X \leq x).$$

Definition 2.1.2 *Suppose X is a discrete random variable, with support S. The* ***probability mass function (pmf)*** *of X is defined by*

$$f_X(x) = \mathbb{P}(X = x), \qquad x \in S.$$

Definition 2.1.3 *Suppose X is a continuous random variable, and there exists a non-negative, integrable function $f_X : \mathbb{R} \to [0, \infty)$ such that,*

$$\mathbb{P}(a \leq x \leq b) = \int_a^b f_X(x)\ dx, \qquad a < b.$$

Then the function f_X is called the ***probability density function (pdf)*** *of X.*

Definition 2.1.4 *If X is a continuous random variable with pdf f_X and support S, then the* ***expectation*** *of X is defined as*

$$\mathbb{E}[X] = \int_S x f_X(x)\ dx.$$

If X is a discrete random variable with pmf f_X, and support S, then the ***expectation*** *of X is defined as*

$$\mathbb{E}[X] = \sum_{x \in S} x f_X(x).$$

The definition of the expectation of a random variable can be extended to the expectation of a function of a random variable. If X is a continuous random variable, then the expectation of $g(X)$ is

$$\mathbb{E}[g(X)] = \int_S g(x) f_X(x)\ dx.$$

If X is a discrete random variable, then the expectation of $g(X)$ is

$$\mathbb{E}[g(X)] = \sum_{x \in S} g(x) f_X(x).$$

The mean, variance, and moments of X are defined in terms of the expectation, as follows.

- The **mean** of a random variable X, denoted as μ or μ_X, is simply the expectation of X,

$$\mathbb{E}[X] = \mu_X.$$

- The **variance** of X, denoted by σ^2, σ_X^2, or $Var(X)$, is given by

$$Var(X) = \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2] - \mu_X^2.$$

- The **standard deviation** of X is the square root of the variance,

$$\sigma = \sqrt{Var(X)}.$$

- The **nth moment of X about the point** a is

$$\mathbb{E}[(X - a)^n].$$

A helpful concept when analysing stochastic processes is that of generating functions. There are several different generating functions that can be useful in different contexts. Here, the *probability generating function (pgf)* is defined for a discrete random variable, X, that takes values on the support $\{0, 1, 2, ...\}$. The pgf can also be defined for discrete random variables that take negative integer values, but since this thesis will focus on modelling biological quantities (e.g. numbers of bacteria and cells), it is only defined here for positive integer values.

Definition 2.1.5 *Suppose X is a discrete random variable, with support $S = \{0, 1, 2, ...\}$. The* ***probability generating function (pgf)*** *of X is defined by*

$$\mathcal{P}_X(z) = \mathbb{E}[z^X] = \sum_{j=0}^{\infty} \mathbb{P}(X = j) z^j,$$

for some $z \in \mathbb{R}$.

The pgf generates the probabilities associated with the distribution by differentiating and setting $z = 0$, to obtain,

$$\mathbb{P}(X = k) = \frac{1}{k!} \frac{d^{(k)}}{dz^{(k)}} \mathcal{P}_X(z) \Big|_{z=0}, \quad k \geq 0.$$

The mean and variance can also be found from the pgf.

$$\mathcal{P}'_X(z) \Big|_{z=1} = \sum_{j=0}^{\infty} j \mathbb{P}(X = j) = \mathbb{E}[X].$$

$$\mathcal{P}''_X(1) + \mathcal{P}'_X(1) - [\mathcal{P}'_X(1)]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = Var(X).$$

2.1.1 Probability distributions

Several well-known discrete distributions will appear in different contexts throughout this thesis. Their pmfs are provided below.

Poisson

$$f_X(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, ..., \\ 0, & \text{otherwise}, \end{cases}$$

where λ is a positive constant.

Bernoulli

$$f_X(x) = \begin{cases} p, & x = 1, \\ 1 - p, & x = 0, \\ 0, & \text{otherwise}, \end{cases}$$

where $0 < p < 1$.

Binomial

The sum of n independent Bernoulli random variables is a binomial random variable. That is, a binomial random variable can be thought of as the number of successes in n independent Bernoulli trials, where p is the probability of success.

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, ..., n, \\ 0, & \text{otherwise}, \end{cases}$$

where n is a positive integer and $0 < p < 1$.

Geometric

$$f_X(x) = p(1-p)^x, \ x = 0, 1, 2, ...,$$

where $0 < p < 1$. A geometric random variable can be interpreted as the number of failed independent Bernoulli trials before the first success, where p is the probability of success.

Negative binomial

$$f_X(x) = \begin{cases} \binom{x+n-1}{n-1} p^n (1-p)^x, & x = 0, 1, 2, ..., \\ 0, & \text{otherwise,} \end{cases}$$

where n is a positive integer and $0 < p < 1$. A negative binomial random variable can be interpreted as the number of failed independent Bernoulli trials before the n^{th} success, where p is the probability of success.

A continuous probability distribution that is very important when analysing continuous-time Markov chain models is the exponential distribution. In these models, the exponential distribution is associated with the time between successive events (typically referred to as the inter-event time).

Definition 2.1.6 *The pdf of an* ***exponentially distributed random variable****, $X \sim Exp(\lambda)$, is given by*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \textit{if } x \geq 0, \\ 0, & \textit{otherwise,} \end{cases}$$

for some rate parameter $\lambda > 0$. The cdf of an exponential distribution is given by

$$F_X(x) = 1 - e^{-\lambda x}, \qquad x \geq 0.$$

The mean and variance are given by

$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad Var(X) = \frac{1}{\lambda^2}.$$

The sum of n independent exponentially distributed random variables, each with rate $\lambda > 0$, is an Erlang(n, λ) distributed random variable, which has pdf given by

$$f_X(x) = \begin{cases} \dfrac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}, & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where n is a positive integer, and is called the *shape* parameter. An extension of this distribution is the Gamma distribution, with pdf given by,

$$f_X(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where the shape parameter α is positive but not necessarily an integer.

Another very well-known continuous probability distribution is the normal distribution.

Definition 2.1.7 *The pdf of a* ***normally distributed random variable****,* $X \sim N(\mu, \sigma^2)$*, is given by*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad x \in \mathbb{R}.$$

The mean and variance are given by

$$\mathbb{E}[X] = \mu, \qquad Var(X) = \sigma^2.$$

2.2 Stochastic processes

This section provides some theory relevant to stochastic processes. In particular, it focuses on continuous-time Markov chains. The main sources for this section are Chapter 6 of Kulkarni (2010) and Chapters 5 and 6 of Allen (2010). Some ideas have also come from Karlin & Taylor (1981).

Definition 2.2.1 *A* ***stochastic process*** *is a collection of random variables,* $\{X(\tau), \tau \in T\}$*, indexed by the parameter* τ *which takes values in the parameter set* T*. The random variables take values in the set* S*, called the state-space of the stochastic process.*

The parameter set T can be discrete, for example $T = \{0, 1, 2, ...\}$, in which case the stochastic process is said to be discrete-time, and usually written $\{X_n, n \geq 0\}$. If studying a population, for example, a stochastic process like this could be used to denote the number of individuals in the n^{th} generation. Alternatively, the parameter set can be continuous, for example $T = [0, \infty)$, in which case the stochastic process is said to be continuous-time, and usually written $\{X(t), t \in [0, \infty)\}$. A stochastic process like this could be used to denote the number of individuals alive at time t. The state-space can also be discrete, for example $S \subseteq \{0, 1, 2, ...\}$, or continuous, for example $S \subseteq (-\infty, \infty)$.

In this thesis, the main stochastic processes considered will be continuous-time Markov chains. The term *chain* implies that the random variables $X(t)$

take values in a discrete state-space. The term *Markov* (named after Andrey Andreyevich Markov) implies that the future of the system only depends on the present state of the system and is independent of the past. This property is called the *Markov property*. The formal definition of a continuous-time Markov chain is as follows.

Definition 2.2.2 *Let $\mathcal{X}$ be a collection of random variables, $\mathcal{X} = \{X(t), t \in [0, \infty)\}$, taking values in a discrete state-space, S. The stochastic process $\mathcal{X}$ is a* ***continuous-time Markov chain (CTMC)*** *if for any sequence of real numbers satisfying $0 \leq t_0 < t_1 < ... < t_n < t_{n+1}$,*

$$\begin{aligned}&\mathbb{P}(X(t_{n+1}) = i_{n+1} \mid X(t_0) = i_0, X(t_1) = i_1, ..., X(t_n) = i_n)\\ &= \mathbb{P}(X(t_{n+1}) = i_{n+1} \mid X(t_n) = i_n),\end{aligned}$$

for any $i_0, i_1, ..., i_{n+1} \in S$.

That is, the probability of the process being in state i_{n+1} at time t_{n+1} only depends on the state of the process at the most recent time, t_n, and does not depend on the history of the process.

2.2.1 Transition probabilities

The *transition probabilities* describe the probability to transition from state i at time s to state j at time t, and are defined as

$$p_{ij}(s, t) = \mathbb{P}(X(t) = j \mid X(s) = i),$$

for $s < t$ and $i, j \in S$.

Definition 2.2.3 *The transition probability, $p_{ij}(s, t)$, is called* ***time-homogeneous*** *if it only depends on the length of the time interval, $t-s$, rather than on the values of s and t themselves. If this is the case, then we can write*

$$p_{ij}(s, t) = p_{ij}(t - s) = \mathbb{P}(X(t - s) = j \mid X(0) = i), \text{ for } s < t.$$

In general, the transition probabilities for continuous-time Markov chains are time-homogeneous and have the following properties, for all $s, t \in [0, \infty)$:

- $0 \leq p_{ij}(t) \leq 1$, for any $i, j \in S$, because they are probabilities.
- $\sum_{j \in S} p_{ij}(t) = 1$, for any $i \in S$, because if the process starts in state i at time 0, it will definitely be in some state, j, at time t, so the summation of the transition probabilities over all possible j will equal 1.
- $p_{ij}(t+s) = \sum_{k \in S} p_{ik}(t) p_{kj}(s)$, for any $i, j \in S$. These are called the *Chapman-Kolmogorov equations.*

To show why the Chapman-Kolmogorov equations hold,

$$\begin{aligned} p_{ij}(t+s) &= \mathbb{P}(X(t+s) = j \mid X(0) = i), \\ &= \sum_{k \in S} \mathbb{P}(X(t+s) = j \mid X(0) = i, X(t) = k) \mathbb{P}(X(t) = k \mid X(0) = i), \\ &= \sum_{k \in S} \mathbb{P}(X(t+s) = j \mid X(t) = k) p_{ik}(t), \qquad \text{(from the Markov property)} \\ &= \sum_{k \in S} p_{ik}(t) p_{kj}(s). \qquad \text{(from time-homogeneity)} \end{aligned}$$

Definition 2.2.4 *The **transition matrix** is the matrix of transition probabilities, denoted as*

$$\boldsymbol{P}(t) = (p_{ij}(t))_{i,j \in S}.$$

In matrix form, the Chapman-Kolmogorov equations are written as

$$\boldsymbol{P}(t+s) = \boldsymbol{P}(t)\boldsymbol{P}(s).$$

2.2.2 Generator matrix

The transition probabilities, $p_{ij}(t)$, of a continuous-time Markov chain, can be used to define the *transition rates*, q_{ij}. Assume that the transition probabilities are continuous and differentiable for $t \geq 0$ and that for $t = 0$ they satisfy,

$$p_{ij}(0) = 0 \;\text{ if }\; i \neq j, \;\text{ and }\; p_{ii}(0) = 1.$$

Definition 2.2.5 *For any $i, j \in S$, the transition rate, q_{ij}, is the derivative of $p_{ij}(t)$, evaluated at $t = 0$. Therefore the* ***transition rates*** *can be defined as,*

$$q_{ij} = \lim_{\Delta t \to 0^+} \frac{p_{ij}(\Delta t) - p_{ij}(0)}{\Delta t} = \lim_{\Delta t \to 0^+} \frac{p_{ij}(\Delta t)}{\Delta t}, \quad i \neq j,$$

$$q_{ii} = \lim_{\Delta t \to 0^+} \frac{p_{ii}(\Delta t) - p_{ii}(0)}{\Delta t} = \lim_{\Delta t \to 0^+} \frac{p_{ii}(\Delta t) - 1}{\Delta t}.$$

By expanding $p_{ij}(\Delta t)$ as a Taylor series around $\Delta t = 0$, it follows from this definition that,

$$p_{ij}(\Delta t) = \delta_{ij} + q_{ij}\Delta t + o(\Delta t), \tag{2.2.1}$$

where δ_{ij} is Kronecker's delta symbol. Then, since $\sum_{j \in S} p_{ij}(\Delta t) = 1$, it follows that,

$$1 - p_{ii}(\Delta t) = \sum_{j \in S, j \neq i} p_{ij}(\Delta t) = \sum_{j \in S, j \neq i} [q_{ij}\Delta t + o(\Delta t)].$$

Thus,

$$q_{ii} = \lim_{\Delta t \to 0^+} \frac{-\sum_{j \in S, j \neq i} [q_{ij}\Delta t + o(\Delta t)]}{\Delta t}$$

$$= -\sum_{j \in S, j \neq i} q_{ij}, \quad i \in S.$$

So we have,

$$\sum_{j \in S} q_{ij} = 0, \quad i \in S.$$

Definition 2.2.6 *The transition rates form a matrix, $\boldsymbol{Q} = (q_{ij})_{i,j \in S}$, known as the* ***infinitesimal generator matrix****. This square matrix can be infinite or finite depending on the cardinality of S, and the sum of each of its rows is 0, by*

the above argument. For example, for $S = \{0, 1, 2...\}$*, we get*

$$\boldsymbol{Q} = \begin{bmatrix} q_{00} & q_{01} & q_{02} & \dots \\ q_{10} & q_{11} & q_{12} & \dots \\ q_{20} & q_{21} & q_{22} & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

$$= \begin{bmatrix} -\sum_{i=1}^{\infty} q_{0i} & q_{01} & q_{02} & \dots \\ q_{10} & -\sum_{i=0, i\neq 1}^{\infty} q_{1i} & q_{12} & \dots \\ q_{20} & q_{21} & -\sum_{i=0, i\neq 2}^{\infty} q_{2i} & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix}.$$

2.2.3 Forward Kolmogorov differential equations

The forward Kolmogorov differential equations describe the rate of change of transition probabilities with respect to time, and can be derived as follows. From the Chapman-Kolmogorov equation, we can write

$$p_{ij}(t + \Delta t) = \sum_{k \in S} p_{ik}(t) p_{kj}(\Delta t). \tag{2.2.2}$$

From Eq. (2.2.1), we also have,

$$p_{kj}(\Delta t) = \delta_{kj} + q_{kj}\Delta t + o(\Delta t),$$

where δ_{kj} is Kronecker's delta symbol. Substituting this into Eq. (2.2.2) gives

$$p_{ij}(t + \Delta t) = \sum_{k \in S} p_{ik}(t)(\delta_{kj} + q_{kj}\Delta t + o(\Delta t)).$$

Subtracting $p_{ij}(t)$ from both sides, dividing by Δt, and letting $\Delta t \to 0$ gives the following differential equation,

$$\frac{dp_{ij}}{dt} = \sum_{k \in S} q_{kj} p_{ik}(t), \qquad i, j \in S, \tag{2.2.3}$$

known as the forward Kolmogorov equation or master equation.

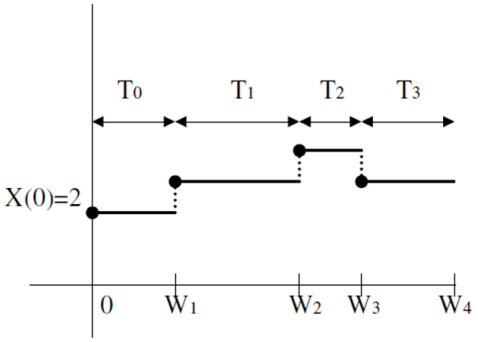

Figure 2.1: A sample path of a CTMC, indicating the waiting times, W_n, which are the random variables for the time of the n^{th} jump, and inter-event times, T_n, which are the random variables for the times between successive jumps. Figure taken from Allen (2010).

2.2.4 Realisations of a CTMC

Since each realisation of a continuous-time Markov chain remains in a particular state for a random amount of time before making a jump to a new state, we can define W_n to be the random variable for the time of the n^{th} jump (let $W_0 = 0$, for convenience). These random variables, $\{W_n\}_{n=0,1,2,\ldots}$, are referred to as the *waiting times* of the process. The random variables $T_n = W_{n+1} - W_n$, $n = 0, 1, 2, \ldots$, are the times between successive jumps and are referred to as the *inter-event times*. The waiting times W_n and inter-event times T_n are illustrated in Figure 2.1.

Let $Y_n = X(W_n)$ be the state that the continuous-time Markov chain jumps to at the n^{th} jump. The discrete-time Markov chain, $\{Y_n\}_{n=0}^{\infty}$, is called the *embedded Markov chain*. Every continuous-time Markov chain has an embedded Markov chain associated to it.

In order to calculate sample paths of continuous-time Markov chains, we need to find the distribution of T_n, as follows. Assume that $Y_n = X(W_n) = i \in S$. Let $\alpha_i \Delta t + o(\Delta t)$ be the probability that the process moves to a different state in the time period Δt, where,

$$\alpha_i = \sum_{j \in S, j \neq i} q_{ij} = -q_{ii}.$$

Then the probability that the process stays at state i for the length of time Δt, is

$$p_{ii}(\Delta t) = 1 - \alpha_i \Delta t + o(\Delta t).$$

Let $G_n(t)$ be the probability that the process remains in state i for a time of length at least $t \geq 0$, which means that $T_n = W_{n+1} - W_n > t$. That is,

$$G_n(t) = \mathbb{P}(T_n > t).$$

For $t = 0$,

$$G_n(0) = \mathbb{P}(T_n > 0) = 1.$$

For Δt sufficiently small,

$$\begin{aligned} G_n(t + \Delta t) &= \mathbb{P}(T_n > t + \Delta t) \\ &= \mathbb{P}(T_n > t + \Delta t \mid T_n > t)\mathbb{P}(T_n > t) \\ &= p_{ii}(\Delta t) G_n(t) \\ &= G_n(t)(1 - \alpha_i \Delta t + o(\Delta t)). \end{aligned}$$

Subtracting $G_n(t)$ from both sides, diving by Δt, and taking the limit as $\Delta t \to 0$, one arrives at the following differential equation,

$$\frac{dG_n(t)}{dt} = -\alpha_i G_n(t),$$

with initial condition, $G_n(0) = 1$. This can be solved by separation of variables to obtain the solution,

$$G_n(t) = \mathbb{P}(T_n > t) = e^{-\alpha_i t}.$$

Therefore the cumulative distribution function (cdf) of the random variable T_n, representing the inter-event time, is,

$$\mathbb{P}(T_n \leq t) = 1 - G_n(t) = 1 - e^{-\alpha_i t}, \qquad \text{for } t \geq 0.$$

This corresponds to an exponential random variable with parameter α_i, so this shows that if there is a positive probability of moving from state i to another state (i.e. $\alpha_i > 0$ so state i is not absorbing), then the inter-event time T_n is exponentially distributed, leading to the following theorem.

Theorem 2.2.1
Let $\mathcal{X} = \{X(t) : t \in [0, \infty)\}$ *be a CTMC with a transition matrix* $\boldsymbol{P}(t) = (p_{ij}(t))_{i,j\in S}$*, such that for* Δt *sufficiently small,*

$$\sum_{j\in S, j\neq i} p_{ij}(\Delta t) = \alpha_i \Delta t + o(\Delta t),$$

and

$$p_{ii}(\Delta t) = 1 - \alpha_i \Delta t + o(\Delta t),$$

where $\alpha_i = \sum_{j\in S, j\neq i} q_{ij} = -q_{ii}$*. Define the inter-event time as* $T_n = W_{n+1} - W_n$*, where* W_n *is time of the nth jump. Given* $X(W_n) = i$*,* T_n *is an exponential random variable with parameter* α_i*, as defined in Definition 2.1.6. The probability density function (pdf) for* T_n *is* $f_{T_n}(t) = \alpha_i e^{-\alpha_i t}$*, for* $t \geq 0$*. The mean and variance are*

$$\mathbb{E}[T_n] = \frac{1}{\alpha_i}, \qquad Var(T_n) = \frac{1}{\alpha_i^2}, \qquad \text{if } \alpha_i > 0.$$

A state i for which $0 \leq \alpha_i < \infty$, is called *stable*. If for state i, we have $\alpha_i = 0$, then this state is called *absorbing* because there are no events that will take the system out of state i, so once state i is entered, the process remains there permanently. Indeed, if $\alpha_i = 0$,

$$\mathbb{P}(X(t) = i \mid X(0) = i) = \mathbb{P}(T_0 > t) = e^{-\alpha_i t} = 1,$$

for all $t \geq 0$. This implies that T_0 is infinite.

On the other hand, if $\alpha_i > 0$, then the time spent in state i is a random variable, exponentially distributed with parameter $\alpha_i > 0$, so therefore some transition out of state i occurs in finite time. Let p_{ij} be the conditional probability of a transition from state i to state j, given that a transition occurs. Define $R_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i, X(t) \neq i)$ for $i \neq j$. Then,

$$p_{ij} = \lim_{\Delta t \to 0^+} R_{ij}(\Delta t).$$

If Δt is small enough so that at most one transition can occur in this length of time, then $R_{ij}(\Delta t)$ is the probability that the process transitions from i to j in the length of time Δt, given that the process does not stay in state i for the length of time Δt. Hence,

$$R_{ij}(\Delta t) = \frac{p_{ij}(\Delta t)}{1 - p_{ii}(\Delta t)}.$$

Dividing the numerator and denominator by Δt, letting $\Delta t \to 0$, and using Definition 2.2.5, we have,

$$p_{ij} = \frac{q_{ij}}{-q_{ii}} = \frac{q_{ij}}{\alpha_i}, \qquad i \neq j.$$

Therefore each sample path of a CTMC starts in a state $X(0) = i_0$ and stays in this state for a random duration of time, $T_0 \sim \text{Exp}(\alpha_{i_0})$, before jumping to a new state $X(W_1) = i_1$ with probability p_{i_0,i_1}, independent of the amount of time spent in state i_0. In general, after the n^{th} transition at time W_n, the process remains in state $X(W_n) = i_n$ for a random duration of time, $T_n \sim \text{Exp}(\alpha_{i_n})$, before moving to state i_{n+1} with probability $p_{i_n,i_{n+1}}$, independent of the history of the process over the time interval $[0, W_{n+1})$.

2.2.5 Gillespie algorithm

The Gillespie algorithm can be used to numerically simulate Markovian stochastic processes. This algorithm works by exploiting the fact that in a continuous-time Markov chain, the time until the next jump is exponentially distributed and independent of the state that the process jumps to. Hence at each step of the algorithm, the inter-event time is sampled from the corresponding exponential distribution, and the state that the stochastic process jumps to at this time is independently sampled from the set of possible jumps according to their relative probabilities. In order to exactly sample the exponentially distributed inter-event times, the algorithm makes use of the following theorem, which is Theorem 5.5 in Allen (2010).

Theorem 2.2.2
Let U be a uniform random variable defined on $[0, 1]$ and T be a continuous random variable defined on $[0, \infty)$. Then $T = F^{-1}(U)$, where F is the cumulative distribution function (cdf) of the random variable T.

Proof. *Since $\mathbb{P}(T \leq t) = F(t)$, we want to show that $\mathbb{P}(F^{-1}(U) \leq t) = F(t)$. Note that $F : [0, \infty) \to [0, 1)$ is strictly increasing, so that F^{-1} exists. Also, since*

U is a uniform random variable on $[0,1]$, $\mathbb{P}(U \leq x) = x$ for $x \in [0,1]$. Therefore, for $t \in [0,\infty)$,

$$\begin{aligned}\mathbb{P}(F^{-1}(U) \leq t) &= \mathbb{P}(F(F^{-1}(U)) \leq F(t)), \quad \textit{since } F \textit{ is strictly increasing,}\\ &= \mathbb{P}(U \leq F(t)),\\ &= F(t).\end{aligned}$$

Suppose that the Markov chain to be simulated is currently in state $\boldsymbol{x}$. Let $r_1, ..., r_n$ be the set of possible events from this state, with each r_k having an associated transition rate, $q_k(\boldsymbol{x})$, and moving the process to potentially different states $\boldsymbol{x}_k$. Then the next inter-event time, T, is an exponential random variable, as stated in Theorem 2.2.1, and has cdf $F(t) = 1 - e^{-\alpha(\boldsymbol{x})t}$, where $\alpha(\boldsymbol{x}) = \sum_{k=1}^{n} q_k(\boldsymbol{x})$. Therefore, by Theorem 2.2.2,

$$\begin{aligned}T &= F^{-1}(U[0,1])\\ &= -\frac{1}{\alpha(\boldsymbol{x})}\log(1 - U[0,1])\\ &= -\frac{1}{\alpha(\boldsymbol{x})}\log(U[0,1]).\end{aligned}$$

Hence one can draw a number from a uniform distribution and use the above transformation to obtain the inter-event time. To select the next event to occur, the unit interval is partitioned into intervals corresponding to the probabilities of each event. Then a number is drawn from a uniform distribution on $[0,1]$ and the interval that it falls into determines which event occurs. For a continuous-time Markov chain, the Gillespie algorithm (Gillespie (2007)) works by iterating these steps at each time increment, until some condition is met, for example reaching a fixed maximum time, T_{max}. This is described in Algorithm 2.1.

2.2.6 Example: Poisson process

One of the simplest continuous-time Markov chains is the *Poisson process*. This process has an infinite state-space, $\{0, 1, 2, ...\}$, and we will assume that the initial state is $X(0) = 0$. In this process, $X(t)$ represents the number of occurrences up

Algorithm 2.1: Gillespie Algorithm

Select T_{max};

Set the initial time to be $t = t_0$;

Set the initial state of the process to be $\boldsymbol{x} = \boldsymbol{x}_0$;

while $t < T_{max}$ **do**

1. Generate two random numbers, u_1 and u_2, from the uniform distribution on the unit interval;
2. Compute the *propensity function*, $\alpha(\boldsymbol{x}) = \sum_{k=1}^{n} q_k(\boldsymbol{x})$, where n is the number of possible events from state $\boldsymbol{x}$;
3. Set the intervent time to be $\Delta t = -\frac{1}{\alpha(\boldsymbol{x})} \log(u_1)$;
4. Select the event that occurs to be r_i where i is the least integer satisfying $\sum_{k=1}^{i} \frac{q_k(\boldsymbol{x})}{\alpha(\boldsymbol{x})} \geq u_2$;
5. Update the time, $t \leftarrow t + \Delta t$ and update the state, $\boldsymbol{x} \leftarrow \boldsymbol{x}_i$;

end

to time t, of an event that occurs randomly, but with a constant rate, λ. For Δt sufficiently small, the transition probabilities satisfy

$$\begin{aligned} p_{ij}(\Delta t) &= \mathbb{P}(X(t+\Delta t) = j \mid X(t) = i), \\ &= \begin{cases} \lambda\Delta t + o(\Delta t), & j = i+1, \\ 1 - \lambda\Delta t + o(\Delta t), & j = i, \\ o(\Delta t), & j \geq i+2, \\ 0, & j < i. \end{cases} \end{aligned}$$

The generator matrix for the process is

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \\ \vdots & \vdots & & \ddots & \ddots \end{bmatrix}.$$

The transition probabilities can be used to derive the system of forward Kolmogorov differential equations satisfied by the probabilities, $p_{0i}(t)$, for $i = 0, 1, 2, ...$ Since the initial condition is assumed to be $X(0) = 0$, I will omit the first index of the transition probabilities when it is equal to the initial state, so that $p_{0i}(t) = p_i(t) = \mathbb{P}(X(t) = i)$. The system of differential equations obtained can be solved to find expressions for the $p_i(t)$ probabilities, which define the state distribution of the process at time t.

We have, by the Chapman-Kolmogorov equation, and for Δt sufficiently small,

$$\begin{aligned} p_0(t+\Delta t) &= p_{00}(t)p_{00}(\Delta t) \\ &= p_0(t)[1 - \lambda\Delta t + o(\Delta t)]. \end{aligned}$$

Subtracting $p_0(t)$ from both sides, dividing by Δt, and taking the limit as $\Delta t \to 0$, gives,

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t).$$

Given the initial condition, $p_0(0) = \mathbb{P}(X(0) = 0) = 1$, the solution to this differential equation is,

$$p_0(t) = e^{-\lambda t}.$$

Similarly, for $i \geq 1$, the Chapman-Kolmogorov equation gives,

$$p_{0i}(t + \Delta t) = p_{0i}(t)p_{ii}(\Delta t) + p_{0,i-1}(t)p_{i-1,i}(\Delta t) + \sum_{k=0}^{i-2} p_{0k}(t)p_{ki}(\Delta t),$$

which is equal to,

$$p_i(t + \Delta t) = p_i(t)[1 - \lambda\Delta t + o(\Delta t)] + p_{i-1}(t)[\lambda\Delta t + o(\Delta t)] + o(\Delta t),$$

for Δt sufficiently small. Subtracting $p_i(t)$ from both sides, dividing by Δt and taking the limit as $\Delta t \to 0$, gives,

$$\frac{dp_i(t)}{dt} = -\lambda p_i(t) + \lambda p_{i-1}(t).$$

We now have the system of forward Kolmogorov differential equations for the state probabilities, $p_i(t)$,

$$\begin{cases} \frac{dp_0(t)}{dt} = -\lambda p_0(t), \\ \frac{dp_i(t)}{dt} = -\lambda p_i(t) + \lambda p_{i-1}(t), \quad \text{for } i \geq 1, \end{cases}$$

with initial conditions $p_0(0) = 1$, $p_i(0) = 0$ for $i \geq 1$. This system can be solved sequentially, and it can be shown by induction that the general solution is,

$$p_i(t) = (\lambda t)^i \frac{e^{-\lambda t}}{i!}, \qquad i = 0, 1, 2, ...$$

Hence, the probability distribution, $\{p_i(t)\}_{i=0}^{\infty}$, represents a Poisson probability distribution with parameter λt. Since $X(t) \sim \text{Poisson}(\lambda t)$, $X(t)$ has mean and variance $\mu(t) = \sigma^2(t) = \lambda t$.

2.2.7 Example: Birth-and-death process

Another well-known CTMC is the birth-and-death process. In this process, the state-space is the set of discrete states, $S = \{0, 1, 2, \dots\}$, and in each jump of the CTMC, the state can only increase or decrease by one, due to a 'birth' event or a 'death' event. The process is specified by birth rates and death rates. When the state of the process is $X(t) = i$, the birth rate is denoted by $\lambda_i \geq 0$ and the death rate is denoted by $\mu_i \geq 0$. We define $\mu_0 = 0$ since a death cannot occur

from state $X(t) = 0$ because it would take the CTMC outside of the state-space. This process is depicted in Figure 2.2.

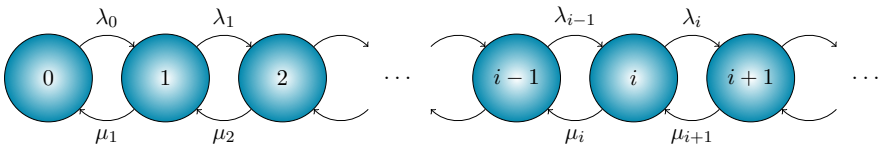

Figure 2.2: The rate diagram for a birth-and-death process. In this process, the size of the population can only increase or decrease by one at a time due to a birth or a death event. The birth and death rates for state i are given by $\lambda_i \geq 0$ and $\mu_i \geq 0$ respectively.

For the *linear* birth-and-death process with infinite state-space, the rates are defined by

$$\lambda_i = i\lambda \qquad \text{and} \qquad \mu_i = i\mu,$$

for some $\lambda > 0,\ \mu > 0$. Hence, $\lambda_i \geq 0,\ \mu_i \geq 0$ for $i = 0, 1, 2, ...$ and $\lambda_0 = \mu_0 = 0$, so 0 is an absorbing state.

The transition rates for a linear birth-and-death process are

$$q_{ij} = \begin{cases} i\lambda, & \text{if } j = i+1, \\ i\mu, & \text{if } j = i-1, \\ -i(\lambda+\mu), & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

Hence the generator matrix is

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \dots \\ 0 & 2\mu & -2(\lambda+\mu) & 2\lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

For Δt sufficiently small, the transition probabilities of a linear birth-and-death

process satisfy

$$\begin{aligned} p_{ij}(\Delta t) &= \mathbb{P}(X(t+\Delta t) = j \mid X(t) = i), \\ &= \begin{cases} i\lambda\Delta t + o(\Delta t), & \text{if } j = i+1, \\ i\mu\Delta t + o(\Delta t), & \text{if } j = i-1, \\ 1 - i(\lambda+\mu)\Delta t + o(\Delta t), & \text{if } j = i, \\ o(\Delta t), & \text{otherwise.} \end{cases} \end{aligned}$$

One can derive the Kolmogorov differential equations for the state probabilities, $p_i(t)$, as follows. For $i = 0$,

$$\begin{aligned} p_0(t+\Delta t) &= p_0(t)p_{00}(\Delta t) + p_1(t)p_{10}(\Delta t) + \sum_{j=2}^{\infty} p_j(t)p_{j0}(\Delta t) \\ &= p_0(t) + p_1(t)[\mu\Delta t + o(\Delta t)] + \sum_{j=2}^{\infty} p_j(t)o(\Delta t). \end{aligned}$$

Subtracting $p_0(t)$ from both sides, dividing by Δt, and taking the limit as $\Delta t \to 0$, one obtains,

$$\frac{dp_0(t)}{dt} = \mu p_1(t).$$

Similarly, for $i > 0$,

$$\frac{dp_i(t)}{dt} = \lambda(i-1)p_{i-1}(t) - (\lambda+\mu)ip_i(t) + \mu(i+1)p_{i+1}(t).$$

For the linear birth-and-death process, the state probabilities, $\{p_i(t)\}_{i=0}^{\infty}$, cannot be calculated easily by sequentially solving the Kolmogorov differential equations, as in the Poisson process example. However they can be easily found by using the probability generating function of the process.

Definition 2.2.7 *The **probability generating function** (pgf) of a CTMC $\mathcal{X}$, with state-space $S = \{0, 1, 2, ...\}$ and initial state $X(0) = k$, is defined by,*

$$G_k(z,t) = \mathbb{E}\left[z^{X(t)}\right] = \sum_{i=0}^{\infty} p_{ki}(t)z^i,$$

where $p_{ki}(t) = \mathbb{P}(X(t) = i \mid X(0) = k)$.

Suppose we have a linear birth-and-death process with initial state $X(0) = 1$. For simplicity, let $G_1(z,t)$ be written as $G(z,t)$. The pgf of the process can be used to calculate the state probabilities, since $p_i(t) = \frac{1}{i!}\frac{\partial^i G}{\partial z^i}\Big|_{z=0}$. To find an explicit formula for $G(z,t)$, one can form and solve a differential equation in G. By considering the possible states that the process can be in at time Δt, we can write,

$$G(z, t+\Delta t) = (1-(\lambda+\mu)\Delta t)G(z,t) + \mu\Delta t + \lambda\Delta t G_2(z,t).$$

It can be seen that $G_2(z,t) = (G(z,t))^2$, since the linear birth-and-death process starting in state 2 can be thought of as the sum of two independent processes starting in state 1. Therefore we have the following ordinary differential equation (ODE) in G,

$$\frac{dG}{dt} = \mu - (\lambda+\mu)G + \lambda G^2,$$

with initial condition $G(0) = z$, since $p_1(0) = 1$ and $p_i(0) = 0$ for $i \neq 1$.
This can be solved to obtain the solution,

$$G(z,t) = \frac{(z\lambda-\mu)e^{(\mu-\lambda)t} - \mu(z-1)}{(z\lambda-\mu)e^{(\mu-\lambda)t} - \lambda(z-1)}, \qquad \text{for } \lambda \neq \mu.$$

We can now write down the state probabilities. For example,

$$p_0(t) = G(0,t) = \frac{\mu(1-e^{(\mu-\lambda)t})}{\lambda - \mu e^{(\mu-\lambda)t}}.$$

Since zero is an absorbing state, $p_0(t)$ is the probability that the population becomes extinct before time t, i.e. it gives the cdf of the extinction time. We can then find the pdf of the extinction time by differentiating,

$$p_{\text{ext}}(t) = \frac{d}{dt}p_0(t) = \frac{\mu(\mu-\lambda)^2 e^{(\mu-\lambda)t}}{(\lambda-\mu e^{(\mu-\lambda)t})^2}.$$

We can find the overall probability of extinction, by taking the limit of $p_0(t)$ as $t \to \infty$, as follows,

$$p_0(\infty) = \lim_{t\to\infty} p_0(t) = \begin{cases} 1, & \text{if } \lambda \leq \mu, \\ \frac{\mu}{\lambda}, & \text{if } \lambda > \mu. \end{cases}$$

Hence, when the death rate is greater or equal to the birth rate, then it is certain that the population will eventually become extinct. Whereas, if the birth rate is greater than the death rate, then the probability of eventual extinction is $\frac{\mu}{\lambda}$.

2.3 Global sensitivity analysis

Sensitivity analysis of a model aims to quantify the relative importance of each input parameter in determining the value of a particular output variable. Local sensitivity analysis quantifies how the parameters affect the model output, by considering the derivative of the model output with respect to each model input parameter. To do this, it is necessary to consider a fixed set of input parameters for the model, $\boldsymbol{\theta}^*$, then vary one parameter, θ_k, at a time, by a small fraction of its nominal value, θ_k^*, and estimate the sensitivity of the model output, u, with respect to θ_k by calculating $\frac{\partial u}{\partial \theta_k} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$. On the other hand, the global sensitivity approach does not specify an input $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, and considers the model over the whole parameter space (i.e. an n-dimensional box if the model has n parameters) (Sobol (2001)). It quantifies how the uncertainty of the input parameters contributes to the output uncertainty, and helps to identify key parameters whose uncertainty affects the output most.

The Sobol method of global sensitivity analysis is explained by Sobol (1993) as follows. Let $\boldsymbol{\theta}$ be a vector of input parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, .., \theta_n)$. Suppose the output of the mathematical model whose sensitivity to the input parameters is to be assessed is described by a scalar function, $f(\boldsymbol{\theta})$. Each parameter is assumed to range over some finite interval, which may be assumed to be [0,1], after rescaling. Hence, assume that $\boldsymbol{\theta}$ takes values on the unit n-dimensional cube,

$$K^n = \{\boldsymbol{\theta} \mid 0 \leq \theta_i \leq 1; i = 1, ..., n\}.$$

The aim is to estimate the sensitivity of $f(\boldsymbol{\theta})$ with respect to its different input parameters or groups of these parameters. Sobol (1993) describes how a function $f(\theta_1, \theta_2, ..., \theta_n)$ can be written as an expansion into terms of increasing dimensions, called the ANOVA representation of $f(\boldsymbol{\theta})$ (Theorem 1 in Sobol (1993)),

$$\begin{aligned} f(\boldsymbol{\theta}) &= f_0 \quad + \quad \sum_{k=1}^{n} \sum_{1 \leq i_1 < ... < i_k \leq n} f_{i_1,...,i_k}(\theta_{i_1}, ..., \theta_{i_k}), \\ &= f_0 \quad + \quad \sum_{i=1}^{n} f_i \quad + \quad \sum_{1 \leq i < j \leq n} f_{i,j} \quad + \quad ... \end{aligned}$$

where

$$\int_0^1 f_{i_1,...,i_k}(\theta_{i_1}, ..., \theta_{i_k}) d\theta_j = 0 \qquad \text{for } j = i_1, ..., i_k.$$

Hence,

$$f_0 = \int_{K^n} f(\boldsymbol{\theta})\ d\boldsymbol{\theta}.$$

Assume $\boldsymbol{\theta}$ is a point randomly uniformly distributed in K^n, so has pdf $f^*(\boldsymbol{\theta}) = 1$. Then the expectation of $f(\boldsymbol{\theta})$ is given by,

$$\int_{K^n} f^*(\boldsymbol{\theta}) f(\boldsymbol{\theta})\ d\boldsymbol{\theta} = \int_{K^n} f(\boldsymbol{\theta})\ d\boldsymbol{\theta} = f_0.$$

Hence the variance of the function $f(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is randomly uniformly distributed is,

$$D = \int_{K^n} f^2(\boldsymbol{\theta})\ d\boldsymbol{\theta} - f_0^2,$$

and the other functions in the ANOVA representation have expectation 0 and variance,

$$D_{i_1,\ldots,i_k} = \int_0^1 \ldots \int_0^1 f^2_{i_1,\ldots,i_k} d\theta_{i_1} \ldots d\theta_{i_k}.$$

The necessary integrals to compute the variances can be estimated by a Monte Carlo integration method (Sobol (2001)). These variances characterise how much the functions vary, so to estimate the sensitivity of the model to certain parameters or groups of parameters one can use the Sobol global sensitivity indices, given by,

$$S_{i_1,\ldots,i_k} = \frac{D_{i_1,\ldots,i_k}}{D} \geq 0,$$

which show how each function in the ANOVA representation varies compared to how the model function varies as a whole. These sensitivity indices all sum to 1. Also $f_{i_1,\ldots,i_k} = 0 \iff S_{i_1,\ldots,i_k} = 0$, so the function $f(\boldsymbol{\theta})$ is independent of θ_i if and only if all the $S_{i_1,\ldots,i_k}$ containing the index i are zero.

Once the sensitivity indices have been estimated, one can order the parameters according to the values $S_1, \ldots, S_n$. These are called the first order sensitivity indices and they describe the contribution to the total variance, D, due to each parameter alone.

To consider sensitivity indices for a subset of m parameters, $y = (\theta_{j_1}, \ldots, \theta_{j_m})$, with $J = \{j_1, \ldots, j_m\}$, the variance corresponding to the subset y is defined as,

$$D_y = \sum_{k=1}^{m} \sum_{(i_1 < \ldots < i_k) \in J} D_{i_1,\ldots,i_k},$$

which sums over all subsets of indices in J. Also the total variance corresponding to the subset y is defined as,

$$D_y^{tot} = D - D_z,$$

where $D_z = \{\theta_1, ..., \theta_n\} \setminus y$. That is, the sum of variances over all subsets of indices where at least one element is a member of J. Finally,

$$S_y = \frac{D_y}{D}, \qquad S_y^{tot} = \frac{D_y^{tot}}{D},$$

where $S_y^{tot} = 1 - S_z$ and $0 \leq S_y \leq S_y^{tot} \leq 1$. If $S_y = S_y^{tot} = 0$, then $f(\boldsymbol{\theta})$ is independent of all parameters in the subset y. If $S_y = S_y^{tot} = 1$, then $f(\boldsymbol{\theta})$ depends only on y and is independent of z.

The total order sensitivity index for parameter θ_i is $S_{T_i} = S_y^{tot}$, where $y = \{\theta_i\}$, and this provides a measure of the total effect of parameter θ_i, including its first order effect and all the effects of interactions with other parameters (Homma & Saltelli (1996)). The difference $S_{T_i} - S_i$ is a measure of how much θ_i is involved in interactions with other parameters. If there is a significant difference between S_{T_i} and S_i, this indicates that there are important interactions involving parameter θ_i (Saltelli *et al.* (2008)).

Usually the full set of first order sensitivity indices and the full set of total order sensitivity indices are computed (Saltelli (2002)). The Python sensitivity analysis library, SALib, is used in this thesis to implement the Sobol method of global sensitivity analysis, to calculate the effects over time of different model parameters on the populations of the intracellular model in Chapter 3.

2.4 Bayesian inference

Throughout this thesis, Bayesian inference will be performed in order to determine estimates for unknown parameters in the mathematical models, using available data. The goal of Bayesian inference is to infer a posterior distribution, $\pi(\boldsymbol{\theta}|X)$, over the model parameters, $\boldsymbol{\theta}$, given the data, X. This is usually done via Bayes' Theorem,

$$\pi(\boldsymbol{\theta}|X) \propto \pi(X|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{2.4.1}$$

which says that the posterior density is proportional to the product of the likelihood, $\pi(X|\boldsymbol{\theta})$, and the prior density, $\pi(\boldsymbol{\theta})$. Hence the likelihood of observing the data under a given model parametrisation is used to update prior information and beliefs about the model parameters, encoded in the prior distribution.

However, the main method of inference used here will be Approximate Bayesian computation (ABC), in which the evaluation of the likelihood function is replaced by simulating data for given parameters and comparing this simulated data with the observed data. Classical ABC is performed by the following steps,

1. sample parameters $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$,
2. simulate data, X^*, from the model, given $\boldsymbol{\theta}$,
3. accept $\boldsymbol{\theta}$ if $d(X^*, X) \leq \varepsilon$,

for some distance function, d, and acceptance threshold, ε. This is repeated until N parameters have been accepted, and these accepted parameters represent a sample of size N from the approximate posterior distribution, $\pi(\boldsymbol{\theta}|d(X^*, X) \leq \varepsilon)$. If ε is chosen small enough, then this will be a good approximation to the posterior distribution. However, decreasing ε can also decrease the acceptance rate.

ABC can be combined with a sequential Monte Carlo scheme (ABC-SMC) (Toni *et al.* (2009)), which can improve acceptance rates by using an iterative reduction of the acceptance thresholds, $\varepsilon_1 \geq \varepsilon_2 \cdots \geq \varepsilon_T \geq 0$. At each iteration, parameter values are sampled from the accepted parameters of the previous iteration and are perturbed with a kernel function. For example, a component-wise uniform perturbation kernel can be used, so that each component of the parameter set is perturbed independently in a uniform interval. The perturbed parameter set is then used to simulate data from the model and is accepted if the distance between the simulated and observed data falls below the acceptance threshold for that iteration. This generates a set of intermediate distributions of accepted parameters, $\pi(\boldsymbol{\theta}|d(X^*, X) \leq \varepsilon_i)$ for $i = 1, \ldots, T$, that gradually converge towards the target posterior. This is described in Algorithm 2.2.

Algorithm 2.2: Algorithm for ABC-SMC

Select prior distribution, $\pi(\boldsymbol{\theta})$;

Select distance function, d;

Select values of $\varepsilon_1 \geq \varepsilon_2 \cdots \geq \varepsilon_T \geq 0$;

Select perturbation kernel(s), K_t;

Set the population indicator to $t = 1$;

while $t \leq T$ **do**

  Set the particle indicator to $i = 1$;

  while $i \leq N$ **do**

    if *t=1* **then**

      Sample $\boldsymbol{\theta}^{**}$ from $\pi(\boldsymbol{\theta})$;

    end

    else

      Sample $\boldsymbol{\theta}^{*}$ from the previous population $\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1}^{N}$ with associated weights $\{w_{t-1}^{(i)}\}_{i=1}^{N}$;

      Perturb the particle to obtain $\boldsymbol{\theta}^{**} \sim K_t(\boldsymbol{\theta}|\boldsymbol{\theta}^{*})$;

    end

    if $\pi(\boldsymbol{\theta}^{**}) > 0$ **then**

      Simulate the data, X^{*}, from the model, given $\boldsymbol{\theta}^{**}$;

      if $d(X^{*}, X) \leq \varepsilon_t$ **then**

        Set $\boldsymbol{\theta}_t^{(i)} = \boldsymbol{\theta}^{**}$;

        Calculate the weight for particle $\boldsymbol{\theta}_t^{(i)}$,

$$w_t^{(i)} = \begin{cases} 1 & \text{if } t = 1, \\ \frac{\pi(\boldsymbol{\theta}_t^{(i)})}{\sum_{j=1}^{N} w_{t-1}^{(j)} K_t(\boldsymbol{\theta}_t^{(i)}|\boldsymbol{\theta}_{t-1}^{(j)})} & \text{if } t > 1. \end{cases}$$

        Set $i = i + 1$;

      end

    end

  end

  Normalise the weights;

  Set $t = t + 1$;

end

Chapter 3

Modelling intracellular anthrax infection with spore germination heterogeneity

Following inhalation of *Bacillus anthracis* spores, these are ingested by alveolar phagocytes. Phagocytosed spores then begin to germinate and divide intracellularly. Meanwhile, the host cell can migrate across the lung lining to transport *B. anthracis* spores and bacteria to the nearby lymph nodes. Some phagocytes successfully eliminate the intracellular bacteria and will recover. On the other hand, phagocytosis of spores can lead to the eventual death of the host cell and the extracellular release of bacterial progeny. These events are referred to as the Trojan Horse model, and although this hypothesis is widely accepted, the fraction of germinated spores that are killed rather than escaping to replicate extracellularly is unknown. More work is needed in order to answer this question and to quantify this key host-pathogen interaction during inhalational anthrax.

In this chapter I will present a stochastic mathematical model of the intracellular infection dynamics of inhalational anthrax, which adapts and extends the deterministic one of Pantha *et al.* (2018). The model by Pantha *et al.* (2018) is a system of ODEs representing the interaction between macrophages and *B. anthracis* spores, and considers two intracellular bacterial populations: newly germinated bacteria, which are susceptible to macrophage killing but unable to replicate, and vegetative bacteria, which are susceptible to macrophage killing and

capable of replicating. Spores germinate into newly germinated bacteria, and the newly germinated bacteria must mature into vegetative bacteria before they can begin to replicate. Similarly to the model by Pantha *et al.* (2018), the model in this chapter considers the germination of spores, replication of bacteria, and killing of bacteria by the host cell. Still, I make use of a stochastic approach, instead of a deterministic one, to describe the population dynamics of spores and bacteria. I follow the methods recently developed by Carruthers *et al.* (2020) for *Francisella tularensis* infection, extended here to include spores and spore germination, since *B. anthracis* is a spore-forming bacteria and *F. tularensis* is not. Thus, the resulting stochastic model is a linear birth-and-death process with killing, extended to account for the mechanisms of spore germination and maturation of *B. anthracis*.

An important mechanism in the stochastic model is the consideration of macrophage rupture, not explicitly considered by Pantha *et al.* (2018). The rupture of host cells and the release of bacteria into the extracellular environment is an important mechanism in the pathogenesis of anthrax. Thus, incorporating this event into the model allows one to better understand both the timescales of macrophage rupture, and the rupture size distribution (*i.e.,* the number of vegetative bacteria released upon rupture). These summary statistics can then play an important role when considering within-host infection dynamics and when linking to dose-response data (Pratt *et al.* (2020)), which is studied in Chapter 4. In the same way as Carruthers *et al.* (2020), I assume that an infected macrophage's rupture probability per unit time is proportional to its bacterial load. Thus, cells with a high bacterial load at a given time are more likely to rupture than those with a lower one. This hypothesis is supported by Ruthel *et al.* (2004), who suggest that the intracellular bacterial load may be a contributing factor to whether a macrophage will survive an infection.

The model is used to explore the potential for heterogeneity in the spore germination rate. Motivation for this comes from the work by Setlow (2003, 2013, 2014), where it is shown that germination rates are highly heterogeneous for the *Bacillus* species spores, with germination times ranging from a few minutes to longer than 24 hours. It is thought that this spore germination heterogeneity

is primarily due to variation in the germinant receptor levels between individual spores. Setlow mentions that spores with very low germinant receptor levels germinate extremely slowly and are termed superdormant (Setlow (2003, 2013, 2014)). Hence, the model is used to explore different hypotheses for this heterogeneity, with the consideration of two extreme cases for the rate distribution: continuous Gaussian and discrete Bernoulli. The first hypothesis is that the germination rate is continuously distributed in a population of spores and follows a truncated normal distribution. The second hypothesis is that the population of spores can be roughly split into two discrete groups, with different germination rates, where one group corresponds to the spores with "average" germinant receptor levels, and the other corresponds to the superdormant spores.

For the linear birth-and-death process with killing including spore germination rate heterogeneity, I show how to compute the probability of either rupture or recovery of the infected cell and the conditional mean times taken to reach these fates. Furthermore, I adapt some of the results from Carruthers *et al.* (2020) in order to compute the probability distribution of rupture times, which is shown to be proportional to the mean number of vegetative bacteria in the cell over time. I also compute the probability distribution of the rupture size, which is the number of bacteria that are eventually released into the extracellular environment from a single infected cell.

Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) (Toni *et al.* (2009)) is used to calibrate the model, using experimental measurements of spore and bacterial counts from *in vitro* infection of murine peritoneal macrophages with spores of the Sterne 34F2 strain of *B. anthracis* (Kang *et al.* (2005)). The calibrated stochastic model is then used to obtain numerical results to quantify the implications of the model. In particular, I compute the probability of rupture, mean time to rupture, and rupture size distribution, of a macrophage that has been infected with one spore. I also obtain the mean spore and bacterial loads over time for a population of cells, each assumed to be initially infected with a single spore. The results support the existence of significant heterogeneity in the germination rate, with a subset of spores expected to germinate much later than the majority. Furthermore, in agreement with experimental evidence, the results suggest that the majority of spores taken up by

macrophages are likely to be eliminated by the host cell, but a few germinated spores may survive phagocytosis and lead to the death of the infected cell.

3.1 Mathematical model

This section introduces the stochastic model for the dynamics of *B. anthracis* spores and bacteria in a single infected phagocyte, starting at the time when the cell phagocytoses a spore, and ending either with rupture and death of the cell and the release of bacteria into the extracellular environment, or with recovery of the cell and the elimination of any intracellular spores or bacteria. When considering low dose exposures, for which the multiplicity of infection (MOI), which represents the ratio between the numbers of spores and phagocytes, will be low, it is reasonable to assume that each phagocyte will only engulf one spore. Therefore, in what follows I only consider infection of a cell that has phagocytosed a single spore. The model presented here includes germination of the spore into a newly germinated bacterium, maturation of the newly germinated bacterium into a vegetative bacterium, replication of vegetative bacteria, death of bacteria, and rupturing of the host cell to release the intracellular bacteria (see Figure 3.1).

The intracellular infection model, depicted in Figure 3.1, corresponds to a continuous-time Markov chain (CTMC), $\mathcal{X} = \{X(t) : t \geq 0\}$, taking values in the state-space $\mathcal{S} = \{1_S, 1_{NGB}, 0, 1, 2, ...\} \cup \{R\}$, where:

- 1_S corresponds to the state where the host cell contains a spore and no bacteria,
- 1_{NGB} corresponds to the state where the host cell contains a single newly germinated bacterium,
- $\{0, 1, 2, ...\}$ represent the number of vegetative bacteria inside the host cell, so that 0 corresponds to the cell's recovery, and
- R corresponds to the state of the host cell having ruptured.

Time $t = 0$ is the time at which the spore is phagocytosed, so the initial state of the process is assumed to be $X(0) = 1_S$. Experimental work using

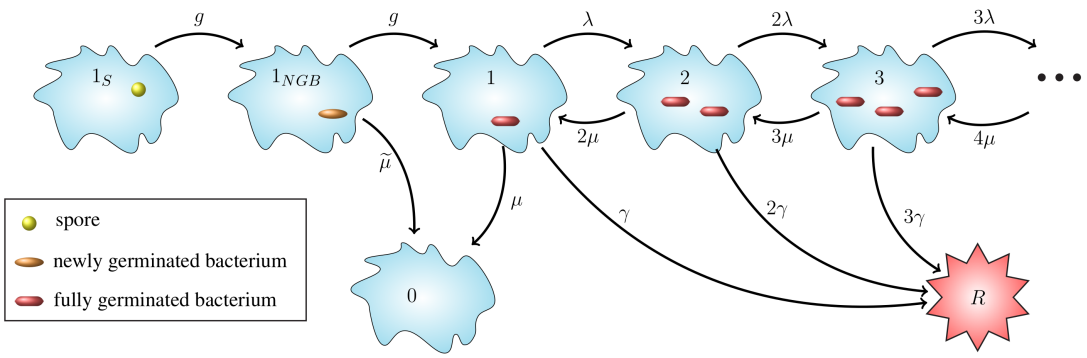

Figure 3.1: Diagram for the intracellular infection model. State 1_S represents a phagocytosed spore and state 1_{NGB} a newly germinated bacterium (NGB). The germination-maturation time is assumed to follow an Erlang(2, g) distribution, which is the simplest approach to consider a non-exponential distribution for the time that it takes the spore to become a vegetative bacterium, while including an intermediate, susceptible state as done by Pantha *et al.* (2018), and keeping the process Markovian. The rate g is assumed to vary between spores. The death rate of the newly germinated bacterium is given by $\widetilde{\mu}$ $hours^{-1}$. States $i \in \mathbb{N} \cup \{0\}$ represent i intracellular bacteria. State 0 represents recovery and state R the rupture of the cell, which are both absorbing states for the stochastic process. Transitions between states $i \in \mathbb{N}$ represent three types of events: transition to state $i+1$ (division of a bacterium), to state $i-1$ (death of a bacterium), and to state R (rupture of the host cell with release of i bacteria). The per bacterium division, death, and rupture rates are $\lambda > 0$, $\mu > 0$, and $\gamma > 0$, respectively, all with units $(bacteria \cdot hours)^{-1}$, leading to a linear birth-and-death process with killing. The infected cell survives for as long as it does not reach state R.

murine macrophages has indicated that the ungerminated spore form of anthrax is not easily eliminated by host cells (Kang *et al.* (2005)), so the model assumes that the initial spore will not die or be killed by the cell. The germination-maturation time is represented as an Erlang(2, g) distribution, as depicted in Figure 3.1, where the rate g has units $hours^{-1}$. As discussed in Section 3.4, a limitation of the model here is that the same rate is considered for each step of the germination-maturation process, since to my knowledge there is no data to properly characterise the difference in the time taken for these two steps. Due to this lack of data, I chose to use an Erlang(2, g) representation because this is the simplest approach to consider a non-exponential distribution for the time that it takes the spore to become a vegetative bacterium, while including an intermediate, susceptible state as done by Pantha *et al.* (2018), and keeping the process Markovian. During the germination process, there is an increase in the hydration of the spore core and the spore cortex is broken down (Setlow (2003)). These changes mean that the spore loses some of its resistance against the anti-microbial environment within the host cell and may be killed by the reactive oxygen and nitrogen species, and anti-microbial peptides within the phagolysosome (Banks *et al.* (2005)). As a result, modelling the germination-maturation process with two stages allows one to incorporate this loss of resistance into the intermediate state 1_{NGB}. The newly germinated bacterium can be killed by the cell, with rate $\widetilde{\mu}\ hours^{-1}$, but cannot replicate unless it matures into a vegetative bacterium, represented by state 1. If the stochastic process reaches state 1, the subsequent replication of bacteria, death of bacteria, and rupture of the host cell is modelled as a birth-and-death process with killing (Karlin & Tavaré (1982)), with state-space $\{0, 1, 2, ...\} \cup \{R\}$ (see Figure 3.1). The per bacterium division, death, and rupture rates are denoted by λ, μ and γ, respectively, all with units $(bacteria \cdot hours)^{-1}$. The stochastic process has two absorbing states: the recovery state, 0, representing elimination of any intracellular spores and bacteria, and the rupture state, R, representing rupture of the infected cell and release of its entire content of bacteria. Table 3.1 lists the parameters of the model, for reference.

In the subsequent sections, the dynamics of the model will be analytically studied. In Section 3.1.1, two hypotheses for the germination rate distribution

Parameter	Description
g	Germination and maturation rate of spores
$\widetilde{\mu}$	Death rate of newly germinated bacteria
λ	Replication rate of vegetative bacteria
μ	Death rate of vegetative bacteria
γ	Rupture rate

Table 3.1: List of parameters for the intracellular infection model in Figure 3.1.

will be introduced. Then in Section 3.1.2, expressions will be found for the mean number of spores and newly germinated bacteria in a single cell over time. The model will also be studied through the calculation of stochastic descriptors, such as the probabilities of recovery and rupture, the distribution and mean of the times to reach recovery and rupture, and the probability distribution of the number of bacteria released from an infected cell (Sections 3.1.3 and 3.1.4, respectively). Finally, Section 3.1.5 will show how to calculate the mean number of intracellular vegetative bacteria over time, which is found to be proportional to the probability density function of the rupture time.

3.1.1 Spore germination heterogeneity

For each spore, its germination-maturation time is assumed to follow an Erlang(2, g) distribution. However, in order to reflect the large heterogeneity in spore germination times as reported in the literature (Setlow (2003, 2013, 2014)), the germination rate g is assumed to vary between spores. Here two different distributions are introduced for the germination rate. These two hypotheses will later be explored and compared, making use of experimental data, in Section 3.2.

Continuous Gaussian distribution

This hypothesis assumes that the germination rate varies continuously among spores, according to a normal distribution on the positive axis. In this case, the germination rate g for a given spore is a realisation of the continuous random variable $G \sim N_{(0,+\infty)}(\mu_g, \sigma_g^2)$, which is normally distributed with mean μ_g $hours^{-1}$

and standard deviation σ_g $hours^{-1}$, and is truncated to the interval $(0, +\infty)$. Therefore, the germination rate has probability density function given by

$$f_G(g) = \frac{1}{Z}\frac{1}{\sigma_g\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{g-\mu_g}{\sigma_g}\right)^2}, \qquad \text{for } g \in (0, +\infty), \tag{3.1.1}$$

where $Z = \Phi\left(\frac{\mu_g}{\sigma_g}\right)$ is a normalisation factor, and Φ is the cumulative distribution function of the standard normal distribution.

Discrete Bernoulli distribution

Setlow (2013, 2014) explains that some spores can be described as superdormant because they have very low germinant receptor levels and germinate extremely slowly, taking many hours or even days. Therefore, a second choice for the germination rate distribution is proposed here, which assumes that the population of spores can be roughly split into two discrete groups, type A and type B, with type A having a faster germination rate than type B. Here, type A corresponds to the spores with "typical" germinant receptor levels, and type B corresponds to spores with significantly lower levels. This is represented here as a discrete Bernoulli distribution with probability mass function as follows

$$f_B(g) = \begin{cases} \varepsilon & \text{if } g = g_A, \\ 1-\varepsilon & \text{if } g = g_B, \end{cases} \tag{3.1.2}$$

for some parameter values $g_A > g_B$ and $\varepsilon \in (0, 1)$.

3.1.2 Number of intracellular spores and newly germinated bacteria

The transition probability from state $i \in \mathcal{S}$ to state $j \in \mathcal{S}$ in time t is defined by,

$$p_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i). \tag{3.1.3}$$

We are interested in the dynamics of the process when the initial state is $X(0) = 1_S$, representing an intracellular spore which was just phagocytosed. Therefore, when the initial state is $i = 1_S$, the first index will be omitted in the notation, so

that the probability that the process is in state j at time t, given that it started with one spore, is denoted by

$$p_j(t) = \mathbb{P}(X(t) = j \mid X(0) = 1_S). \tag{3.1.4}$$

To describe the dynamics of the germination-maturation process, one can consider the transient state probabilities, $p_j(t)$, for each of the states, $j \in \{1_S, 1_{NGB}\}$. These probabilities obey the following Kolmogorov differential equations, with initial conditions $p_{1_S}(0) = 1$ and $p_{1_{NGB}}(0) = 0$,

$$\frac{dp_{1_S}}{dt} = -gp_{1_S}, \tag{3.1.5}$$

$$\frac{dp_{1_{NGB}}}{dt} = gp_{1_S} - (g + \widetilde{\mu})p_{1_{NGB}}. \tag{3.1.6}$$

If a phagocyte engulfs a spore with germination rate g at time $t = 0$, then the cell will contain one spore at time t, if $X(t) = 1_S$, and zero spores if $X(t) \neq 1_S$. Hence, the mean number of intracellular spores at time t is equal to $p_{1_S}(t)$. Similarly, the mean number of intracellular newly germinated bacteria at time t is equal to $p_{1_{NGB}}(t)$. These are given by the solutions to Eqs. (3.1.5) and (3.1.6),

$$p_{1_S}(t) = e^{-gt}, \tag{3.1.7}$$

$$p_{1_{NGB}}(t) = \frac{g}{\widetilde{\mu}}\left(e^{-gt} - e^{-(g+\widetilde{\mu})t}\right), \tag{3.1.8}$$

for any $t \geq 0$.

Given the truncated normal distribution for the germination rate, the mean number of spores inside the infected cell at time $t \geq 0$ is given by

$$\begin{aligned} S(t) &= \int_0^{+\infty} f_G(g)p_{1_S}(t)\, dg \\ &= \int_0^{+\infty} f_G(g)e^{-gt}\, dg \\ &= \frac{1}{Z}\exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t\right)\Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g}\right). \end{aligned} \tag{3.1.9}$$

Similarly, the mean number of newly germinated bacteria inside the infected cell

at time $t \geq 0$ is given by

$$\begin{aligned} B_{NGB}(t) &= \int_0^{+\infty} f_G(g) p_{1_{NGB}}(t)\, dg \\ &= \int_0^{+\infty} f_G(g) \frac{g}{\widetilde{\mu}} \Big(e^{-gt} - e^{-(g+\widetilde{\mu})t} \Big)\, dg \\ &= \frac{1}{\widetilde{\mu} Z} (1 - e^{-\widetilde{\mu} t}) \left[\frac{\sigma_g}{\sqrt{2\pi}} e^{-\frac{\mu_g^2}{2\sigma_g^2}} + (\mu_g - \sigma_g^2 t) \exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t \right) \Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g} \right) \right]. \end{aligned} \tag{3.1.10}$$

For the Bernoulli distribution, the mean number of spores and newly germinated bacteria inside the infected cell at time $t \geq 0$ are, respectively,

$$\begin{aligned} S(t) &= \varepsilon e^{-g_A t} + (1 - \varepsilon) e^{-g_B t}, \\ B_{NGB}(t) &= \frac{\varepsilon g_A}{\widetilde{\mu}} \Big[e^{-g_A t} - e^{-(g_A + \widetilde{\mu})t} \Big] + \frac{(1-\varepsilon) g_B}{\widetilde{\mu}} \Big(e^{-g_B t} - e^{-(g_B + \widetilde{\mu})t} \Big). \end{aligned} \tag{3.1.11}$$

.

3.1.3 Probabilities and times to reach absorbing states

The CTMC in Figure 3.1 has two absorbing states, R and 0, which denote rupture and recovery of the host cell, respectively. In this section I show how to compute the probability of either rupture or recovery of the cell, and the probability density functions of the recovery and rupture times. I also explain how to compute the conditional mean times taken to reach each of the two cellular fates.

The transient probabilities for the states $j \in \mathbb{N} \cup \{0, R\}$, obey the following system of Kolmogorov differential equations

$$\begin{aligned} \frac{dp_1}{dt} &= g p_{1_{NGB}} + 2\mu p_2 - (\lambda + \mu + \gamma) p_1, \\ \frac{dp_j}{dt} &= \lambda(j-1) p_{j-1} + \mu(j+1) p_{j+1} - (\lambda + \mu + \gamma) j p_j, \qquad \text{for } j \geq 2, \\ \frac{dp_0}{dt} &= \widetilde{\mu} p_{1_{NGB}} + \mu p_1, \\ \frac{dp_R}{dt} &= \sum_{j=1}^{+\infty} \gamma j p_j. \end{aligned} \tag{3.1.12}$$

The long-term probabilities of recovery or rupture for an infected cell, starting from any state $i \in \mathcal{S}$ can be denoted, respectively, by

$$r_i^0 = \lim_{t \to +\infty} p_{i,0}(t), \qquad r_i^R = \lim_{t \to +\infty} p_{i,R}(t). \tag{3.1.13}$$

These probabilities can also be expressed in terms of the times it takes the process $\mathcal{X}$ to reach states 0 and R, respectively. In particular, one can denote the times taken for the process to travel from state i to states 0 or R, respectively, by

$$T_i^0 = \inf\{t \geq 0 : X(t) = 0 \mid X(0) = i\}, \quad T_i^R = \inf\{t \geq 0 : X(t) = R \mid X(0) = i\}. \tag{3.1.14}$$

Since there is a choice of two possible absorbing fates (recovery or rupture), the random variables T_i^0 and T_i^R may be infinite with non-zero probability. That is, the time to recovery, T_i^0, will be infinite if the process ends in the rupture state, and vice versa. Thus, we can write

$$p_{i,0}(t) = \mathbb{P}(T_i^0 \leq t), \qquad p_{i,R}(t) = \mathbb{P}(T_i^R \leq t), \tag{3.1.15}$$

and

$$r_i^0 = \mathbb{P}(T_i^0 < +\infty) = 1 - \mathbb{P}(T_i^R < +\infty) = 1 - r_i^R, \tag{3.1.16}$$

where r_i^0 and r_i^R represent the recovery and rupture probabilities, respectively. Carruthers *et al.* (2020) studied a very similar process to the one described here, for the non-sporulating bacteria *F. tularensis*. This section makes use some of their results. For instance, survival analysis allowed Carruthers *et al.* (2020) to show that the probability density function of the random variable T_1^R (the rupture time starting from state 1, representing a single fully vegetative bacterium), is given by

$$f_{T_1^R}(t) = \frac{\gamma(b-a)^2 e^{-\lambda(b-a)t}}{[b - 1 + (1-a)e^{-\lambda(b-a)t}]^2}, \quad t \geq 0, \tag{3.1.17}$$

with

$$\begin{aligned} a &= \frac{(\lambda + \mu + \gamma) - \sqrt{(\lambda + \mu + \gamma)^2 - 4\mu\lambda}}{2\lambda}, \\ b &= \frac{(\lambda + \mu + \gamma) + \sqrt{(\lambda + \mu + \gamma)^2 - 4\mu\lambda}}{2\lambda}. \end{aligned} \tag{3.1.18}$$

They also showed that the probability that a cell eventually ruptures, starting in state 1 with one vegetative bacterium, is $r_1^R = 1 - a$, and the probability that a cell eventually recovers is $r_1^0 = a$. One can adapt these results to the model studied here with the help of first-step analysis, to find the probabilities of rupture and recovery starting with one initial spore in state 1_S. If a phagocyte is infected with a spore at time $t = 0$, then at some time point, the initial spore will germinate, transitioning to the intermediate state of newly germinated bacterium (NGB). At some later time, the NGB will either die, with probability $\tilde{\mu}/(\tilde{\mu} + g)$, or will mature into a vegetative bacterium with probability $g/(\tilde{\mu} + g)$. Since it is certain that the first step of the process will be the transition from spore to newly germinated bacterium, the probability of eventual recovery or rupture starting from state 1_S is the same as the probability of recovery or rupture starting from state 1_{NGB}. The only way that the process will eventually reach the rupture state is if the newly germinated bacterium matures into a vegetative bacterium, and then the cell eventually ruptures starting from state 1. On the other hand, the cell can recover if either the newly germinated bacterium dies before it matures into a vegetative bacterium, or the newly germinated bacterium matures and the cell eventually recovers starting from state 1. In particular, given a germination rate g for the initial spore, the probabilities for rupture and recovery, starting from state 1_S, are given by

$$\begin{aligned} r^R_{1_S}(g) &= r^R_{1_{NGB}}(g) = \frac{g}{\tilde{\mu}+g} r_1^R = \frac{g(1-a)}{\tilde{\mu}+g}, \\ r^0_{1_S}(g) &= r^0_{1_{NGB}}(g) = \frac{\tilde{\mu}}{\tilde{\mu}+g} + \frac{g}{\tilde{\mu}+g} r_1^0 = \frac{\tilde{\mu}+ga}{\tilde{\mu}+g}. \end{aligned} \tag{3.1.19}$$

One can also derive the probability density functions of the time to rupture and the time to recovery. These shed light on the distribution of times to rupture and recovery across cells that have been infected with one spore. In what follows, the random variable for the time to transition from state i to state j is denoted by $T_i^j = \inf\{t \geq 0 : X(t) = j \mid X(0) = i\}$, and the probability density function for this random variable is denoted by $f_{T_i^j}(t)$.

First, let us consider the total time for the initial intracellular spore to germinate and mature into a vegetative bacterium, given by the random variable $T^1_{1_S}$,

with probability density function, $f_{T^1_{1_S}}(t)$. This function will be needed later to calculate the probability density function of the times to rupture and recovery. To find the function $f_{T^1_{1_S}}(t)$, let $F_{T^1_{1_S}}(t) = \mathbb{P}(T^1_{1_S} \leq t)$ be the probability that the cell has reached state 1 by time t, given that the germination rate of the initial spore is equal to g. Let us consider a small time step Δt, such that only one transition can occur in the interval $(t,\ t + \Delta t)$. If the cell has not entered state 1, representing a vegetative bacterium, before time t, then it will only be possible to enter state 1 before time $t + \Delta t$ if the cell is in state 1_{NGB} at time t. Furthermore, if the phagocyte contains a newly germinated bacterium at time t, then the probability to transition to a vegetative bacterium between time t and $t + \Delta t$ is $g\Delta t$. Hence,

$$F_{T^1_{1_S}}(t + \Delta t) = F_{T^1_{1_S}}(t) + p_{1_{NGB}}(t) g \Delta t. \tag{3.1.20}$$

The function $F_{T^1_{1_S}}(t)$ is the cumulative distribution function of the random variable, $T^1_{1_S}$, for the time that the process takes to reach a vegetative bacterium, starting from one spore. Therefore, the probability density function for this random variable is

$$f_{T^1_{1_S}}(t; g) = \frac{dF_{T^1_{1_S}}(t)}{dt} = g\ p_{1_{NGB}}(t) = \frac{g^2}{\widetilde{\mu}}\Big(e^{-gt} - e^{-(g+\widetilde{\mu})t}\Big), \quad t \geq 0\ , \tag{3.1.21}$$

where it is written explicitly that $f_{T^1_{1_S}}$ is a function of the germination rate g. It can be verified that

$$\mathbb{P}(T^1_{1_S} < +\infty) = \int_0^{+\infty} f_{T^1_{1_S}}(t; g)\ dt = \frac{g}{g + \widetilde{\mu}}, \tag{3.1.22}$$

which is the probability that the process will eventually reach state 1, or equivalently, the probability that a spore will mature into a vegetative bacterium instead of being cleared by the infected cell.

Time to recovery

Here I show how to compute the probability density function, $f_{T^0_{1_S}}(t)$, for the time to recovery of an infected cell starting with one spore. Following the same

approach as above, the probability density function for the random variable $T^0_{1_S}$, given that the spore has germination rate g, is

$$\begin{aligned} f_{T^0_{1_S}}(t;g) = \frac{dF_{T^0_{1_S}}(t)}{dt} &= \widetilde{\mu} p_{1_{NGB}}(t) + \mu p_1(t) \\ &= g\left(e^{-gt} - e^{-(g+\widetilde{\mu})t}\right) + \mu \int_0^t f_{T^1_{1_S}}(s;g) p_{1,1}(t-s)\, ds, \quad t \geq 0, \end{aligned} \tag{3.1.23}$$

where $p_{1,1}(t)$ is the probability that the process, starting in state 1, is in state 1 at time t, and can be derived from results of Carruthers *et al.* (2020) as follows,

$$p_{1,1}(t) = \frac{(a-b)^2 e^{-\lambda(b-a)t}}{(ae^{-\lambda(b-a)t} - b)^2}, \tag{3.1.24}$$

with a and b defined in Eq. (3.1.18). Note that the probability density function for the time to recovery from state 1 can be written as

$$f_{T^0_1}(t) = \mu p_{1,1}(t). \tag{3.1.25}$$

When the germination rate follows a continuous Gaussian distribution (see Eq. (3.1.1)), the probability density of the recovery time starting from state 1_S is given by

$$f_{T^0_{1_S}}(t) = \int_0^{+\infty} f_G(g) f_{T^0_{1_S}}(t;g)\, dg. \tag{3.1.26}$$

Alternatively, if the germination rate follows a discrete Bernoulli distribution (see Eq. (3.1.2)), the probability density of recovery times is given by

$$f_{T^0_{1_S}}(t) = \varepsilon f_{T^0_{1_S}}(t;g_A) + (1-\varepsilon) f_{T^0_{1_S}}(t;g_B). \tag{3.1.27}$$

The probability density of recovery times yields the distribution of recovery times across cells, since each phagocyte is assumed to be initially infected by a single spore. To gain insights into the expected time of recovery, one can also compute the conditional mean time to recovery of an infected cell, which is the expected time to recovery, given that the cell eventually recovers. This is denoted by $\mathbb{E}[T^0_{1_S} \mid T^0_{1_S} < +\infty]$, given that the eventual recovery of a cell is equivalent to its recovery time being finite. For any initial state $i \in \mathcal{S}$, one can define the restricted mean time to recovery as $\tau^0_i = \mathbb{E}[T^0_i \cdot \delta_{T^0_i < +\infty}]$, where δ_A is equal to 1

if A is satisfied and 0 otherwise. Then the conditional mean time to recovery, starting from state $i \in \mathcal{S}$, is defined by

$$\mathbb{E}[T_i^0 \mid T_i^0 < +\infty] = \frac{\tau_i^0}{r_i^0}. \tag{3.1.28}$$

Hence, in order to calculate the conditional expectation, $\mathbb{E}[T_{1_S}^0 \mid T_{1_S}^0 < +\infty]$, one must restrict the sample space of $T_{1_S}^0$ to finite values, and divide by the probability that the recovery time is finite. The set of finite recovery times can be partitioned into the set where $T_{1_S}^1 = +\infty$ and the set where $T_{1_S}^1 < +\infty$. In other words, the cell can either recover without ever containing vegetative bacteria, or the cell can recover after having contained at least one vegetative bacterium. Therefore, the restricted mean time to recovery can be written as follows

$$\tau_{1_S}^0 = \mathbb{E}[T_{1_S}^0 \cdot \delta_{T_{1_S}^1=+\infty}] + \mathbb{E}[T_{1_S}^0 \cdot \delta_{T_{1_S}^1<+\infty} \cdot \delta_{T_1^0<+\infty}].$$

Using the fact that $T_{1_S}^0 \cdot \delta_{T_{1_S}^1<+\infty} \cdot \delta_{T_1^0<+\infty} = (T_{1_S}^1 + T_1^0) \cdot \delta_{T_{1_S}^1<+\infty} \cdot \delta_{T_1^0<+\infty}$, and that $T_{1_S}^1$ and T_1^0 are independent, one finds that the restricted mean time to recovery for a cell infected with one spore with germination rate g, is

$$\begin{aligned}
\tau_{1_S}^0(g) &= \left(\frac{1}{g} + \frac{1}{\widetilde{\mu}+g}\right)\frac{\widetilde{\mu}}{\widetilde{\mu}+g} + \mathbb{E}[T_{1_S}^1 \cdot \delta_{T_{1_S}^1<+\infty}]\mathbb{P}(T_1^0 < +\infty) \\
&\quad + \mathbb{E}[T_1^0 \cdot \delta_{T_1^0<+\infty}]\mathbb{P}(T_{1_S}^1 < +\infty) \\
&= \frac{\widetilde{\mu}(\widetilde{\mu}+2g)}{g(\widetilde{\mu}+g)^2} + a\int_0^{+\infty} t f_{T_{1_S}^1}(t;g)\, dt + \frac{g}{g+\widetilde{\mu}}\int_0^{+\infty} t f_{T_1^0}(t)\, dt \\
&= \frac{g}{\widetilde{\mu}+g}\left[\frac{(\widetilde{\mu}+2g)(ga+\widetilde{\mu})}{g^2(\widetilde{\mu}+g)} + \frac{1}{\lambda}\log\left(\frac{b}{b-a}\right)\right],
\end{aligned} \tag{3.1.29}$$

where I have made use of Eqs. (3.1.21), (3.1.22), and (3.1.25). The values a and b are defined in Eq. (3.1.18).

With this restricted mean time at hand, and the probability of recovery in Eq. (3.1.19), one can use Eq. (3.1.28) to find the conditional mean time until recovery for the two different distributions of the germination rate. In particular, when the germination rate follows the continuous Gaussian distribution in

Eq. (3.1.1), the conditional mean time to recovery for an infected cell starting with one spore is given by

$$\mathbb{E}[T^0_{1_S} \mid T^0_{1_S} < +\infty] = \frac{\int_0^{+\infty} f_G(g)\tau^0_{1_S}(g)\ dg}{\int_0^{+\infty} f_G(g)r^0_{1_S}(g)\ dg}. \tag{3.1.30}$$

Alternatively, if one considers the discrete Bernoulli distribution for the germination rate, the conditional mean time to recovery is given by

$$\mathbb{E}[T^0_{1_S} \mid T^0_{1_S} < +\infty] = \frac{\varepsilon\tau^0_{1_S}(g_A) + (1-\varepsilon)\tau^0_{1_S}(g_B)}{\varepsilon r^0_{1_S}(g_A) + (1-\varepsilon)r^0_{1_S}(g_B)}. \tag{3.1.31}$$

Time to rupture

The time taken for the initial phagocytosed spore to transition into a vegetative bacterium is given by the random variable $T^1_{1_S}$ and the time from vegetative bacterium to rupture is denoted by T^R_1. Thus, the total time between the cell engulfing a spore, and the time of rupture, is $T^R_{1_S} = T^1_{1_S} + T^R_1$. The corresponding probability density function for $T^1_{1_S}$ was given by $f_{T^1_{1_S}}(t;g)$ in Eq. (3.1.21), and the probability density function for the rupture time starting from one vegetative bacterium was given by $f_{T^R_1}(t)$ in Eq. (3.1.17). One can convolve these two functions to find the probability density function for the total time to rupture, giving

$$f_{T^R_{1_S}}(t;g) = \int_0^t f_{T^1_{1_S}}(s;g) f_{T^R_1}(t-s)\ ds. \tag{3.1.32}$$

When the germination rate across spores follows the continuous Gaussian distribution, the density of rupture times is given by

$$f_{T^R_{1_S}}(t) = \int_0^{+\infty} f_G(g) f_{T^R_{1_S}}(t;g)\ dg. \tag{3.1.33}$$

Alternatively, in the discrete Bernoulli case, the density of rupture times is given by

$$f_{T^R_{1_S}}(t) = \varepsilon f_{T^R_{1_S}}(t;g_A) + (1-\varepsilon) f_{T^R_{1_S}}(t;g_B). \tag{3.1.34}$$

As done previously for recovery, one can also calculate the conditional mean time to rupture, denoted $\mathbb{E}[T^R_{1_S} \mid T^R_{1_S} < +\infty]$ and defined similarly to Eq. (3.1.28) with 0 replaced by R. Since the random variables $T^1_{1_S}$ and T^R_1 are independent, it can

be shown that the restricted mean time to rupture, for a cell initially infected with a spore with germination rate g, is

$$\begin{aligned}\tau_{1_S}^R(g) &= \mathbb{E}[T_{1_S}^1 \cdot \delta_{T_{1_S}^1<+\infty}]\mathbb{P}(T_1^R < +\infty) + \mathbb{E}[T_1^R \cdot \delta_{T_1^R<+\infty}]\mathbb{P}(T_{1_S}^1 < +\infty) \\ &= (1-a)\int_0^{+\infty} t f_{T_{1_S}^1}(t;g)\ dt + \frac{g}{g+\widetilde{\mu}}\int_0^{+\infty} t f_{T_1^R}(t)\ dt \\ &= \frac{g}{\widetilde{\mu}+g}\left[\frac{(\widetilde{\mu}+2g)(1-a)}{g(\widetilde{\mu}+g)} + \frac{1}{\lambda}\log\left(\frac{b-a}{b-1}\right)\right],\end{aligned} \tag{3.1.35}$$

where I have made use of Eqs. (3.1.21), (3.1.22) and (3.1.17). The values a and b are defined in Eq. (3.1.18).

Given this restricted mean time, and the probability of rupture from Eq. (3.1.19), one can use Eq. (3.1.28), with 0 replaced by R, to find the conditional mean time until rupture, for the two different distributions of the germination rate. In particular, when the germination rate across spores follows a continuous Gaussian distribution, the conditional mean time to rupture is of the same form as Eq. (3.1.30), with 0 replaced by R. Similarly, if one considers the discrete Bernoulli distribution for the germination rate, the conditional mean time to rupture is of the same form as Eq. (3.1.31), with 0 replaced by R.

3.1.4 Rupture size distribution

For an infected cell described by the CTMC $\mathcal{X}$ it is possible to find the probability distribution of its *rupture size*, which is the number of bacteria released into the extracellular environment from the infected cell. If the time for the process to enter state 0 is finite, then the rupture size is equal to 0, indicating that the host cell recovers and does not release any bacteria. On the other hand, if the time to reach state R, denoted by $T_{1_S}^R$, is finite, and $X(T_{1_S}^R - \Delta t) = n$ for small and positive Δt, this means that the process transitions into the rupture state from state $n \in \mathbb{N}$. This corresponds to the death and rupture of the host cell, and the release of n bacteria into the extracellular environment. Let R_i^n denote the probability that the cell will release n bacteria in total, given that the process

starts at state $i \in \mathcal{S}$. This is defined as

$$R_i^n = \begin{cases} \mathbb{P}(T_i^0 < +\infty), & \text{for } n = 0, \\ \mathbb{P}((T_i^R < +\infty) \text{ and } (X(T_i^R - \Delta t) = n)), & \text{for } n \in \mathbb{N}. \end{cases} \tag{3.1.36}$$

With this definition, R_i^0 is the probability that the cell recovers, so $r_i^0 = R_i^0$. For $n \in \mathbb{N}$, R_i^n is the probability that the cell ruptures and releases n bacteria, so the overall probability of rupture is $r_i^R = \sum_{n=1}^{+\infty} R_i^n$. For states $i \in \mathbb{N}$, the probabilities R_i^n do not depend on the germination rate, g. However for $i \in \{1_S, 1_{NGB}\}$, these probabilities do depend on the germination rate, and so will be denoted by $R_{1_S}^n(g)$ and $R_{1_{NGB}}^n(g)$.

One can use the method of Karlin & Tavaré (1982) to find the probabilities R_1^n, as follows. If the cell begins with a vegetative bacterium, so that $X(0) = 1$, then for a small time interval $\Delta t \to 0$, one has,

$$\begin{aligned} \mathbb{P}((X(t) = n) \text{ and } (t < T_1^R \leq t + \Delta t)) &= \mathbb{P}(X(t) = n)\mathbb{P}(t < T_1^R \leq t + \Delta t \mid X(t) = n) \\ &= p_{1,n}(t) n\gamma \Delta t, \end{aligned} \tag{3.1.37}$$

since if the cell contains n bacteria at time t, the probability of rupture between time t and $t + \Delta t$ is $n\gamma\Delta t$. An expression for $p_{1,n}(t)$, which is the probability that a cell contains n bacteria at time t, given that it contains one bacterium at time 0, was given by Carruthers *et al.* (2020),

$$p_{1,n}(t) = \frac{(b-a)^2 e^{-\lambda(b-a)t}(e^{-\lambda(b-a)t} - 1)^{n-1}}{(ae^{-\lambda(b-a)t} - b)^{n+1}}, \qquad n \geq 1. \tag{3.1.38}$$

Hence for an infected phagocyte starting with one vegetative bacterium, the probability that the cell releases $n \geq 1$ bacteria, R_1^n, is then

$$R_1^n = \int_0^{+\infty} p_{1,n}(t) n\gamma \, dt = \frac{(1-a)(b-1)}{b^n}, \tag{3.1.39}$$

with a and b defined in Eq. (3.1.18). Moreover, a first-step argument allows one to obtain the probability $R_{1_S}^n(g)$ from R_1^n. For $n \geq 1$ and germination rate g, one has

$$R_{1_S}^n(g) = R_{1_{NGB}}^n(g) = \frac{g}{\widetilde{\mu} + g} R_1^n. \tag{3.1.40}$$

When the germination rate across spores follows the continuous Gaussian distribution, the probability that the rupture size of a cell initially infected with one spore is equal to $n \in \mathbb{N}$ bacteria, is given by

$$R_{1_S}^n = \begin{cases} \displaystyle\int_0^{+\infty} f_G(g) \frac{\widetilde{\mu} + ga}{\widetilde{\mu} + g} \, dg, & n = 0, \\ \displaystyle R_1^n \int_0^{+\infty} f_G(g) \frac{g}{\widetilde{\mu} + g} \, dg, & n \geq 1. \end{cases} \tag{3.1.41}$$

Alternatively, if one considers the discrete Bernoulli distribution for the germination rate, the probability that the rupture size of a cell initially infected with one spore is equal to $n \in \mathbb{N}$ bacteria, is given by

$$R_{1_S}^n = \begin{cases} \dfrac{\varepsilon(\widetilde{\mu} + g_A a)}{\widetilde{\mu} + g_A} + \dfrac{(1-\varepsilon)(\widetilde{\mu} + g_B a)}{\widetilde{\mu} + g_B}, & n = 0, \\ \displaystyle R_1^n \left(\frac{\varepsilon g_A}{\widetilde{\mu} + g_A} + \frac{(1-\varepsilon) g_B}{\widetilde{\mu} + g_B} \right), & n \geq 1. \end{cases} \tag{3.1.42}$$

3.1.5 Number of intracellular vegetative bacteria

Given a particular germination rate g for the phagocytosed spore, the mean number of intracellular vegetative bacteria at time t is denoted here by $B_v(t; g)$, where

$$B_v(t; g) = \sum_{j=1}^{+\infty} j p_j(t) = \frac{1}{\gamma} \frac{dp_R(t)}{dt}, \tag{3.1.43}$$

with the second equality arising from Eq. (3.1.12). Since $p_R(t)$ represents the cumulative distribution function of the rupture time starting with one spore, this means that the average number of vegetative bacteria is proportional to the probability density function of the rupture time. That is, the mean number of vegetative bacteria at time t, given germination rate g, is

$$B_v(t; g) = \frac{f_{T_{1_S}^R}(t; g)}{\gamma}. \tag{3.1.44}$$

Once this is averaged over the possible values of the germination rate, g, for either germination rate distribution hypothesis, the mean number of vegetative bacteria inside a cell at time t, is given by

$$B_v(t) = \frac{f_{T_{1_S}^R}(t)}{\gamma}, \tag{3.1.45}$$

where $f_{T_{1_S}^R}(t)$ is defined in Eqs. (3.1.33) and (3.1.34) for the two germination rate distributions.

3.2 Parameter calibration

This section makes use of experimental data from *in vitro* studies to calibrate the intracellular model and to compare the two hypotheses of germination heterogeneity in their ability to describe the data. I have performed Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) (Toni *et al.* (2009)) to estimate the model parameters.

3.2.1 Experimental data

The main experimental data set used for the calibration comes from an *in vitro* study by Kang *et al.* (2005), that was discussed by Pantha *et al.* (2018) and used to calibrate their ODE model. In the experiment 10^6 murine peritoneal macrophages were incubated with different numbers of Sterne 34F2 strain spores for 30 minutes, during which time phagocytosis occurred (Kang *et al.* (2005)). The ratio of spores to cells in the solution at the start of the incubation period is called the multiplicity of infection (MOI) and in this case the four MOIs considered were spore to macrophage ratios of 1:1, 1:2, 1:10 and 1:20, corresponding to the initial number of spores in the solution of 10^6, 5×10^5, 10^5, and 5×10^4. At the end of 30 minutes, the solutions were washed, so no extracellular spores remained, and no more spores were phagocytosed after this time. Then the solutions were incubated with an antibacterial agent called gentamicin for 30 minutes to remove any extracellular bacteria. At various time points after this, samples of parallel replicates of the experiment were washed and the number of intracellular spores and bacteria determined.

While in reality cells could phagocytose more than one spore each in this experiment, this is less likely to happen when the average number of spores per cell in the solution is low. Therefore only the data for MOI 1:2, 1:10 and 1:20 have been used to perform the parameter calibration, since these low MOIs will be more consistent with the model assumption in Section 3.1 that each macrophage

engulfs at most one spore, leading to the initial condition for the model in Figure 3.1. Still, once the posterior samples of the parameters are obtained, the model predictions will be compared to the MOI 1:1 data, as a qualitative validation.

The experiment described above was also performed using a germination-deficient strain of *B. anthracis* spores, in which spore germination is inhibited. The average spore counts at one hour from two duplicate samples that used the germination-deficient strain are provided in Table 3.2. The number of spores of the germination-deficient strain would have remained unchanged between 0.5 hours and 1 hour, because they will not have germinated, and all extracellular spores were removed by washing at 0.5 hours, so there would have been no more phagocytosis after this time. Thus, if one assumes that there is no difference in spore phagocytosis rates between the germination-deficient and Sterne strains, one concludes that the spore counts for the germination-deficient strain are a good representation of the total number of Sterne strain spores that would have been phagocytosed during the first 0.5 hours of the experiment for each MOI. In the parameter calibration for their Phase II subsystem model, Pantha *et al.* (2018) used these numbers of intracellular spores from the germination-deficient experiment as the initial condition for the intracellular dynamics. The justification given is that germination does not seem to be a dominating process at 0.5 hours, so the number of spores of the Sterne strain at 0.5 hours will be similar to the number of germination-deficient spores at the same time point. Therefore, I have made the same assumption, that no germination of the Sterne strain spores has occurred before 0.5 hours. Because of this, the estimates for g_A and g_B in the discrete Bernoulli model should be interpreted with this 30 *min* delay in mind, and in particular the estimated germination rates might be slightly overestimated as a result. However, this delay could possibly be interpreted as a time-lag after phagocytosis for the activation of germination to occur. Furthermore, after learning about the parameters with ABC-SMC inference, it can be seen that even the spores with a quicker germination rate (type A) take on average longer than an hour to germinate, so it seems (*a posteriori*) reasonable to assume that germination does not happen in the first 30 minutes of the experiment.

In the data from the experiment, time $t = 0$ corresponds to the start of the incubation period of cells and spores (Kang *et al.* (2005)). On the other hand,

Intracellular germination-deficient spore count at one hour

MOI 1:2	139000
MOI 1:10	30500
MOI 1:20	13925

Table 3.2: Data taken from Table 2 of Pantha *et al.* (2018), giving the average number of intracellular spores counted at 1 hour from two replicates of the experiment using spores of a germination-deficient strain of *B. anthracis*. The number of intracellular spores of the germination-deficient strain should have remained unchanged between 0.5 hours and 1 hour, because they cannot germinate, and all extracellular spores were removed by washing at 0.5 hours, so there would have been no more phagocytosis after that time. Note that the value for MOI 1:10 reported in Table 2 of Pantha *et al.* (2018) was inconsistent with that observed in their Figure 2, so the second one is used here, since it is more consistent with the trajectory over time for the spore counts in Figure 2 of Pantha *et al.* (2018) for MOI 1:10, and also so that my predictions are comparable with those of Pantha *et al.* (2018).

	Time (hours)	0	0.5	2.5	4.5	23.5
MOI 1:2	Number of spores	139000	105000	11400	10250	29500
	Number of bacteria	0	23000	67100	52250	20000
MOI 1:10	Number of spores	30500	27000	12000	9100	2750
	Number of bacteria	0	9500	14000	7650	2500
MOI 1:20	Number of spores	13925	7900	6450	3100	300
	Number of bacteria	0	6000	2250	1750	1000

Table 3.3: Data for the number of intracellular spores and bacteria present at different time points, which have been used to perform ABC-SMC. These data have been taken from Tables 2 and 3 of Pantha *et al.* (2018), although time points are shifted by 30 *min* to account for the first phase of the experiment where phagocytosis occurs. The initial conditions ($t = 0$) correspond to the values reported in Table 3.2 from the germination-deficient experiment. The counts at each time point are averages of two experimental replicates.

the intracellular model described in Section 3.1 considers a single host cell that begins with one intracellular spore, and $t = 0$ is assumed to be the start of the germination process of this spore. Since phagocytosis only occurs during the first 0.5 hours of the experiment, and it is assumed that germination does not occur until after the first 0.5 hours of the experiment, phagocytosis has not been explicitly included in the mathematical model in Section 3.1. Instead, the time points are modified so that $t = 0.5$ in the experiment corresponds to $t = 0$ in the model. That is, the number of intracellular spores from the germination-deficient experiment, given in Table 3.2, is taken to be the initial conditions for $t = 0$, and the data is used as it is shown in Table 3.3.

Some of the model parameters are not identifiable from the measurements of intracellular spore and bacterial counts alone, since, for instance, it is difficult to determine whether a reduction in the number of intracellular bacteria is due to bacterial death or macrophage rupture. Therefore, I have also used a small amount of data regarding the rupture time of cells from a study by Akoachere *et al.* (2007), in order to gain preliminary knowledge about some of the model parameters. Akoachere *et al.* (2007) examined macrophage-spore interactions by fluorescence microscopy. To do this, they used two different dyes, SYTO 13 and PI, which both bind to DNA. SYTO 13 is a green dye that is cell permeable, whereas PI is a red dye that is impermeable to cell membranes. Therefore, when staining macrophages with these dyes, the nucleus of healthy cells shows green fluorescence and the nucleus of dead cells shows red fluorescence. After infecting murine macrophages with *B. anthracis* Sterne strain spores at a spore to macrophage ratio (MOI) of 20:1, they observed 20% PI-positive (ruptured) macrophages at 3.5 hours after exposure, and 90% at 7 hours after exposure, respectively. I fit the intracellular model to these data and used the results to inform the selection of some of the prior distributions to be used in the fitting with the data in Table 3.3.

3.2.2 Sensitivity analysis

Before performing parameter calibration, it is useful to investigate the effects that the model parameters have on the model output. To do this, I have used the

Sobol approach to global sensitivity analysis (Sobol (1993)). This allows one to quantify how variation in the parameter values affects the variation in the model output. I have focused on the model outputs of the expected number of spores and bacteria present inside a single cell over time, making use of the expressions in Sections 3.1.2 and 3.1.5. The sensitivity indices have been evaluated for times up to 24 hours, corresponding to the length of the time-course of the data in Table 3.3, which will be used in the parameter calibration. First-order and total-order sensitivity indices are presented. The first-order indices measure the contribution to the variance of the model output from varying each parameter alone. The total-order indices measure the total contribution to the output variance caused by each parameter, including both from the effect of varying the parameter alone, and the effect of all higher order interactions, which are found from varying the parameter along with groups of other parameters.

Model with continuous Gaussian germination rate distribution

The top row of plots in Figure 3.2 shows the first-order (S_1) and total-order (S_T) sensitivity indices of each model parameter, for the model with a truncated Gaussian germination rate distribution, when we consider the model output as the mean number of intracellular spores. It can be seen that variance in the parameter μ_g, corresponding to the most likely value for the germination rate of a given spore, plays a large part in contributing to the output variance. Variance in the standard deviation, σ_g, of the Gaussian distribution also has some effect on the output variance. The sensitivity indices of μ_g and σ_g are time-dependent. At early times the expected number of spores in the cell mostly depends on the mode of the germination rate distribution. Then at later times, the expected number of spores becomes gradually more dependent on the variance of the germination rate distribution. Apart from μ_g and σ_g, none of the other parameters have any influence on the output variance since they do not appear in the expression for the mean number of spores.

The sensitivity indices corresponding to the model output of the mean number of intracellular bacteria are shown in the bottom row of Figure 3.2. The parameters λ and γ affect the bacteria model output most significantly. These

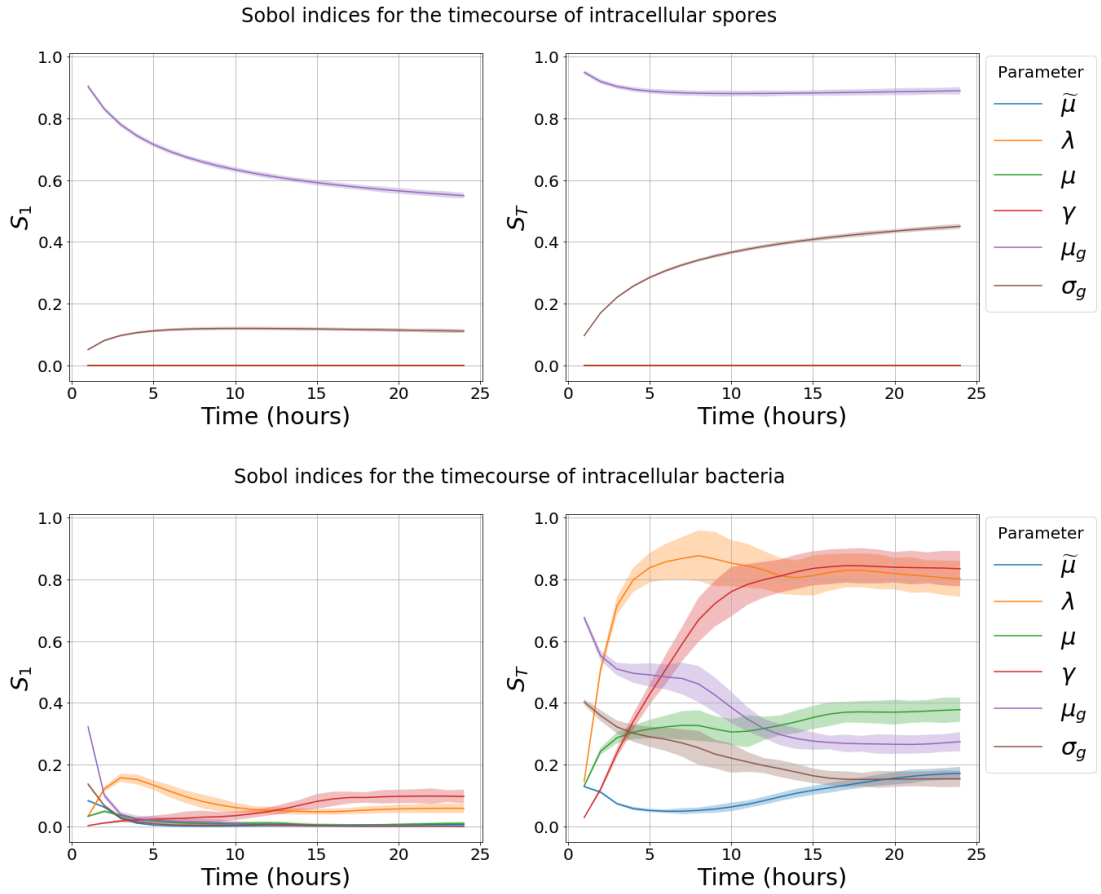

Figure 3.2: Plots of the first-order (S_1) and total-order (S_T) Sobol sensitivity indices for each parameter in the model with the truncated Gaussian germination rate distribution. The sensitivity indices indicate the importance of each parameter in describing the dynamics of intracellular spores and bacteria, on the top and bottom row of plots respectively, during the first 24 hours of infection. The solid lines represent the Sobol indices of each parameter over time, and the shaded regions indicate 95% confidence intervals. The ranges over which each parameter is varied are: $\mu_g \in [10^{-2}, 10]$, $\sigma_g \in [10^{-2}, 10^{0.15}]$, $\widetilde{\mu} \in [10^{-4}, 10]$, $\lambda \in [10^{-1.5}, 10]$, $\mu \in [10^{-4}, 10]$, and $\gamma \in [10^{-4}, 1]$.

Spores			Bacteria		
Parameter	Units	S_T mean	Parameter	Units	S_T mean
μ_g	h^{-1}	0.89	λ	$(bacteria \cdot h)^{-1}$	0.78
σ_g	h^{-1}	0.36	γ	$(bacteria \cdot h)^{-1}$	0.67
λ	$(bacteria \cdot h)^{-1}$	0.00	μ_g	h^{-1}	0.37
γ	$(bacteria \cdot h)^{-1}$	0.00	μ	$(bacteria \cdot h)^{-1}$	0.33
μ	$(bacteria \cdot h)^{-1}$	0.00	σ_g	h^{-1}	0.22
$\widetilde{\mu}$	h^{-1}	0.00	$\widetilde{\mu}$	h^{-1}	0.11

Table 3.4: Means of the total-order Sobol sensitivity indices (S_T) over the whole time-course for each model parameter and variable in the model with the truncated Gaussian germination rate distribution. Parameters are listed in order of most to least important in the columns for each model variable. Values are given to 2 decimal places.

parameters have an increasing effect at later times, because the probability that the spore has germinated into a bacterium increases over time and the population of bacteria begins to be more affected by replication and the rupture of cells. The parameters concerning the germination of spores, μ_g and σ_g, provide a large contribution to the variance of the model output at the initial stages, when the initial spore begins to germinate, but the effect of these parameters diminishes at time progresses. The parameter μ also plays a fairly significant role in the model output, but less so than λ and γ. On the other hand, the death rate of newly germinated bacteria, $\widetilde{\mu}$, has very small sensitivity indices over the entire time-course, indicating that it has very little effect on the model output. Since the uncertainty in the value of $\widetilde{\mu}$ does not have a large contribution to the uncertainty of the model output, the need to reduce the uncertainty in the value of this parameter through performing the parameter calibration is less important than for other parameters.

For both the spores and bacteria model outputs, the plots show that the total-order indices are generally larger than the first-order indices, indicating that higher order interactions are having an effect on the output variance. The means of S_T over the whole time-course for each parameter and model variable are reported in Table 3.4, to indicate which parameters are the most influential overall for each model variable.

Model with discrete Bernoulli germination rate distribution

The top row of plots in Figure 3.3 shows the first-order and total-order Sobol indices for the model with a Bernoulli germination rate distribution, when we consider the model output as the mean number of intracellular spores. It can be seen that the variance in the germination rates, g_A and g_B, contributes the most to the variance in the model output, since these parameters determine how quickly the spores germinate into bacteria. When calculating these sensitivity indices, g_A and g_B have been varied over the same ranges and it has not been specified that g_A is larger than g_B. Therefore the two germination rates can be interchanged without changing the model. Thus, due to this symmetry, the first-order sensitivity index for ε is zero. When interactions between parameters are considered, the sensitivity index for ε is non-zero, since this parameter affects the proportion of spores associated with each germination rate. Hence varying ε will change how quickly spores are expected to germinate, when the germination rates are also allowed to vary. Interestingly, the Sobol indices in the top row of Figure 3.3 are time-independent, in contrast with those in Figure 3.2 for the truncated Gaussian distribution. This is because the values of g_A, g_B, and ε determine how quickly the majority of spores germinate, as well as how spread out the germination rate distribution is. Therefore, even though the total variance of the spore output decreases with time, the portion of this variance that is explained by the uncertainty in each parameter stays the same, so that the sensitivity indices remain constant with time. As with the previous model, the sensitivity indices for all other parameters are exactly zero, since they do not affect the germination of spores, which is the only process that changes the size of the intracellular spore population.

For the model output of the expected number of intracellular bacteria, the sensitivity indices are shown in the bottom row of Figure 3.3. As in the previous model, the most influential parameters are the replication rate, λ, and the rupture rate, γ. In addition to λ and γ, the germination rates, g_A and g_B, are important for describing the mean number of intracellular bacteria, since these rates determine how quickly the bacteria population grows initially, as the spore germinates into a newly germinated bacterium and then a vegetative bacterium.

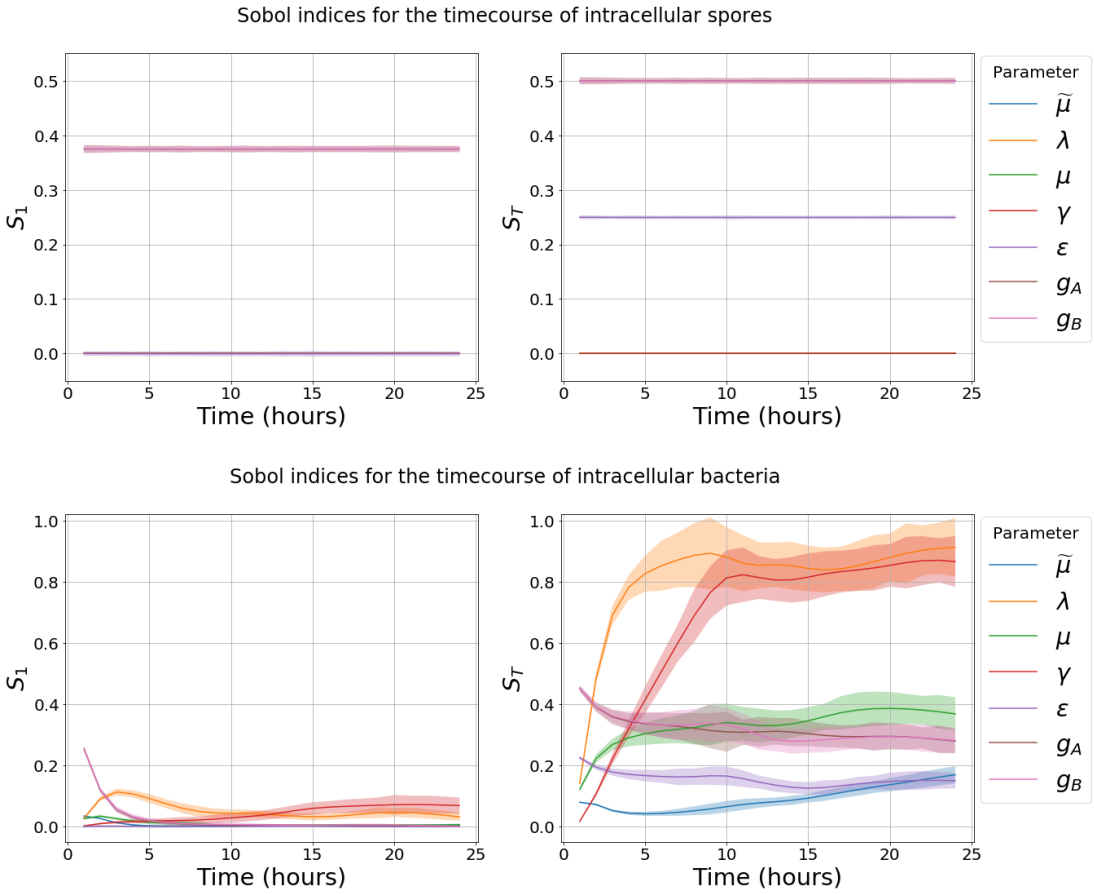

Figure 3.3: Plots of the first-order (S_1) and total-order (S_T) Sobol sensitivity indices for each parameter in the model with the Bernoulli germination rate distribution. These sensitivity indices indicate the importance of each parameter in describing the dynamics of intracellular spores and bacteria, on the top and bottom row of plots respectively, during the first 24 hours of infection. The solid lines represent the Sobol indices of each parameter over time, and the shaded regions indicate 95% confidence intervals. The ranges over which each parameter is varied are: $\varepsilon \in [0, 1]$, $g_A \in [10^{-4}, 10]$, $g_B \in [10^{-4}, 10]$, $\widetilde{\mu} \in [10^{-4}, 10]$, $\lambda \in [10^{-1.5}, 10]$, $\mu \in [10^{-4}, 10]$, and $\gamma \in [10^{-4}, 1]$.

Spores			Bacteria		
Parameter	Units	S_T mean	Parameter	Units	S_T mean
g_A	h^{-1}	0.50	λ	$(bacteria \cdot h)^{-1}$	0.81
g_B	h^{-1}	0.50	γ	$(bacteria \cdot h)^{-1}$	0.67
ε	-	0.25	μ	$(bacteria \cdot h)^{-1}$	0.33
λ	$(bacteria \cdot h)^{-1}$	0.00	g_A	h^{-1}	0.32
γ	$(bacteria \cdot h)^{-1}$	0.00	g_B	h^{-1}	0.32
μ	$(bacteria \cdot h)^{-1}$	0.00	ε	-	0.16
$\widetilde{\mu}$	h^{-1}	0.00	$\widetilde{\mu}$	h^{-1}	0.09

Table 3.5: Means of the total-order Sobol sensitivity indices (S_T) over the whole time-course for each model parameter and variable in the model with the Bernoulli germination rate distribution. Parameters are listed in order of most to least important in the columns for each model variable. Values are given to 2 decimal places.

Together with ε and $\widetilde{\mu}$, they also determine with what probability a spore will survive the germination process to become a fully vegetative bacteria. The sensitivity analysis shows that the uncertainty in the death rate of newly germinated bacteria, $\widetilde{\mu}$, does not have as much effect on the variance of the model output as the other parameters.

The means of S_T over the whole time-course for each parameter and model variable are reported in Table 3.5. Overall, the parameters that seem to be the most important are λ, γ, g_A and g_B. These are the parameters that one would hope to learn most about when performing parameter calibration.

3.2.3 Approximate Bayesian Computation Sequential Monte Carlo

Determination of prior distributions with data from Akoachere *et al.*

I have carried out parameter calibration for the intracellular model by means of Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) (Toni *et al.* (2009)), and by making use of the spore and bacterial counts measured in experiments by Kang *et al.* (2005). However, I first leveraged rupture time

data from an experiment by Akoachere *et al.* (2007) to estimate potential prior distributions for some of the model parameters.

In order to compare the model with the observations from Akoachere *et al.* (2007) that 20% of cells had ruptured by 3.5 hours and 90% had ruptured by 7 hours, it is necessary to introduce a delay for phagocytosis into the model, because time $t = 0$ in this experiment represents the time when the macrophages and spores were placed in contact, rather than the time when the spores were phagocytosed. The process with a delay for phagocytosis is shown in Figure 3.4.

In order to estimate this delay for phagocytosis, I have used maximum likelihood estimation (MLE) to find an estimate for the phagocytosis rate of spores, using the data reported in Table 3.2. As described in Section 3.2.1, the values in Table 3.2 are assumed to be the numbers of spores that were phagocytosed within the first 0.5 hours of the experiment by Kang *et al.* (2005), for the three multiplicities of infection (MOI). Since there were the same number of macrophages in the experiment for each MOI, one can consider the same per spore phagocytosis rate, and obtain a single estimate for this rate. If we assume that each spore is independently phagocytosed at a rate $\rho\ h^{-1}$, then the probability that a given spore has been phagocytosed before time t hours is $1 - e^{-\rho t}$. The initial number of extracellular spores was 5×10^5, 10^5, and 5×10^4 for MOI 1:2, 1:10, and 1:20, respectively. Therefore, the likelihood that the intracellular spore counts at 0.5 hours are equal to the values in Table 3.2, given phagocytosis rate ρ, is

$$\begin{aligned} L(\rho) = &\binom{5 \times 10^5}{139000} (1 - e^{-\frac{\rho}{2}})^{139000} (e^{-\frac{\rho}{2}})^{5\times 10^5 - 139000} \\ &\times \binom{10^5}{30500} (1 - e^{-\frac{\rho}{2}})^{30500} (e^{-\frac{\rho}{2}})^{10^5 - 30500} \\ &\times \binom{5 \times 10^4}{13925} (1 - e^{-\frac{\rho}{2}})^{13925} (e^{-\frac{\rho}{2}})^{5\times 10^4 - 13925}, \end{aligned}$$

and the log-likelihood is given by

$$\log L(\rho) = 183425 \log(1 - e^{-\frac{\rho}{2}}) - 466575\frac{\rho}{2},$$

which is maximised for a value of the per spore phagocytosis rate equal to $\rho = 0.66311\ h^{-1}$. Figure 3.5 shows the estimated mean number of intracellular spores

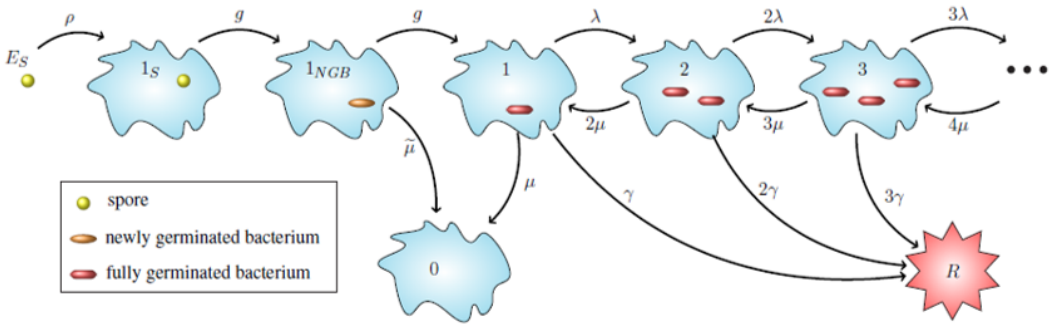

Figure 3.4: Intracellular infection model with a delay for phagocytosis. This is the same as the model in Figure 3.1, with the addition of state E_S, which is the initial state here, representing one extracellular spore. The per spore phagocytosis rate is ρ $hours^{-1}$.

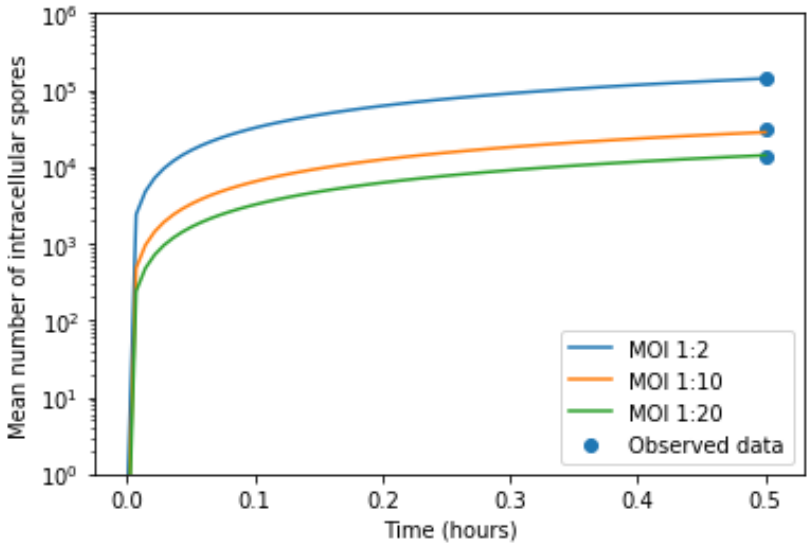

Figure 3.5: A comparison between model predictions of the mean number of intracellular spores for the first 0.5 hours of the experiments by Kang *et al.* (2005), and the observed mean number of intracellular spores provided in Table 3.2. These predictions have been obtained using the estimated per spore phagocytosis rate of $\rho = 0.66311$ h^{-1}.

over the first 0.5 hours of the experiment, with this value of the phagocytosis rate, ρ, for each of the three MOIs.

The process in Figure 3.4 starting with one extracellular spore has initial condition $X(0) = E_S$. For $i \in \{E_S, 1_S, 1_{NGB}\} \cup \mathbb{N}$, let $p_i(t)$ be the probability that the CTMC is in state i at time t, given that it started in state E_S. That is,

$$p_i(t) = \mathbb{P}(X(t) = i \mid X(0) = E_S).$$

One can write the system of Kolmogorov differential equations for these state probabilities, given by

$$\begin{aligned}
\frac{dp_{E_S}}{dt} &= -\rho p_{E_S}, \\
\frac{dp_{1_S}}{dt} &= \rho p_{E_S} - g p_{1_S}, \\
\frac{dp_{1_{NGB}}}{dt} &= g p_{1_S} - (g + \widetilde{\mu}) p_{1_{NGB}}, \\
\frac{dp_1}{dt} &= g p_{1_{NGB}} + 2\mu p_2 - (\lambda + \mu + \gamma) p_1, \\
\frac{dp_i}{dt} &= \lambda(i-1) p_{i-1} + \mu(i+1) p_{i+1} - (\lambda + \mu + \gamma) i p_i, \qquad \text{for } i \geq 2.
\end{aligned}$$

The first three equations can be easily solved to obtain, for $g \neq \rho$,

$$\begin{aligned}
p_{E_S}(t) &= e^{-\rho t}, \\
p_{1_S}(t) &= \frac{\rho}{g - \rho}(e^{-\rho t} - e^{-g t}), \\
p_{1_{NGB}}(t) &= \frac{g\rho}{(g-\rho)(g+\widetilde{\mu}-\rho)} e^{-\rho t} - \frac{g\rho}{(g-\rho)\widetilde{\mu}} e^{-g t} + \frac{g\rho}{\widetilde{\mu}(g+\widetilde{\mu}-\rho)} e^{-(g+\widetilde{\mu})t}.
\end{aligned}$$

The time to rupture of a cell following the process in Figure 3.4 is a random variable, $T^R_{E_S}$. The probability density function for this random variable, given that the spore has germination rate g, is obtained by convolving the function $f_{T^1_{E_S}}(t; g)$, which is the density function for the time to reach state 1, representing

a fully vegetative bacteria, with $f_{T_1^R}(t)$, which is the density for the time to reach rupture from state 1. Hence, one has

$$f_{T_{E_S}^R}(t;g) = \int_0^t f_{T_{E_S}^1}(s;g) f_{T_1^R}(t-s)\ ds,$$

where $f_{T_{E_S}^1}(t;g) = g p_{1_{NGB}}(t)$, and $f_{T_1^R}(t)$ was given in Eq. (3.1.17).

Taking into account the distribution of the germination rate, the overall density function of the rupture time is given by

$$f_{T_{E_S}^R}(t) = \int_0^{+\infty} f_G(g) f_{T_{E_S}^R}(t;g)\ dg,$$

in the case where the germination rate follows a truncated normal distribution, and,

$$f_{T_{E_S}^R}(t) = \varepsilon f_{T_{E_S}^R}(t;g_A) + (1-\varepsilon) f_{T_{E_S}^R}(t;g_B),$$

in the case of the discrete Bernoulli distribution. Then, for both hypotheses, the survival function of a cell starting with one extracellular spore can be written as

$$S^1(t) = \mathbb{P}(X(t) \neq R \mid X(0) = E_S) = 1 - \int_0^t f_{T_{E_S}^R}(t)\ dt.$$

However, since the MOI in the experiment by Akoachere *et al.* (2007) was a spore to macrophage ratio of 20:1, with many more spores than host cells, the assumption that each cell only phagocytoses one spore might be invalid. Therefore, we can consider that at the beginning of the experiment by Akoachere *et al.* (2007), each macrophage is surrounded by a "pool" of spores, and that each spore can only be phagocytosed by its nearest cell at the beginning of the experiment. If the suspension of spores and cells is well mixed, then the cell that is closest to each spore will be completely random. Suppose that the cells are numbered $i = 1, ..., m$, and the spores are randomly distributed between host cells. Then for each spore, a cell is sampled with replacement so that cell i is chosen with probability $1/m$. Hence, the distribution of spores for all host cells is multinomial, and the number of spores in a given pool (for a fiducial cell) is binomially distributed with parameters $n = 20m$ and $p = 1/m$, since an MOI of 20:1 indicates there are 20 times as many spores as cells. If we assume that there are large numbers of spores and cells in the experiment, then n will be large and p will be very small. In this limit, a Poisson distribution with mean $np = 20$ provides

a good approximation to the binomial distribution for the number of spores in a fiducial pool. Hence, we can assume the number of spores in a host cell pool to be Poisson distributed with mean $np = 20$. If we consider one such cell in the experiment, then each of its k surrounding spores will follow the process in Figure 3.4, and given all rates are linear, the stochastic process for each of the k spores will be independent, up until the time when the first stochastic process enters the rupture state.

Following the argument above, the probability that a cell starting with k surrounding spores has not ruptured by time t is the same as the probability that k processes each starting with one extracellular spore have all not reached the rupture state by time t. That is, $S^k(t) = (S^1(t))^k$, where $S^k(t)$ is the survival function of a cell starting with k extracellular spores in its pool. Thus, we can define the probability that a given cell in this experiment has not ruptured before time t by,

$$S(t) = \sum_{k=0}^{+\infty} \frac{20^k e^{-20}}{k!} [S^1(t)]^k.$$

In practise I have only included values of k up to $k = 40$ in this sum because $\mathbb{P}(k > 40) < 10^{-4}$.

I used ABC-SMC to calibrate the model parameters with the experimental measurements from Akoachere *et al.* (2007); namely, 20% of cells had ruptured before 3.5 hours and 90% had ruptured before 7 hours, or equivalently, 80% of cells were still alive at 3.5 hours, and 10% of the cells were alive at 7 hours. The per spore phagocytosis rate was fixed to be the maximum likelihood estimate $\rho = 0.66311\ h^{-1}$. Uniform prior distributions were considered for the remaining parameters, as reported in Table 3.6. These parameters are log-transformed because the prior range spans multiple orders of magnitude. For the model with two types of spores, I fixed $g_A > g_B$ to represent that, without any loss of generality, type A spores have a faster germination rate than those of type B. In order to sample parameter values g_A and g_B with prior distributions reported in Table 3.6, and under the constraint $g_A > g_B$, I followed the ideas from Goggans *et al.* (2014). I used the Euclidean distance function,

$$[S(3.5) - 0.8]^2 + [S(7) - 0.1]^2,$$

Parameter	Description	Prior distribution
Model with continuous germination rate distribution		
μ_g	Mean of the normal distribution for G	$\log_{10}\mu_g \sim$ U(-2, 1)
σ_g	Standard deviation of the normal distribution for G	$\log_{10}\sigma_g \sim$ U(-2, 0.15)
$\widetilde{\mu}$	Death rate of newly germinated bacteria	$\log_{10}\widetilde{\mu} \sim$ U(-4, 1)
λ	Replication rate of vegetative bacteria	$\log_{10}\lambda \sim$ U(-4, 1)
μ	Death rate of vegetative bacteria	$\log_{10}\mu \sim$ U(-4, 1)
γ	Rupture rate	$\log_{10}\gamma \sim$ U(-4, 1)
Model with discrete Bernoulli germination rate distribution		
ε	Probability that a given spore is of type A	$\varepsilon \sim$ U(0, 1)
g_A	Germination and maturation rate of spores of type A	$\log_{10}g_A \sim$ U(-4, 1)
g_B	Germination and maturation rate of spores of type B	$\log_{10}g_B \sim$ U(-4, 1)
$\widetilde{\mu}$	Death rate of newly germinated bacteria	$\log_{10}\widetilde{\mu} \sim$ U(-4, 1)
λ	Replication rate of vegetative bacteria	$\log_{10}\lambda \sim$ U(-4, 1)
μ	Death rate of vegetative bacteria	$\log_{10}\mu \sim$ U(-4, 1)
γ	Rupture rate	$\log_{10}\gamma \sim$ U(-4, 1)

Table 3.6: Prior distributions used in the ABC-SMC with data from Akoachere *et al.* (2007), for the model with continuous heterogeneity of germination rate (top) and the model with two types of spores (bottom). For the ABC-SMC using the data from Kang *et al.* (2005), all prior distributions remain as reported here, apart from those for λ and γ. For these two parameters, the marginal posterior distributions obtained from the Akoachere *et al.* (2007) calibration (shown in blue in Figure 3.6) are used as prior distributions in the calibration with the Kang *et al.* (2005) data. Parameters μ_g, σ_g, g_A, g_B, and $\widetilde{\mu}$ all have units h^{-1}, and λ, μ, and γ have units $(bacteria \cdot h)^{-1}$.

to compare the expected fraction of cells still alive, predicted by the analytical expression for the survival function, to the fraction of cells still alive in the experiment at the two time points.

Histograms of the posterior distributions from the final iteration of ABC-SMC are shown in Figure 3.6 for every model parameter and for both hypotheses. By using this small amount of data regarding the rupture time of cells, it was possible to learn significantly about the intracellular replication rate λ and the rupture rate γ, since in the model the rupture rate is proportional to the size of the bacterial population, which in turn depends on the replication rate of the bacteria. However, we are unable to learn much about the other model parameters in this way, since only two data points based on a single experiment are available. Therefore, only the posterior distributions obtained for λ and γ will be used as prior distributions in the subsequent application of ABC-SMC, with experimental data from Kang *et al.* (2005) for the time-course of the number of intracellular spores and bacteria. This second round of ABC-SMC will allow us to learn about the rest of the model parameters.

The distributions obtained for λ and γ seem to agree well for the two different germination hypotheses. The medians of the distributions for λ are around 0.9 $(bacteria \cdot h)^{-1}$, which is consistent with the doubling time of 0.78 hours, measured by Kalns *et al.* (2002) and used as an estimate in the within-host model by Day *et al.* (2011). Furthermore, these preliminary estimates lead to a good representation of the rupture dynamics, as shown in Figure 3.7, where the prediction of the function $1 - S(t)$, giving the expected fraction of cells to rupture before time t, is plotted together with the two data points from Akoachere *et al.* (2007), for each of the two germination heterogeneity hypotheses. The solid line shows the model prediction using the accepted parameter set that gave the smallest distance in the ABC-SMC, while the shaded region shows the pointwise 95% credible interval of the predictions from the complete posterior sample.

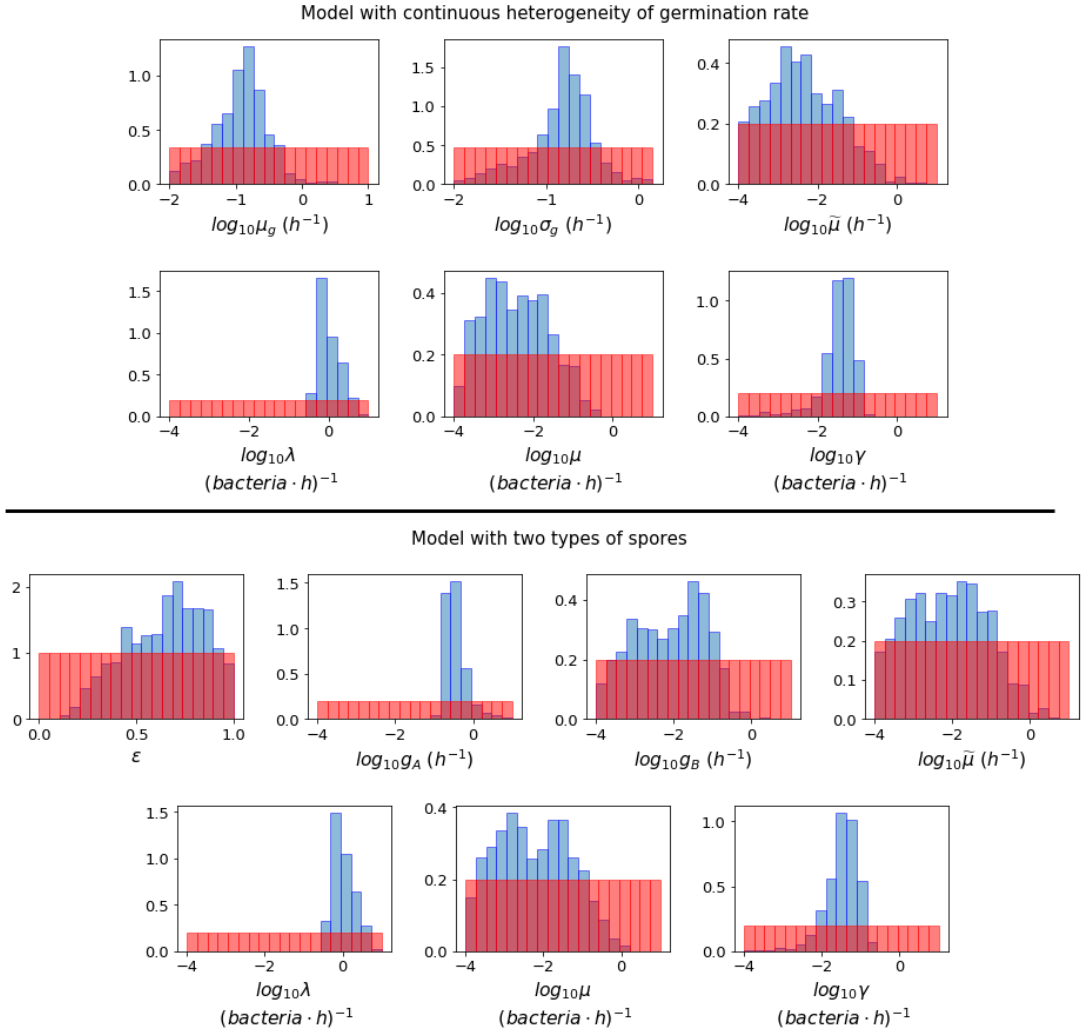

Figure 3.6: Prior distributions considered (red) and posterior histograms obtained (blue) when performing ABC-SMC with data from Akoachere *et al.* (2007) of the proportion of dead macrophages at two different time points, for the model with continuous heterogeneity in the germination rate (top), and the model with two types of spores (bottom).

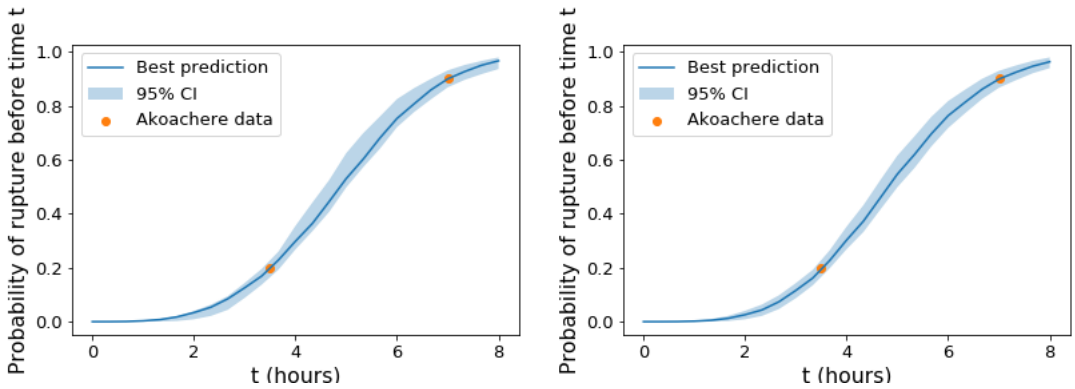

Figure 3.7: Best predictions (solid lines) and pointwise 95% credible intervals (shaded regions) of the fraction of cells that would be expected to rupture before time t in an experiment with MOI 20:1, compared to data from Akoachere *et al.* (2007), for the model with two types of spores (left), and the model with continuous distribution for the germination rate (right). The best prediction is the model output obtained by using the accepted parameter set with the smallest distance from the data. The pointwise 95% credible intervals show the uncertainty in the predictions given the posterior distributions for the model parameters.

ABC-SMC with experimental data of numbers of intracellular spores and bacteria over time

This section presents the results obtained from using the MOI 1:2, 1:10 and 1:20 data sets from Table 3.3 to obtain posterior parameter distributions, for the continuous model of germination rate, and for the discrete Bernoulli model. Pantha *et al.* (2018) used these data to estimate different parameter sets for each MOI, mentioning that for lower MOIs the smaller average intracellular burden could give a better environment for spores to germinate and bacteria to replicate, leading to larger values of the parameters. However, since I am assuming that every infected cell begins with only a single intracellular spore, this means that the cellular MOI is assumed to be identical across all data sets, and in this way the MOI is simply a measure of the system size. Hence the parameters considered in the stochastic model may not depend on the initial conditions given by the MOI. Thus, I have used the three data sets together to obtain a single set of estimates for the parameters, aiming to give a reasonable fit to the data sets with

a significantly smaller number of parameters in the model (compared to Pantha *et al.* (2018)).

In the experiment, the samples were washed each time before counting the number of intracellular spores and bacteria. Therefore, any bacteria released from a ruptured macrophage would be removed during the washing process and would not contribute to the number of bacteria observed in the data. Thus, to compare the model with the data, the model is used to calculate the per cell mean number of intracellular spores and bacteria over time. For the model with continuous distribution for the germination rate, Eqs. (3.1.9), (3.1.10), and (3.1.45) define the expected number of spores and bacteria in a cell at time t, given that it contained a spore at time $t = 0$. Since each infected cell is assumed to be independent, these can be multiplied by the number of initial spores, to obtain the mean number of spores and bacteria present inside the population of cells at time t. Let S_0 be the total number of initial intracellular spores for the population of cells at time $t = 0$. Then for the hypothesis of continuous heterogeneity in the germination rate, the total expected number of spores in all cells at time t is given by

$$S^*(t) = \frac{S_0}{Z} \exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t\right) \Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g}\right). \tag{3.2.1}$$

The total (in all cells) expected number of intracellular bacteria at time t hours, including newly germinated bacteria and vegetative bacteria, is given by

$$\begin{aligned} B^*(t) =& S_0 \frac{1}{\widetilde{\mu} Z}(1 - e^{-\widetilde{\mu} t}) \left(\frac{\sigma_g}{\sqrt{2\pi}} e^{-\frac{\mu_g^2}{2\sigma_g^2}} + (\mu_g - \sigma_g^2 t) \exp\left(\frac{\sigma_g^2 t^2}{2} - \mu_g t\right) \Phi\left(\frac{\mu_g - \sigma_g^2 t}{\sigma_g}\right)\right) \\ &+ S_0 B_v(t), \end{aligned} \tag{3.2.2}$$

where $B_v(t)$ is the expected number of vegetative bacteria in one cell at time t, defined in Eq. (3.1.45). Similarly, for the model with two discrete germination rates, the model predictions for number of spores and bacteria are given by

$$S^*(t) = S_0 \left[\varepsilon e^{-g_A t} + (1 - \varepsilon) e^{-g_B t}\right], \tag{3.2.3}$$

and

$$B^*(t) = S_0\left[\frac{\varepsilon g_A}{\widetilde{\mu}}\left(e^{-g_A t} - e^{-(g_A+\widetilde{\mu})t}\right) + \frac{(1-\varepsilon)g_B}{\widetilde{\mu}}\left(e^{-g_B t} - e^{-(g_B+\widetilde{\mu})t}\right) + B_v(t)\right], \tag{3.2.4}$$

where $B_v(t)$ is the expected number of vegetative bacteria in one cell at time t.

In the ABC-SMC, the numbers of intracellular spores and bacteria over time from the experiment by Kang *et al.* (2005) are compared to the model outputs given by $S^*(t)$ and $B^*(t)$, with the initial number of spores equal to $S_0 = 139000$ for MOI 1:2, $S_0 = 30500$ for MOI 1:10 and $S_0 = 13925$ for MOI 1:20. The model outputs are compared with the data via a distance function. I have used the Euclidean distance for the logarithm of the predicted values and observed data, given by

$$d(Model\ prediction,\ Data) = \sqrt{\sum_{i\in\{2,10,20\}}\sum_{t\in T}\left(log\left(\frac{S_i^*(t)}{s_i(t)}\right)\right)^2 + \left(log\left(\frac{B_i^*(t)}{b_i(t)}\right)\right)^2}, \tag{3.2.5}$$

where $T = \{0.5, 2.5, 4.5, 23.5\}$, $S_i^*(t)$ and $B_i^*(t)$ are the respective model predictions for the number of spores and bacteria at time t for MOI $1:i$, and $s_i(t)$, $b_i(t)$ are the respective observed number of spores and bacteria at time t, given by the data for MOI $1:i$.

In the model with continuous germination rate distribution, where the germination rate follows a truncated normal distribution, $G \sim N_{(0,+\infty)}(\mu_g, \sigma_g^2)$, the parameters characterising germination are the mean germination rate μ_g and its standard deviation σ_g. In the discrete Bernoulli model, the parameters characterising germination are the probability ε that a given spore is of type A, and the two germination rates, g_A and g_B. The rest of the parameters to be estimated are common to both versions of the model: the death rate of newly germinated bacteria, $\widetilde{\mu}$, the replication rate of vegetative bacteria, λ, the death rate of vegetative bacteria, μ, and the rupture rate, γ.

To perform ABC-SMC, one needs to choose prior distributions from which to sample the parameter values at the first iteration. The distributions obtained for λ and γ from the fitting to the Akoachere *et al.* (2007) data are used as the prior distributions for these two parameters. Uniform prior distributions are considered for the remaining parameters, as reported in Table 3.6.

Figure 3.8 shows the posterior histograms obtained by performing ABC-SMC for the two hypotheses considered, using the Kang *et al.* (2005) data, while summary statistics for these posteriors are reported in Table 3.7. By comparing the posterior histograms in blue with the prior distributions in red, one can see that it has been possible to learn significantly about most of the parameters for both hypotheses. For the model with continuous germination rate distribution, the value of μ_g, corresponding to the most likely value for the germination rate of a given spore, is likely to be between 10^{-2} and 10^{-1} h^{-1}. For the model with two germination rates, the value of ε is likely to be between 0.5 and 1, so that the majority of the spores will germinate with rate g_A, which is likely to be of the order of 10^{-1} h^{-1}, and the rest will germinate with rate g_B, which is likely to be of the order of 10^{-2} h^{-1}. For both hypotheses we learn that the death rate of newly germinated bacteria, $\tilde{\mu}$, is likely to be very small. For the model with continuous heterogeneity in the germination rate, the posterior histograms for λ and γ show that the value of these parameters that produce a good match between this model and the Kang *et al.* (2005) data, are similar to the values that gave a good fit to the data from Akoachere *et al.* (2007) that was used to inform the prior distributions for these parameters. For the model with two types of spores, the posterior histogram for λ is shifted slightly to the left from the prior distribution. For both hypotheses we have been able to learn significantly about the death rate of bacteria, μ, and these accepted values are usually larger than the corresponding values for the replication rate, λ. This is shown in the posterior histograms for the ratio between the birth and death rate of bacteria, λ/μ, which mostly contain values less than 1, indicating that the bacteria are likely to die more quickly than they replicate.

Model predictions were obtained for each accepted parameter set from the ABC-SMC. Figure 3.9 shows the pointwise 95% credible intervals of these time-courses, which indicate the uncertainty in the mean number of intracellular spores and bacteria from the model, due to the range of accepted parameter sets. The solid lines show the model output for the parameter sets with the smallest distance to the data, referred to as the best model predictions. For the model with continuous germination rate distribution, the predictions are close to the data at some time points, but overall this model struggles to capture the data, since the

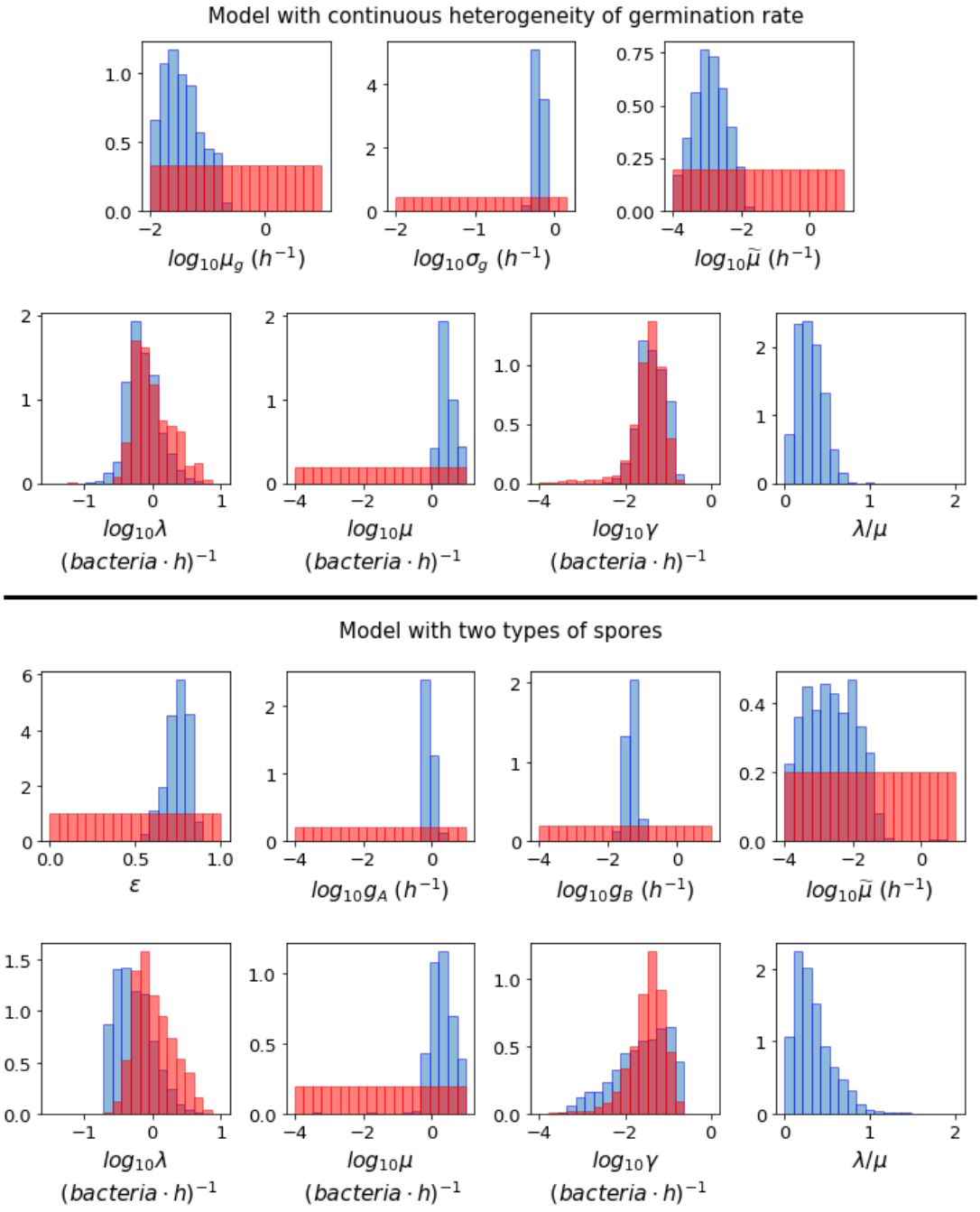

Figure 3.8: Prior distributions (red) and posterior histograms (blue) when performing ABC-SMC for the model with continuous heterogeneity of germination rate (top), and the model with two types of spores (bottom), using data from Kang *et al.* (2005) of the number of intracellular spores and bacteria at different time points for MOIs 1:2, 1:10, and 1:20.

Parameter	Min	Median	Mean	Max	95% Credible Interval
Model with continuous germination rate distribution					
μ_g	1.0×10^{-2}	3.3×10^{-2}	3.6×10^{-2}	2.4×10^{-1}	$(1.2 \times 10^{-2},\ 1.6 \times 10^{-1})$
σ_g	4.4×10^{-1}	6.3×10^{-1}	6.2×10^{-1}	8.4×10^{-1}	$(5.0 \times 10^{-1},\ 7.7 \times 10^{-1})$
$\widetilde{\mu}$	1.0×10^{-4}	1.2×10^{-3}	1.2×10^{-3}	1.7×10^{-2}	$(1.5 \times 10^{-4},\ 9.6 \times 10^{-3})$
λ	1.4×10^{-1}	6.9×10^{-1}	7.3×10^{-1}	5.0×10^{0}	$(2.8 \times 10^{-1},\ 2.4 \times 10^{0})$
μ	9.7×10^{-1}	2.6×10^{0}	2.8×10^{0}	9.7×10^{0}	$(1.3 \times 10^{0},\ 7.5 \times 10^{0})$
γ	3.6×10^{-3}	4.0×10^{-2}	4.2×10^{-2}	1.8×10^{-1}	$(1.1 \times 10^{-2},\ 1.3 \times 10^{-1})$
Model with discrete Bernoulli germination rate distribution					
ε	5.3×10^{-1}	7.5×10^{-1}	7.5×10^{-1}	9.1×10^{-1}	$(5.9 \times 10^{-1},\ 8.5 \times 10^{-1})$
g_A	4.9×10^{-1}	8.2×10^{-1}	8.5×10^{-1}	2.7×10^{0}	$(5.4 \times 10^{-1},\ 1.7 \times 10^{0})$
g_B	1.2×10^{-3}	4.9×10^{-2}	4.7×10^{-2}	1.0×10^{-1}	$(2.2 \times 10^{-2},\ 8.8 \times 10^{-2})$
$\widetilde{\mu}$	1.0×10^{-4}	2.2×10^{-3}	2.3×10^{-3}	5.0×10^{0}	$(1.3 \times 10^{-4},\ 4.3 \times 10^{-2})$
λ	2.0×10^{-1}	4.9×10^{-1}	5.3×10^{-1}	4.3×10^{0}	$(2.1 \times 10^{-1},\ 2.0 \times 10^{0})$
μ	5.0×10^{-4}	1.9×10^{0}	2.0×10^{0}	1×10^{1}	$(5.8 \times 10^{-1},\ 8.2 \times 10^{0})$
γ	4.1×10^{-4}	3.0×10^{-2}	2.3×10^{-2}	2.2×10^{-1}	$(9.6 \times 10^{-4},\ 1.9 \times 10^{-1})$

Table 3.7: Summary statistics for the posterior sample of each parameter, shown in blue in Figure 3.8, for the model with continuous heterogeneity of germination rate (top) and two types of spores (bottom). Parameters μ_g, σ_g, g_A, g_B, and $\widetilde{\mu}$ all have units h^{-1}, and λ, μ, and γ have units $(bacteria \cdot h)^{-1}$.

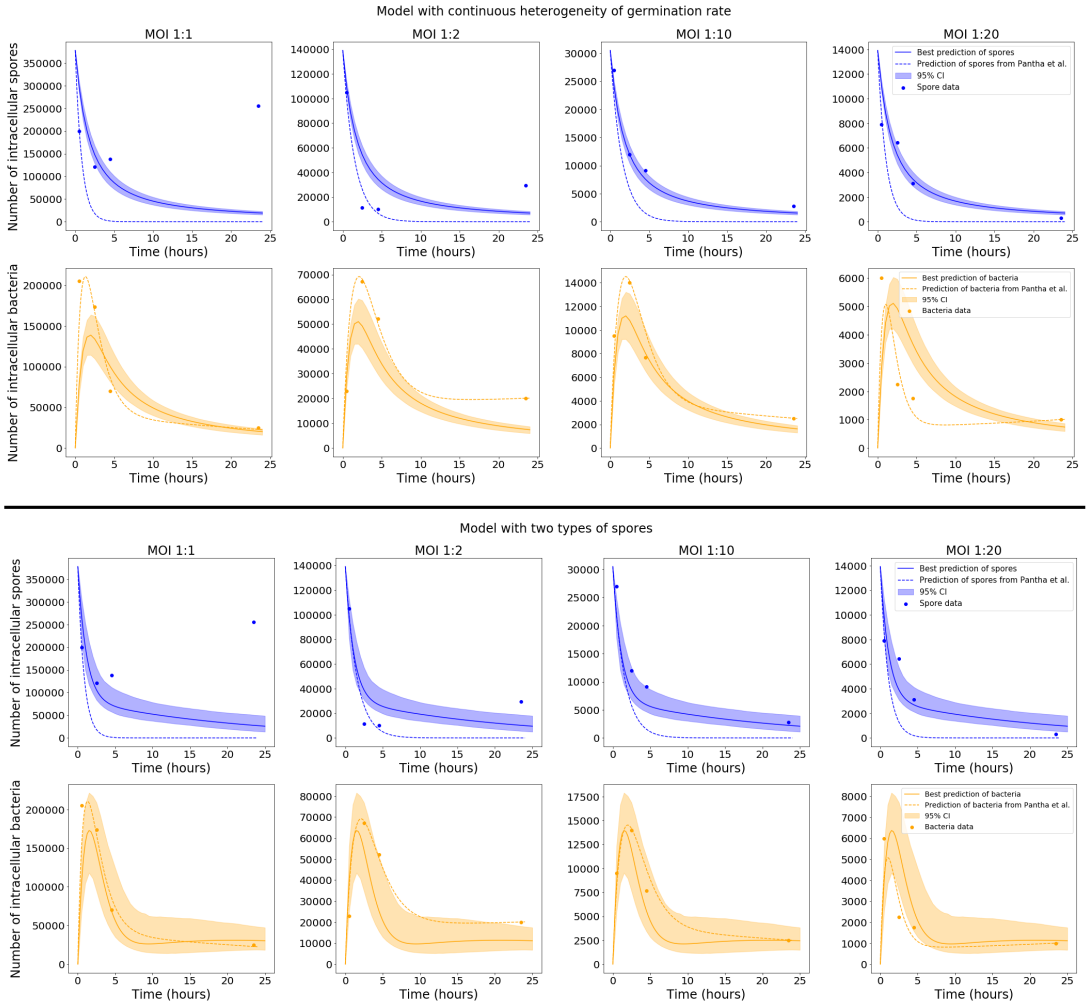

Figure 3.9: Time-courses for the mean number of intracellular spores and bacteria, for the model with continuous heterogeneity of germination rate (top), and the model with two types of spores (bottom), compared to data from Kang *et al.* (2005). The best predictions (solid lines) are the model outputs obtained by using the accepted parameter set with the smallest distance to the data, and the pointwise 95% credible intervals (shaded regions) show the uncertainty in those predictions, given the range of parameter values in the 10^3 accepted parameter sets from the final iteration of ABC-SMC. The predictions from the model by Pantha *et al.* (2018), using separate sets of parameter estimates for each MOI, are shown as dashed lines. For my model, only the data for MOI 1:2, 1:10, and 1:20 were used in the ABC-SMC to calibrate model parameters. The comparison of the model predictions with the MOI 1:1 data is shown here as a qualitative model validation.

peak of intracellular bacteria in the model predictions seems to be lower than the peak indicated by the data. On the other hand, the predictions of the model with two types of spores show a fairly good agreement with all data sets. It seems that this latter model, with two discrete germination rates, is better able to describe the pattern of biphasic decay in the number of spores seen in the data. This model also explains the bacterial data significantly better than the hypothesis of continuous germination rate distribution.

Only the data for MOI 1:2, 1:10, and 1:20 were used in the ABC-SMC analysis to calibrate model parameters. Still, the calibrated model predictions are also compared with the MOI 1:1 data, by making use of the corresponding initial condition for this MOI, $S_0 = 377500$. For the model with two types of spores, the best prediction is very close to the MOI 1:1 data at some time points, which demonstrates a possibility that the model could be used to extrapolate between MOIs. However, the assumption that each cell will only phagocytose a single spore is less viable at MOI 1:1, so the model may need to be adapted slightly in order to explain the dynamics at higher MOIs. Note that the long-term behaviour of spores for MOI 1:1, where a significant unexpected increase in the number of spores over time is observed, was not explained by Kang *et al.* (2005) or Pantha *et al.* (2018), and cannot be mechanistically explained by the intracellular infection model proposed here either, regardless of the choice of parameters.

Comparing models using Akaike's information criterion

Figure 3.9 also shows, as dashed lines, the predictions of the model by Pantha *et al.* (2018) using their parameter estimates. When these are compared to the predictions of my model, it seems that with the consideration of heterogeneity in the germination rate, my model provides a better explanation for the number of intracellular spores seen in the data, especially for MOI 1:10 and MOI 1:20. However, the fit of my model to the bacterial data does not look quite as good as the predictions by Pantha *et al.* (2018). This is not surprising, since Pantha *et al.* (2018) allow different parameter values for each data set (i.e. for each MOI), meaning that their model has more parameters and higher complexity, giving it more freedom to fit the data.

To compare the goodness-of-fit of the models, I have used Akaike's information criterion (AIC), which penalises models with a higher number of parameters if there is not enough improvement in the goodness-of-fit to warrant the additional complexity. Since there are only 32 data points in total to compare the model predictions to, I used a form of the AIC that is corrected for small sample sizes. The formula for this is given by Burnham & Anderson (2002), as follows

$$\text{AIC}_C = -2\log(\mathcal{L}(\hat{\underline{\theta}}, \hat{\sigma})) + \frac{2Kn}{n-K-1}.$$

In this formula, $\log(\mathcal{L}(\hat{\underline{\theta}}, \hat{\sigma}))$ is the maximised log-likelihood, n is the number of data points, and $K = r+1$, where r is the dimensionality of the model parameter space. Roughly speaking, lower values of the AIC_C indicate a better fit to the observed data.

I make the assumption that once one log-transforms the observed data and the predicted values from the model, the errors are normally distributed. Hence, let

$$\log(d(t)) = \log(m(\underline{\theta}, t)) + \varepsilon_t \qquad \text{for } t \in T,$$

where $d(t)$ is the experimental data point at time t, and $m(\underline{\theta}, t)$ is the predicted model output at the same time, for each of the n time points in $T = \{t_1, t_2, \ldots, t_n\}$. The errors, ε_t, are assumed to be independent, normally distributed with a constant variance σ^2. Therefore, these residuals have the following joint probability distribution

$$f(\underline{\varepsilon} \mid \underline{\theta}, \sigma) = \prod_{t \in T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{\varepsilon_t}{\sigma})^2},$$

and hence the likelihood is,

$$\mathcal{L}(\underline{\theta}, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2}\sum_{t \in T}(\frac{\varepsilon_t}{\sigma})^2}.$$

Then the log-likelihood is given by

$$\log(\mathcal{L}(\underline{\theta}, \sigma)) = -n\log(\sqrt{2\pi}) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{t \in T}\varepsilon_t^2.$$

The Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) algorithm was used for the parameter calibration to obtain parameter estimates,

$\hat{\underline{\theta}}$, which minimise $\sum \hat{\varepsilon}_t^2$, and hence maximise the log-likelihood with respect to the errors ε_t. To maximise the log-likelihood with respect to σ, one can differentiate with respect to σ and equate to 0, giving the MLE of σ^2 to be

$$\hat{\sigma}^2 = \frac{\sum_{t\in T} \varepsilon_t^2}{n}.$$

Thus, the maximised log-likelihood is

$$\log\Big(\mathcal{L}(\hat{\underline{\theta}}, \hat{\sigma})\Big) = -\frac{1}{2} n \log(\hat{\sigma}^2) - \frac{n}{2}\log(2\pi) - \frac{n}{2}.$$

The additive constants can be discarded because they will be equal for all models with the same data set. Thus, we obtain

$$\log\Big(\mathcal{L}(\hat{\underline{\theta}}, \hat{\sigma})\Big) = -\frac{1}{2} n \log(\hat{\sigma}^2).$$

Then the final formula that I used for the AIC is

$$\mathrm{AIC}_C = n\log\left(\frac{\mathrm{SSR}}{n}\right) + \frac{2Kn}{n-K-1}, \tag{3.2.6}$$

where n is the number of data points and K the number of parameters to be estimated, including the r model parameters and the variance of the residuals, σ^2. In this formula, SSR is the sum of the squared residuals, given by,

$$\mathrm{SSR} = \sum_{t\in T} \Big[\log(d(t)) - \log(m(\hat{\underline{\theta}}, t))\Big]^2.$$

I have calculated the value of AIC_C for the model by Pantha *et al.* (2018) and the two versions of my model with different distributions for the germination rate. For my models the value of AIC_C was calculated using the parameter set that gave the smallest distance in the ABC-SMC, obtaining a value of $\mathrm{AIC}_C = 1.42$ for the model in which the germination rate is a continuous random variable, and a value of $\mathrm{AIC}_C = -3.8$ for the one with two types of spores. For the model by Pantha *et al.* (2018), the value of AIC_C was calculated using the parameter estimates reported in Table 7 of their paper, giving a value of $\mathrm{AIC}_C = 176.73$. My models have a lower AIC_C than the model by Pantha *et al.* (2018), mainly because I used the same parameter estimates for each MOI, whereas Pantha *et al.* (2018) obtained separate estimates for each MOI, meaning that they have many more

free parameters. Even with the inclusion of heterogeneity in the germination rate, the version of my model with continuously distributed germination rates is not able to properly capture all the data. Instead, a bi-modal model of heterogeneity in the germination rate is needed to explain both the spore and the bacterial data. This is supported by the fact that the AIC_C value is lower for the discrete Bernoulli hypothesis, even though it has one more parameter. This indicates that this model may explain the data better than the model with continuously distributed germination rates.

3.3 Results

The results of the parameter calibration suggest that the hypothesis of germination rate heterogeneity with two discrete types of spores is better supported by the data than the model with continuous heterogeneity. The model with two types of spores is able to describe the biphasic decay seen in the spore data, especially for MOIs 1:10 and 1:20, as well as the behaviour of the bacterial data. Therefore, this section will only focus on the model with two types of spores. The set of parameter values that gave the smallest distance to the data in the ABC-SMC is provided in Table 3.8. In this section, these parameter values are used to calculate the various descriptors of the model discussed in Section 3.1. The effect of the uncertainty in the parameter values indicated by the posterior distributions obtained from ABC-SMC is also investigated.

Section 3.1.2 gave expressions for the probabilities, $p_{1_S}(t)$ and $p_{1_{NGB}}(t)$, that an infected macrophage will contain a spore or newly germinated bacterium, respectively, at time $t \geq 0$, given that the macrophage contained a spore with germination rate g at time 0. Note that since in the model the macrophage is assumed to only contain at most one spore or newly germinated bacterium at any one time, these probabilities are equal to the mean number of spores and newly germinated bacteria inside the macrophage at time $t \geq 0$. Section 3.1.5 explained how to calculate the mean number of vegetative bacteria, $B_v(t; g)$, in an infected macrophage at time $t \geq 0$, given that the macrophage contained a spore with germination rate g at time 0. One can then consider a population of S_0 independent infected cells, each containing a single spore at time 0. Assuming

Parameter	Description	Value
ε	Probability that a given spore is of type A	0.778846
g_A	Germination and maturation rate of spores of type A	0.894274
g_B	Germination and maturation rate of spores of type B	0.046794
$\widetilde{\mu}$	Death rate of newly germinated bacteria	0.003502
λ	Replication rate of vegetative bacteria	0.643111
μ	Death rate of vegetative bacteria	1.637989
γ	Rupture rate	0.043792

Table 3.8: Parameter values that gave the smallest distance between the two types of spores model and the data from Kang *et al.* (2005) in the ABC-SMC analysis. Parameters g_A, g_B, and $\widetilde{\mu}$ all have units h^{-1}, and λ, μ, and γ have units $(bacteria \cdot h)^{-1}$.

that each initial spore can have one of two possible germination rates (either rate g_A with probability ε, or rate g_B with rate $1-\varepsilon$), one can calculate a time-course for the total mean number of intracellular spores, newly germinated bacteria, and vegetative bacteria, for a population of infected cells, split into the populations arising from each of the two types of spores. This is depicted in Figure 3.10, where the first column corresponds to the populations arising from spores of type A, the second column corresponds to the populations arising from spores of type B, and the third column shows the sum of the populations from both types of spores. The solid lines indicate the means for the estimated parameter values in Table 3.8, while the shaded regions indicate the pointwise 95% credible intervals for these means, when the uncertainty in the parameter values from the posterior is considered. The first two plots show two very different timescales for the dynamics of each kind of spore, and when these populations are added together in the third plot, one can observe the biphasic behaviour in the number of spores that is observed in the data from Kang *et al.* (2005). The blue curve here indicates the prediction from the model for the total mean number of intracellular spores over time. The other two curves indicate the predictions for the mean number of intracellular newly germinated bacteria (orange), and vegetative bacteria (green), so that when these are added together, one obtains the prediction for the total number of intracellular bacteria, as shown in the predictions from the parameter calibration in Figure 3.9. The top row of plots in Figure 3.10 corresponds to an

initial condition of $S_0 = 30500$ spores, equal to the initial condition from the MOI 1:10 data used in the ABC-SMC. On the other hand, the bottom row corresponds to an initial condition of $S_0 = 100$. In both cases, the coloured dots indicate the sizes of the different populations over time, from a stochastic simulation of the model starting with S_0 spores. The results from the simulations show that when there are many initial spores (top row), the behaviour is very deterministic, but when the number of initial spores is relatively small (bottom row), there is much more stochasticity. This stochasticity could be relevant in *in vivo* settings, where infection might depend on a small group of spores germinating and producing a relatively small number of bacteria.

Section 3.1.3 explained how to calculate the probability density functions for the times to recovery and rupture of a macrophage initially infected with one spore. The probability density functions for the time to rupture are plotted along the top row of Figure 3.11, for the inferred parameter values in Table 3.8. From left to right, the first two plots show the density functions for the time to rupture of an infected cell containing a spore of type A and type B, respectively. Very different rupture timescales are observed for each kind of spore. The third plot shows these densities on the same plot, when they are scaled by the relative frequencies of each germination rate, so that the sum of these two densities gives the overall probability density function for the rupture time of a cell infected with a single spore. Note that this probability density function does not integrate to 1, but instead the probability of rupture. Nevertheless, one can divide the density by the probability of rupture, giving the conditional density function of rupture time, shown as a solid line in the fourth plot. Also shown on the fourth plot is a histogram of the finite rupture times from 10^6 stochastic simulations of the model in Figure 3.1. On the bottom row of Figure 3.11 are the analogous functions for the recovery time. Interestingly, the conditional probability density functions for the rupture and recovery times are almost identical. This is likely due to the fact that these timescales are heavily dominated by the germination time of the spore, and once the spore has germinated, rupture or recovery of the phagocyte happens relatively quickly. One can also compute the conditional mean times to rupture or recovery of a macrophage infected with a single spore, which are the means of the rightmost histograms in Figure 3.11. For the parameter values in Table 3.8,

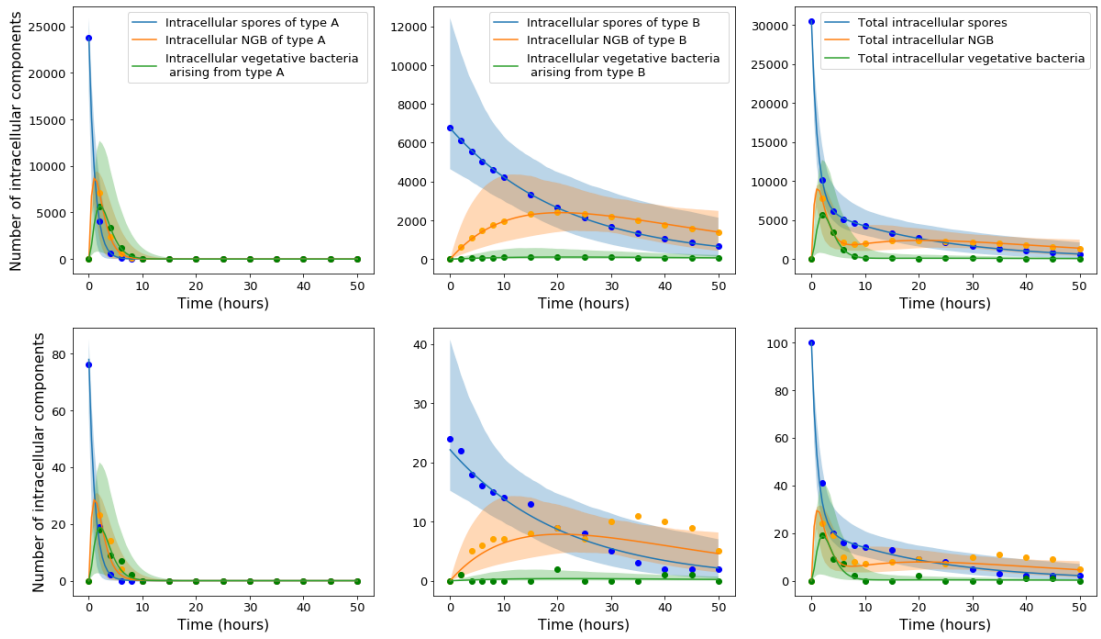

Figure 3.10: The top row of plots corresponds to a population of $S_0 = 30500$ infected cells, each containing a single spore at time 0, whereas the bottom row corresponds to an initial condition of $S_0 = 100$. **Left:** The mean number of type A spores, $\varepsilon S_0 p_{1_S}(t; g = g_A)$, type A newly germinated bacteria, $\varepsilon S_0 p_{1_{NGB}}(t; g = g_A)$, and vegetative bacteria, $\varepsilon S_0 B_v(t; g_A) = \varepsilon S_0 \sum_{i=1}^{\infty} i p_i(t; g = g_A)$, arising from the type A spores in the infected macrophages. **Centre:** The analogous functions for the populations arising from the initial spores with germination rate g_B. **Right:** The overall mean number of spores, $S_0 S(t)$, newly germinated bacteria, $S_0 B_{NGB}(t)$, and vegetative bacteria, $S_0 B_v(t)$, obtained by adding together the populations for each type of spore. The solid lines indicate the means for the estimated parameter values in Table 3.8, while the shaded regions indicate the pointwise 95% credible intervals for these means, when the uncertainty in the parameter values from the posterior distributions is taken into account. The equations used to compute these curves were Eqs. (3.1.7), (3.1.8), (3.1.11), (3.1.44), and (3.1.45). The dots show values for the size of the different populations over time from a single stochastic simulation beginning with S_0 spores.

the conditional mean rupture and recovery times are approximately 11.3 hours and 11.6 hours, respectively. However, the uncertainty in the parameter values, shown in the posterior distributions, leads to uncertainty in these timescales. For instance, this is indicated by the range of conditional mean times until rupture shown in the plot on the right of Figure 3.12.

The rupture time distribution is a key quantity that will be used to link with the within-host model in Chapter 4, in order to simulate multiple cells rupturing, without having to simulate the dynamics inside each individual cell. In order for this rupture process to be Markovian, which simplifies stochastic simulations and calculations, we can make use of the fact that the rupture time distribution can be well approximated by an Erlang distribution. To obtain parameters of an Erlang distribution that give a good approximation to the true rupture time distribution of the intracellular model, I minimized the euclidean distance between the first two moments of each distribution. From this, I found that the conditional rupture time distribution of cells infected with a spore of type A is well approximated by an Erlang distribution with shape parameter $n = 3$, and mean 3.17 hours. Similarly, the conditional rupture time distribution of cells infected with a spore of type B is well approximated by an Erlang distribution with shape parameter $n = 2$ and mean 41.67 hours. These approximations are shown on the relevant plots in Figure 3.11.

The probability distribution for the number of bacteria released by an infected macrophage, for the parameter values in Table 3.8, is provided in Figure 3.12 (left). The probability that no bacteria are released by the macrophage is predicted to be 0.96, which suggests that most macrophages will be able to recover and eliminate the intracellular infection, and that one would expect only 4% of infected macrophages to eventually rupture and release bacteria. The results also indicate that when macrophages do rupture, they will only release a few bacteria, with an average of 1.6 bacteria released from a macrophage that ruptures. This is consistent with the fact that a high dose of spores is required for infection, reported to be between 8×10^3 and 5×10^4 spores for humans (Oyston *et al.* (2004)). Cote *et al.* (2008) explain that although macrophages kill most of the germinated bacteria that they encounter, a low percentage of bacteria survive the antimicrobial environment in the macrophage and escape to begin the

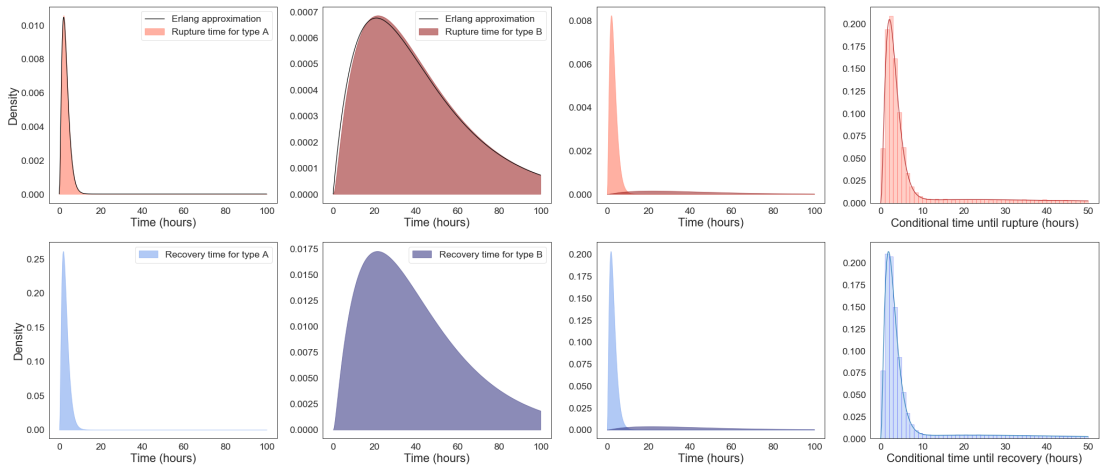

Figure 3.11: Top row: From left to right, the first two plots show the probability density functions for the rupture time of a macrophage infected with a spore of type A and type B, respectively, given by $f_{T^R_{1_S}}(t; g_A)$ and $f_{T^R_{1_S}}(t; g_B)$. Also shown are approximations of these rupture time distributions with Erlang distributions. The third plot shows these densities on the same plot, when they are scaled by the relative frequencies of each germination rate: $\varepsilon f_{T^R_{1_S}}(t; g_A)$ and $(1-\varepsilon) f_{T^R_{1_S}}(t; g_B)$. The fourth plot shows as a solid line the probability density function for the rupture time of a macrophage infected with a single spore, conditioned on rupture occurring, which is given by $f_{T^R_{1_S}}(t)/r^R_{1_S}$. Also shown on the fourth plot is a histogram of the finite rupture times from 10^6 stochastic simulations of the model in Figure 3.1. **Bottom row:** Plots correspond to the analogous densities for the time to recovery of an infected macrophage. The estimated parameter values in Table 3.8 were used to compute these functions.

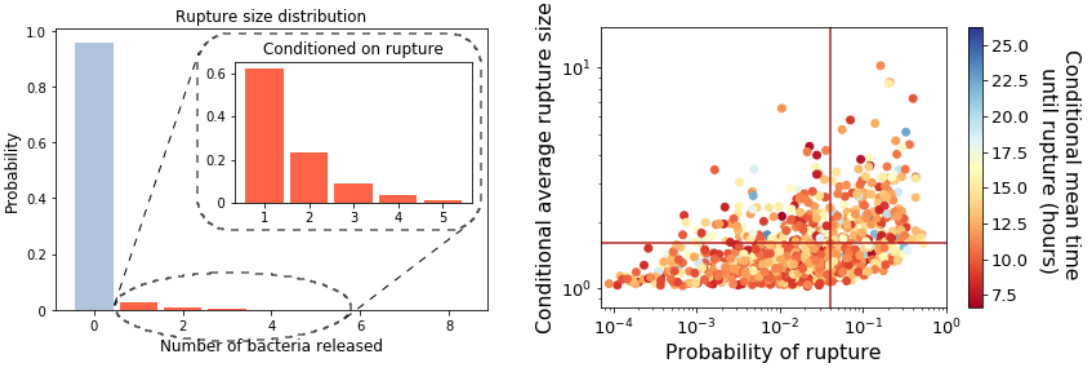

Figure 3.12: Left: The best predicted rupture size distribution for the model with two types of spores, computed using Eq. (3.1.42), with the estimated parameter values from Table 3.8. Inset is the conditional rupture size distribution, for the number of bacteria released by a macrophage infected with a single spore, given that it ruptures rather than recovers. **Right:** Scatter plot of the probability of rupture against the expected rupture size (conditioned on rupture occurring), for each parameter set in the posterior distribution. Lines indicate the values which correspond to the parameter set from Table 3.8. The colours of the points indicate the conditional mean time to rupture for each parameter set considered.

extracellular infection. This is further supported by findings from Jones *et al.* (1967), who observed that after infection of a guinea pig with 10^7 spores, 99% of the germinated spores were killed within an hour, but the 1% that survived managed to replicate extracellularly and ultimately reached a concentration of 10^8 bacteria/ml in the blood at the time of death.

Table 3.8 reports the best prediction for the parameter values according to the distance in the ABC-SMC. Yet, the advantage of a Bayesian approach in the parameter calibration is that it quantifies the uncertainty in the parameter values, which translates into the uncertainty in the descriptors of the model. The scatter plot on the right of Figure 3.12 shows the probability of rupture plotted against the average of the conditional rupture size distribution, for each parameter set in the posterior sample shown in Figure 3.8 (bottom), with the lines indicating the corresponding values for the parameter set in Table 3.8. There is a positive correlation between these two descriptors, indicating that if macrophages are more likely to rupture, they are also likely to release more bacteria when they

do rupture. Parameter sets leading to a very small probability of rupture will likely correspond to death rates much larger than the replication rate. Conversely, parameter sets leading to a larger probability of rupture correspond to bacterial death rates closer to the replication rate. This would allow for greater bacterial replication on average before rupture, and in turn a larger average rupture size. The colours of the points on this scatter plot indicate the conditional mean time until rupture for each parameter set. This illustrates the uncertainty of the timescale for rupture, with the possible mean rupture times from the posterior distributions ranging from around 7.5 to 25 hours. Furthermore, it is possible to find pairs of parameter sets that give differing average rupture times, but with similar probabilities of rupture and conditional average rupture sizes.

3.4 Discussion

I have proposed a stochastic model for the dynamics of *B. anthracis* spores and bacteria inside an infected phagocyte. One of the main features of the model is the consideration of heterogeneity in the germination rate of spores. Two hypotheses were considered to characterise this heterogeneity. The first hypothesis was that the germination rate is continuously distributed in a population of spores and follows a truncated normal distribution. The second hypothesis was that the spore population can be split into two kinds that germinate at different rates. I carried out parameter calibration, for each hypotheses, by means of Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) (Toni *et al.* (2009)). The results suggest that the discrete germination rates hypothesis is better supported by the data, since the model with this distribution of germination rates accounts for the biphasic decline seen in the spore counts, as well as the observed behaviour of the bacterial counts. This assumption of two types of spores also agrees with experimental evidence showing that in some *Bacillus* spore populations, a subset of the spores germinate much more slowly than the average spore, and are termed superdormant (Setlow (2013)). This leads to qualitatively different predictions for the mean number of spores over time in a population of *in vitro* cells, compared to previous theoretical predictions made by Pantha *et al.* (2018), as shown in Figure 3.9. Although the posterior estimated values

of the germination rate of spores of type A are similar to the germination rates predicted in by Pantha *et al.* (2018), the model predicts that a subset of spores will germinate much more slowly than this. The discrete distribution with two rates, g_A and g_B, considered here, is the most simple way to define a bi-modal distribution of the germination rate, but more complex descriptions of the germination rate heterogeneity could be considered. Experimental quantification of germination times would allow the exploration and validation of more complex models. A limitation of the model here is that the same rate is considered for each step of the germination-maturation process, compared to the model by Pantha *et al.* (2018), where a separate rate, m, was considered for the maturation step. However, Pantha *et al.* (2018) did not calibrate this rate m, and instead performed a sensitivity analysis. Since the available experimental data is limited, it would be difficult to calibrate different rates for each germination-maturation step. However, if further data were to become available which allowed one to distinguish between newly germinated and vegetative bacteria, then a separate rate could be incorporated for the maturation step of the germination process. This would also allow for more complicated distributions to account for heterogeneity in the germination and maturation rates.

Another important feature of the model is the consideration of rupture of infected phagocytes. This means that different behaviours can be described by the model compared to the model by Pantha *et al.* (2018), since in my model there is a chance that some intracellular bacteria will survive the microbicidal environment of the phagocyte and cause the cell to rupture. In the experiment by Kang *et al.* (2005), if bacteria were released into the extracellular medium then they would have been washed away before intracellular numbers of spores and bacteria were measured, and hence the decrease of intracellular bacteria seen in the data may not have been purely due to macrophage-induced killing of bacteria but may have been due to the release of intracellular bacteria from dying cells. Further data including information about macrophage rupture versus survival would be needed in order to disentangle these processes.

The stochastic model in Figure 3.1 allows one to compute the probability that an infected cell will eliminate the infection and recover, and the probability that an infected cell will rupture and release its bacterial content. I have also computed

the mean time for an infected cell to reach each of these two fates, conditioned on the event occurring. The probability distribution for the number of bacteria released by an infected macrophage has also been calculated. By calibrating the parameters using *in vitro* experimental data, the rupture size distribution shown in Figure 3.12 is able to capture the fact that the majority of spores taken up by macrophages are likely to be eliminated by the host cell, releasing no bacteria, but a few germinated spores may survive phagocytosis, leading to death of the host cell and release of a small number of bacteria. This is in agreement with previous experimental work (Banks *et al.* (2005)).

Although the model has been parametrised with data from a study that used macrophages, there is also evidence to suggest that dendritic cells play a role in the early infection stages of anthrax (Cleret *et al.* (2007)). Therefore, this is another important cell type to consider. If *in vitro* infection data for dendritic cells become available, it would be relevant to re-parametrise the model with such data to investigate the differences between the roles of the two host cell types in anthrax disease. For example, work by Shetron-Rama *et al.* (2010) indicates that dendritic cells may not be as capable as macrophages in their abilities to reduce bacterial numbers.

A restrictive assumption of the intracellular model introduced here is that it only considers a macrophage infection by a single phagocytosed spore. However at higher exposures, it is possible that some macrophages may phagocytose more than one spore. For example, in an experiment by Ruthel *et al.* (2004) an average of 4 spores were engulfed per macrophage, with a range of 1-11 spores per macrophage. Thus, when modelling scenarios with higher MOIs, it would be necessary to include this consideration in the intracellular model. Furthermore, a linear death rate of intracellular bacteria was proposed, in order to keep the model analytically tractable. There is evidence that macrophages with a low intracellular bacterial burden are much more efficient at killing bacteria than those with a higher burden (Cote *et al.* (2008, 2011); Kang *et al.* (2005); Ruthel *et al.* (2004)). The model could be generalised to include a non-linear, density-dependent death rate of intracellular bacteria, similar to the burden-dependent killing function used by Pantha *et al.* (2018). However, the model predicts that the intracellular burden of a cell initially infected with a single spore will remain

very low, so the inclusion of a density-dependent death rate may be more appropriate when considering higher multiplicities of infection, when each macrophage will phagocytose more than one spore.

To calibrate the parameters of the model, I made use of experimental measurements from a study by Kang *et al.* (2005), of mouse peritoneal macrophage infection with the attenuated non-capsule-producing Sterne strain of *B. anthracis*. This enabled a mathematical description of a system which can be characterised easily in the laboratory. In fact, one of the more commonly used animal models of anthrax is the AJ mouse model infected with Sterne strain (Twenhafel (2010)). Ideally, for modelling human inhalational anthrax, spores from a fully virulent strain and cells more similar to a human alveolar macrophage would be used. However, this type of data for anthrax is extremely limited. The more clinically relevant alveolar macrophage is complex to isolate and culture, so in the same way that mice are used as a surrogate for primates, peritoneal macrophages are used as a surrogate for the lung's resident phagocytes. Moreover, the Sterne strain is often used in laboratory settings since it poses a reduced infection risk to laboratory workers, and research with virulent strains of *B. anthracis*, such as the Ames strain, requires enhanced biosafety laboratories (Twenhafel (2010)). However, it can be generally agreed that its value has limitations when modelling disease (Twenhafel (2010)).

The capsule is known to protect extracellular bacteria from phagocytosis (Jang *et al.* (2011); Sharma *et al.* (2020)), and thus should be considered when modelling the extracellular dynamics of anthrax infection. It is possible that the capsule also plays a role in protecting emerging intracellular bacteria from the antimicrobial environment of the host cell, since germinating spores are able to quickly produce and coat themselves in the capsule (Ezzell Jr & Welkos (1999); WHO (2008)). However, it has been shown that macrophages are still able to kill intracellular bacteria even when they are from a strain that is coated in an antiphagocytic capsule, like the Ames strain (Welkos *et al.* (2002)), and the capsule does not seem to be fully protective against the bactericidal activity of macrophages (Gimenez *et al.* (2004)). Therefore, re-parametrising the intracellular model using data from a fully virulent strain would be extremely useful in determining whether

the capsule has a significant effect on the intracellular dynamics and fate of a phagocytosed spore.

In conclusion, I have developed and analysed a novel stochastic mathematical model of the intracellular bacterial dynamics of a macrophage infected with a single anthrax spore. By calibrating the model with experimental data, I have found support for a discrete Bernoulli distribution of the spore germination rate, which provides independent evidence for the role of superdormant spores (Setlow (2013, 2014)). This is both of clinical and biological interest. From a clinical perspective, it indicates the importance to maintain antibiotic dosing for long periods, given the potential for the slow germinating spores to contribute to the characteristic persistence of spores in the lungs after inhalational exposure (Jenkins & Xu (2013)). From a biological perspective, it demonstrates that there might be selective pressure for spores to distribute their germination rates in a heterogeneous manner. This might protect spore populations by ensuring that a reservoir of spores is maintained in case of accidental germination in environments not suitable for growth (Titball & Manchee (1987)). The results of the calibrated model also predict, in agreement with experimental findings, that many macrophages may be able to recover and resolve the bacterial infection, provided their initial intracellular burden is low. Yet, the results predict a low but non-zero risk of cellular rupture, leading to the release of bacteria from the cell.

Chapter 4

Within-host model of inhalational anthrax

In this chapter, the intracellular stochastic model proposed in Chapter 3 is extended towards an *in vivo* infection setting. A mechanistic within-host model of inhalational anthrax infection is presented, which describes the dynamics of *B. anthracis* within the lungs and lymph nodes of an individual, following inhalation of some initial dose of spores.

Several types of mathematical models have previously been developed to study inhalational anthrax. Statistical dose-response models can be used to describe the probability of response for different initial inhaled doses of spores, at the population level. However, these provide limited information about the within-host disease mechanisms, or the timescale of disease. A standard approach in dose-response assessment is the use of single-hit models. These models assume that when an individual is infected with a pathogen, the organisms act independently in the host so that the probability that any one organism in the initial dose produces an eventual infection is independent of the size of the dose, and the probability of infection is equivalent to the probability that at least one of the organisms in the initial dose will lead to an infection. An example is the exponential dose-response model, which involves a single parameter to denote the probability that an individual organism will produce a response. The competing-risks model for inhalational anthrax was proposed by Brookmeyer *et al.* (2005) as an extension of the exponential dose-response model. This is also a single-hit

model, but allows some insight into disease mechanisms because it includes parameters describing the germination rate of spores and the rate at which spores are cleared from the lungs. However, it is still a very simple model. The hypothesis of the competing risks model for anthrax is that if a single spore survives ingestion by a macrophage and successfully germinates without getting cleared, then the resulting bacterium will be certain to establish an infection. However, some immune system cells may be able to kill vegetative *B. anthracis* bacteria (Liu *et al.* (2018)), which means that once a spore in an infected cell has germinated and a bacterium is released, there is no guarantee that the bacterium will survive and cause infection. Another simplification used in the competing risks model is that germination and clearance of spores are both assumed to be exponential processes. However, this might not be the case. For instance, in the intracellular model presented in Chapter 3, as well as the one by Pantha *et al.* (2018), the consideration of the newly germinated bacterium means that the total germination-maturation time is non-exponential.

Deterministic models have also been developed for inhalational anthrax, consisting of differential equations to model variables of the within-host disease dynamics. These models can capture and quantify key biological mechanisms. Some deterministic models are very complex, for example the model by Day *et al.* (2011) includes variables for spores, bacteria, toxins, different types of immune system cells, and antibiotics. It models the dynamics in two compartments representing the lungs and the lymph nodes. The model parameters are not calibrated with specific data, but the model aims to provide a qualitative description of the disease in humans. On the other hand, Gutting (2014) has developed a simple model to describe the bacterial kinetics in a rabbit model of inhalational anthrax, and used various datasets to calibrate the model parameters. However, this model only describes the amount of bacteria within the lung airways and the rabbit body as a whole, rather than dividing the body into multiple compartments.

Here, a novel model of the within-host dynamics of inhalational anthrax is presented, which falls between the one by Day *et al.* (2011) and Gutting (2014) in terms of model complexity. Furthermore, the stochastic nature of the within-host model presented here allows it to capture the variability and randomness of the biological processes, and to explain dose-response data for anthrax infection.

For example, the model considers inter-phagocyte variability in rupture size by incorporating the rupture size distribution from the intracellular model into the within-host infection dynamics. Heterogeneity of the rupture size has been shown to be important in a similar model for the pathogen *F. tularensis* (Carruthers *et al.* (2018)). The stochastic modelling approach, together with dose-response data sets, is used to investigate the individual infection risk given an initial inhaled dose. As well as computing the probability of a host response to infection, the approach also allows quantification of the timescale of symptom onset.

The within-host model incorporates the role played by the anthrax toxins during infection, since the toxins are a key component of anthrax pathogenesis. The role of toxins was included in the model by Day *et al.* (2011), but the toxin level was modelled as a number between 0 and 1, with arbitrary units. Here I instead make use of *in vitro* data on toxin production in order to obtain biologically relevant parameter values that can be used in the within-host model. Parameter calibration of the within-host model is then performed by comparing the mean-field approximation of the stochastic model to *in vivo* data from rabbit and guinea pig infections, as well as linking the model to dose-response data in order to ensure that the model reliably describes the probability of response following exposure to different doses of spores.

This chapter is organised into two main sections. Section 4.1 outlines the model, and a simple discrete-time version of the model at early times is described, which is used to obtain a mechanistic expression for the dose-response parameter of the exponential dose-response model. Section 4.2 reports the results of the parameter calibration.

4.1 Mathematical model

This section outlines the within-host model of inhalational anthrax. It is a hybrid model comprising both discrete-valued and continuous variables. A multi-dimensional continuous-time Markov chain is used to describe the numbers of extracellular spores and bacterial colony-forming units (CFU) within various compartments of the body following exposure, whereas ODEs describe the amount of the protective antigen (PA) component of the anthrax toxin present in the

compartments. Since PA is thought to be the limiting toxin component and is the component that the therapeutic antibody targets, the model only includes a variable for the amount of PA, and neglects the other two toxin components, EF and LF. The compartments considered here are the airways of the lungs (A), the mediastinal lymph nodes (LN), and the blood/circulation (C). However, other compartments could easily be included, if relevant *in vivo* data sets were available.

The within-host model, depicted in Figure 4.1, contains the following variables:

- $S_A(t)$ is the number of extracellular spores in the airways of the lungs.
- $P(t)$ is the number of infected phagocytes migrating from the lungs to the mediastinal lymph nodes.
- $B_{LN}(t)$ is the number of extracellular vegetative bacterial CFU in the lymph nodes compartment.
- $B_C(t)$ is the number of extracellular vegetative bacterial CFU in the blood compartment.
- $T_{LN}(t)$ is the amount of PA (units ng) in the lymph nodes compartment.
- $T_C(t)$ is the amount of PA in the blood compartment.

After inhalation, spores are transported through the respiratory system and can become deposited in different areas, or may be exhaled without becoming deposited. Hence the initial dose of spores that is delivered to the lungs is generally smaller than the original inhaled dose (Weir & Haas (2011)). In the model, time $t = 0$ represents the time at which inhaled spores become deposited in the lungs. If an individual is exposed through inhalation to a dose of D spores, the initial number of spores deposited in the airways will be sampled from a binomial distribution, $S_A(0) \sim Bin(D, \hat{\phi})$, where $\hat{\phi}$ is the probability that a given inhaled spore is deposited in the lungs (Pratt *et al.* (2020)).

In the airways of the lungs, spores can be mechanically cleared due to ciliated epithelial cells that beat and propel material up the airways to be expelled. Alternatively, spores that reach the alveoli of the lungs have the chance to cause an

infection and can be engulfed by phagocytes and transported from the airways into the lung tissue. Similarly to the model by Gutting (2014), it is assumed here that spores in the lung airways are cleared with rate k_c, or phagocytosed and transported into the lung tissue with rate k_p. Thus there is a competition between the two mechanisms of removal of spores from the airways, and the overall rate of removal is denoted by,

$$\rho = k_c + k_p.$$

Furthermore, we are mainly interested in the spores that have the possibility to go on to cause an infection, namely the ones that are phagocytosed and transported into the lung tissue. Therefore, let us define a second parameter,

$$\phi = \hat{\phi}\frac{k_p}{k_c + k_p},$$

which is the probability for a given inhaled spore to be deposited in the lungs and ultimately phagocytosed rather than mechanically cleared. The phagocytosed spores are then transported intracellularly to the lymph nodes.

The within-host model proposed here is in the form of a compartmental model, and does not explicitly consider the dynamics inside each individual infected cell as in the intracellular model presented in Chapter 3. Instead, the intracellular dynamics are implicitly incorporated by using key quantities calculated from the intracellular model to link with the within-host model. These key quantities are the rupture size distribution and the rupture time distribution. With these quantities, the intracellular dynamics can be implicitly modelled under the assumption that each infected phagocyte ruptures after an amount of time sampled from the rupture time distribution, and subsequently releases some number of bacteria, which is sampled from the rupture size distribution. In order for the within-host model to remain Markovian, the rupture time distribution is approximated by an Erlang distribution. Specific Erlang approximations for the rupture time distributions were given in Section 3.3 for the best parameter estimates of the intracellular model. However, since the intracellular model was calibrated with data from an *in vitro* experiment that used mouse peritoneal macrophages, the timing of rupture from this *in vitro* system may not be realistic in a model of *in vivo* dynamics. Thus, in the within-host model, it is assumed that the time taken

for phagocytes to migrate to the lymph nodes and rupture follows an Erlang distribution with shape parameter $n = 3$ and unknown rate parameter δ. This is represented in the model by partitioning the infected phagocytes into three stages, giving $P(t) = \sum_{i=1}^{3} P_i(t)$, where $P_i(t)$ represents the number of infected phagocytes in stage i. Each phagocyte passes through the three stages, and releases some number $n \in \mathbb{N} \cup \{0\}$ of vegetative bacteria into the lymph nodes upon leaving the third stage, with probability $R_{1_S}^n$. The time spent in each stage is exponentially distributed with rate δ.

In the lymph nodes compartment, extracellular bacteria replicate with per bacterium rate λ_{LN}, and die with per bacterium rate μ_{LN}. Death of extracellular bacteria in the lymph nodes with rate μ_{LN} occurs due to the action of the host immune defences, and this single term accounts for multiple mechanisms of extracellular bacterial death, such as from neutrophils, macrophages, dendritic cells, lymphocytes, and soluble components of the innate immune system (Day *et al.* (2011)). This single term is used to represent all of the mechanisms since there is not the necessary data available at the moment to distinguish between the contribution to bacterial death from each of the different mechanisms.

If an infection becomes established in the lymph nodes and the number of bacteria continues to grow, this will eventually result in an inflammatory response. This can cause damage to the lymph nodes, and the tissues and blood vessels can become leaky in order to facilitate an influx of immune system cells. However, this can also allow bacteria to escape from the lymph nodes and find their way into the circulation. This is included in the model by assuming that once the number of bacteria in the lymph nodes reaches a certain threshold, M (representing a certain level of damage to the lymph nodes), bacteria can begin to migrate to the blood compartment with per bacterium rate m_B. In the blood compartment, extracellular bacteria proliferate with per bacterium rate λ_C. This does not necessarily represent replication within the blood itself, but it reflects the accumulation of bacteria in the blood due to replication and migration from numerous colonised organs throughout the body, which are not explicitly modelled here in terms of different compartments. It would be preferable to also include possible death of bacteria in the circulation compartment. However, it would not be possible to identify separate birth and death rates from the data that is used

Transition			Rate
S_A	$\longrightarrow$	$S_A - 1$	$k_c S_A$
$(S_A,\ P_1)$	$\longrightarrow$	$(S_A - 1,\ P_1 + 1)$	$k_p S_A$
$(P_i,\ P_{i+1})$	$\longrightarrow$	$(P_i - 1,\ P_{i+1} + 1)$	$\delta P_i,\ i = 1, 2$
$(P_3,\ B_{LN})$	$\longrightarrow$	$(P_3 - 1,\ B_{LN} + n)$	$R^n_{1_S} \delta P_3,\ n \in \mathbb{N} \cup \{0\}$
B_{LN}	$\longrightarrow$	$B_{LN} + 1$	$\lambda_{LN} B_{LN}$
B_{LN}	$\longrightarrow$	$B_{LN} - 1$	$\mu_{LN} B_{LN}$
$(B_{LN},\ B_C)$	$\longrightarrow$	$(B_{LN} - 1,\ B_C + 1)$	$m_B B_{LN} \mathbf{1}_{B_{LN} > M}$
B_C	$\longrightarrow$	$B_C + 1$	$\lambda_C B_C$

Table 4.1: The transitions and their corresponding rates in the Markov chain part of the within-host model in Figure 4.1. $\mathbf{1}_A$ is an indicator function, which is equal to 1 if A is true, and 0 otherwise.

to calibrate the model parameters in Section 4.2.2. In the deterministic version of the model, given in Section 4.2.2, λ_C can be interpreted as a single parameter representing the net growth rate (i.e., division rate minus death rate).

The model is depicted in Figure 4.1, and the possible transitions considered in the CTMC part of the within-host model are shown in Table 4.1. Additionally, the amount of PA in the lymph nodes and blood compartments are real numbers given by $T_{LN}(t)$ and $T_C(t)$ respectively. These follow the equations,

$$\frac{dT_{LN}}{dt} = \beta B_{LN} - (\mu_T + \nu B_{LN}) T_{LN},$$

$$\frac{dT_C}{dt} = \beta B_C - (\mu_T + \nu B_C) T_C.$$

PA increases from production at a rate proportional to the number of extracellular bacterial CFU, where β is the production rate of PA per bacterial CFU. Three factors will contribute to the removal of PA *in vivo*. It will naturally break down at a rate perhaps similar to the degradation that is observed *in vitro* in the absence of bacteria. Secondly, the PA will interact with host cell membranes

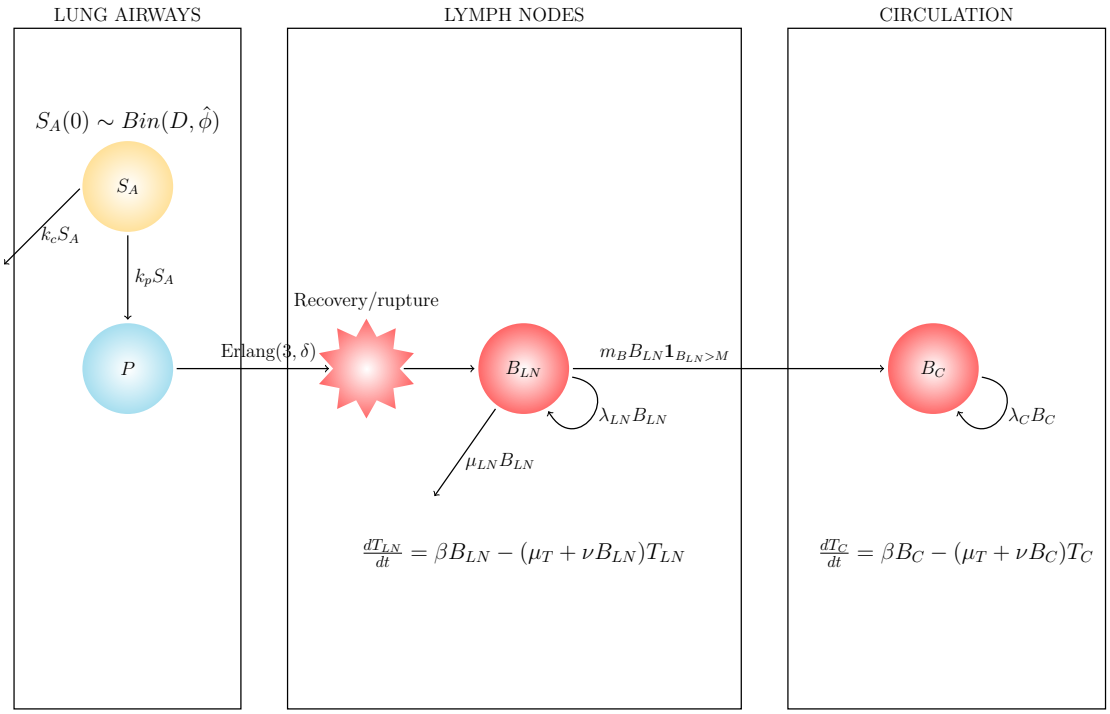

Figure 4.1: Diagram of the within-host model for inhalational anthrax. In the airways of the lungs, spores are cleared (with rate k_c) by ciliated epithelial cells that beat and propel material up the airways to be expelled. Some spores are instead engulfed by phagocytes and are transported into the lung tissue (with rate k_p). It is assumed that each phagocyte only phagocytoses one spore. The infected phagocyte then migrates to the lymph nodes and after an Erlang-distributed time will either recover, releasing zero bacteria, or rupture, releasing some positive number of bacteria. The number of bacteria released is determined by the rupture size distribution of the intracellular model, given in Eq. (3.1.42). In the compartments containing extracellular bacteria, the bacteria proliferate with a linear replication rate, and a linear death rate accounts for multiple mechanisms of extracellular bacterial death in the lymph nodes. Migration of extracellular bacteria from the lymph nodes to the circulation compartment occurs with rate $m_B B_{LN}$ if the number of bacteria in the lymph nodes is greater than the migration threshold, M.

and be cleaved by the host furin, to create a binding site for the active toxin components, LF and EF. To model these removal mechanisms, the PA decreases at a rate proportional to the current amount of PA, where μ_T is the rate of decrease from natural decay and binding. Thirdly, some of the PA will actively be digested due to proteases produced by the bacteria, at a rate proportional to the bacterial CFU (See Section 4.2.1). When the within-host model is simulated using the Gillespie algorithm, the values of T_{LN} and T_C are updated via the Euler method. For example, given a small time-step, Δt, and the value of $T_{LN}(t)$, this variable is updated to

$$T_{LN}(t + \Delta t) = T_{LN}(t) + \Delta t \left[\beta B_{LN}(t) - (\mu_T + \nu B_{LN}(t))T_{LN}(t)\right].$$

4.1.1 Dose-response relationship

If an individual inhales a dose of *B. anthracis* spores, it is possible that their immune system will be capable of clearing the pathogen, without resulting in a detectable infection. However, in some cases the infection cannot be contained, and an infection will become established. The within-host model can be used to investigate the probability that infection becomes established, given a known initial dose of inhaled spores.

In order to find the probability of infection, one can consider a simplification of the model in Figure 4.1 in which the timings of the events are ignored. We only need to focus on the chain of events that occur, leading to a discrete-time process that represents the probabilities of different outcomes at each initial stage of the infection. Furthermore, due to the presence of the migration threshold, M, it can be assumed that migration of bacteria to the blood will not occur until the infection is established in the lymph nodes, and therefore migration is neglected in the very early stages; that is, the probability of a response for a given dose depends on the early-time dynamics in the airways and lymph nodes, but not in the blood. This approximation leads to the simple discrete-time model shown in Figure 4.2, which can be used to predict the infection risk from exposure to different doses of spores.

Parameter	Units	Description
ϕ	$-$	Probability that an inhaled spore will become deposited in the lungs and will then become phagocytosed rather than mechanically cleared
ρ	$(spore \cdot h)^{-1}$	Rate of clearance of spores from the airways (by either phagocytosis or mucociliated clearance)
δ	$(phagocyte \cdot h)^{-1}$	Rate parameter of Erlang-distributed time for migration to the lymph nodes and rupture of host cells
λ_{LN}	$(CFU \cdot h)^{-1}$	Growth rate of extracellular bacteria in the lymph nodes
λ_C	$(CFU \cdot h)^{-1}$	Growth rate of extracellular bacteria in circulation
μ_{LN}	$(CFU \cdot h)^{-1}$	Killing rate of extracellular bacteria in lymph nodes
m_B	$(CFU \cdot h)^{-1}$	Migration rate of extracellular bacteria from the lymph nodes into the circulation
M	CFU	Threshold number of extracellular bacteria needed in the lymph nodes before migration into the circulation can occur
β	$ng\ (CFU \cdot h)^{-1}$	Production rate of PA by extracellular bacteria
μ_T	h^{-1}	Removal rate of PA due to natural degradation, and binding to cells
ν	$(CFU \cdot h)^{-1}$	Removal rate of PA due to degradation by proteases produced by the bacteria

Table 4.2: List of parameters for the within-host infection model. Units given here correspond to the stochastic model, but some of these units will change when incorporating these parameters into the ODEs provided in Section 4.2.2.

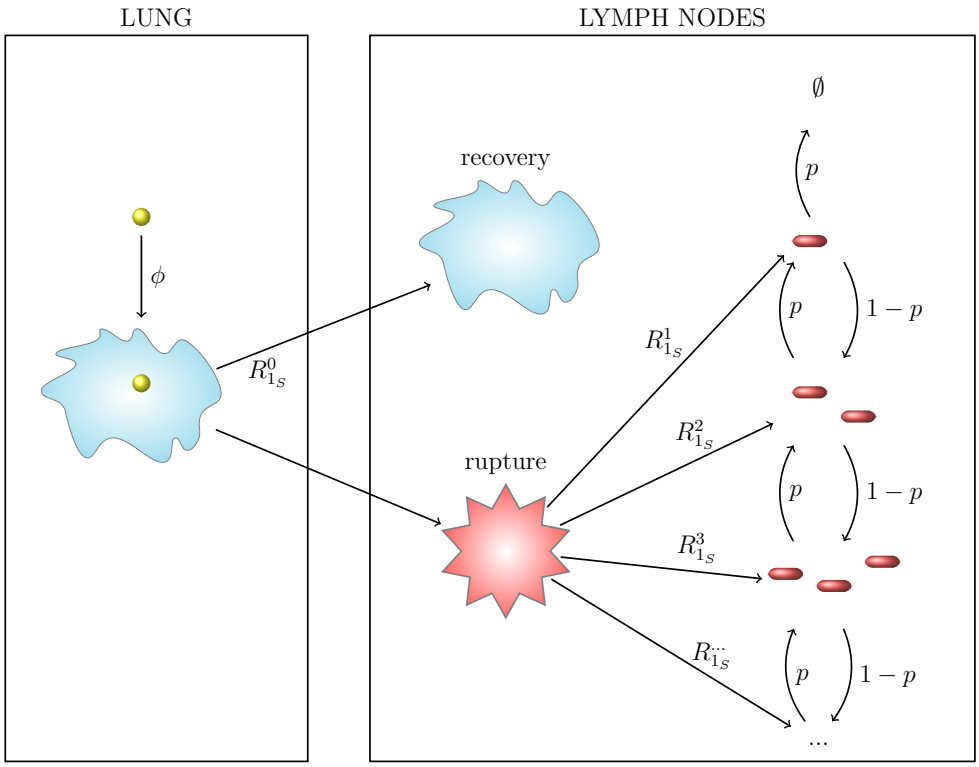

Figure 4.2: Discrete-time model for the possible fates of a single spore during the very early stages of inhalational anthrax infection. Within the lung, an inhaled spore (represented by the yellow ball) becomes deposited in the alveoli and ingested by a host phagocyte with probability ϕ. The intracellular spore germinates and the phagocyte might kill the germinated bacterium or the bacterium may survive the antimicrobial environment, replicate and cause the phagocyte to rupture, as described by the intracellular model in Chapter 3. The infected phagocyte migrates to the lymph nodes and either recovers, or ruptures and releases some bacteria, according to the probabilities calculated from the intracellular model, R^n_{1S}. The population of bacteria released follows a birth and death process, where an extracellular bacterium may be killed by host immune cells with probability p or a bacterium will replicate extracellularly with probability $1-p$. In this birth and death process, $\emptyset$ represents the state where there are no extracellular bacteria remaining.

Figure 4.2 shows a representation of the possible fates of each inhaled spore in the early stages of inhalational anthrax infection in the lungs and lymph nodes. It is assumed that only the inhaled spores that are deposited in the lungs and phagocytosed have the ability to cause an infection. As mentioned previously, the parameter ϕ gives the probability for an inhaled spore to be deposited in the alveoli of the lungs and ultimately phagocytosed. The infected host phagocyte then migrates to the mediastinal lymph nodes and releases $n \in \mathbb{N} \cup \{0\}$ bacteria with probability $R_{1_S}^n$ calculated from the intracellular model. Once released into the lymph nodes, each extracellular bacterium can be killed by host immune cells with probability $p = \frac{\mu_{LN}}{\mu_{LN}+\lambda_{LN}}$, or proliferate extracellularly with probability $1-p$. It is assumed that $p < 0.5$, otherwise the bacterial population would be certain to go extinct.

Each inhaled spore has two possible fates: established infection or clearance. Starting with a single inhaled spore, an infection will become established if all three of the following occur. First, the spore must be deposited in the lungs and be phagocytosed. Then this infected host cell must release some positive number, $n \geq 1$, of bacteria into the lymph nodes. Finally, the population of bacteria in the lymph nodes, starting at size n, must not become extinct. The size of the bacterial population in the lymph nodes follows a simple one-dimensional random walk, with zero the only absorbing state, and probabilities p and $1-p$ of moving from state j to $j-1$ and $j+1$, respectively. If the population starts at state n after a rupture event, it is well known that the probability of ultimate extinction of this process is $\left(\frac{p}{1-p}\right)^n$. Therefore the probability that infection is established by a single inhaled spore is given by,

$$\begin{aligned} r &= \phi \sum_{n=1}^{\infty} R_{1_S}^n \left(1 - \left(\frac{p}{1-p}\right)^n\right), \\ &= \phi(1 - R_{1_S}^0)\left[1 - \frac{p}{1-p}\left(1 - \frac{1}{b}\right)\sum_{n=0}^{\infty}\left(\frac{p}{(1-p)b}\right)^n\right], \\ &= \phi(1 - R_{1_S}^0)\left[1 - \frac{p}{1-p}\left(1 - \frac{1}{b}\right)\frac{1}{1-\frac{p}{(1-p)b}}\right], \\ &= \phi(1 - R_{1_S}^0)\frac{1-\frac{p}{1-p}}{1-\frac{p}{(1-p)b}}, \end{aligned} \tag{4.1.1}$$

where b is a combination of parameters from the intracellular model (see Eq. (3.1.18)) that is involved in the definition of the rupture size distribution, given by Eq. (3.1.39) and Eq. (3.1.42).

If multiple spores are initially inhaled, the probability of an infection becoming established is equal to the probability that not all of the inhaled spores are cleared, i.e. at least one of the inhaled spores leads to an established infection. Since the initial spores are assumed to be independent of each other, the dose-response curve can therefore be described using the exponential dose-response model. That is, for a Poisson-distributed dose with mean D, the probability of infection is given by,

$$I(D;r) = 1 - e^{-rD}, \tag{4.1.2}$$

where r is the probability that a single spore will produce a response, given by Eq. (4.1.1).

Similar to the competing risks model by Brookmeyer *et al.* (2005), the model presented here is an extension of the exponential dose-response model in which mechanistic detail has been incorporated into the dose-response parameter, r. The competing risks model involves parameters for the two competing processes of spore germination and spore clearance but does not explicitly consider macrophage rupture or intracellular bacterial dynamics. Here, a fully mechanistic model has been proposed, which goes beyond the simple competing risks assumptions. A formula for r has been obtained that takes into account the deposition probability of inhaled spores, the stochasticity of rupture events, and the possibility that even if a few bacteria are released from a rupturing cell, infection might not occur if these few bacteria are killed by the host immune defences before they are able to proliferate to a sufficient number to cause a response. This novel methodology could potentially also be applied to other pathogens.

4.2 Parameter calibration

4.2.1 *In vitro* toxin production

In this section *in vitro* data of *B. anthracis* bacterial growth and toxin production are used to obtain preliminary estimates for the replication rate of extracellular

bacteria, and the rate at which the bacteria produce PA, to help with the calibration of the within-host model. The dataset used comes from an experiment by Charlton *et al.* (2007). In this experiment, they simulated the UK anthrax vaccine manufacturing process, which uses the Sterne 34F2 strain, and obtained *in vitro* data on anthrax toxin production. Specifically, Thompson bottles containing 450 ml of basal medium were sterilized by autoclaving at 121°C for 15 min and then warmed to 37°C prior to inoculation with 50 ml of a spore suspension that had a concentration of 2×10^4 CFU/ml, giving an initial spore concentration in the Thompson bottles of 2×10^3 CFU/ml. These Thompson bottles were then incubated statically at 37°C for up to 32 hours. At various time points, three Thompson bottles were sacrificed, and the numbers of spore and bacterial CFU in the culture were measured. Individual bottle sacrificing was used because repeated sampling from the same bottle was found to disturb the growing cultures.

It was found that the spore counts of each bottle remained fairly constant throughout the 32 hours, at around 30% of the number used to inoculate each bottle. This is because the spores were heat activated prior to inoculation, so would have germinated extremely quickly on contact with the glucose and amino acids of the culture media, and therefore the only remaining spores by the time the first CFU measurements were obtained were those spores that would not go on to germinate during the timescale of the experiment. Therefore the spore data are not used in the subsequent model calibration presented in this section.

As well as the CFU measurements, the protective antigen (PA) and lethal factor (LF) levels in the culture supernatants of each sacrificed bottle were also determined, using antigen-capture ELISA. Only the PA concentration data is used in the following analysis, since this is the toxin component included in the within-host model.

The data of PA concentration (ng/ml) and bacterial growth can be used to calibrate a deterministic model for the number of newly desporulated bacteria, $N(t)$, the number of vegetative bacterial CFU, $V(t)$, and the amount of PA, $T(t)$, over time *in vitro*. The proposed model of *in vitro* bacterial growth and PA

production takes the form of the following system of differential equations.

$$\begin{aligned}\frac{dN}{dt} &= -mN,\\ \frac{dV}{dt} &= mN + \lambda_e V\left(1 - \frac{V}{K}\right),\\ \frac{dT}{dt} &= \beta V - (\nu_0 + \nu V)T.\end{aligned} \tag{4.2.1}$$

The total bacterial CFU is given by $B(t) = N(t) + V(t)$. The inclusion of the newly desporulated bacteria helps to explain the lag period in the bacterial growth curve observed in the data (see Figure 4.6), because these newly desporulated bacteria must mature into vegetative bacteria with rate m before they are able to proliferate. The proliferation of the vegetative bacterial CFU is modelled by logistic growth. It is assumed that PA is produced at a constant rate by each vegetative bacterial CFU, but is not produced by the newly desporulated bacteria.

Collaborators at Dstl have conducted an experiment which aimed to find how quickly PA decays. To do this they incubated a culture of bacteria for 24 hours, during which the bacteria were proliferating and producing PA. After these 24 hours of growth, the bacteria were filtered out. The supernantant was then incubated again and at various time points the supernatant was sampled to quantify how much PA remained. From this, it was found that the rate that PA naturally breaks down in the absence of bacteria is around $\nu_0 = 0.0365\ h^{-1}$ (unpublished data). However, this is much slower than the rapid rate of PA degradation observed in some *in vitro* scenarios where bacteria are present throughout (Zai *et al.* (2016), see Section 1.1.1). It is known that *B. anthracis* produces proteases that contribute to the pathology of anthrax through the degradation of host tissues and modulation of host defences (Chung *et al.* (2006)). However, it has been found that these proteases are also able to degrade PA and LF, since knocking out genes encoding certain proteases, or adding protease inhibitors, has been found to increase yields of the toxin components (Pflughoeft *et al.* (2014); Pomerantsev *et al.* (2011); Zai *et al.* (2016)). This implies that the production of these proteases by the bacteria may be inadvertently self-inhibiting in some way. Due to the large discrepancy between the rate of PA degradation in the presence and

absence of bacteria, there is reason to believe that proteases produced by the bacteria are the dominant mechanism of PA degradation *in vitro*. Hence, a second mechanism of PA degradation is included in the *in vitro* model, in which the rate of PA removal is assumed to be proportional to the vegetative bacterial CFU. The natural degradation rate of PA is fixed to $\nu_0 = 0.0365\ h^{-1}$, and the value of ν is estimated.

I have used the model in Eq. (4.2.1), along with the *in vitro* data from Charlton *et al.* (2007), to obtain estimates for the rate of PA production per bacterial CFU, the rate of PA degradation, and the parameters determining the logistic growth of the bacteria. All bacterial CFU are assumed to be initially in the newly desporulated state, and thus the initial conditions of the model are taken to be $N(0) = 10^{2.29}$ (the first data point used), $B(0) = 0$, and $T(0) = 0$. I have performed a Bayesian approach to the parameter calibration, via ABC-SMC. During the model simulation step of the ABC-SMC algorithm, noise is added to each simulated data point, to take into account measurement errors in the observed data (Alahmadi *et al.* (2020)). These added errors are independent Gaussian with zero mean and standard deviation equal to the standard deviation of the experimental data at the corresponding time-point.

Since the model is simultaneously being fitted to two datasets of different types (bacterial CFU measurements and PA measurements) with different units (CFU vs ng/ml), it is necessary to ensure that the calculation of the distance between the model and the data is not affected by scale differences between the data types. For example, if one simply added together the squared errors between the model and data for each data point, the bacteria fit would be prioritised much more than the toxin fit in the calibration because the numbers are so much higher. Therefore I have defined two distance functions,

$$d_1(Model\ prediction,\ Data) = \sum_{t \in Times} \left(log_{10}\left(\frac{B^*(t)}{B(t)} \right) \right)^2, \tag{4.2.2}$$

$$d_2(Model\ prediction,\ Data) = \sum_{t \in Times} \left(T^*(t) - T(t) \right)^2, \tag{4.2.3}$$

where $B^*(t)$ and $T^*(t)$ are the respective model predictions for the amount of bacterial CFU and PA at time t, $B(t)$ is the geometric mean of the observed number

Parameter	Units	Prior
m	h^{-1}	$\log_{10} m \sim U(-3, 1)$
λ_e	h^{-1}	$\log_{10} \lambda_e \sim U(-1, 1)$
K	$CFU\ ml^{-1}$	$\log_{10} K \sim U(5, 9)$
β	$ng\ (CFU \cdot h)^{-1}$	$\log_{10} \beta \sim U(-7, -2)$
ν	$(CFU \cdot h)^{-1}$	$\log_{10} \nu \sim U(-15, -5)$

Table 4.3: List of model parameters and corresponding prior distributions used in the ABC-SMC for the model of *in vitro* bacterial growth and PA production.

of bacterial CFU across the three Thompson bottles sacrificed at time t in the experiment by Charlton *et al.* (2007), and $T(t)$ is the arithmetic mean amount of PA observed at time t in the experiment. The values of the bacterial CFU predictions and data are log-transformed since they span a large range of magnitudes. At each iteration of the ABC-SMC algorithm, two distance thresholds are used, ε_1 and ε_2, and parameter sets are only accepted if $d_1(Model\ prediction,\ Data) < \varepsilon_1$ and $d_2(Model\ prediction,\ Data) < \varepsilon_2$.

Uniform priors were used for each (log-transformed) parameter over the ranges in Table 4.3. From the posterior histograms in Figure 4.3 one can see that it has been possible to learn a lot about most of the parameters. Furthermore, the parameter estimates obtained seem realistic. For example, the estimates for λ_e are similar to the one used by Day *et al.* (2011) in their within-host model, which was 0.8 h^{-1}. They are also consistent with the best estimate for the intracellular bacterial replication rate from the calibration of the intracellular model in Chapter 3, which was 0.64 h^{-1}.

The scatter plots in Figure 4.5 show correlations between some pairs of parameters. For example, there is a strong negative correlation between λ_e and m, because if the desporulated spores mature quickly, then there needs to be a slower replication rate of vegetative bacterial CFU in order to explain the data, and vice versa. There is a strong positive correlation between β and ν, and a fairly wide range of values of these parameters were accepted. However, we can learn more about the ratio $\beta K/(\nu_0 + \nu K)$, which determines the steady state of PA in the model and is estimated to be around 8×10^3 ng/ml, as shown in

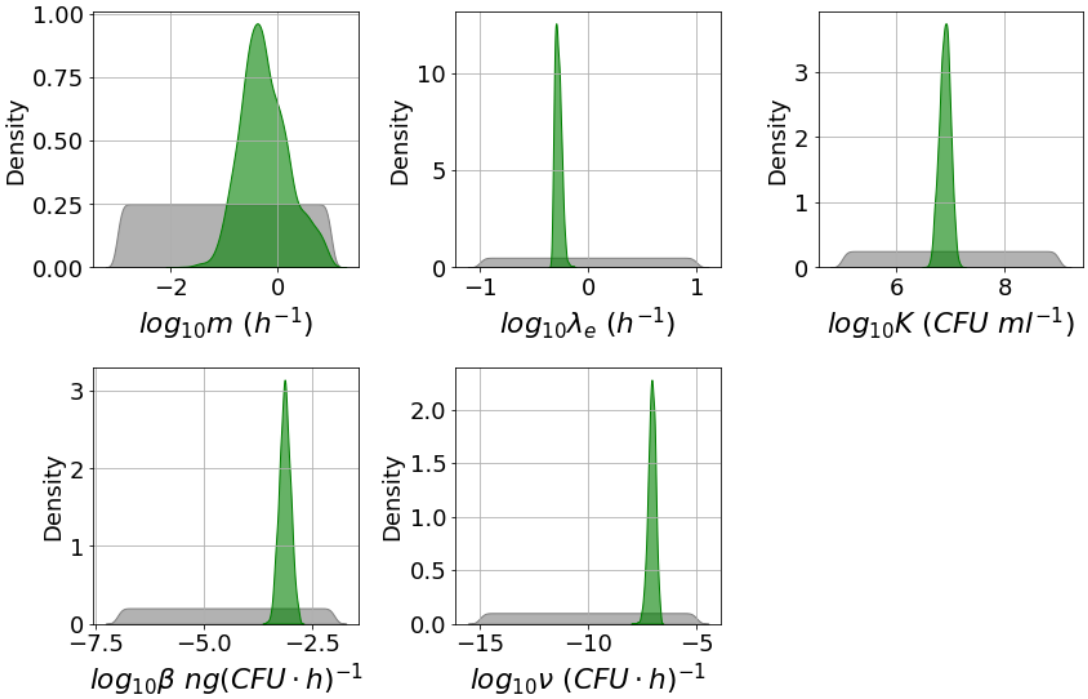

Figure 4.3: Kernel density estimates for the prior distribution for each parameter in grey and the marginal posterior distribution for each parameter in green, from calibration of the model of *in vitro* bacterial growth and PA production.

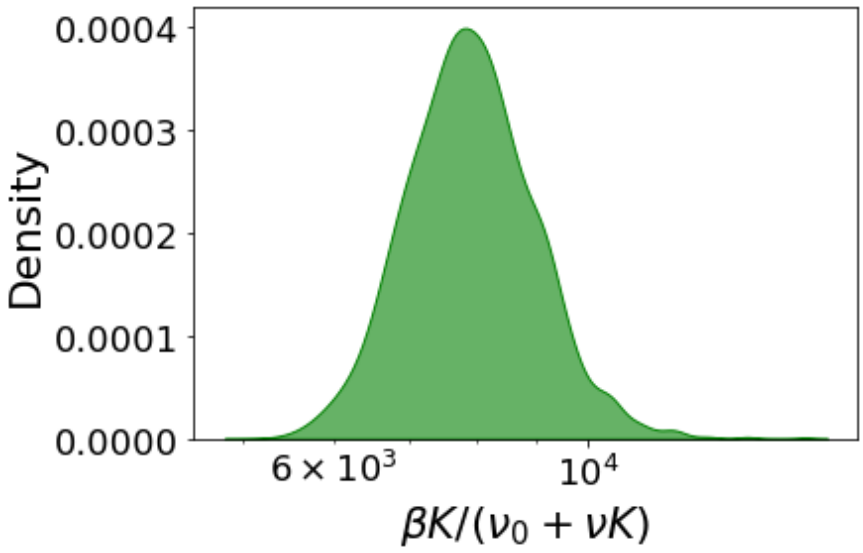

Figure 4.4: Kernel density estimate for the posterior distribution of the transformation $\beta K/(\nu_0 + \nu K)$, which determines the steady state of PA in the model of *in vitro* bacterial growth and PA production. The units of this quantity are ng.

Figure 4.4. It may be helpful in future to use data from experiments where an effort has been made to reduce the degradation of PA by bacterial proteases, for example through the use of protease inhibitors (Zai *et al.* (2016)), in order to more accurately estimate the production rate of PA and help to disentangle the pair of parameters, β and ν.

Figure 4.6 shows the predicted amount of bacteria and PA versus the *in vitro* observations (dots showing the mean of three runs and bars showing the standard deviation). The solid lines represent the pointwise median of the model predictions from all parameter estimates in the posterior sample obtained via ABC-SMC. These predicted curves seem to follow the data quite well.

4.2.2 Dose-response and *in vivo* dynamics

The exponential dose-response model, given by the function in Eq. (4.1.2), is widely used in microbial risk assessment, and has been previously used many times to establish dose-response curves for inhalational anthrax. For example, Gutting *et al.* (2015) have previously fitted the model to dose-response data sets

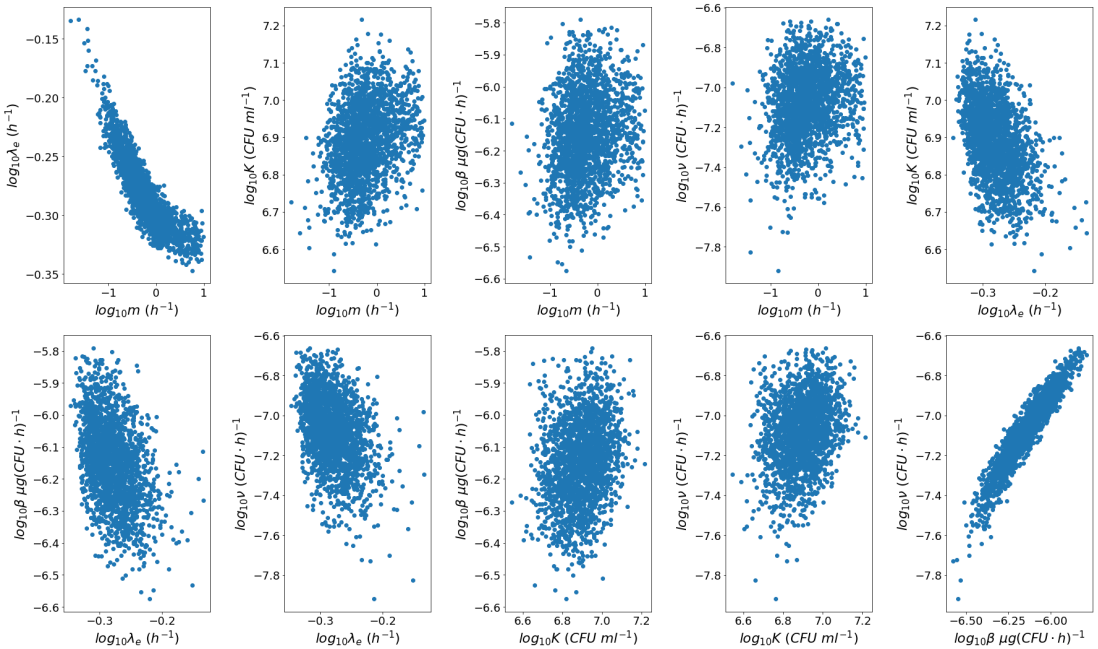

Figure 4.5: Scatter plots showing the relationships between the posterior distributions of pairs of parameters from the model of *in vitro* bacterial growth and PA production.

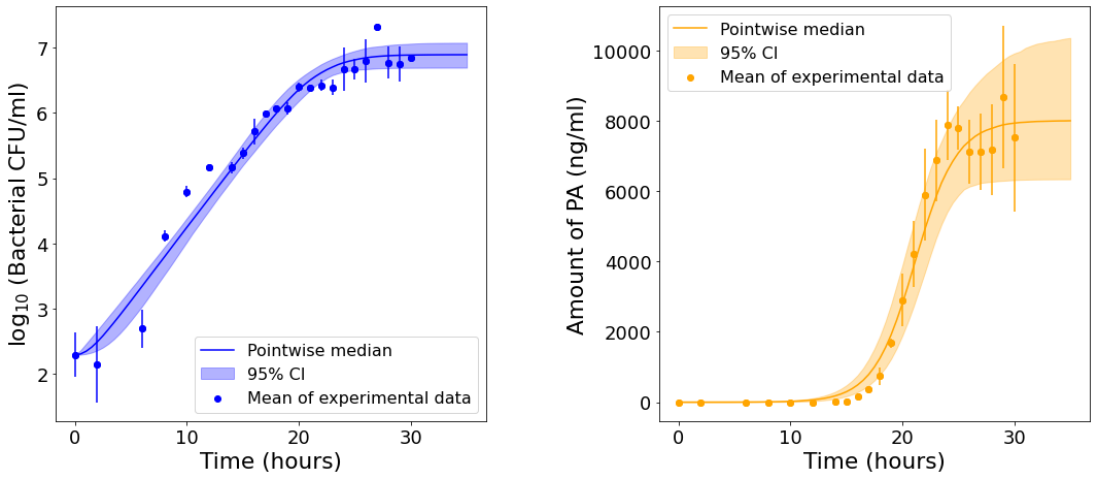

Figure 4.6: Predictions for bacterial CFU (left) and PA concentration (right) over time obtained by using ABC-SMC to fit the model in Eq. (4.2.1) to *in vitro* experimental data from Charlton *et al.* (2007). Solid lines show pointwise medians of the set of predictions using all the parameter sets in the posterior sample, and the shaded regions represent the pointwise 95% credible intervals of these predictions.

for a number of different animal species, in order to estimate the parameter r for each species. Gutting (2014) has also separately calibrated a simple deterministic within-host model of inhalational anthrax, using bacterial counts from infected rabbits. This section will focus on calibrating the model in Figure 4.1 using data from various studies of rabbits and guinea pigs, as well as making use of some of the parameters calibrated using the *in vitro* data in Section 4.2.1.

Dose-response data

When fitting the exponential model to pooled dose-response data for rabbits infected with the Ames strain of *B. anthracis*, Gutting *et al.* (2015) found the value of r that gave the best fit to be $r = 6.75 \times 10^{-6}$. Gutting *et al.* (2015) also fit the exponential model to dose-response data for guinea pigs, and obtained a parameter estimate of $r = 9.79 \times 10^{-6}$. Figure 4.7 shows these exponential dose-response curves compared to the rabbit and guinea-pig data sets. Using the model in Figure 4.1, a formula for the probability r has been obtained instead in terms of parameters (ε, g_A, g_B, $\widetilde{\mu}$, λ, μ, γ, $\hat{\phi}$, k_c, k_p, λ_{LN}, μ_{LN}) with mechanistic interpretations, given in Eq. (4.1.1). However, many combinations of these parameters can produce the same value of r. The parameter calibration of the intracellular model in Chapter 3 incorporates parameter uncertainty, encoded in the posterior distributions (see Figure 3.8). Thus, different parameter sets will lead to different predicted rupture size distributions, which is one of the main components of the mechanistic description of r in Eq. (4.1.1). Depending on the rupture size distribution used, the same dose-response curve can be obtained using a large number of combinations of the parameters $\phi = \hat{\phi}\frac{k_p}{k_c+k_p}$ and $p = \frac{\mu_{LN}}{\mu_{LN}+\lambda_{LN}}$. The rightmost plot of Figure 4.7 shows a wide range of parameter sets involving the value of p, the probability of deposition and phagocytosis ϕ, and the average rupture size derived from the intracellular model parameters, which all lead to the same value of r obtained by fitting the exponential dose-response curve to the rabbit dose-response data. In general, as the average rupture size or deposition/phagocytosis probability ϕ decrease, the required value of p decreases, since more replication will be required to establish a population of bacteria in the lymph nodes.

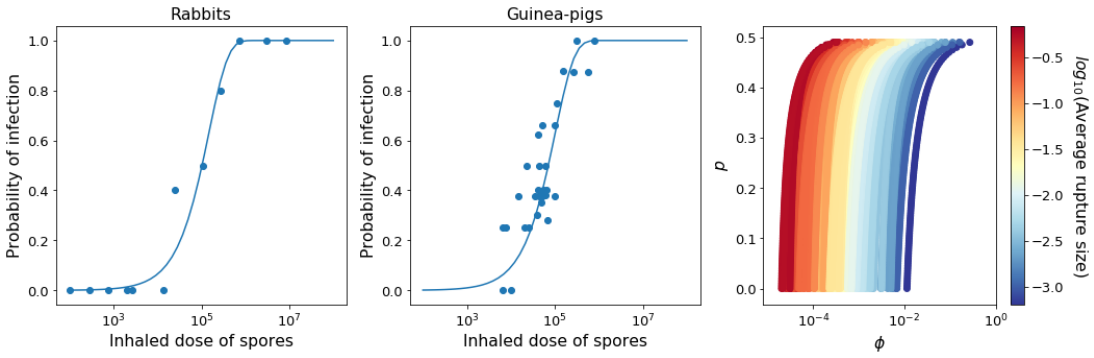

Figure 4.7: Left: Exponential dose-response model fit to pooled rabbit Ames strain dose-response data, by Gutting *et al.* (2015). The probability that one inhaled spore will cause a response is estimated as $r = 6.75 \times 10^{-6}$. **Centre:** Exponential model fit to pooled guinea-pig dose-response data, by Gutting *et al.* (2015), obtaining an estimate of $r = 9.79 \times 10^{-6}$. **Right:** Combinations of parameter values for p and ϕ that give $r = 6.75 \times 10^{-6}$ according to Eq. (4.1.1), for different rupture size distributions sampled from the posterior distribution of the intracellular model. The average of each rupture size distribution is indicated by the colour of the points.

In vivo data

Since different combinations of parameter values can provide the same value of r and hence the same dose-response curve, it is not possible to estimate specific values of the mechanistic parameters using dose-response data alone. Further experimental data, such as measurements of CFU in different compartments over time in infected animals, is needed to calibrate the parameters in the within-host model of inhalational anthrax. I have therefore made use of *in vivo* data sets from infection studies of rabbits and guinea pigs in order to find estimates for the model parameters that are listed in Table 4.2. For each species, the model is fitted to measurements of CFU from animals that have been exposed to a high dose of spores, as well as mortality rates from dose-response data sets, simultaneously. In this way, it is possible to ensure that the model accurately describes both the *in vivo* dynamics of infection, as well as the probability of response after exposure to various doses of spores.

I have used data from a study by Gutting *et al.* (2012) in which rabbits were infected with the highly virulent Ames strain of *B. anthracis* in order to

Time (hours)	CFU in TBLN	CFU in blood	Number of rabbits
1	0.0 ± 0.0	0.0 ± 0.0	4
6	0.0 ± 0.0	0.0 ± 0.0	5
12	10 ± 3.8	0.0 ± 0.0	5
24	$3.167 \pm 2.467 (\times 10^3)$	$4.31 \pm 1.66 (\times 10^5)$	5 (TBLN), 10 (blood)
36	$3.057 \pm 1.486 (\times 10^7)$	$1.18 \pm 0.97 (\times 10^7)$	5 (TBLN), 8 (blood)

Table 4.4: Data for the number of CFU (mean ± SEM) present in the tracheobronchial lymph nodes (TBLN) and blood of groups of rabbits at different times after exposure to a mean inhaled dose of 4.428×10^7 spores. The number of rabbits sacrificed to obtain the measurements at each time point is specified in the final column. The CFU in the TBLN could not be quantified for all rabbits at 24h and 36h, which is why the number of rabbits is fewer for the TBLN than the blood at these times. These data were obtained from the study published by Gutting *et al.* (2012), and have been provided by Brad Gutting via private communication.

characterise bacterial dissemination following exposure to a high aerosol dose of spores. The mean inhaled dose given was 4.428×10^7 spores. Data was provided (mean ± SEM) for the number of bacterial CFU in the tracheobronchial lymph nodes (TBLN) and blood of rabbits at times 1, 6, 12, 24, and 36 hours after exposure. For the measurements of CFU in the TBLN, Gutting *et al.* (2012) pooled all lymph nodes that they could find from the tracheal and bronchial regions, to obtain a number of CFU in the TBLN of each animal. However, the lymph nodes varied in size and number from animal to animal. For the blood measurements, Gutting *et al.* (2012) used a conversion factor of 56 ml blood/kg body weight to convert CFU/ml blood to total CFU counts in the circulation. For some of the time points, data was provided for two separate experiments. I have combined the data into a single mean ± SEM for each compartment at each time point, provided in Table 4.4. Gutting (2014) used similar data to parametrise a simple deterministic model for inhalational anthrax in rabbits, but the data from Gutting *et al.* (2012) is used here because it reports numbers of CFU in the lymph nodes and blood separately, rather than the numbers in the whole rabbit body, as reported by Gutting (2014).

I have also made use of data from a study by Savransky *et al.* (2013) in which

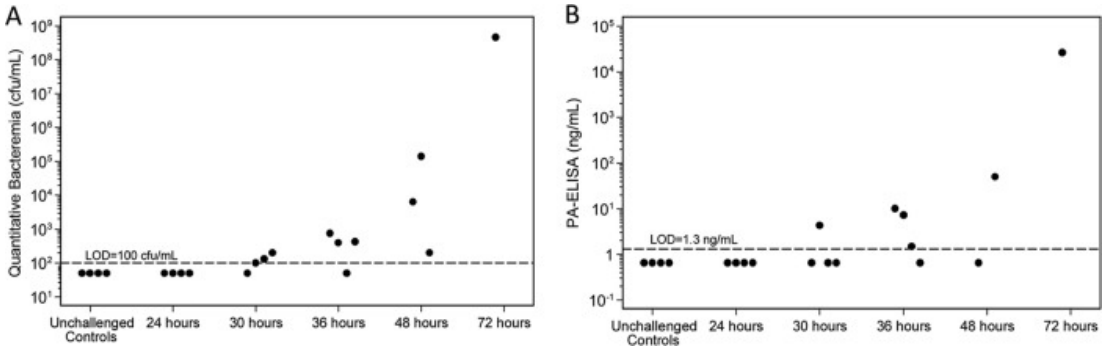

Figure 4.8: (A) Bacterial CFU/ml and (B) PA levels (ng/ml) post-challenge, measured in the blood of guinea pigs exposed to a dose of 2×10^7 Ames spores. Each dot represents a measurement taken from an individual animal. Figure taken from Savransky *et al.* (2013).

guinea pigs were infected with the Ames strain of *B. anthracis* using nose-only aerosol exposure. The exposed dose given was calculated to be 2×10^7 spores. They measured the number of CFU per ml of blood, for which the lower limit of detection was 100 CFU/ml. They also measured the amount of PA in sera. The lower limit of detection for the PA measurements was 1.3 ng/ml. The bacteria and PA levels are shown in Figure 4.8, taken from Savransky *et al.* (2013). I have used the means of the measurements for each time point from 30 hours onward, since all measurements at 24 hours were below the limits of detection. To convert the values of CFU/ml and PA concentration to the total CFU counts and PA amount in the circulation, a volume of 70ml is assumed for the circulating blood of a guinea pig (Williams & Kendall (2015)). The data used is given in Table 4.5.

Deterministic equations

The stochastic nature of the within-host model in Figure 4.1 allows the variability of the infection processes to be captured, which is especially important at low doses of spores. However, the animals in the studies by Gutting *et al.* (2012) and Savransky *et al.* (2013) were exposed to very large numbers of spores. In the case of a large initial dose of spores, there is less stochasticity in the model realisations for each particular set of parameter values. Therefore, a deterministic version of the model can be used to describe the dynamics of infection after a

Time (hours)	CFU in blood	PA in blood (ng)
30	8.14×10^3	1.06×10^2
36	2.65×10^4	3.30×10^2
48	3.16×10^6	1.73×10^3
72	3.07×10^{10}	1.87×10^6

Table 4.5: Data for the mean number of CFU and amount of PA present in the blood of groups of guinea pigs at different times after exposure to a mean dose of 2×10^7 spores. These data have been obtained from Figure 4.8 and converted to the amounts in the whole blood compartment by assuming a volume of 70 ml for the blood of a guinea pig.

high initial dose of spores. Here, a system of ODEs is formed using a mean-field approximation of the expected population levels over time for the model in Figure 4.1. This deterministic version of the model is used to compare with the *in vivo* average measurements of CFU counts from groups of animals infected with large doses of spores in the studies by Gutting *et al.* (2012) and Savransky *et al.* (2013). Note that in these ODEs the number of spores, phagocytes, and bacterial CFU are seen as dimensionless quantities, so that the parameters ρ, δ, λ_{LN}, λ_C, μ_{LN}, and m_B will now have units h^{-1}, instead of the units given in Table 4.2 for the stochastic model.

As described in Section 4.1, the model assumes that spores in the lung airways are removed either by physical clearance with rate k_c, or are phagocytosed and transported into the lung tissue with rate k_p. Thus, the average number of spores remaining in the airways over time can be described by the following differential equation,

$$\frac{dS_A}{dt} = -(k_c + k_p)S_A,$$

with initial condition, $S_A(0) = \hat{\phi}D$, since this is the average number of spores to be initially deposited in the lungs, given an inhaled dose D, and a deposition probability $\hat{\phi}$. This ODE can be easily solved, giving

$$S_A(t) = \hat{\phi}De^{-(k_c+k_p)t} = \hat{\phi}De^{-\rho t},$$

where $\rho = k_p + k_c$ is the overall removal rate of spores from the lung airways.

The time until rupture of infected cells is assumed to follow an Erlang(3, δ) distribution, which is implemented by introducing three infected phagocyte compartments, P_i, $i \in \{1, 2, 3\}$. The mean number of cells in each compartment can be described by the differential equations,

$$\begin{aligned} \frac{dP_1}{dt} &= k_p S_A - \delta P_1, \\ \frac{dP_2}{dt} &= \delta P_1 - \delta P_2, \\ \frac{dP_3}{dt} &= \delta P_2 - \delta P_3, \end{aligned} \tag{4.2.4}$$

with initial conditions $P_1(0) = P_2(0) = P_3(0) = 0$. The equation for $P_1(t)$ can be rewritten as

$$\frac{dP_1}{dt} = \hat{\phi} D k_p e^{-\rho t} - \delta P_1 = \phi D \rho e^{-\rho t} - \delta P_1,$$

where $\phi = \hat{\phi} \frac{k_p}{k_p + k_c}$. Solving the equations for the mean numbers of infected phagocytes in each compartment gives, for $\delta \neq \rho$,

$$\begin{aligned} P_1(t) &= \frac{\phi D \rho}{\delta - \rho} \left(e^{-\rho t} - e^{-\delta t} \right), \\ P_2(t) &= \frac{\phi D \rho \delta}{\delta - \rho} \left[\frac{1}{\delta - \rho} \left(e^{-\rho t} - e^{-\delta t} \right) - t e^{-\delta t} \right], \\ P_3(t) &= \frac{\phi D \rho \delta^2}{\delta - \rho} \left[\frac{1}{(\delta - \rho)^2} \left(e^{-\rho t} - e^{-\delta t} \right) - \frac{t e^{-\delta t}}{\delta - \rho} - \frac{t^2 e^{-\delta t}}{2} \right]. \end{aligned} \tag{4.2.5}$$

If $\delta = \rho$, the solutions are

$$\begin{aligned} P_1(t) &= \phi D \rho t e^{-\delta t}, \\ P_2(t) &= \frac{\phi D \rho \delta t^2 e^{-\delta t}}{2}, \\ P_3(t) &= \frac{\phi D \rho \delta^2 t^3 e^{-\delta t}}{6}. \end{aligned} \tag{4.2.6}$$

Then the average release rate of bacteria into the lymph nodes is given by,

$$\begin{aligned} R(t) &= \delta \mathbb{E}[R_{1_S}] P_3(t), \\ &= \begin{cases} \dfrac{\phi D \rho \delta^3 \mathbb{E}[R_{1_S}]}{\delta - \rho} \left[\dfrac{1}{(\delta - \rho)^2} \left(e^{-\rho t} - e^{-\delta t} \right) - \dfrac{t e^{-\delta t}}{\delta - \rho} - \dfrac{t^2 e^{-\delta t}}{2} \right], & \delta \neq \rho, \\ \dfrac{\phi D \rho \delta^3 \mathbb{E}[R_{1_S}] t^3 e^{-\delta t}}{6}, & \delta = \rho, \end{cases} \end{aligned} \tag{4.2.7}$$

where R_{1_S} is the random variable for the number of bacteria released from a phagocyte initially infected with a single spore, and $\mathbb{E}[R_{1_S}] = \left(1 - R^0_{1_S}\right) \dfrac{b}{b-1}$.

Finally, the differential equations for the bacterial and PA levels in the lymph nodes and circulation are given by (see Figure 4.1),

$$\begin{aligned} \frac{dB_{LN}}{dt} &= R(t) + (\lambda_{LN} - \mu_{LN}) B_{LN} - m_B B_{LN} \mathbf{1}_{B_{LN} > M}, \\ \frac{dB_C}{dt} &= m_B B_{LN} \mathbf{1}_{B_{LN} > M} + \lambda_C B_C, \\ \frac{dT_{LN}}{dt} &= \beta B_{LN} - (\mu_T + \nu B_{LN}) T_{LN}, \\ \frac{dT_C}{dt} &= \beta B_C - (\mu_T + \nu B_C) T_C. \end{aligned} \tag{4.2.8}$$

The set of ODEs in Eq. (4.2.8) describes the number of CFU in the lymph nodes and circulation, given by $B_{LN}(t)$ and $B_C(t)$ respectively, and the amount of PA in the lymph nodes and circulation, given by $T_{LN}(t)$ and $T_C(t)$ respectively. Initially all four populations are zero (i.e. $B_{LN}(0) = T_{LN}(0) = B_C(0) = T_C(0) = 0$). Bacteria arrive in the lymph nodes due to transportation from the lungs, described by the rate $R(t)$, which was defined in Eq. (4.2.7). These bacteria then proliferate in the lymph nodes with a linear growth rate, can be killed by the immune response, or can migrate into the blood (once a threshold of bacteria is reached, representing damage to the lymph nodes resulting in the migration of bacteria). In the blood compartment, the bacterial population also grows with a linear rate, reflecting the general proliferation of bacteria within various organs of the body. Bacteria are assumed to produce PA at a constant rate β per CFU, and multiple mechanisms of PA removal are modelled. PA is removed by natural degradation, and through binding with host cell membranes, at a rate proportional to the current amount of PA. PA can also be actively digested due to proteases produced by the bacteria, which occurs at a rate proportional to the bacterial CFU.

Rabbit model calibration

Gutting (2014) has estimated a deposition probability for inhaled *B. anthracis* spores in the tracheobronchial and pulmonary regions of rabbit lungs of $\hat{\phi} = 0.092$, using a combination of particle deposition data and Regional Deposited Dose Ratio (RDDR) software. A similar software is the Multiple-Path Particle Dosimetry (MPPD) model, which is a computational model that can be used for estimating the deposition and clearance of aerosols in the respiratory tract of humans and various laboratory animal species. I used MPPD with similar inputs to those used by Gutting (2014), and found very consistent values for the deposition fractions. Therefore, I am fairly confident in the deposition probability used by Gutting (2014), and have chosen to fix its value in the model to $\hat{\phi} = 0.092$. Furthermore, Gutting (2014) used data for the decrease of spores in the airways to estimate values for the mucociliated (physical) clearance rate of spores from rabbit airways, $k_c = 0.0628\ h^{-1}$, and the rate of transport of spores from the

lung airways into the lung tissue, $k_p = 0.0107\ h^{-1}$. These two parameters can be combined to obtain a probability for a given spore to be phagocytosed in the lungs rather than cleared from the lungs, given by $\frac{k_p}{k_c+k_p}$. When I have previously attempted to calibrate a value for this phagocytosis probability as well as the probability for the infected phagocytes to rupture and release bacteria once they enter the lymph nodes, I found an extremely strong correlation between these two probabilities, with a correlation coefficient very close to -1. This introduces an issue of identifiability, since only the product of the probabilities can be identified with the available data. For instance, multiplying the phagocytosis probability by some amount and dividing the rupture probability by the same amount will not change the estimated number of bacteria entering the lymph nodes. Therefore, I have decided to make use of the estimates from Gutting (2014) to set values for

$$\phi = \hat{\phi}\frac{k_p}{k_c + k_p} = 0.013,$$

and

$$\rho = k_c + k_p = 0.0735\ h^{-1}.$$

The rupture size distribution depends on multiple parameters from the intracellular model. It is not practical to re-calibrate all these parameters with the available *in vivo* data. Therefore, I have fixed the conditional rupture size distribution to the best estimate from the intracellular model, shown in the inset in Figure 3.12. This distribution has a mean of 1.6, meaning that if a phagocyte engulfs a spore and eventually ruptures, it will release 1.6 bacteria on average. This is actually very similar to the (per spore) rupture size in the model by Day *et al.* (2011), where it was assumed that a phagocyte would release 5 bacteria into the lymph nodes after engulfing 3 spores in the lungs. There was large variation in the probabilities of rupture from the posterior sample of the intracellular model, spanning a wide range of orders of magnitude, from as small as 9×10^{-5} to 0.53 (see Figure 4.9). Therefore I have allowed the rupture probability, $1 - R^0_{1S}$, to be estimated as a parameter in the calibration of the within-host model with the *in vivo* data and dose-response data.

For calibrating the model with the rabbit data from Table 4.4, I found that it was necessary to include a carrying capacity for the growth of bacteria in

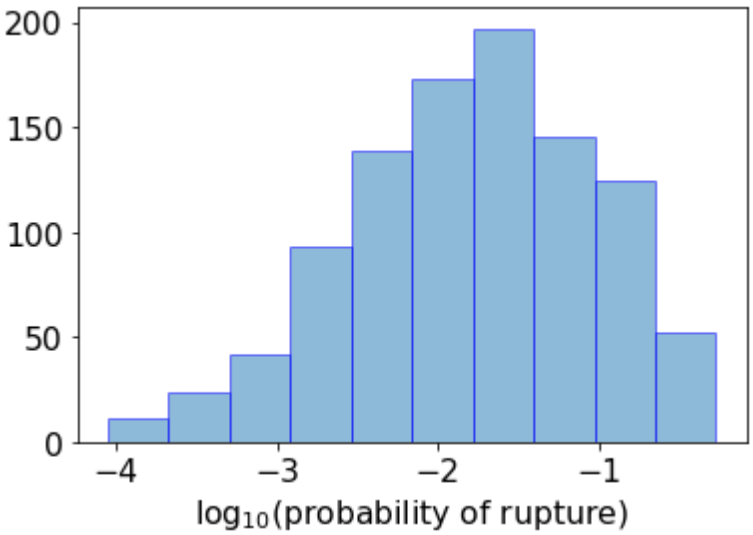

Figure 4.9: Histogram for the probabilities of rupture obtained from the posterior sample of the intracellular model calibration in Chapter 3.

the circulation compartment. This is because the bacteria in the blood is not observed to grow much above 10^7 CFU, indicating that there must be limiting growth at some stage. Hence, the equation describing the bacterial dynamics in the circulation compartment for rabbits is modified to incorporate a logistic growth term,

$$\frac{dB_C}{dt} = m_B B_{LN} \mathbf{1}_{B_{LN}>M} + \lambda_C B_C \left(1 - \frac{B_C}{K_C}\right),$$

where K_C is the introduced carrying capacity. The same saturation is not observed in the lymph node data during the time period of observation, so it is not necessary to include a carrying capacity in the lymph nodes, since it would not be possible to identify a value for it with the available data.

The model in Figure 4.1 has been simultaneously fitted to the rabbit dose-response data shown in Figure 4.7 and the *in vivo* numbers of bacteria in Table 4.4, using ABC-SMC. To do this, values of most of the (log-transformed) parameters were sampled from uniform prior distributions, with the exception of λ_{LN}. For the replication rate in the lymph nodes, λ_{LN}, the posterior sample for the replication rate in the *in vitro* model in Eq. (4.2.1) (obtained from fitting the model to the data from Charlton *et al.* (2007), see Figure 4.3) was used as a prior distribution

here. For each sampled parameter set, the exponential dose-response curve was calculated according to Eqs. (4.1.1) and (4.1.2), to obtain values of the probability of infection $p_D = I(D; r)$ for each dose D in the dose-response data set. Then for each dose D, a number of infections was sampled from a Binomial(N_D, p_D) distribution, where N_D is the size of the exposed group for dose D in the data set. While the endpoint of the dose-response studies was lethality, and p_D has been referred to as a probability of infection, it is assumed here that infection with inhalational anthrax will invariably lead to death in the absence of treatment, so that the number of animals that die in the study is assumed to be equivalent to the number infected (Toth *et al.* (2013)). Therefore, the Euclidean distance was computed between the set of sampled numbers of infections from the model and the observed numbers of deaths in each group from the data, giving a distance measure between the model and the dose-response data for a particular parameter set. To obtain a distance measure between the model and the *in vivo* bacterial counts from Table 4.4, the output from the system of ODEs in Eq. (4.2.8) (with adapted equation for B_C to include logistic growth) was compared to the data of mean numbers of CFU in the TBLN and blood. For this, only the equations for B_{LN} and B_C are needed, since we do not have any data for toxin amount in this dataset, and the numbers of bacteria do not depend on the amount of PA in this model. Therefore, values of β, μ_T, and ν were not calibrated using this data.

The model predictions from this calibration are shown in Figure 4.10, where the top row shows the predicted amount of bacteria versus the *in vivo* observations (dots showing the mean CFU measurements from groups of rabbits and bars showing the SEM). The solid lines represent the pointwise median of the model predictions from all parameter estimates in the posterior sample obtained via ABC-SMC. The plot on the bottom row shows the predictions of the dose-response curve compared to the observed dose-dependent mortality rates in groups of rabbits exposed to different doses. Figure 4.11 shows the binomial distributions for the number of infections, given by the model with the median parameter values from the posterior sample, for each dose in the rabbit data set. The observed number of deaths is indicated as a vertical line, for each dose.

From the posterior histograms in Figure 4.12 one can see that it has been possible to learn a lot about most of the parameters. One parameter with a

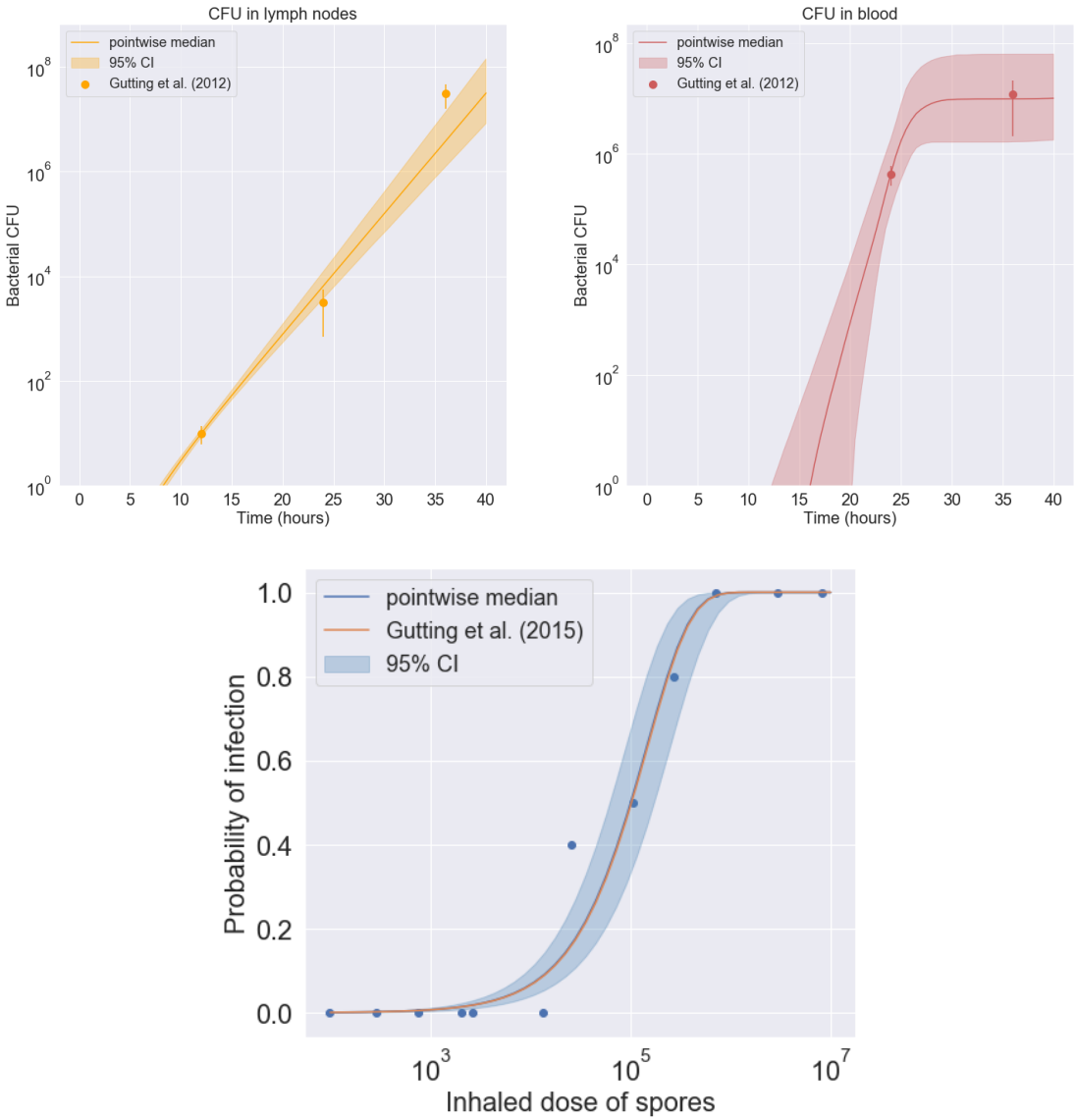

Figure 4.10: Model predictions compared to the data used in the ABC-SMC for the rabbit model calibration. **Top row:** Predictions from the system of ODEs in Eq. (4.2.8) compared to the mean CFU loads in the TBLN and blood from the rabbit data in Table 4.4. **Bottom:** Prediction of the mechanistic exponential dose-response model given by Eqs. (4.1.1) and (4.1.2) compared to rabbit dose-response data from Gutting *et al.* (2015).

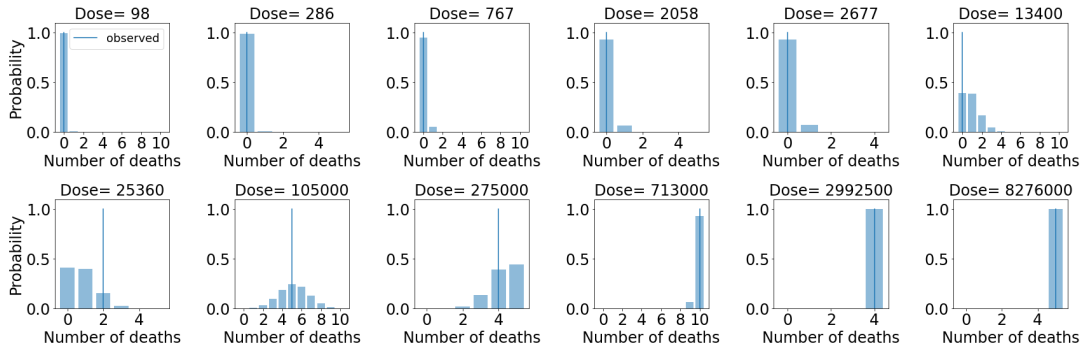

Figure 4.11: Binomial distributions for the number of deaths, given by the model with the median parameter values from the posterior sample, for each dose in the rabbit data set. The observed number of deaths is indicated as a vertical line, for each dose.

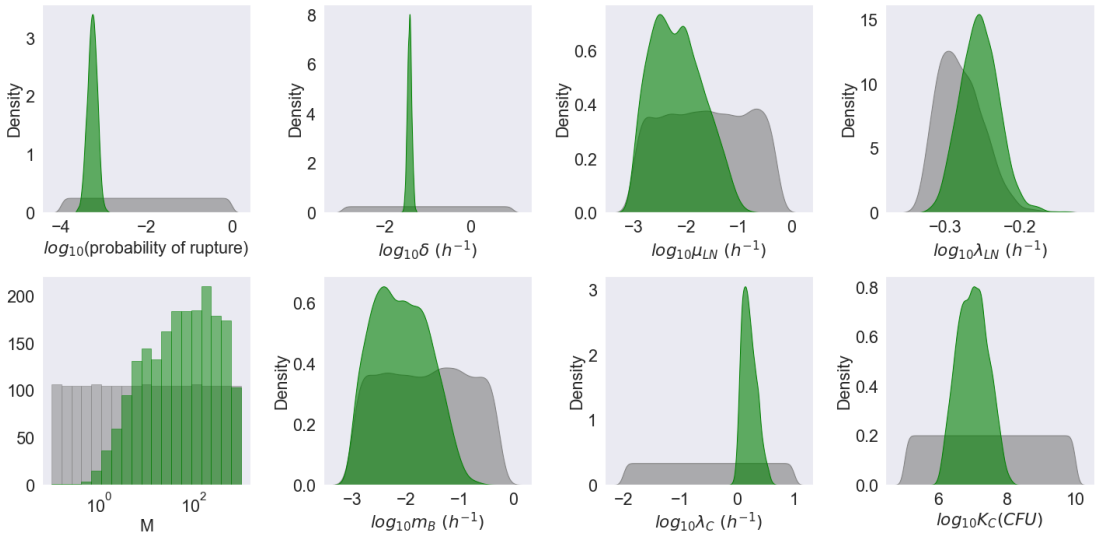

Figure 4.12: Kernel density estimates for the prior distribution for each parameter in grey and the marginal posterior distribution for each parameter in green. These posterior distributions were obtained from fitting the mechanistic dose-response model given by Eqs. (4.1.1) and (4.1.2) to the rabbit dose-response data from Gutting *et al.* (2015), and simultaneously fitting the model in Eq. (4.2.8) to the mean CFU loads in the TBLN and blood from the rabbit data in Table 4.4.

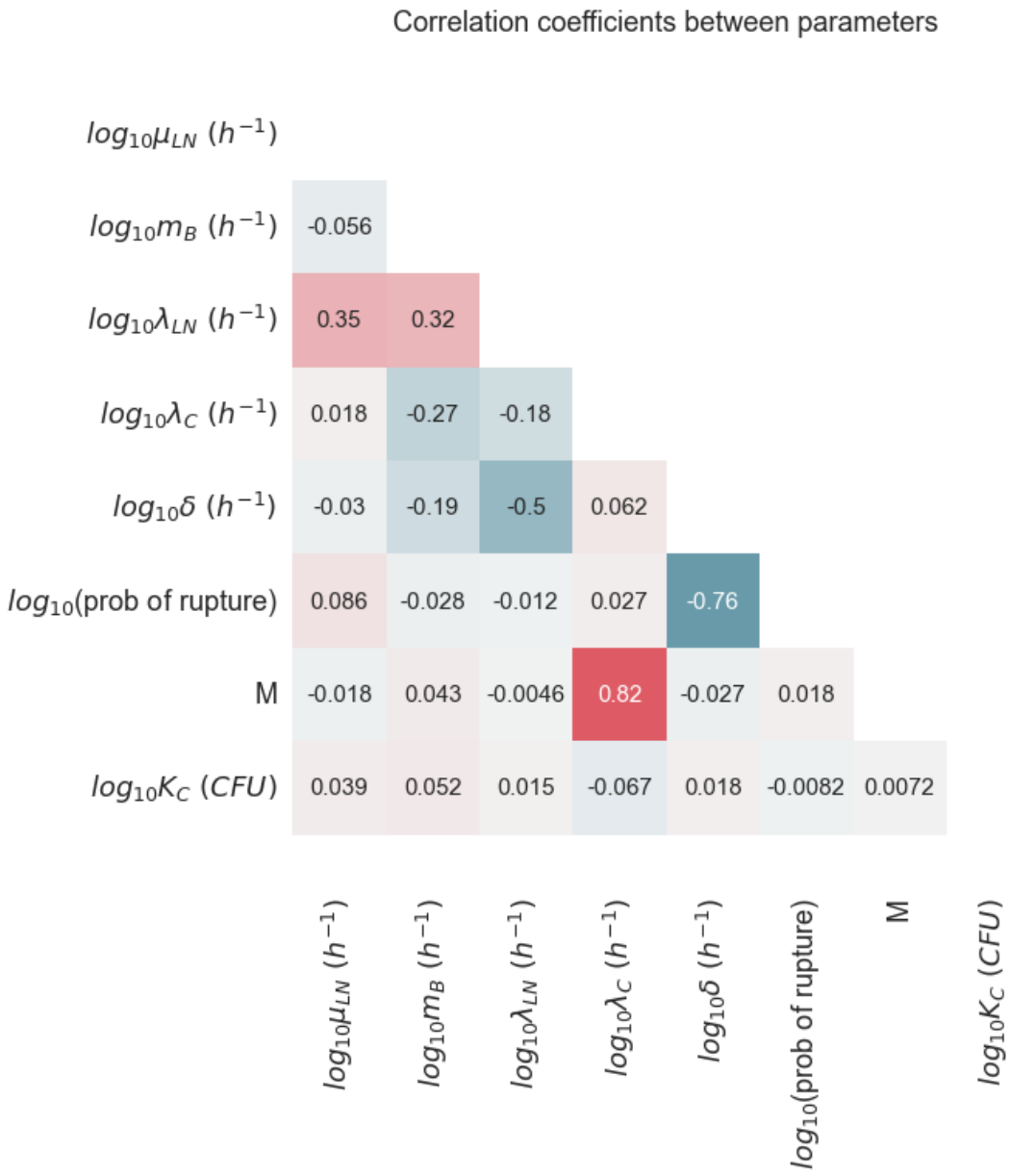

Figure 4.13: Correlation coefficients between the posterior samples of pairs of parameters in the rabbit model calibration.

particularly narrow posterior distribution is δ, which determines how quickly infected phagocytes migrate from the lungs to the lymph nodes and rupture. With the median value of $\delta = 0.04\ h^{-1}$, the Erlang(3, δ) distribution of the time for phagocytes to rupture has a mean of 75 hours. Compared to the average rupture time of 20 hours used in the model by Day *et al.* (2011), rupture is occurring very slowly in the model predictions in Figure 4.10, and only a fairly small proportion of the infected phagocytes will be expected to rupture during the 40 hour time period shown in the plots. However, once a few phagocytes have ruptured and released bacteria into the lymph nodes, the dynamics in the lymph nodes begins to be dominated by the extracellular replication of bacteria, and the subsequent rupture events do not have a great impact on the dynamics.

The probability of rupture is estimated to be around 5×10^{-4} which is towards the lower end of the prior range. However, this is still consistent with the estimates from the intracellular model, since around 7% of the rupture probabilities in the histogram in Figure 4.9 are below 10^{-3}. Since the intracellular model was calibrated with *in vitro* data, this difference in the estimated probability of rupture could indicate that macrophages are better at killing bacteria *in vivo*. Also, the cells used in the *in vitro* experiment were murine peritoneal macrophages, so rabbit alveolar macrophages and dendritic cells may react quite differently. Note that parameter estimates from Gutting (2014) were used to fix the probability for spores to be phagocytosed in the lungs. If the true phagocytosis probability was actually smaller than the one used here, one would expect the estimated rupture probability to increase. This is because these two probabilities are highly related, in the sense that only the product of the two probabilities can be identified from the data used here. Fortunately, when previously including the phagocytosis probability in the list of parameters to be estimated, there was no substantial correlation observed between the phagocytosis probability and any parameter other than the rupture probability. Hence, it seems that the other parameter estimates will be robust to changes in the phagocytosis probability.

Some parameters have fairly wide posterior distributions, such as the killing rate of bacteria in the lymph nodes, μ_{LN}, and the migration rate of bacteria from the lymph nodes to the blood, m_B. Both of these parameters are estimated to be very small. In fact, when μ_{LN} was set to zero before the calibration, the posterior

distributions of the other parameters were not affected. This is consistent with the competing risks hypothesis, for example, that if a bacterium is released from a phagocyte, then it will be almost certain to proliferate and cause an infection.

The replication rate in the lymph nodes, λ_{LN}, is estimated to be very similar to, but slightly larger on average than the values in the posterior sample from the previous fitting to the data from Charlton *et al.* (2007). Here the median value λ_{LN} is 0.56 h^{-1}, which is also similar to the value of 0.8 h^{-1} used by Day *et al.* (2011) in their within-host model, and the estimate of 0.64 h^{-1} for the intracellular bacterial replication rate from the calibration of the intracellular model in Chapter 3.

The threshold, M, for the number of bacteria present in the lymph nodes before migration into the circulation compartment can occur is estimated to be around 10^1 or 10^2, and values of zero were very rarely accepted into the posterior sample. This small migration threshold is needed in order to explain the counts of zero CFU up to 12 hours in the blood (see Table 4.4). Once bacteria start to arrive in the blood, the population of bacteria in the blood grows very quickly, reflected by the large values of λ_C, giving a doubling time of around 30 minutes. The values of λ_C are much larger than the replication rate of bacteria in the lymph nodes, but this is because the growth rate in the blood is representing migration into the blood from multiple infected organs and tissues, where proliferation of bacteria will be happening, rather than direct replication in the blood. There is a strong positive correlation between M and λ_C in the posterior sample (see Figure 4.13). This is because the later that bacteria start migrating into the blood, the larger the growth rate in the blood needs to be in order to be consistent with the observed data points at 24h.

Guinea-pig model calibration

Here I have used ABC-SMC to obtain some estimates for the parameter values of the within-host model that describe the guinea-pig model of infection. As with the rabbit data, I have simultaneously fitted the model in Figure 4.1 to the guinea-pig dose-response data from Gutting *et al.* (2015) and the *in vivo* numbers of bacterial CFU and PA amount from Savransky *et al.* (2013), shown in Figure 4.8

and Table 4.5. I compared the exponential dose-response curve, computed using Eqs. (4.1.1) and (4.1.2), to the guinea-pig dose-response data in Figure 4.7, in the same way as described for the rabbit data. I also compared the deterministic model outputs from the set of ODEs in Eq. (4.2.8) to the means of the data in Figure 4.8 at each time point from 30 hours onward, since all measurements at 24 hours were below the limits of detection.

For guinea pigs, Gutting *et al.* (2015) estimated that the deposition probability was $\hat{\phi} = 0.3$. I have therefore used this value, along with the estimates from rabbits for the physical clearance rate of spores out of the lungs, and the transport rate of spores from the lung airways into the lung tissue. Thus for guinea pigs, values are fixed to $\phi = 0.0437$, and $\rho = 0.0735\ h^{-1}$.

With only the data from the blood, it is difficult to quantitatively estimate what is happening in the lymph nodes. Therefore, when I previously tried to calibrate all parameters using this data, confidence in the values of most of the parameters was limited, as indicated by wide posterior distributions. Because of this, I have aimed to reduce the parameter space by assuming that some parameters determining the dynamics in the lymph nodes will be similar to the ones for rabbits. In this way, I have fixed the values of some parameters to those estimated for the rabbit model of infection, while still allowing important differences between the two models of infection. In particular, I have fixed the values for $\delta = 0.04\ h^{-1}$, $\lambda_{LN} = 0.56\ h^{-1}$, and $m_B = 0.008\ h^{-1}$ to the median values estimated for rabbits. However, ϕ is different between the two animal models, since it has been estimated that the proportion of inhaled spores deposited in the lungs is larger in guinea pigs than rabbits (Gutting *et al.* (2015)). I also allow values for μ_{LN} and λ_C to be different between the two animal models, since guinea pigs and rabbits may have different strengths of immune response against the extracellular bacteria in the lymph nodes and in other organs, as well as different sizes of organs.

The model predictions from this calibration are shown in Figure 4.14. For the rabbit model, a carrying capacity was included in order to explain the data that was observed for rabbits, whereas for the guinea-pig model it was not necessary to include logistic growth in the blood, since no obvious saturation was observed in the data during the time period of observation. However, it is possible that

the saturation observed in the rabbit blood could be an artefact, and if there was data for a longer timescale, more growth might be observed in the blood, similar to the guinea-pig data. Overall, the growth of bacteria in the blood is slower in the guinea-pig data than for the rabbits. For example, at 36 hours post infection, the mean number of CFU in the blood of guinea pigs is around 400 CFU/ml, whereas for rabbits it is around 7×10^4 CFU/ml, assuming a volume of blood of 168 ml for rabbits. This has an effect on the posterior distributions, shown in Figure 4.16. For example, the growth rate in the blood, λ_C, is estimated to be much smaller for guinea pigs than it was for rabbits. Since the amount of bacteria in the blood depends strongly on the amount of bacteria in the lymph nodes in this model, the guinea-pig model also needs much slower growth in the lymph nodes in order to explain the slower growth in the blood compartment. Therefore, the death rate of bacteria in the lymph nodes, μ_{LN}, is estimated to be larger for guinea pigs than rabbits, and the probability of rupture is estimated to be slightly smaller. Since the amount of bacteria in the lymph nodes is predicted to be smaller, the value of the migration threshold, M, is also estimated to be smaller for guinea pigs.

For the production rate of PA, β, the posterior distribution from the previous calibration with the *in vitro* data from Charlton *et al.* (2007) was used as the prior distribution here for this parameter. Some values from this prior distribution also work well for the *in vivo* data, but the posterior is shifted slightly to the right, indicating that the bacteria are predicted to be producing PA at a slightly faster rate *in vivo*.

With a smaller probability of rupture, and a larger death rate of extracellular bacteria in the lymph nodes, this means that each spore that is deposited in the lungs has a smaller probability to cause an infection in the guinea-pig model compared with the rabbit model. However, each inhaled spore has a higher chance of becoming deposited in the guinea-pig model. Figure 4.15 shows the binomial distributions for the number of infections, given by the model with the median parameter values, for some of the doses in the guinea-pig dose-response data set used here. The observed number of deaths is indicated for each dose shown. There is quite a lot of variability in the dose-response data used here. If each animal in an exposed group has an identical probability of infection for

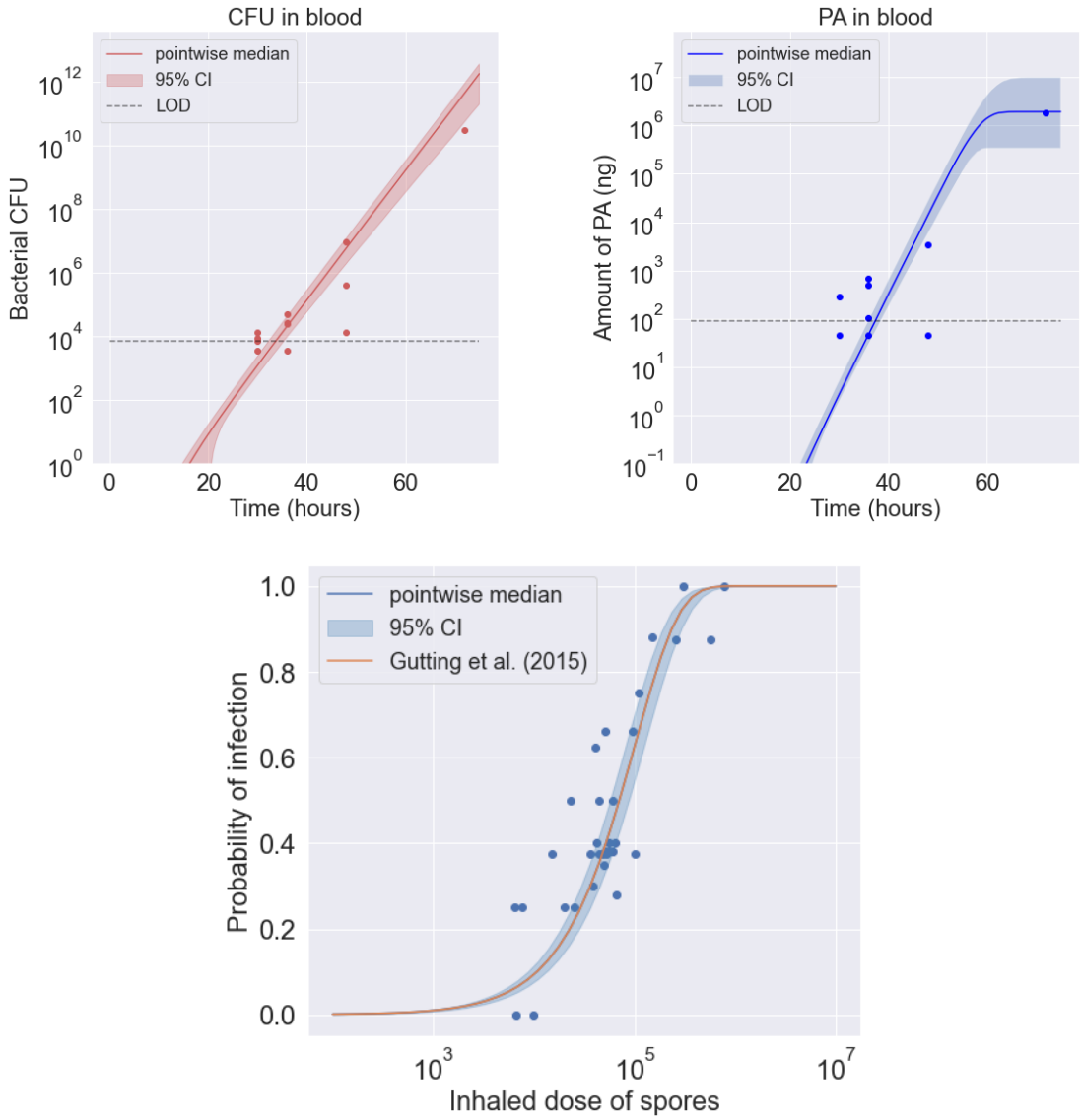

Figure 4.14: Model predictions compared to the guinea-pig data used in the ABC-SMC. **Top row:** Predictions from the model in Eq. (4.2.8) compared to the CFU loads and PA amounts in the blood from the guinea-pig data in Figure 4.8. **Bottom:** Prediction from the mechanistic exponential dose-response model given by Eqs. (4.1.1) and (4.1.2), compared to guinea-pig dose-response data from Gutting *et al.* (2015).

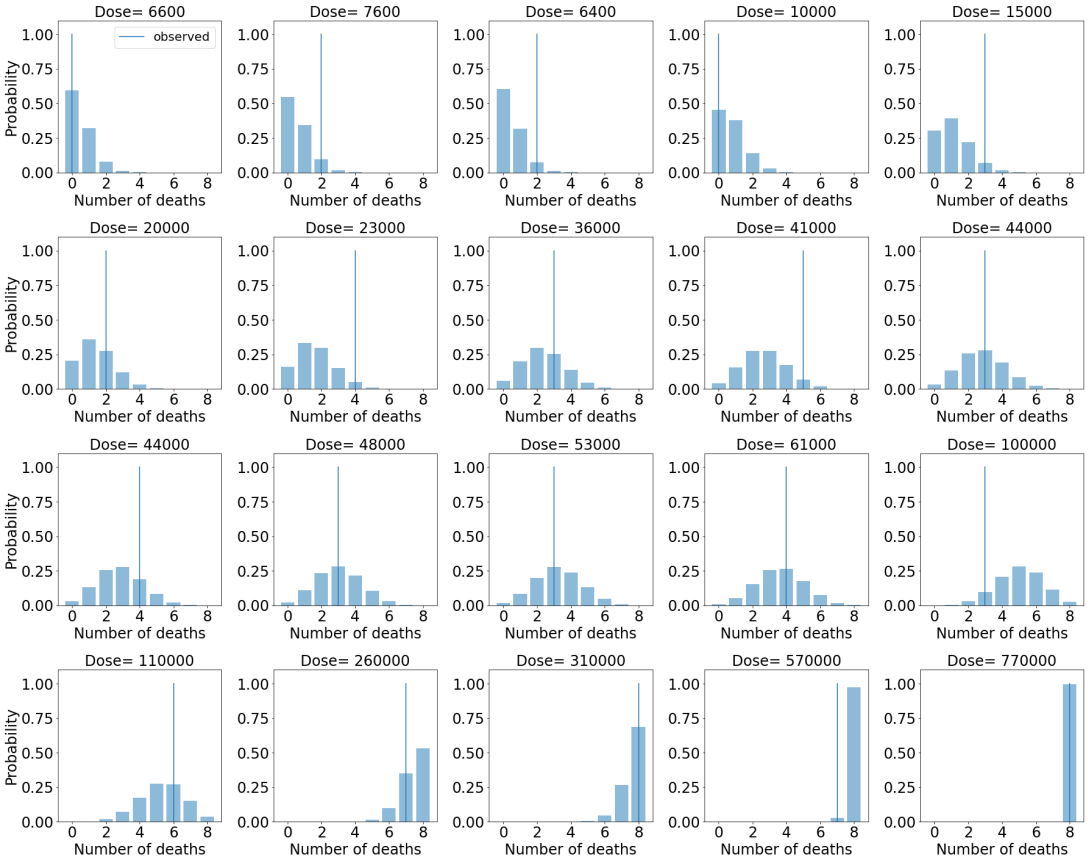

Figure 4.15: Binomial distributions for the number of deaths, given by the model with the median parameter values from the posterior sample, for some of the doses in the guinea-pig dose-response data set used here. The observed number of deaths is indicated as a vertical line, for each dose shown.

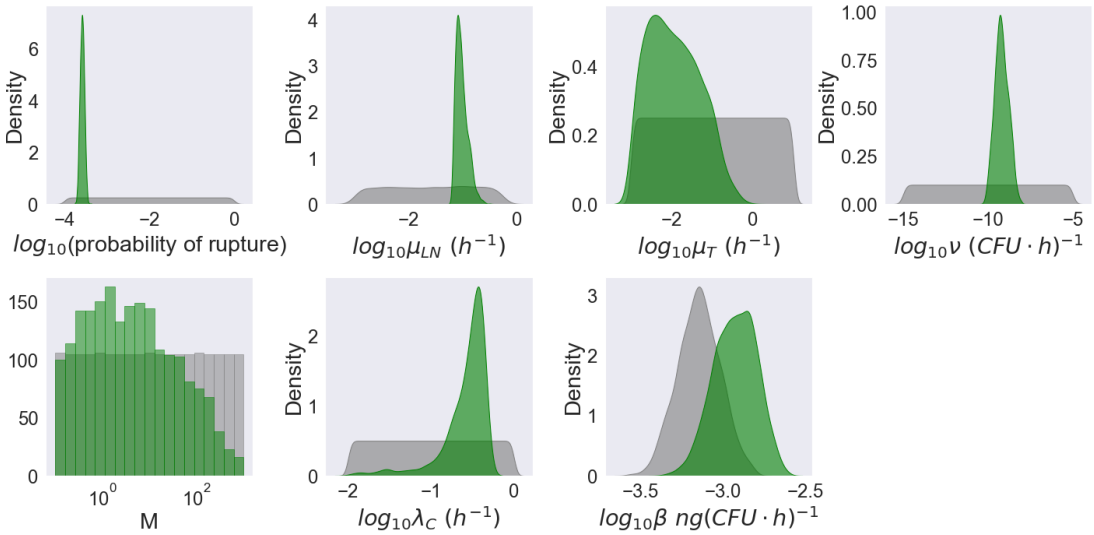

Figure 4.16: Kernel density estimates for the prior distribution for each parameter in grey and the marginal posterior distribution for each parameter in green. These posterior distributions were obtained from fitting the mechanistic dose-response model given by Eqs. (4.1.1) and (4.1.2) to the guinea-pig dose-response data from Gutting *et al.* (2015), and simultaneously fitting the model in Eq. (4.2.8) to the mean CFU loads and PA amounts in the blood from the guinea-pig data in Figure 4.8.

Figure 4.17: Correlation coefficients between the posterior samples of pairs of parameters in the guinea-pig model calibration.

a given dose, as assumed by the model, then the variability would purely be due to the variance of the binomial distribution that determines the number of observed infections in this model. However, individual animals are likely to have slightly different susceptibilities to infection, so this could also be a reason for the observed variability. For example, exposed groups with a slightly higher or lower mortality rate than expected may include animals with an unusually high or low susceptibility to infection, respectively. Furthermore, the exposed dose and deposited dose of spores received by each animal may also be highly stochastic, since the deposition probability, $\hat{\phi}$, depends on a number of different factors, such as breathing rate, which will vary between individual animals. Therefore, this could also be introducing a lot of variability.

4.3 Discussion

In this Chapter, I have proposed a stochastic within-host model of inhalational anthrax infection, which incorporates inter-phagocyte variability in rupture size by making use of the rupture size distribution from the intracellular model of Chapter 3. This rupture size probability distribution plays an important role when using the model to determine the probability of infection following inhalation of some dose of spores. The within-host model proposed here also incorporates the dynamics of PA, which is the anthrax toxin component that the therapeutic antibody targets, so this will be an important aspect of the model if it is used to model certain treatments in future work. *In vitro* data of bacterial growth and PA production by *B. anthracis* Sterne strain was leveraged to obtain preliminary estimates for the replication rate of extracellular bacteria, and the production rate of PA. Making use of these preliminary estimates as prior distributions, the within-host model was then calibrated by comparing the mean-field approximation of the stochastic model with *in vivo* bacterial counts from rabbits and guinea pigs infected with the highly virulent Ames strain of *B. anthracis*. At the same time, the probability of infection resulting from the stochastic model was fitted to dose-response data sets from infection of rabbits with the Ames strain, and infection of guinea pigs with the M36 strain or Ames strain. This resulted in a single model structure that is able to describe both the *in vivo* dynamics of infection, as

well as the probability of response, in two different animal species with slightly different values of the parameters. Due to lack of data, it has been necessary to merge data corresponding to different strains of *B. anthracis* (Sterne, Ames, M36) and different animals; for example, the cells used in the *in vitro* experiment that informed the conditional rupture size distribution used in the within-host model were from mice, whereas the within-host model was eventually calibrated with data from rabbits and guinea pigs.

During the calibration with rabbit and guinea-pig data, some of the parameters were set equal between the two animal models, due to a lack of data for the lymph nodes compartment in guinea pigs. It is clear from the data that with the current structure of the model, there needs to be lower numbers of CFU in the lymph nodes in the guinea-pig model compared with what is predicted for the rabbit model. However, since we do not have data for the CFU in the lymph nodes of guinea pigs, I set most of the parameters determining these dynamics to be equal to their median estimates from rabbits. The only one of these parameters that I did not fix was the death rate of bacteria in the lymph nodes, which specified that it must be a difference in the death rate that is accounting for differences in the lymph nodes compartment dynamics between the two species. These assumptions might change if more data became available.

Although I have followed a Bayesian approach of model calibration, which allows for the uncertainty in parameter estimates to be quantified, I have not considered the parameter values to vary between individual animals. This is a simplification, and could be improved with approaches such as mixed effects modelling, which allow a probability distribution to be estimated to describe how the parameters vary across individuals. However, this would be difficult in this particular situation, since we do not have more than one data measurement for each animal. Some parameters may be easier to assign a probability distribution to than others. For example, in a dose-response model for Q fever (a bacterial infection caused by *Coxiella burnetii*), Heppell *et al.* (2017) approximated a distribution for the probability of deposition, $\hat{\phi}$, using the MPPD model software package.

As the model stands, it cannot explain the existence of spores in the lymph nodes, since the model assumes that at most one spore is engulfed by each phago-

cyte in the lungs and only bacteria are released from rupturing phagocytes. However, in future the model could be extended to allow the uptake of multiple spores by individual phagocytes, which may be likely at high inhaled doses. This would also allow the consideration that cells may release a mixture of spores and bacteria when they rupture, which would explain observations such as spores identified in the lymph nodes of mice (Loving *et al.* (2007)). The model may then need to also include the consideration that these spores may be engulfed by further phagocytes in the lymph nodes and lead to further rounds of intracellular replication, which is not currently considered in the model.

A benefit of the mechanistic modelling approach proposed here is that the underlying mechanisms of the model can be extended and modified as new scientific knowledge and data is generated. Furthermore, because the parameters of the model have mechanistic interpretations, the model can be used to investigate the effects of different biological parameters, and changes can be made to model different situations, such as incorporating additional mechanisms to represent treatments. In future work, the aim will be to include medical treatments into this stochastic multi-scale model of *B. anthracis* infection, to quantify their efficacy. This will be done using pharmacokinetic (PK) data that describes how the within-host concentration of the treatment will change through time, as well as pharmacodynamic (PD) data that could, for example, be used to define an extracellular killing rate of bacteria as a function of antibiotic concentration (Foerster *et al.* (2016)). It will then be possible to use the resulting model to quantify the effect that the initial timing, frequency, and dosage of treatment has on reducing the probability of infection. In this way, the multi-scale framework could provide useful information for the design and optimisation of medical treatments for inhalational anthrax.

Chapter 5

Reproduction number probability distributions in stochastic models of viral dynamics with non-exponential infectious period

The basic reproduction number, R_0, was introduced in an epidemiological context to denote the mean number of secondary infections caused by an initial infected individual in an otherwise susceptible population. In this chapter I will consider the cellular-level equivalent in the context of within-host and *in vitro* viral dynamics (Burg *et al.* (2009); Nowak & May (2000); Pearson *et al.* (2011); Wodarz & Nowak (2002); Yan *et al.* (2020)). In this chapter the cellular-level basic reproduction number will be understood as a random variable, R, representing the number of new cells infected by one initial infected cell in an otherwise susceptible cell population. The expected value of the basic reproduction number will be denoted by $\bar{R}$.

Viral dynamics models are often deterministic and formulated as sets of ordinary differential equations. For these models, the mean value of the basic reproduction number determines the outcome of infection. If $\bar{R} < 1$ then the virus-free steady state is stable and the infection will rapidly die out. In stochastic models, if the expected number of secondary infections caused by each infected cell is smaller than one, the infection can grow, but will also certainly die out

eventually. Conversely, when the mean basic reproduction number is greater than one, deterministic models predict that an infection will be established. However, for stochastic models, even if $\bar{R} > 1$, there can be a positive probability that the virus and infected cells will be eliminated before an established infection can occur. This probability depends on the *probability distribution* of the reproduction number. By exclusively focusing on the expected value of the reproduction number, one may fail to capture important dynamics during the early events in infection, which are inherently stochastic, particularly if an individual is only exposed to a low level of virus.

When a free virus particle is taken up by a host cell, the genome is replicated and used to produce viral proteins. New virus particles are then assembled in the cell. For some viruses, progeny virions accumulate inside the host cell and eventually numerous virus particles exit at once in a burst, killing the cell. On the other hand, for most enveloped viruses, new virus particles are released throughout the lifetime of the infected cell via a process called budding. The total number of virions produced by an infected cell during its lifetime is usually referred to as the "burst size", even if the virus is released continuously by budding (Liao *et al.* (2020); Rong *et al.* (2013)). The progeny virus particles produced by an infected cell may infect further host cells, or may degrade before they manage to do this. Hence, the reproduction number, which is the number of secondary infections caused by an initial infected cell, will depend on the burst size of the infected cell and the likelihood that these progeny virions will go on to infect new cells before they degrade. The stochasticity of the intracellular processes of the viral lifecycle, coupled with variation in the lifetime of infected cells, can lead to large variation in the number of virions released by individual cells. Bacsik *et al.* (2022) measured progeny production from single influenza-virus-infected cells and found that this was extremely heterogeneous. This variation in the burst size is likely to have a large impact on the overall variability in the reproduction number between cells.

Finding the reproduction number distribution is particularly important in order to characterise the viral dynamics at the very early stages of infection, when there may only be small numbers of virions and infected cells. This is because stochastic effects are very important at this stage, and the distribution

of the reproduction number determines the probability that the virus will be cleared before a major infection is established. Even if the expected reproduction number is fairly large, the initial infection may die out by chance, due to a positive probability that an infected cell will cause zero secondary infections. Previous studies have focussed on calculating the probability of viral extinction during the early stages of infection for different models. For example, Pearson *et al.* (2011) calculated the probability of viral extinction in models with either a geometric or Poisson distribution for the reproduction number.

In this chapter I will study stochastic, Markov chain versions of two simple deterministic viral infection models, with non-exponential infectious period distributions. I will derive the probability distributions for the burst size and reproduction number in these models. The first model considered is one of the simplest deterministic models for viral infection, consisting of a system of ODEs for the populations of target cells, infected cells, and virions, in which infected cells are assumed to release virus at a constant rate while they are productively infectious (Perelson (2002)). Pearson *et al.* (2011) studied the stochastic version of this model, with one infected cell compartment, but I will consider a more generalised model in which the infected cell state is split into multiple compartments, creating an Erlang-distributed instead of exponentially-distributed infectious period. Models often incorporate this multistage representation (Beauchemin *et al.* (2017); Liao *et al.* (2020); Yan *et al.* (2020)). It has been shown that distributions like the Erlang may be more appropriate than an exponential distribution to model the infectious period, and can lead to different estimates of the key infection parameters, such as the basic reproduction number (Holder & Beauchemin (2011)). The second model considered is one developed by Guedj *et al.* (2013) for Hepatitis C virus (HCV) infection, which includes an additional equation for the intracellular viral genome. This model accounts for the increase in intracellular genome level due to replication, and its depletion due to decay or release from the cell via budding in the form of progeny virions (Guedj *et al.* (2013)). In this model, the inclusion of these simple intracellular dynamics means that the viral production rate from infected cells is not constant but depends on the intracellular state of the cell. Deterministic models like these are very successful in describing viral infections when the populations of infected cells and virions

are large, and have been used to estimate values for viral kinetic parameters. In this chapter I use some such parameter values estimated in previous modelling studies for Ebola virus, Influenza, Human Immunodeficiency virus-1 (HIV-1), and HCV (Guedj *et al.* (2013); Liao *et al.* (2020); Yan *et al.* (2020); Yuan & Allen (2011)) to present numerical results for the burst size and reproduction number distributions of the models.

5.1 Model with constant viral production rate

Deterministic model

This section considers a model of viral dynamics described by the following ODEs (Bai *et al.* (2019); Liao *et al.* (2020); Yan *et al.* (2020); Yuan & Allen (2011)),

$$\begin{aligned}
\frac{dT}{dt} &= -\beta TV, \\
\frac{dE_1}{dt} &= \beta TV - \left(\frac{n_E}{\tau_E} + \nu\right) E_1, \\
\frac{dE_i}{dt} &= \frac{n_E}{\tau_E}(E_{i-1} - E_i) - \nu E_i, \qquad i = 2, ..., n_E, \\
\frac{dI_1}{dt} &= \frac{n_E}{\tau_E} E_{n_E} - \left(\frac{n_I}{\tau_I} + \nu\right) I_1, \\
\frac{dI_i}{dt} &= \frac{n_I}{\tau_I}(I_{i-1} - I_i) - \nu I_i, \qquad i = 2, ..., n_I, \\
\frac{dV}{dt} &= p\sum_{i=1}^{n_I} I_i - cV - \beta TV,
\end{aligned} \tag{5.1.1}$$

where the variable T is the number of uninfected target cells, $\sum_{i=1}^{n_E} E_i + \sum_{i=1}^{n_I} I_i$ is the number of infected cells, and V is the amount of free (i.e. extracellular) infectious virus. A diagram of the model is provided in Figure 5.1.

The model equations in Eq. (5.1.1) are based on those of Yan *et al.* (2020) and Liao *et al.* (2020), with a modification to represent removal of infected cells by the immune response (Bai *et al.* (2019); Yuan & Allen (2011)). This additional

mechanism is included by specifying that infected cells can be cleared with rate ν. The model without immune response can be recovered by setting $\nu = 0$, for example in order to represent *in vitro* dynamics (Liao *et al.* (2020); Yan *et al.* (2020)). Some models similar to the one in Eq . (5.1.1) also include dynamics of production and death of uninfected target cells, however these dynamics can be neglected when only considering a short timescale (Pearson *et al.* (2011)). Here, only the transition of uninfected target cells to an infected state is included, as is usually the case for models describing *in vitro* viral dynamics (Liao *et al.* (2020); Yan *et al.* (2020)), or dynamics of *in vivo* acute infections, where the viral load increases to a viral peak and then declines due to the depletion of target cells.

In the model, the uninfected target cells are infected by virus with rate β. The population of infected cells is partitioned into eclipse phase cells, $\sum_{i=1}^{n_E} E_i$, and infectious phase cells, $\sum_{i=1}^{n_I} I_i$. When target cells become infected, they enter the eclipse phase, which is a period after infection of the cell but before the cell has begun to release virus particles. Eventually the cell can transition into the infectious phase in which it begins to release virus. The model incorporates a multistage representation of the eclipse and infectious periods in order to represent an Erlang distribution for the time that a cell spends in each phase. The eclipse phase is split into n_E stages, with rate $\delta_E = \frac{n_E}{\tau_E}$ of moving to the next stage. When the cell exits the last state of the eclipse phase, it enters the infectious phase. Similarly, the infectious phase is divided into n_I stages, with rate $\delta_I = \frac{n_I}{\tau_I}$ of moving to the next stage. Thus, when $\nu = 0$, the eclipse and infectious periods follow Erlang distributions with means equal to τ_E and τ_I, and shape parameters n_E and n_I, respectively. When $\nu > 0$, these Erlang distributions compete with the possibility of cell death due to the immune response. That is, there is the possibility for the eclipse or infectious period to "complete early" because of the possibility of the cell dying due to the immune response, which occurs with rate ν. It is assumed that infected cells may be cleared by the immune system before beginning active viral production (Bai *et al.* (2019)). For example, eclipse phase cells have been found to express viral peptides, so $CD8^+$ T cells can recognise and kill them (Baral *et al.* (2019); Sacha *et al.* (2010)). In the model considered here, $CD8^+$ T cells are assumed to kill eclipse phase cells and infectious cells at the same rate, ν. However, it would be straight forward to adapt the results

obtained for the case of differing death rates between cells in the eclipse and infectious phases.

The infectious cells produce virus at a constant rate p per cell. Free virus is cleared with rate c. In the equation for V, the term corresponding to loss of virus due to infection is explicitly included here, however this is sometimes neglected in the literature because the term βTV is usually small compared to cV (Perelson & Nelson (1999)). Also, if the number of target cells T is assumed to stay fairly constant, both viral loss terms can be incorporated into one constant, $c' = c + \beta T$ (Canini & Perelson (2014)). Interestingly, the units for β differ between the equations for T and I, and the equation for V because βTV is in units of virions in the equation for V, whereas in the other two equations, βTV is in units of cells. However, since the model assumes that one virion infects one cell with rate βTV, the rate can be taken to be the same for all three equations.

Some proportion of the virus released from infected cells is usually non-infectious and is not able to infect new cells. However it will be assumed throughout this chapter that p denotes the production rate of infectious virus, so that the variable V represents the amount of extracellular infectious virus rather than total virus.

Stochastic model

I will consider a stochastic version of the model in Eq. (5.1.1) in order to find the probability distribution of the burst size and reproduction number for this model. In particular, the model can be formulated as a continuous-time Markov chain (CTMC) where $T(t)$, $V(t)$, $E_i(t)$, and $I_j(t)$ (for $i \in \{1, ..., n_E\}$ and $j \in \{1, ..., n_I\}$) are discrete random variables with values in the set of non-negative integers, for $t \in [0, \infty)$. The possible transitions considered in the stochastic model are shown in Table 5.1.

Average burst size and reproduction number

The burst size is defined as the total number of virions produced by an infected cell during its lifetime. If $\nu = 0$, then the average time that an infected cell

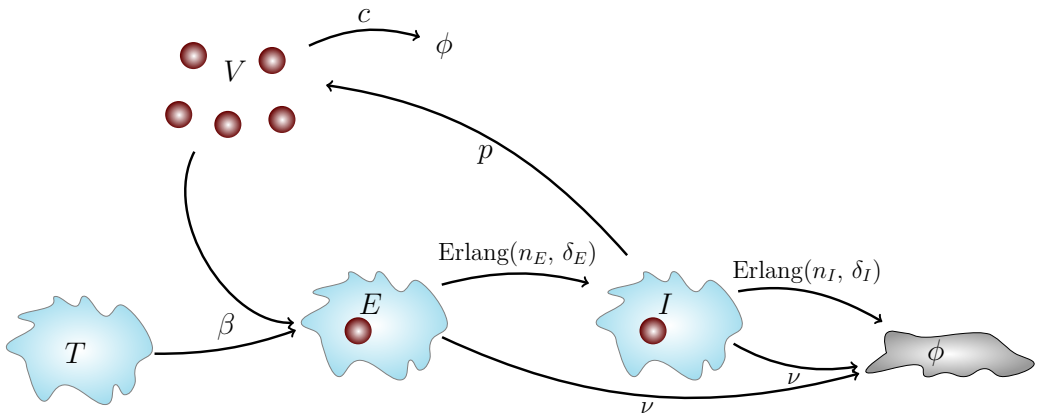

Figure 5.1: Diagram for the model with constant viral production rate. T represents a target cell, E represents an eclipse phase cell, I represents an infectious phase cell, and V represents free virus. ϕ denotes the empty set and represents clearance of infected cells or virus. The arrows show the possible transitions and their corresponding rates, or the distribution of the time taken for the event in the case of non-exponentially (Erlang) distributed transition times. Target cells are infected by virus with rate β and enter the eclipse phase. Eclipse phase cells can be cleared by the immune response with rate ν or they can eventually transition into the infectious phase and begin to release virus at rate p. Infectious cells can also be cleared by immune system cells with rate ν, or eventually suffer virus-induced cell death. Free virus is cleared with rate c.

Transition			Rate
$(V,\ T,\ E_1)$	$\longrightarrow$	$(V-1,\ T-1,\ E_1+1)$	βTV
$(E_i,\ E_{i+1})$	$\longrightarrow$	$(E_i-1,\ E_{i+1}+1)$	$\frac{n_E}{\tau_E}E_i,\ i\in\{1,...,n_E-1\}$
$(E_{n_E},\ I_1)$	$\longrightarrow$	$(E_{n_E}-1,\ I_1+1)$	$\frac{n_E}{\tau_E}E_{n_E}$
E_i	$\longrightarrow$	E_i-1	$\nu E_i,\ i\in\{1,...,n_E\}$
$(I_i,\ I_{i+1})$	$\longrightarrow$	$(I_i-1,\ I_{i+1}+1)$	$\frac{n_I}{\tau_I}I_i,\ i\in\{1,...,n_I-1\}$
I_i	$\longrightarrow$	I_i-1	$\nu I_i,\ i\in\{1,...,n_I-1\}$
I_{n_I}	$\longrightarrow$	$I_{n_I}-1$	$(\frac{n_I}{\tau_I}+\nu)I_{n_I}$
V	$\longrightarrow$	$V+1$	$p\sum_{i=1}^{n_I}I_i$
V	$\longrightarrow$	$V-1$	cV

Table 5.1: The transitions and their corresponding rates in the Markov chain version of the model in Eq. (5.1.1).

spends infectious is τ_I, and the rate of virion production is p. Thus the average burst size in the model with $\nu = 0$ is given by $p\tau_I$.

The expected value of the basic reproduction number can be calculated from Eq. (5.1.1) using the next generation matrix approach. It can also be derived from the Jacobian matrix of the system of ODEs, as a threshold quantity determining the stability of the infection-free steady state. For the model with $\nu = 0$, the mean basic reproduction number is given by,

$$\bar{R} = \frac{p\tau_I\beta T_0}{c+\beta T_0}, \tag{5.1.2}$$

where T_0 denotes the size of the target cell population in the infection-free steady state. It should be noted that the basic reproduction number can be defined differently if the infected cells and extracellular virus are instead assumed to be separate infected classes. In that case, the basic reproduction number is defined as the average number of new infectives produced per infective in any class per generation, and becomes the square root of the $\bar{R}$ defined in Eq. (5.1.2), since there are two infected classes in the infection cycle (Heffernan *et al.* (2005)).

The reproduction number considered here is sometimes instead called a type reproduction number because it gives the average number of infectives *of the same type* produced by a single infective of that type (Yuan & Allen (2011)).

Now let us consider the case when the removal of infected cells due to the immune response is included ($\nu > 0$). In this case, it is possible for cells to be killed by the immune system before passing through all n_I infectious stages, and hence infectious phase cells will progress through some (unknown) number of stages, $k \in \{1, ..., n_I\}$, before dying from stage k. In each infectious stage, the rate of death of the cell due to the immune response is ν, and the rate of progressing to the next stage is $\delta_I = n_I/\tau_I$. Thus, the time taken to exit each stage will be the minimum of two competing exponential random variables, $T_1 \sim \text{Exp}(\nu)$ and $T_2 \sim \text{Exp}(\delta_I)$. It is well known that the minimum of T_1 and T_2 is an exponential random variable with rate $\delta_I + \nu$, and this is independent of whether the cell is killed by the immune response or progresses to the next stage. That is, the rate of leaving a particular infectious stage will always be $\delta_I + \nu$, even if we condition on progression to the next stage. Thus, given that a cell dies from infectious stage k, the infected cell will pass through k infectious stages until it stops producing virus, and the time spent in each stage is exponentially distributed with rate $\delta_I + \nu$. So the distribution of time that the cell spends producing progeny virus is equivalent to the sum of k exponential distributions, each with rate $\delta_I + \nu$, which is an Erlang(k, $\delta_I + \nu$) distribution. Therefore, the overall distribution of the infectious period, T_I (i.e., the random variable representing the time since a cell starts releasing virions until it dies), will be a weighted sum of Erlang distributions with rate $\delta_I + \nu$ and shape parameters $k = 1, ..., n_I$, where n_I is the maximum number of stages that an infected cell can pass through before it dies. The weights of each of these distributions will be given by the probability that an infectious cell progresses to stage k and dies from stage k, which can be found as follows. For an infected cell in infectious stage $i \in \{1, ..., n_I - 1\}$, the exit rate of the cell out of this stage is $\delta_I + \nu$, and when the cell exits this stage it will either die with probability $\frac{\nu}{\delta_I+\nu}$, or move to the next infectious stage with probability $\frac{\delta_I}{\delta_I+\nu}$. Note that for the final infectious stage, n_I, the cell will always die when exiting this stage. Therefore the probability that an infectious cell progresses to

stage k and dies from stage k is given by

$$\begin{cases} \left(\frac{\delta_I}{\delta_I+\nu}\right)^{k-1}\left(\frac{\nu}{\delta_I+\nu}\right) & \text{for } k=1,...,n_I-1, \\ \left(\frac{\delta_I}{\delta_I+\nu}\right)^{n_I-1} & \text{for } k=n_I. \end{cases}$$

Thus, the density function of the infectious period is given by,

$$f_{T_I}(t) = \sum_{k=1}^{n_I-1}\left[\left(\frac{\delta_I}{\delta_I+\nu}\right)^{k-1}\left(\frac{\nu}{\delta_I+\nu}\right)\frac{(\delta_I+\nu)^k t^{k-1}e^{-(\delta_I+\nu)t}}{(k-1)!}\right] + \left(\frac{\delta_I}{\delta_I+\nu}\right)^{n_I-1}\frac{(\delta_I+\nu)^{n_I}t^{n_I-1}e^{-(\delta_I+\nu)t}}{(n_I-1)!}.$$

Then the mean infectious period is

$$\mathbb{E}[T_I] = \int_0^{+\infty} t f_{T_I}(t)\, dt = \frac{1}{\nu}\left(1-\left(\frac{n_I}{n_I+\tau_I\nu}\right)^{n_I}\right),$$

which tends to τ_I as $\nu \to 0$, by l'Hôpital's rule. Furthermore, virus particles will only be released from an infected cell if it survives all the stages of the eclipse phase and progresses to the infectious phase. For an infected cell in any given eclipse stage, the exit rate of the cell out of this stage is $(\frac{n_E}{\tau_E}+\nu)$, and when the cell exits this stage it will either die with probability $\frac{\tau_E\nu}{n_E+\tau_E\nu}$, or move to the next stage with probability $\frac{n_E}{n_E+\tau_E\nu}$. Hence the probability that the cell will survive all n_E eclipse stages and progress to the infectious phase is $\left(\frac{n_E}{n_E+\tau_E\nu}\right)^{n_E}$. Thus, the mean basic reproduction number when $\nu > 0$ is given by,

$$\bar{R} = \frac{p\beta T_0}{\nu(c+\beta T_0)}\left(\frac{n_E}{n_E+\tau_E\nu}\right)^{n_E}\left(1-\left(\frac{n_I}{n_I+\tau_I\nu}\right)^{n_I}\right).$$

5.1.1 Burst size probability distribution

Let B be the random variable representing the viral burst size, i.e., the total number of virions released from an infected cell during its lifetime. The burst size does not depend on the whole process shown in Figure 5.1 but only on the dynamics of viral production from a single infected cell. Therefore to find the distribution of B one can simply consider a discrete-time Markov chain for an individual infected cell and the amount of virus produced.

As mentioned previously, an infected cell will only produce virus if it survives all the stages of the eclipse phase and progresses to the infectious phase, which occurs with probability $\left(\frac{n_E}{n_E+\tau_E\nu}\right)^{n_E}$. On the other hand, the infected cell will die before becoming infectious with probability $1-\left(\frac{n_E}{n_E+\tau_E\nu}\right)^{n_E}$, in which case the burst size will certainly be zero. If the cell does enter the infectious phase, it will progress through some (unknown) number of stages, $k \in \{1, ..., n_I\}$, before dying from stage k, and as explained in the previous section, the probability that an infected cell progresses to stage k and dies from stage k is given by

$$\begin{cases} \left(\frac{\delta_I}{\delta_I+\nu}\right)^{k-1}\left(\frac{\nu}{\delta_I+\nu}\right) & \text{for } k = 1, ..., n_I - 1, \\ \left(\frac{\delta_I}{\delta_I+\nu}\right)^{n_I-1} & \text{for } k = n_I. \end{cases}$$

I will first consider the burst size distribution for an infectious cell, conditioned on the value of k, and then compute a weighted sum over the possible values of k.

Assume that an infectious cell progresses through exactly k stages and dies from stage k. Then consider the two dimensional discrete-time Markov chain, $\mathcal{X} = \{\boldsymbol{X}_n = (I_n, V_n) : n \in \{0, 1, 2, ...\}\}$, where $I_n \in \{1, ..., k, k+1\}$ denotes the infectious stage of the infected cell, and $V_n \in \mathbb{N} \cup \{0\}$ represents the number of virions produced by the infected cell by step n. Here the state $I = k + 1$ is an absorbing state representing the death of the cell, when no more virus can be produced. The initial state, when the cell first becomes infectious, is $(I_0, V_0) = (1, 0)$, since the infected cell will begin in stage 1 of the infectious period, and 0 infectious virions have been produced by the infected cell. During the time that it is infected, the cell is assumed to produce virus at a constant rate, p, so that the amount of virus produced follows a Poisson process, with an exponentially distributed time between the production of each new virion. The cell will also progress to each new stage of the infectious phase with rate $\delta_I + \nu$. Hence, one can think of the infected cell as progressing through a series of events, where each event is either the production of a virion with probability $p/(p + \delta_I + \nu)$, or progression to the next infectious stage with probability $(\delta_I+\nu)/(p+\delta_I+\nu)$. That

is, the Markov chain $\mathcal{X}$ can be defined by the one-step transition probabilities,

$$\begin{aligned} p_{(i,j),(m,n)} =& \mathbb{P}(\boldsymbol{X}_{n+1} = (m,n) \mid \boldsymbol{X}_n = (i,j)) \\ =& \begin{cases} \frac{\delta_I+\nu}{p+\delta_I+\nu} & \text{if } (m,n) = (i+1,j), \\ \frac{p}{p+\delta_I+\nu} & \text{if } (m,n) = (i,j+1), \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

for states $(i,j) \in \{1,...,k\} \times (\mathbb{N} \cup \{0\})$. The total number of progeny virions produced by the infected cell is the number of virions produced before the cell dies from stage k and stops producing virus, V_∞. This can be thought of as the number of "successes" before the k^{th} "failure", where a "success" is the production of a virion with probability $p/(p+\delta_I+\nu)$ and a "failure" is progression to the next infectious stage with probability $(\delta_I+\nu)/(p+\delta_I+\nu)$. Therefore, given that the cell starts in infectious stage 1 and dies from stage k, the burst size B follows a negative binomial distribution, with parameters k and $p/(p+\delta_I+\nu)$.

Once we sum over the possible values of k, the probability mass function (p.m.f.) for the burst size of a cell starting in infectious stage 1 is

$$\begin{aligned} &\mathbb{P}(B = b \mid \text{cell starts in stage } I_1) \\ &= \sum_{k=1}^{n_I} \mathbb{P}(B = b \mid \text{cell dies from stage } k)\mathbb{P}(\text{cell dies from stage } k) \\ &= \sum_{k=1}^{n_I-1} \left[\binom{k+b-1}{b} \left(\frac{p}{p+\delta_I+\nu} \right)^b \frac{\delta_I^{k-1}\nu}{(p+\delta_I+\nu)^k} \right] \\ &\quad + \binom{n_I+b-1}{b} \left(\frac{p}{p+\delta_I+\nu} \right)^b \frac{\delta_I^{n_I-1}(\delta_I+\nu)}{(p+\delta_I+\nu)^{n_I}}, \qquad b \in \{0,1,2,...\}. \end{aligned}$$

If the infectious period is assumed to be exponentially distributed ($n_I = 1$), then this is a geometric distribution.

Taking into account the probability of death during the eclipse phase, the overall p.m.f. for the burst size distribution of a cell that enters into the eclipse phase is

$$\mathbb{P}(B=0) = \left[1 - \left(\frac{n_E}{n_E+\tau_E\nu} \right)^{n_E}\right] + \left(\frac{n_E}{n_E+\tau_E\nu} \right)^{n_E} \mathbb{P}(B = 0 \mid \text{cell starts in stage } I_1),$$

and for $b > 0$,

$$\mathbb{P}(B=b) = \left(\frac{n_E}{n_E+\tau_E\nu} \right)^{n_E} \mathbb{P}(B = b \mid \text{cell starts in stage } I_1). \tag{5.1.3}$$

For $\nu = 0$, the p.m.f. for the number of virions released from an infected cell is

$$\mathbb{P}(B = b) = \binom{n_I + b - 1}{b} \left(\frac{p}{p + \delta_I}\right)^b \left(\frac{\delta_I}{p + \delta_I}\right)^{n_I}, \qquad b \in \{0, 1, 2, \dots\}, \quad (5.1.4)$$

with

$$\mathbb{E}[B] = \frac{pn_I}{\delta_I} = p\tau_I,$$

and

$$\mathrm{Var}[B] = \mathbb{E}[B]\left(1 + \frac{\mathbb{E}[B]}{n_I}\right) = \frac{pn_I}{\delta_I} + \frac{p^2 n_I}{\delta_I^2}.$$

Weighted production rates

Although the model in Eq. (5.1.1) with constant viral production rate has often been used to describe viral dynamics (with $\nu = 0$), this model could be extended slightly by allowing the viral production rate of infected cells to be weighted according to their stage in the infectious phase. That is, the equation for V would become,

$$\frac{dV}{dt} = p\sum_{i=1}^{n_I} \gamma_i I_i - cV - \beta TV,$$

for some weights $\gamma_i \geq 0$, $i \in \{1, \dots, n_I\}$. For this extended model, and in the particular case $\nu = 0$, the burst size, B, would be defined as a sum of independent random variables, $X_1 + X_2 + \dots + X_{n_I}$, where X_i is geometrically distributed with mean $p\gamma_i/\delta_I$.

For example, if one wished to model the situation in which the viral production rate increases linearly each time the infected cell moves to a new stage, one could define,

$$\gamma_i = \frac{2i}{n_I + 1}.$$

In this case, the average burst size would remain as

$$\mathbb{E}[B] = \frac{p\sum_{i=1}^{n_I} \gamma_i}{\delta_I} = \frac{pn_I}{\delta_I},$$

whereas the variance of the burst size would increase to

$$\mathrm{Var}[B] = \frac{p\sum_{i=1}^{n_I} \gamma_i(\delta_I + p\gamma_i)}{\delta_I^2} = \frac{pn_I}{\delta_I} + \frac{2p^2 n_I(2n_I + 1)}{3(n_I + 1)\delta_I^2}.$$

Of course the weights γ_i could be defined in many different ways depending on what is known about the particular virus being modelled.

5.1.2 Reproduction number probability distribution

Let R be the random variable representing the cellular-level basic reproduction number, i.e., the number of secondary infections produced by a single infected cell in an otherwise susceptible cell population. The p.m.f. of R may be written in terms of the burst size probabilities as

$$\mathbb{P}(R = r) = \sum_{b=0}^{\infty} \mathbb{P}(R = r \mid B = b)\mathbb{P}(B = b).$$

Hence, I will find the probabilities $\mathbb{P}(R = r \mid B = b)$, defining the probability distribution for the number of secondary infections produced, given that the initial infected cell releases $B = b$ virions throughout its lifetime. Then I will use these together with the probability distribution for B in Eq. (5.1.3) to find the probability distribution for R.

The only possible outcomes for each released virion are that it eventually infects another cell, or it eventually decays and does not infect another cell, as shown in Figure 5.2. Usually, calculations of the basic reproduction number assume that the number of infectious virus particles is low enough so that the number of uninfected target cells is not strongly affected by the transformation to infected cells. Thus the number of target cells is usually assumed to remain constant (Czuppon *et al.* (2021)). However, if the number of available target cells at the start of infection is fairly small and the infection process is much faster than the dynamics of production and death of uninfected cells, it can become important to consider the transition of target cells to infected cells. This is because newly infected cells are produced with rate βTV, which depends on the number of available target cells, T. Therefore the depletion of uninfected target cells will affect the ability for virus particles produced from an initially infected cell to infect new cells. Hence the number of secondary infections caused by a single infected cell will be affected by, (i) competition between virus produced by the initial infected cell for available target cells, and (ii) competition for target cells from virions being produced by secondary infections. Here I address the former

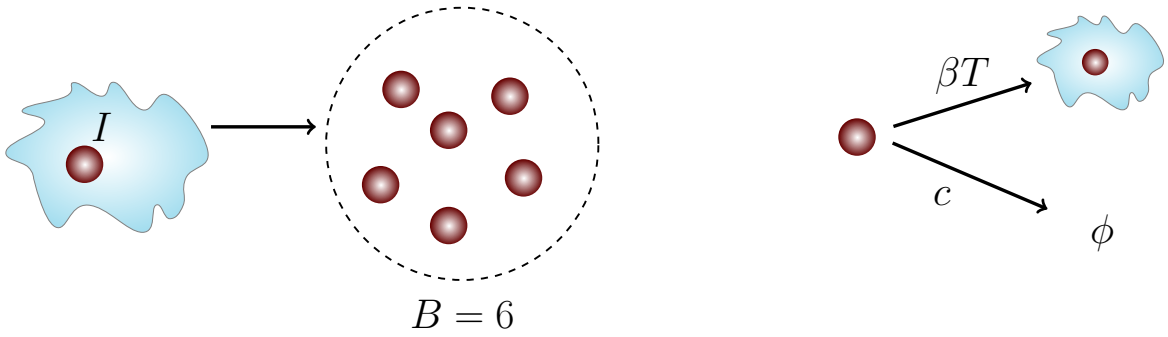

Figure 5.2: I represents an infected cell that has produced $B = 6$ virions over its lifetime. The free virions that have been released can then infect a new cell with rate βT, or are cleared with rate c. The per virion infection rate, βT, will change each time a secondary infection is produced, since this will reduce the number of available target cells. If this reduction in the number of uninfected target cells is ignored, then the number of uninfected target cells will remain constant and so will the infection rate. Then each virion will have probability $\theta = \beta T_0/(c+\beta T_0)$ of infecting a new cell. In this case, all the virions are independent of each other, so the number of secondary infections is the sum of independent and identical Bernoulli random variables, and hence follows a binomial distribution.

source of competition, while ignoring the second. That is, I consider depletion of the number of target cells due to secondary infections caused by virus particles released from the initially infected cell, but do not consider depletion of target cells due to virus particles released from subsequently infected cells.

Limited number of target cells (Case 1)

The probability of a given virion successfully infecting a cell before it decays depends on the number of available target cells for it to infect. Thus, the virions will not be independent, but their probability of successfully infecting a cell will depend on how many target cells are still available after previous virions have infected some of the target cells. Let us assume that there are T_0 target cells in the population to begin with, and $B = b$ infectious virions are released while the initial infected cell is in the infectious phase. Since we are only interested in the number of secondary infections produced, and not the timings of these infections, we can imagine that at discrete time steps, one of these virions is chosen uniformly at random to either decay or infect an available target cell. A Markov chain can be set up to calculate the probability distribution of the number of secondary infections. Consider the discrete-time Markov chain, $\mathcal{Y} = \{Y_n : n \in \{0, 1, 2, ..., b\}\}$, where Y_n denotes the number of secondary infections that have occurred after n of the infectious extracellular virions have either decayed or infected an available target cell. The initial state of the process will be $Y_0 = 0$ and the state space of $\mathcal{Y}$ will be given by $\mathcal{S}_{\mathcal{Y}} = \{0, 1, 2, ..., M\}$, where $M = \min\{b, T_0\}$ is the maximum number of secondary infections that can be produced.

The Markov chain $\mathcal{Y}$, depicted in Figure 5.3, can be defined by the following one-step transition probabilities, for states $i, j \in \mathcal{S}_{\mathcal{Y}}$. For $i \neq M$,

$$p_{i,j} = \mathbb{P}(Y_{n+1} = j \mid Y_n = i) = \begin{cases} \frac{\beta(T_0 - i)}{\beta(T_0 - i) + c} & \text{if } j = i + 1, \\ \frac{c}{\beta(T_0 - i) + c} & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

$$p_{M,j} = \begin{cases} 1 & \text{if } j = M, \\ 0 & \text{otherwise.} \end{cases}$$

The probability that r secondary infections will occur, given that the cell released b infectious virions, will be given by $\mathbb{P}(R = r \mid B = b) = \mathbb{P}(Y_b = r \mid$

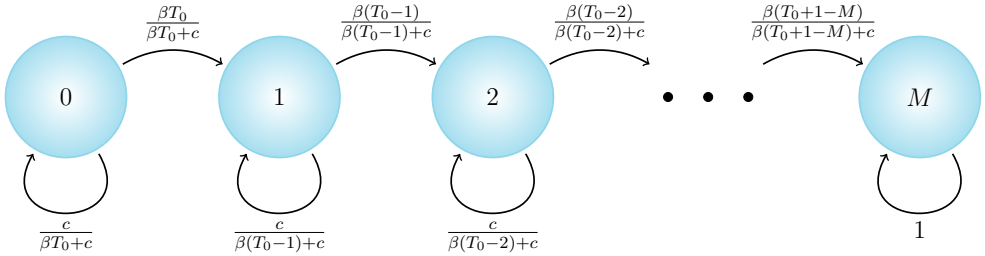

Figure 5.3: A depiction of the discrete-time Markov chain, $\mathcal{Y}$, used to model the number of secondary infections produced by a single infected cell that has released $B = b$ virions during its lifetime. State i represents i secondary infections. The process begins in state $Y_0 = 0$ and takes b steps (one for each virion). The final state of the process after b steps is the number of secondary infections produced. The maximum number of secondary infections that can be produced is $M = \min\{b, T_0\}$, where T_0 is the initial number of available uninfected target cells.

$Y_0 = 0$), which is the probability that the process $\mathcal{Y}$ is in state r after the b infectious extracellular virions have either decayed or infected new cells. For ease of notation, let $p_i(n) = \mathbb{P}(Y_n = i \mid Y_0 = 0)$.

Using first step analysis, these probabilities can be calculated recursively, as follows. We have $p_0(0) = 1$. Then for $n \geq 1$,

$$p_i(n) = \begin{cases} p_0(n-1)p_{0,0}, & \text{if } i = 0, \\ p_{i-1}(n-1)p_{i-1,i} + p_i(n-1)p_{i,i}, & \text{if } 0 < i \leq M. \end{cases}$$

Thus, substituting the one-step transition probabilities from above,

$$p_i(n) = \begin{cases} \left(\frac{c}{\beta T_0+c}\right)^n & \text{if } i = 0, \\ p_{i-1}(n-1)\frac{\beta(T_0+1-i)}{\beta(T_0+1-i)+c} + p_i(n-1)\frac{c}{\beta(T_0-i)+c} & \text{if } 0 < i \leq \min(n, T_0), \\ 0 & \text{if } i > \min(n, T_0). \end{cases}$$

These can be solved sequentially to obtain $p_r(b)$ for all required values of r and b. Then we have,

$$\mathbb{P}\Big(R = r\Big) = \sum_{b=0}^{\infty} p_r(b)\mathbb{P}(B = b). \tag{5.1.5}$$

Number of target cells remains constant (Case 2)

If the number of target cells is small and infected cells are not quickly replenished by new target cells, it is important to take into account the reduction of uninfected

target cells as they become infected, as considered above. However, if the number of uninfected target cells is assumed to be very large, things can be simplified by neglecting the loss of uninfected target cells due to infection, so that the population of uninfected target cells is assumed to be constant. In this case, all virions are independent of each other, in the sense that they will have the same probability of successfully infecting a new cell, regardless of the fate of other virions. For each infectious virion released from the infected cell during the infectious phase, the probability that it will go on to infect another cell is

$$\theta = \frac{\beta T_0}{c + \beta T_0},$$

where β is the rate at which extracellular virus infects target cells, T_0 is a constant number of target cells in the population, and c is the rate of degradation of extracellular virus. Therefore the number of secondary infections will be the sum of b independent and identical Bernoulli random variables. Hence, $R \mid B = b$ follows a binomial distribution with parameters (b, θ). Then we have,

$$\mathbb{P}\Big(R = r\Big) = \sum_{b=r}^{\infty} \binom{b}{r} \theta^r (1-\theta)^{b-r} \mathbb{P}(B = b).$$

Substituting in the p.m.f of the burst size distribution from Eq. (5.1.3), it can be found that the p.m.f of the reproduction number has the same form as the burst size p.m.f., but with the parameter p replaced by θp, which is equivalent to assuming that secondary infections are produced from infected cells at a constant rate, θp. This makes sense since each virion produced from an infected cell is assumed to infect a new cell with probability θ, and when focusing only on the number of secondary infections produced, the timescale of these events can be ignored.

If $\nu = 0$ and the burst size, B, follows a negative binomial distribution as in Eq. (5.1.4), then the p.m.f. of the reproduction number can be simplified to

$$\mathbb{P}(R = r) = \binom{n_I + r - 1}{r} \left(\frac{\theta p}{\theta p + \delta_I}\right)^r \left(\frac{\delta_I}{\theta p + \delta_I}\right)^{n_I}, \tag{5.1.6}$$

so the number of secondary infections also follows a negative binomial distribution.

This has a correspondence to the analysis of the epidemic reproduction number at the population level, since if infected individuals are assumed to infect others at a constant rate while they are infectious, the same distribution can be obtained for the epidemic setting. Lloyd-Smith *et al.* (2005) investigated the variation in infectiousness for various disease outbreaks, and compared the diseases through their dispersion parameter, k. Assuming a negative binomial distribution for the number of new infections caused by an infectious individual, the dispersion parameter, k, is the shape parameter of this distribution, and represents the degree of transmission heterogeneity. If k is much larger than one, there is a fairly low degree of dispersion, whereas a low value of k corresponds to a higher level of dispersion in the distribution and suggests that a small number of infected individuals, known as "superspreaders", may trigger many secondary infections. The value of the dispersion parameter is particularly important for the early dynamics of an epidemic when there are only a few infected individuals. For example Lloyd-Smith *et al.* (2005) investigated the probability of stochastic extinction for an outbreak beginning with one infected individual, for different values of k. Smaller values of k were shown to increase the probability of stochastic extinction, but small values of k also mean that an epidemic can quickly take off due to the possibility of superspreading events, leading to infrequent but explosive epidemics. In the case studied here, for the cellular reproduction number, the dispersion parameter of the negative binomial distribution in Eq. (5.1.6) is given by n_I, and will be important during the early dynamics of a viral infection. Due to the way the model is defined, n_I must always be an integer greater than or equal to one, since it is the shape parameter of the Erlang distribution for the infectious period. As $n_I \to \infty$, the infectious period distribution approaches a Dirac delta distribution, i.e. a fixed infectious period (when $\nu = 0$). Therefore, a value of $n_I \gg 1$ means that there will be little variation in the number of virions released from each cell, and in the reproduction number between cells, when $\nu = 0$. An Erlang distribution is chosen in order for the model to be Markovian. However if the infectious period was assumed to be Gamma distributed with shape parameter smaller than 1, then there would be much more variation in the infectious period and consequently in the distribution of virions released

from infected cells. This would decrease the chance of successful establishment of disease, but would also introduce the potential for superspreader cells.

5.1.3 Bursting versus budding

I have so far focused on a model in which virus particles are produced and released from infectious cells gradually via budding. However for some viruses, progeny virions accumulate inside the host cell and are released in a burst upon cell lysis. Yuan & Allen (2011) have previously considered a model of the bursting strategy and made the assumption that if the death of an infected cell is virus-induced then virions will be released in a burst, but if the infected cell is cleared due to the action of the immune response, no virions will be released. They also assumed that each infected cell that bursts will have the same burst size.

To include into the ODEs in Eq. (5.1.1) the assumption that when an infected cell bursts (on leaving the final infectious stage with rate $\frac{n_I}{\tau_I}$) it releases exactly N virions, this would result in changing the equation for V to,

$$\frac{dV}{dt} = N\frac{n_I}{\tau_I}I_{n_I} - cV - \beta TV. \tag{5.1.7}$$

Eq. (5.1.7) also has the interpretation that cells release virions at a constant rate Nn_I/τ_I during their time in infectious stage n_I, so it is not possible to model a burst at a discrete time using ODEs, since the dynamics are averaged across time. However, for larger values of n_I or smaller values of τ_I, one can achieve something almost like a burst at a discrete time, related to a very short stay in infectious stage n_I. On the other hand, in the Markov chain version of the model it is possible to accurately describe bursts at discrete times, and this corresponds to changing the event,

$$V \longrightarrow V + 1,$$

which happens with rate $p\sum_{i=1}^{n_I} I_i$ in the budding model, to the event,

$$V \longrightarrow V + N,$$

which happens with rate $\frac{n_I}{\tau_I}I_{n_I}$.

Yuan & Allen (2011) assumed that in the bursting case, if an infected cell bursts, its burst size would always be $N = p\tau_I$ and would not depend on the time

that the cell bursts. For example, this might be the case if the accumulation of virions inside the cell and the burst of the cell are coupled in such a way that the burst of the cell is caused by the number of intracellular virions reaching a particular level of N virions. On the other hand, one could consider a case in which the accumulation of virions and the burst of the cell are assumed to be independent processes. In this case, cells that burst earlier may not produce as many virus particles as those that survive longer. This would introduce variability in the burst size, related to the timings of the bursts.

As well as viruses, the bursting strategy is also used by some intracellular bacteria. For example, the stochastic model in Chapter 3 of this thesis was used to predict the number of *B. anthracis* bacteria released from a host phagocyte. In this intracellular anthrax model, the bursting process depended on the intracellular dynamics. An alternative approach is that by Carruthers *et al.* (2018), where the burst size distribution was predicted for the bacterium *Francisella tularensis* using a model in which the rate that the burst occurs was assumed to be independent of the number of bacteria within the cell. This was done by considering a stochastic model for the intracellular accumulation of bacteria, so that the number of bacteria released from an infected cell when it bursts depends on the stochastic dynamics of this intracellular process, as well as on the time that the burst takes place. A similar method to the one used by Carruthers *et al.* (2018) for *Francisella tularensis* could be applied to viruses that are released by bursting in order to estimate the burst size distribution. For example, instead of a fixed burst size for all bursting cells, one could assume that in the bursting case, virus particles are produced inside the cell at a constant rate p during the "infectious" period, but are not released unless the cell bursts. Note that under this assumption, the eclipse phase would represent a period in which the cell begins to synthesise the viral proteins and replicate the viral genome, but progeny virions have not yet started to be assembled.

To include this into the ODE model, let us introduce variables P_i, $i \in \{1, ..., n_I\}$ to represent the total number of virions inside all cells in infectious stage i, via

the following equations,

$$\frac{dP_1}{dt} = pI_1 - \left(\frac{n_I}{\tau_I} + \nu\right) P_1,$$

$$\frac{dP_i}{dt} = \frac{n_I}{\tau_I} P_{i-1} + pI_i - \left(\frac{n_I}{\tau_I} + \nu\right) P_i, \qquad i = 2, ..., n_I,$$

$$\frac{dV}{dt} = \frac{n_I}{\tau_I} P_{n_I} - cV - \beta TV.$$

Each P_i increases due to viral production at a rate proportional to the number of cells in infectious stage i and due to cells entering stage i, and decreases when cells exit stage i. The rate that a cell in stage i will transition to stage $i+1$ is $\frac{n_I}{\tau_I} I_i$ and, in the deterministic model, the P_i virions are equally shared between all cells in stage i. Therefore, the number of virions that will transition to the next stage when an event like this happens will be P_i/I_i. Thus, the overall rate that virions transition from compartment P_i to P_{i+1} is

$$\frac{n_I}{\tau_I} I_i \frac{P_i}{I_i} = \frac{n_I}{\tau_I} P_i.$$

Similarly, νP_i is the rate at which virions in stage i are lost when infectious cells are killed due to the action of the immune response.

In the corresponding Markov process, each infected cell is independent, with its own number of intracellular virions that increases at a constant rate, until the infected cell dies. If the infected cell reaches stage n_I and the death of the cell is virus-induced (with rate $\frac{n_I}{\tau_I}$), then the virions present in the infected cell at its time of death will be released in a burst into the extracellular environment. On the other hand, if the infected cell is cleared due to the action of the immune response (with rate ν), it is assumed that no virions will be released.

When $\nu = 0$, corresponding to the model without the action of the immune response, the distribution of the burst size is equal for the two strategies of budding and bursting, since the number of virus particles produced inside the cell has the same distribution in each case, but they are either released from the cell gradually in the case of budding, or remain in the cell until the infected cell dies in the case of bursting. However, when the action of the immune response is included by modelling the clearing of infected cells by killer cells or cytotoxic

lymphocytes (CTLs), the two strategies produce different distributions for the burst size. This is because in the budding strategy some virions may bud from the infected cell before the cell is detected and destroyed by CTLs. Whereas in the bursting strategy, if a CTL detects and destroys an infected cell before it bursts, then all intracellular virions are also destroyed in the process and the burst size is zero. Hence, in the case of the bursting strategy, virions will only be released from the infected cell if the cell is not cleared by CTLs from any of the eclipse phase stages or infectious phase stages, and exits the final infectious stage by bursting. Given that a burst does occur, the burst size will be negative binomially distributed with parameters n_I and $\frac{p}{p+\delta_I+\nu}$, as shown in Section 5.1.1. Thus the probability of a burst size of zero is given by

$$\mathbb{P}(B=0) = 1 - \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E}\left(\frac{\delta_I}{\delta_I+\nu}\right)^{n_I}\left[1-\left(\frac{\delta_I+\nu}{p+\delta_I+\nu}\right)^{n_I}\right],$$

and for $b > 0$,

$$\mathbb{P}(B=b) = \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E}\binom{n_I+b-1}{b}\left(\frac{p}{p+\delta_I+\nu}\right)^{b}\left(\frac{\delta_I}{p+\delta_I+\nu}\right)^{n_I}.$$

In the case of a constant number of uninfected target cells where each virion has probability θ to infect a new cell, the reproduction number distribution can be obtained by taking the p.m.f. of the burst size distribution and replacing p by θp, so that the probability of zero secondary infections is given by

$$\mathbb{P}(R=0) = 1 - \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E}\left(\frac{\delta_I}{\delta_I+\nu}\right)^{n_I}\left[1-\left(\frac{\delta_I+\nu}{\theta p+\delta_I+\nu}\right)^{n_I}\right],$$

and for $r > 0$,

$$\mathbb{P}(R=r) = \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E}\binom{n_I+b-1}{b}\left(\frac{\theta p}{\theta p+\delta_I+\nu}\right)^{b}\left(\frac{\delta_I}{\theta p+\delta_I+\nu}\right)^{n_I}. \tag{5.1.8}$$

This gives an expected reproduction number of,

$$\bar{R} = \tau_I \theta p \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E}\left(\frac{n_I}{n_I+\tau_I\nu}\right)^{n_I+1}. \tag{5.1.9}$$

5.1.4 Probability of viral extinction

It is useful to be able to calculate the probability of viral extinction during the early stages of infection. When the number of uninfected target cells is assumed to stay constant, the initial dynamics of the models considered here can be described by branching processes in which each infected cell produces a number of new infected cells according to the reproduction number distribution. In such a branching process, the probability of extinction starting with one infected cell is equal to the smallest fixed point of the probability generating function (p.g.f.) of the offspring distribution (Csernica (2015)). If the branching process is counting the number of infected cells at each generation, then the offspring distribution will be the reproduction number distribution.

In the case of budding, the p.g.f. of the reproduction number distribution (when the number of uninfected target cells remains constant) is

$$\begin{aligned}
\pi_{bud}(s) &= \sum_{r=0}^{\infty} s^r \mathbb{P}(R=r), \\
&= \mathbb{P}(R=0) + \sum_{r=1}^{\infty} s^r \mathbb{P}(R=r), \\
&= 1 - \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E} \left[1 - \sum_{k=1}^{n_I-1} \frac{\delta_I^{k-1}\nu}{z^k} - \frac{\delta_I^{n_I-1}(\delta_I+\nu)}{z^{n_I}}\right] \\
&\quad + \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E} \sum_{r=1}^{\infty} s^r \left[\sum_{k=1}^{n_I-1} \left[\binom{k+r-1}{r} \left(\frac{\theta p}{z}\right)^r \frac{\delta_I^{k-1}\nu}{z^k}\right]\right. \\
&\quad \left. + \binom{n_I+r-1}{r} \left(\frac{\theta p}{z}\right)^r \frac{\delta_I^{n_I-1}(\delta_I+\nu)}{z^{n_I}}\right], \qquad (5.1.10) \\
&= 1 - \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E} \left[1 - \sum_{k=1}^{n_I-1} \left[\sum_{r=0}^{\infty} s^r \binom{k+r-1}{r} \left(\frac{\theta p}{z}\right)^r \frac{\delta_I^{k-1}\nu}{z^k}\right]\right. \\
&\quad \left. - \sum_{r=0}^{\infty} s^r \binom{n_I+r-1}{r} \left(\frac{\theta p}{z}\right)^r \frac{\delta_I^{n_I-1}(\delta_I+\nu)}{z^{n_I}}\right], \\
&= 1 - \left(\frac{\delta_E}{\delta_E+\nu}\right)^{n_E} \left[1 - \sum_{k=1}^{n_I-1} \frac{\tau_I \nu n_I^{k-1}}{(n_I+\tau_I\nu+(1-s)\tau_I\theta p)^k}\right. \\
&\quad \left. - \frac{n_I^{n_I-1}(n_I+\tau_I\nu)}{(n_I+\tau_I\nu+(1-s)\tau_I\theta p)^{n_I}}\right],
\end{aligned}$$

where $z = \theta p + \delta_I + \nu$. In the last step I have used $\delta_I = n_I/\tau_I$ to remove δ_I from the formula, and have also used the fact that the p.g.f. of a negative binomial random variable, X, with p.m.f

$$\mathbb{P}(X = x) = \binom{n + x - 1}{x} q^x (1 - q)^n$$

is

$$\pi(s) = \left(\frac{1 - q}{1 - qs}\right)^n.$$

In the case of bursting, considered in Section 5.1.3, the p.g.f. of the reproduction number distribution in Eq. (5.1.8) is

$$\pi_{burst}(s) = 1 - \left(\frac{\delta_E}{\delta_E + \nu}\right)^{n_E} \left(\frac{n_I}{n_I + \tau_I \nu}\right)^{n_I} \left[1 - \left(\frac{n_I + \tau_I \nu}{n_I + \tau_I \nu + (1 - s)\tau_I \theta p}\right)^{n_I}\right]. \tag{5.1.11}$$

When $n_I = 1$, corresponding to an exponentially distributed time until infected cell death, the fixed points of these functions can be found explicitly. For $n_I = 1$, the solutions to $\pi_{bud}(s) = s$ are $s_1 = 1$ and

$$s_2 = 1 - \left(\frac{\delta_E}{\delta_E + \nu}\right)^{n_E} \left(1 - \frac{1}{\bar{R}}\right),$$

where $\bar{R} = \left(\frac{\delta_E}{\delta_E + \nu}\right)^{n_E} \frac{\theta p}{\delta_I + \nu}$ is the average reproduction number. The smallest fixed point of the p.g.f. gives the probability of viral extinction starting with one infected cell, so in this case extinction will be certain if $s_2 \geq 1$ and will occur with probability s_2 if $s_2 < 1$. Similarly, the solutions to $\pi_{burst}(s) = s$, when $n_I = 1$, are $s_1 = 1$ and

$$s_2 = 1 - \left(\frac{\delta_E}{\delta_E + \nu}\right)^{n_E} \frac{\delta_I}{\delta_I + \nu} \left(1 - \frac{1}{\bar{R}}\right),$$

where $\bar{R} = \left(\frac{\delta_E}{\delta_E + \nu}\right)^{n_E} \frac{\theta p \delta_I}{(\delta_I + \nu)^2}$ is the average reproduction number. In each case, the probability of extinction starting with one infected cell is $s_1 = 1$ if $\bar{R} \leq 1$, and s_2 if $\bar{R} > 1$.

For $n_I > 1$, the smallest fixed points of the p.g.f.s can be found numerically, giving the probability of viral extinction starting with one infected cell. Since each initial infected cell is independent, the probability of extinction starting with one cell can easily be used to find the probability of extinction starting with multiple cells.

Parameter	Units	Value
p	virions $(\text{cell} \cdot \text{ h})^{-1}$	$m \cdot 10^{1.12}$
τ_I	h	83.2
n_I	-	14
τ_E	h	30.5
n_E	-	13
β	$(\text{cell} \cdot \text{virion} \cdot \text{h})^{-1}$	$10^{-6.48}/m$
c	$(\text{virion} \cdot \text{h})^{-1}$	6.14×10^{-2}

Table 5.2: Parameter values that have been used to obtain some numerical results applicable to Ebola virus (EBOV). These are the modes of posterior samples that Liao *et al.* (2020) obtained using data of EBOV infection *in vitro*. The parameters p and β are transformed into units of virions using a ratio of m infectious virus particles per TCID_{50}.

5.1.5 Numerical results

Example: Ebola virus

Liao *et al.* (2020) fitted a mathematical model similar to Eq. (5.1.1) with $\nu = 0$ to data of Ebola virus (EBOV) infection *in vitro*. In this section I present some numerical results for the burst size and reproduction number distributions, using parameter estimates which are the modes of posterior samples obtained by Liao *et al.* (2020). In order to measure the infectious viral load, the study by Liao *et al.* (2020) used the 50% Tissue Culture Infectious Dose (TCID_{50}) assay, in which varying virus dilutions are added to cell populations with the same number of cells and incubated until a cytopathic effect can be seen. The TCID_{50} value represents the amount of virus dilution required to induce cytopathic effects in 50% of wells containing the inoculated cell culture. Therefore the data measures infectious virus in units of TCID_{50}, where one unit of TCID_{50} is the dose of virus for which there is approximately a 50% chance to induce cytopathic effects. It is likely that one TCID_{50} unit corresponds to multiple infectious virus particles. Hence, the parameter values need to be transformed into units of virions rather than TCID_{50} in order to be used in the calculations of the reproduction number distribution. For the values presented in Table 5.2, $m \geq 1$ is defined to be the

number of infectious virus particles per unit of TCID_{50}, and the transformed parameters are left in terms of m. It will later be shown how the value of m affects the predictions for the reproduction number distribution.

Burst size

B is the random variable representing the amount of virus released by an infected cell during its lifetime. When an *in vitro* situation is considered (corresponding to $\nu = 0$ in the model), B follows a negative binomial distribution with mean $p\tau_I$ and shape parameter n_I, given in Eq. (5.1.4). The only parameter involved in this distribution that is affected by the value of m is the viral production rate, p, which is proportional to m. So as m increases, the distribution of the burst size remains a negative binomial distribution with shape parameter n_I, but with a mean proportional to m. For the parameters in Table 5.2, with $\nu = 0$, the expected amount of infectious virus produced over an infected cell's lifetime is

$$\mathbb{E}[B] = p\tau_I \approx 1097m \text{ virions.}$$

The p.m.f. for B is shown in Figure 5.4, for $m = 1$ and $\nu = 0$. Figure 5.5 shows how the cumulative distribution function for the burst size changes for different values of the immune killing rate ν.

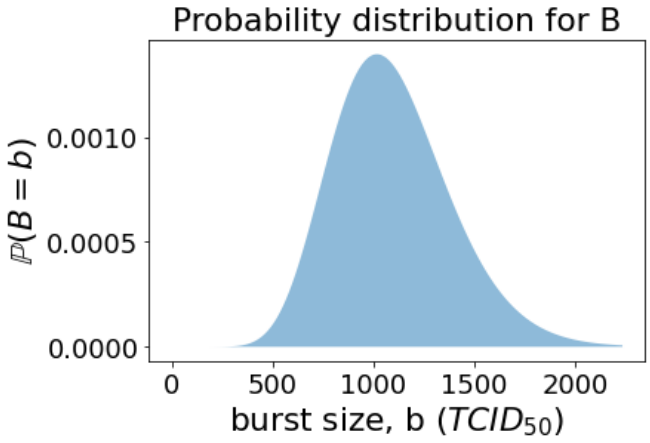

Figure 5.4: Probability distribution for B, the random variable for the amount of virus released by a single infected cell during its lifetime. The parameter values in Table 5.2 have been used to calculate this distribution, with $m = 1$ and $\nu = 0$.

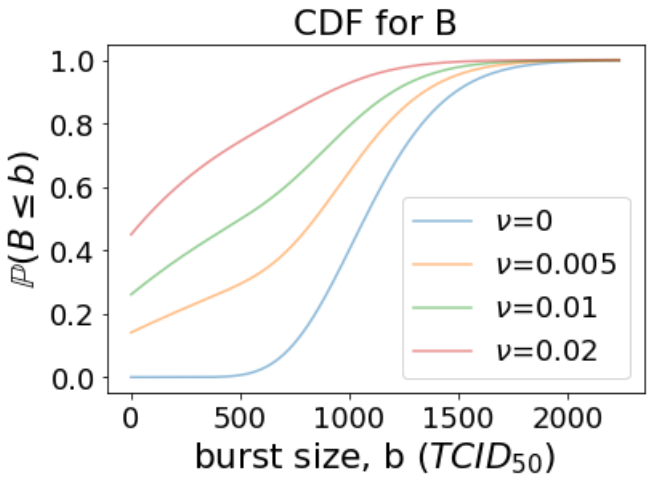

Figure 5.5: Cumulative distribution functions for B, for different values of ν. Other parameters are set to their values from Table 5.2, with $m = 1$.

Reproduction number

R is the random variable representing the number of secondary infections caused by a single infected cell in an otherwise susceptible population. Section 5.1.2 described how to calculate the distribution of the reproduction number, R. If the number of target cells is assumed to be a limiting factor for the number of secondary infections that can be produced, the reproduction number distribution can be calculated using Eq. (5.1.5) (Case 1). Alternatively, when the number of target cells is not limited because the population is very large or is constantly replenished (Case 2), the p.m.f. of the reproduction number has the same form as the burst size p.m.f. given in Eq. (5.1.3), but with p substituted by θp. When $\nu = 0$, this simplifies to the negative binomial distribution in Eq. (5.1.6). For the latter case, in which the target cells are assumed to be in excess, the mean number of secondary infections produced will be

$$\bar{R} = \mathbb{E}[R] = \theta\mathbb{E}[B] = \frac{\beta T_0}{\beta T_0 + c}\mathbb{E}[B] = \frac{\beta T_0}{\beta T_0 + c} p\mathbb{E}[T_I], \tag{5.1.12}$$

where T_I is the random variable for the amount of time that an infected cell will spend infectious. Note that $T_I = 0$ if the infected cell is killed while in the eclipse phase. When the value of $\bar{R}$ is much smaller than T_0, and only a small proportion

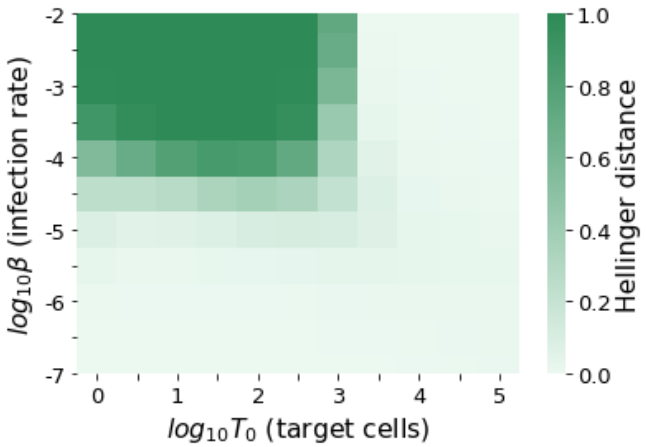

Figure 5.6: Heatmap showing changes in the Hellinger distance between the two distributions of the reproduction number, R, calculated using methods Case 1 and Case 2, for different values of β and T_0. All other parameter values were fixed to the values in Table 5.2, with $m = 1$ and $\nu = 0$.

of the available cells are likely to become infected, the probability distributions obtained for R are similar when using either the Case 1 or Case 2 method. Note that $T_0 \gg \bar{R}$ will be true if $T_0 \gg \mathbb{E}[B]$, but if $T_0 < \mathbb{E}[B]$ it is still possible to have $T_0 \gg \bar{R}$ if

$$\beta \ll \frac{c}{\mathbb{E}[B] - T_0}. \tag{5.1.13}$$

The smaller the difference is between the number of available cells and the expected number to become infected (i.e. $T_0 - \bar{R}$), the more different the distributions obtained from the two methods become. The heatmap in Figure 5.6 illustrates how the difference between the two distributions changes as the values of β and T_0 are varied, but with all other parameters fixed to the values in Table 5.2, with $m = 1$ and $\nu = 0$. The difference between the distributions is quantified by calculating the Hellinger distance, giving a value between 0 and 1, where 0 indicates that the two distributions are identical. In general, for larger values of the infection rate, β, and smaller values of T_0, the distributions become more different, since these parameter values will mean that the virus is likely to infect a higher proportion of the available target cells.

Figure 5.7 shows some examples of the probability distributions for R calcu-

lated with the Case 1 and Case 2 methods, for different values of the infection rate, β, and number of available cells, T_0. The Case 1 method assumes that there is a population of T_0 susceptible cells to begin with, but each secondary infection reduces this number by one, meaning that it is not possible for more than T_0 secondary infections to occur. On the other hand, Case 2 assumes that there is a constant population of T_0 susceptible cells that does not decrease as cells become infected. The two plots in the top row of Figure 5.7 have parameter values of β and T_0, such that $\theta = 0.016$ and the average reproduction number is $\bar{R} = 17.6$ (for Case 2, from Eq. (5.1.12)). However for the plot on the left, $T_0 = 10 < \bar{R}$, making the distribution for R very different, depending on whether it is calculated using the method for Case 1 or Case 2. In this case, the Hellinger distance between the two distributions is 0.81. When $T_0 = 10^3 \gg \bar{R}$, the two distributions become very similar, with a much smaller Hellinger distance of 0.01. For the plots on the bottom row of Figure 5.7, the values of β and T_0 give $\theta = 0.62$ and $\bar{R} = 680$. For the plot on the left, where $T_0 = 10^3$ is only slightly larger than $\bar{R}$, there is clear difference between the two distributions, and the Hellinger distance is 0.33. When T_0 is increased to 10^5 in the plot on the right, this difference becomes much smaller and the Hellinger distance is 0.002.

The value for β that was estimated by Liao *et al.* (2020) was $10^{-6.48}$ $(\text{cell} \cdot \text{TCID}_{50} \cdot \text{h})^{-1}$. To transform this into units of virions, I have set $\beta = 10^{-6.48}/m$, where $m \geq 1$ is the number of infectious virus particles in a single unit of TCID_{50}. For this value of β, with any $m \geq 1$, the average reproduction number will be much smaller than T_0, for any $T_0 \in \mathbb{N}$, due to Eq. (5.1.13) with c, p, and τ_I set to their values in Table 5.2. Therefore the distribution of R will be very similar when obtained using either method. Nevertheless, the distribution will strongly depend on the assumption of the value for T_0, which is the number of target cells available to be infected. In the *in vitro* experiment described by Liao *et al.* (2020), there was an initial population of $T_0 = 10^5$ susceptible cells/ml. However in an *in vivo* situation, the number of target cells at the site of initial infection may be smaller than this, and this will affect the overall probability of an infection becoming established in the host. The reproduction number distribution is also very sensitive to the rate ν at which the immune system clears infected cells *in vivo*. The distribution for R is shown in Figure 5.8 for a few different values

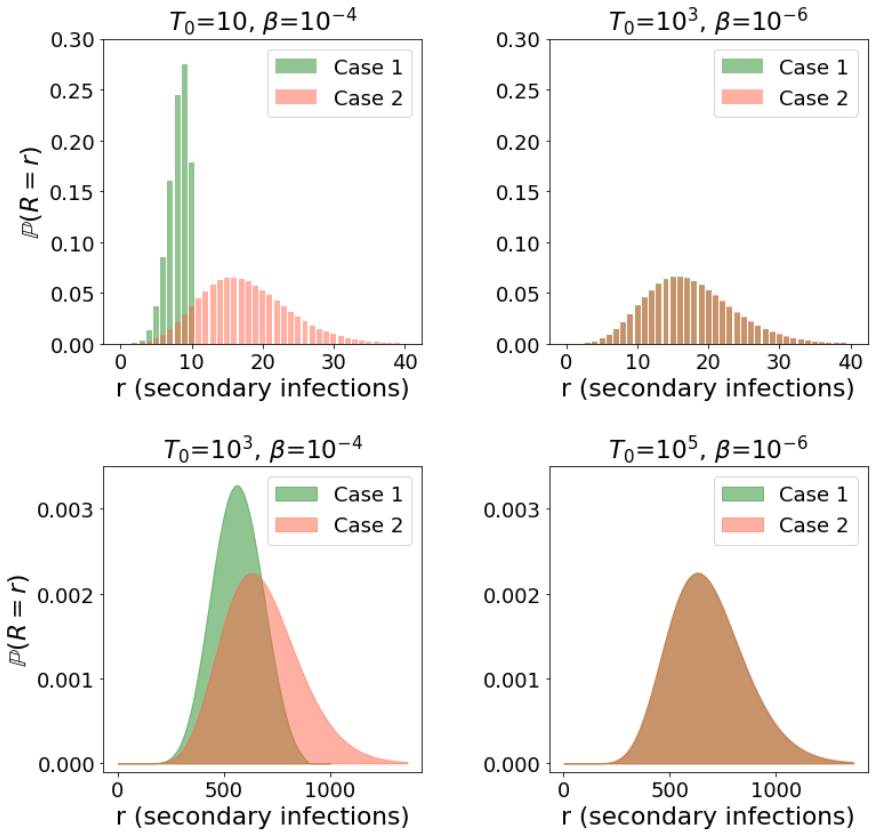

Figure 5.7: Histograms for the probability distribution of the reproduction number, for different values of β and T_0. For each pair of parameter values, the distributions obtained from methods of Case 1 (number of target cells decreases as they become infected) and Case 2 (number of target cells stays constant) are shown. All other parameter values are fixed to the values in Table 5.2, with $m = 1$ and $\nu = 0$.

of ν and T_0. The distributions shown have been calculated according to the assumption that the number of uninfected target cells remains constant and thus infectious cells are assumed to produce new infections at a constant rate θp. The Hellinger distances between these distributions and the ones obtained using the method in Eq. (5.1.5) are on the order of 10^{-3}.

The value of the mean basic reproduction number calculated by Liao *et al.* (2020) differs from the one calculated here. This is because their model did not consider the loss of virus due to infection of cells, so that the $\bar{R}$ of their model has a slightly different formula. For the $\bar{R}$ defined in Eq. (5.1.12), as $T_0 \to \infty$, we have $\bar{R} \to p\mathbb{E}[T_I]$, which is the average number of virions released from an infected cell. However for models that neglect the term corresponding to loss of virus due to infection, the mean basic reproduction number is given by

$$\bar{R} = \frac{\beta T_0}{c} p\mathbb{E}[T_I], \tag{5.1.14}$$

which becomes infinite as $T_0 \to \infty$. This means that there is no finite limit on the expected number of secondary infections produced by an infected cell, so that as the initial target cell population becomes larger, the expected number of secondary infections will eventually become larger than the expected number of virions released from the infected cell. This is not really consistent because each virion can only infect at most one cell.

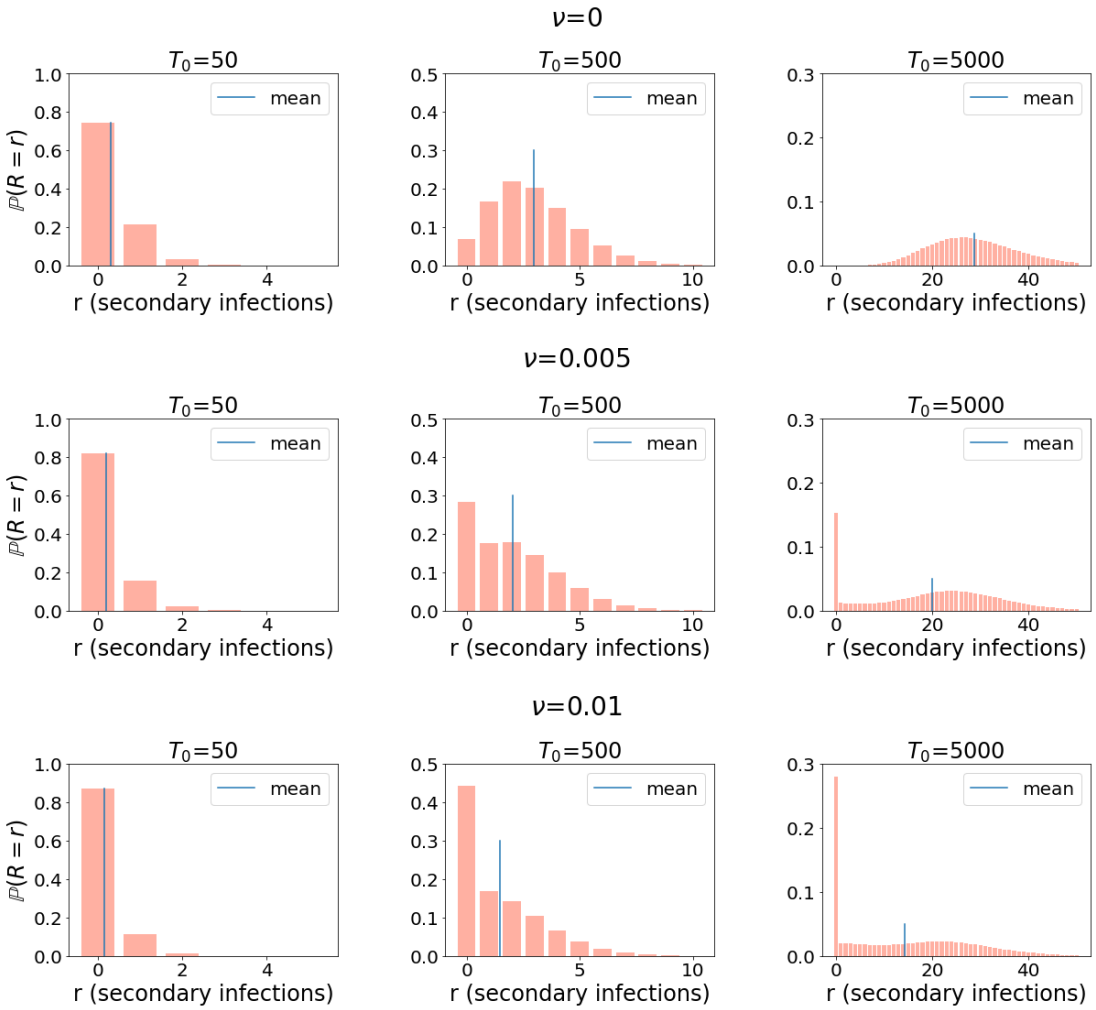

Figure 5.8: Probability distributions for the reproduction number, for different values of ν and T_0. Each row corresponds to a different value of ν and each column to a different value of T_0. The three values of ν used are 0, 0.005, and 0.01, from top to bottom row, respectively. The three values of T_0 used are 50, 500, and 5000, corresponding to values of θ of 3×10^{-4}, 3×10^{-3}, and 3×10^{-2}, respectively. The parameter values in Table 5.2 have been used to calculate these distributions (using the method of Case 2), with $m = 1$.

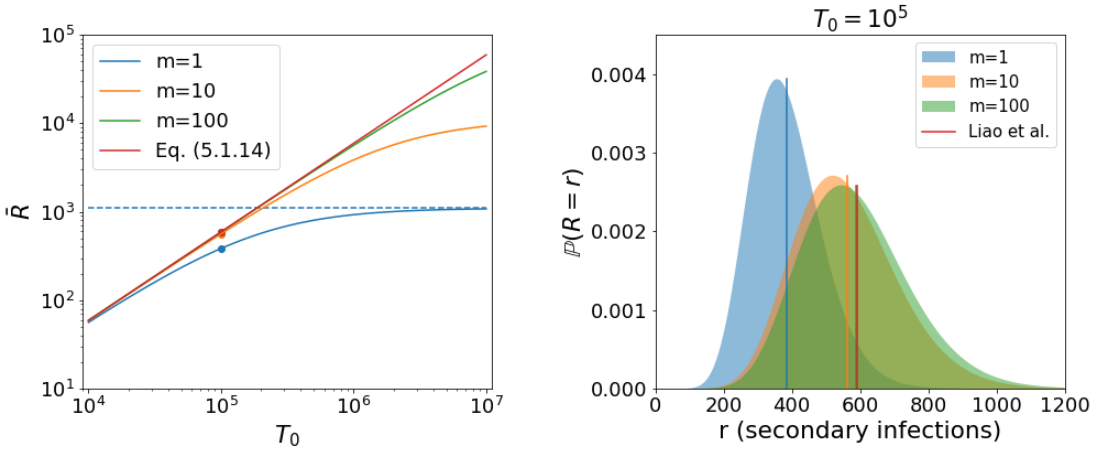

Figure 5.9: Left: Plot to show how the value of the mean basic reproduction number, $\bar{R}$, changes depending on the size of the target cell population, T_0, for different values of m. The solid blue, orange, and green lines correspond to the $\bar{R}$ defined in Eq. (5.1.2), for the model considering loss of virus due to infection of cells. The red line corresponds to the $\bar{R}$ defined in Eq. (5.1.14), for the model that neglects this process in the equations, which is independent of m. The dashed blue line indicates the expected number of virions released by an infected cell, when $m = 1$ and $\nu = 0$. The dots highlight the values of the mean basic reproduction numbers when $T_0 = 10^5$, which are also indicated as vertical lines on the right histograms. **Right:** Histograms for the negative binomial distribution of the reproduction number, for different values of m, when $T_0 = 10^5$ and $\nu = 0$. The parameter values in Table 5.2 have been used to obtain these plots.

Figure 5.8 shows that the reproduction number distribution can depend strongly on the number of available target cells. In addition to this, the reproduction number distribution will also be affected by the value of m, which is defined here to be the number of infectious virus particles in a single unit of TCID_{50}. As m increases, the infection rate per virion, denoted by β, will decrease, but the production rate of virions from the infected cell, p, will increase, meaning that the product βp will remain unchanged. Thus, the mean basic reproduction number in Eq. (5.1.14), for the model neglecting loss of virions due to infection, does not depend on the value of m. However, the mean basic reproduction number defined in Eq. (5.1.12) will increase with m, and will in fact converge to the mean basic reproduction number in Eq. (5.1.14) as $m \to \infty$. The value of $\bar{R}$ from Eq.

(5.1.12) as a function of T_0 is shown in the left plot of Figure 5.9 for a few values of m, and is compared to the $\bar{R}$ formula from Eq. (5.1.14), when $\nu = 0$. The right of Figure 5.9 shows the whole reproduction number distribution for the same values of m, when $T_0 = 10^5$. These plots demonstrate that the average reproduction number is very sensitive to m for small values of m, whereas when m is large (for example larger than 100) the results will become much less sensitive to this parameter and the average reproduction number can be fairly well estimated by using Eq. (5.1.14). In reality, it is challenging to estimate the value of m, but it is likely to be quite high. For example Yan *et al.* (2020) estimated this value to be between around 10 and 300, indicating that it may be reasonable to use the average reproduction number from Eq. (5.1.14).

Example: Influenza

In a recent study, Yan *et al.* (2020) fitted a model to *in vitro* data and compared estimates of growth rate, reproduction number, and generation time, for six influenza A strains. In this section I use estimated parameter values from Yan *et al.* (2020) for two of these strains, to identify the distributions of the number of virions released from a single infected cell, and the number of secondary infections produced. The parameter values used are presented in Table 5.3, for the strain A/Canada/RV733/2003 (seasonal H1N1), and A/Mexico/INDRE4487/2009 (pandemic H1N1). Infectious virus was again measured in units of TCID_{50}, but values for the conversion between amount of TCID_{50} and number of infectious virus particles were estimated by Yan *et al.* (2020) during the process of fitting the model to the data. Therefore I have used these estimated values to transform the necessary parameter values into units of virions rather than TCID_{50}.

B is the random variable representing the amount of virus released by an infected cell during its lifetime. The probability distribution for B for the two strains is shown in Figure 5.10, for the case where $\nu = 0$.

The distribution for the number of secondary infections, R, has been calculated according to the negative binomial distribution in Eq. (5.1.6), for different values of the initial population of susceptible cells, T_0, as shown in Figure 5.11.

Parameter	Units	sH1N1	pH1N1
p	virions $(\text{cell} \cdot \text{ h})^{-1}$	1082.57	119.88
τ_I	h	7.40	41.32
n_I	-	10	10
β	$(\text{cell} \cdot \text{virion} \cdot \text{h})^{-1}$	7.81×10^{-10}	2.41×10^{-8}
c	$(\text{virion} \cdot \text{h})^{-1}$	5.15×10^{-2}	4.69×10^{-2}

Table 5.3: Parameter values that have been used to obtain some numerical results applicable to influenza. These are estimates that Yan *et al.* (2020) obtained using data of *in vitro* influenza infection.

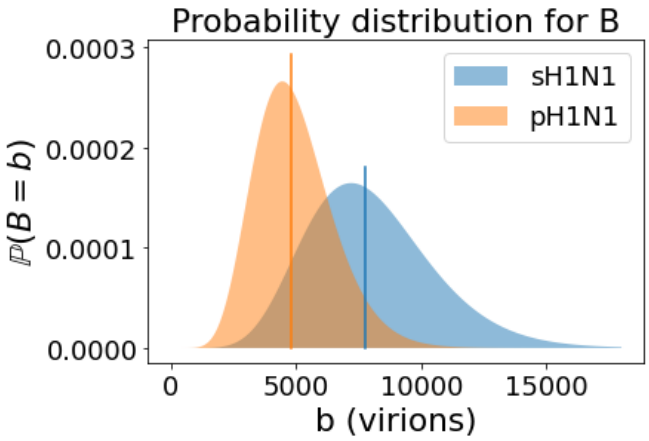

Figure 5.10: Probability distributions for B, the random variable for the number of virions released by a single infected cell during the infectious phase. The medians of each distribution are indicated by the vertical lines. The parameter values in Table 5.3 are used for the two strains, with $\nu = 0$.

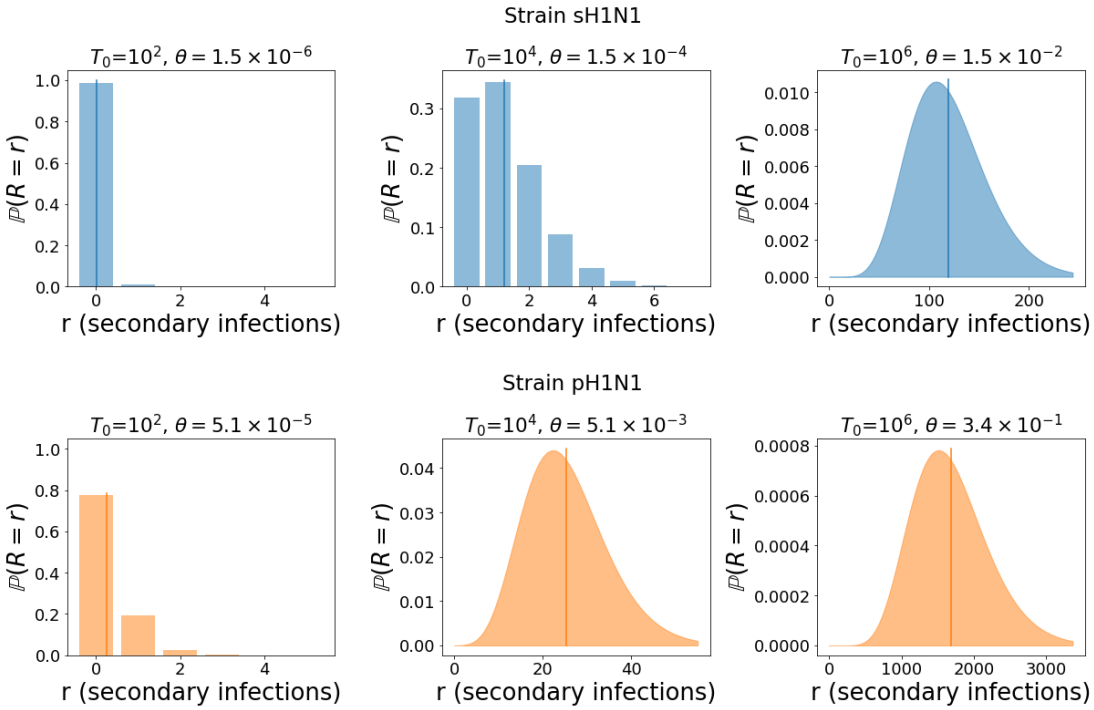

Figure 5.11: Probability distributions for R, the random variable for the number of secondary infected cells due to an initial cell in an otherwise susceptible population. Three different values of T_0 are used: 10^2, 10^4, and 10^6. The means of each distribution are indicated by the vertical lines. The parameter values in Table 5.3 are used for the two strains, with $\nu = 0$.

In the *in vitro* experiment described by Yan *et al.* (2020), there was an initial population of $T_0 = 10^6$ susceptible cells, however it is unknown how many cells would initially be available to the virus in an *in vivo* situation. In each column of Figure 5.11, the value of T_0 is set equal for the two strains, but the resulting value of $\theta = \frac{\beta T_0}{c+\beta T_0}$ differs because the values of β and c are different for each strain.

The expected value of virus released from an infected cell is much larger for the seasonal strain than the pandemic strain. However, as can be seen in the plots of the distribution for the reproduction number, the pandemic strain always has a higher mean basic reproduction number due to a higher infectivity rate and lower clearance rate.

Parameter	Units	Value
p	virions $(\text{cell} \cdot \text{day})^{-1}$	100
τ_I	days	1.25
n_E	-	0
ν	$(\text{cell} \cdot \text{day})^{-1}$	1.6
β	$(\text{cell} \cdot \text{virion} \cdot \text{day})^{-1}$	2×10^{-6}
c	$(\text{virion} \cdot \text{day})^{-1}$	7.3
T_0	cells	10^6

Table 5.4: Parameter values that have been used to obtain some numerical results applicable to HIV. These parameter values are the same as ones used by Yuan & Allen (2011), apart from the value of ν, which has been reduced.

Example: HIV

Yuan & Allen (2011) have considered a similar model to Eq. (5.1.1) but without an eclipse phase and with only one infectious stage. They analysed this model using parameters applicable to Human Immunodeficiency Virus-1 (HIV-1) and studied the difference in the probability of viral extinction when comparing a model of the budding strategy with a geometrically distributed burst size, and the bursting strategy with a fixed burst size. They showed that the bursting strategy was more successful for viral invasion when a model without immune response was considered, but that the more successful strategy switched to budding when the immune response was included in the model. In each case, the average burst size and reproduction number in the budding case were equal to their values in the bursting case, indicating that the distributions of the burst size and reproduction number are important in determining the probability of viral extinction; the mean alone is not enough.

In this section I use the parameter values in Table 5.4 to study the probability of viral extinction for the model of bursting considered in Section 5.1.3. I focus on the effect of changing the number of infectious phase stages in the model, given by the parameter n_I.

Yuan & Allen (2011) assumed a constant burst size of $N = p\tau_I$ virions from bursting cells, whereas the model considered here allows variation in the burst

size. It assumes that virus particles are produced inside the cell at a constant rate p and are released when the cell bursts. Furthermore, it assumes that the timing of the burst is independent of the number of intracellular virions. When the immune response is not included in the model ($\nu = 0$), the average burst size in the model considered here is the same as that of Yuan & Allen (2011). However, in the model considered here, the average number of virions released from bursting cells changes when the immune response is included in the model. The reason for this is that once the immune response is included in the model, a competition is introduced between the two mechanisms of death of infected cells, so that the cells that do burst will do so more quickly on average than they would have if the immune response was not included. This is because cells that are destined to burst quickly are more likely to burst before they are killed by the immune response, but those that would burst later are more likely to be cleared by immune system cells before they have a chance to burst. Cells that burst earlier will not have produced as many virus particles on average as those that survive longer. Hence when the immune response is included in the model, the average number of virions released from bursting cells is reduced, reflecting the reduced average time until these cells burst. Due to this, the parameter values considered by Yuan & Allen (2011) result in average burst sizes of less than one when used with these model assumptions. This means that the probability of viral extinction is certain, which is not a relevant case for the results I would like to illustrate. For the purpose of this illustration, I have therefore reduced the value of ν from 10 to 1.6. The chosen value of $\nu = 1.6$ means that the immune system clearance of infected cells occurs on average two times faster than virus-induced cell death.

As shown in Section 5.1.3, the reproduction number distribution of the bursting model is given by a zero-inflated negative binomial distribution. This is because if an infected cell does burst and release virions, then the number of secondary infections caused by the cell follows a negative binomial distribution with shape parameter n_I and probability $\frac{\theta p}{\theta p+\delta_I+\nu}$. However, the probability of zero secondary infections is inflated, since if the initial infected cell is killed by immune system cells before it bursts (either in the eclipse phase or infectious phase), then zero virions will be released and zero secondary infections can be

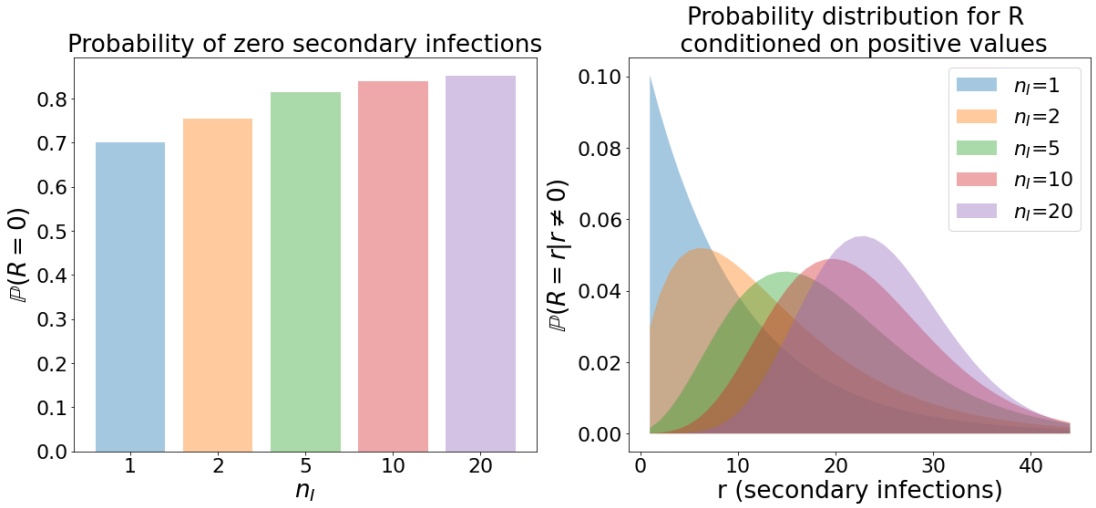

Figure 5.12: Reproduction number distributions for the bursting model, for different numbers of infectious phase stages, n_I. For each value of n_I, the mean of the Erlang-distributed time until cell burst is kept fixed to the value of τ_I in Table 5.4. All other parameters are also set to their values from Table 5.4. The probabilities of zero secondary infections for each n_I are shown on the left, and the distributions conditioned on positive values of the reproduction number are shown on the right.

produced. Figure 5.12 shows the reproduction number distribution for different values of n_I, when the parameters in Table 5.4 are used in the bursting model. Note that as n_I is varied, $\delta_I = n_I/\tau_I$ also changes, since it depends on n_I. As n_I increases, the initial infected cell must survive more infectious stages in order for a burst to occur, so there is a higher chance that it will be killed before it bursts, leading to a higher probability of zero secondary infections produced. However, if the infected cell does burst, it will release a higher number of virions on average when the value of n_I is higher. Therefore, when conditioned on positive numbers of secondary infections, the reproduction number distribution moves to the right with increasing n_I.

Figure 5.13 presents some interesting results about the probability of viral extinction, which is found by numerically calculating the smallest fixed point of the p.g.f in Eq. (5.1.11). The number of infectious stages, n_I, is varied, but with the average of the Erlang-distributed time until cell burst kept constant. The probability that a given extracellular virion goes on to infect a new cell rather than

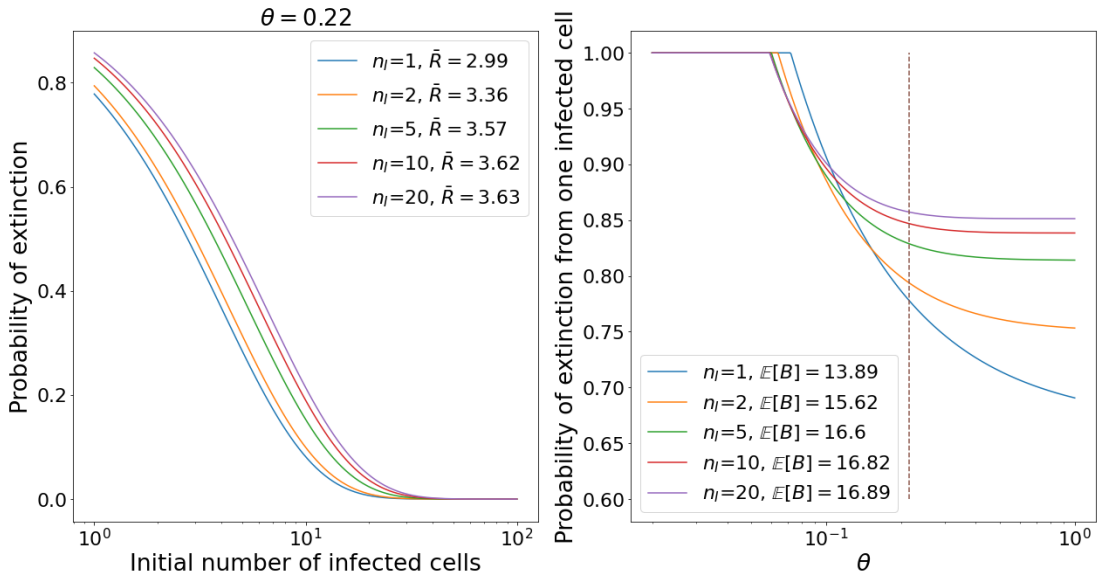

Figure 5.13: Left: Plot to show how the probability of viral extinction depends on the initial number of infected cells, when virus is assumed to be released in a burst only upon viral-induced cell death. The curves are shown for different numbers of infectious phase stages, n_I. For each value of n_I, the mean of the Erlang-distributed time until cell burst is kept fixed to the value of τ_I in Table 5.4. All other parameters are also set to their values from Table 5.4. **Right:** Plot to show how the probability of viral extinction starting from one infected cell changes as a function of θ, for different values of n_I. The value of θ used for the left plot is indicated by the dashed line.

be cleared is given by θ and this is also varied in Figure 5.13. For the parameter values in Table 5.4 and the values of n_I considered, the expected reproduction number in the bursting case, given in Eq. (5.1.9), increases with the value of n_I. For the model without immune response ($\nu = 0$), the chance of viral extinction reduces as the number of infectious stages increases. This is also the case when the immune response is included in the model of the budding strategy. However, for large enough θ, the opposite is true when the bursting strategy is considered. That is, for some values of θ, the probability of extinction is an increasing function of n_I. This is surprising, considering the expected reproduction number increases with n_I, and gives further evidence that focusing only on the mean reproduction number can be very misleading.

The plot on the left of Figure 5.13 shows how the probability of extinction

depends on the initial number infected cells, for the fixed value of $\theta \approx 0.22$ which comes from the parameter values in Table 5.4. For this value of θ, it can be seen that the probability of extinction is an increasing function of n_I, even though the average reproduction number $\bar{R}$ grows with n_I. The plot on the right of Figure 5.13 shows how the probability of extinction starting from one infected cell changes as a function of θ, for different values of n_I. When θ is very small, extinction is certain for all values of n_I because the average reproduction numbers are less than 1. When θ grows large enough to increase the average reproduction number above 1, initially there is a larger probability of extinction for smaller n_I, but as θ grows, the ordering changes and eventually the probability of extinction is largest for the largest value of n_I.

5.2 Model with age-dependent viral production rate

Many models of viral dynamics assume a constant budding rate of virus from infected cells, for example the model considered in Section 5.1. However some models of viral dynamics have instead used an age-dependent viral production rate, since it is realistic to assume that the rate of production and release of new virus is not constant but depends on the length of time the cell has been infected and the internal state of the cell (Heldt *et al.* (2013); Nelson *et al.* (2004); Quintela *et al.* (2018)). This section focuses on an example of such a model by Guedj *et al.* (2013), which includes basic intracellular dynamics.

Macroscopic model

Guedj *et al.* (2013) have modelled hepatitis C virus (HCV) infection with a deterministic multi-scale model in order to study the potential effects of direct-acting antiviral agents (DAAs) on intracellular viral RNA (vRNA) production, degradation, and secretion of progeny virus into the extracellular environment. The model was originally presented as a set of PDEs, but it has been shown by Kitagawa *et al.* (2018) that the model can be transformed into a set of ODEs. The model by Guedj *et al.* (2013) assumes an exponentially-distributed infected cell

lifetime. However, as mentioned previously, other distributions may be more appropriate. In fact, for HCV, some models have actually neglected infected cell death altogether in the time frame of their study, since HCV is thought to be non-cytolytic (Ribeiro *et al.* (2012)). Since cells infected with HCV are relatively long lived, it seems unrealistic to assume that a significant fraction of cells would die almost immediately after infection, which is an assumption of the exponential distribution. Hence, I will generalise this model by assuming that the time that a cell spends in the infectious phase has an Erlang(n_I, δ_I) distribution. The exponential distribution is a particular case, with $n_I = 1$, allowing an easy comparison with larger values of n_I. In the case of an Erlang-distributed infectious period, the model can be written as follows.

$$\begin{aligned}
\frac{dT}{dt} &= s - \lambda T - \beta TV, \\
\frac{dI_1}{dt} &= \beta TV - \delta_I I_1, \\
\frac{dI_i}{dt} &= \delta_I (I_{i-1} - I_i), \qquad\qquad i = 2, ..., n_I, \\
\frac{dP_1}{dt} &= \beta TV + \alpha I_1 - (\mu + \rho + \delta_I) P_1, \\
\frac{dP_i}{dt} &= \delta_I P_{i-1} + \alpha I_i - (\mu + \rho + \delta_I) P_i, \qquad i = 2, ..., n_I, \\
\frac{dV}{dt} &= \rho \sum_{i=1}^{n_I} P_i - cV - \beta TV.
\end{aligned} \tag{5.2.1}$$

In this model, infected cells must pass through n_I stages before cell death occurs, and the rate of moving between each stage is equal. The variable I_i denotes the number of infected cells in stage i, and the variable P_i denotes the total amount of intracellular positive HCV RNA in all infected cells in stage i. The intracellular RNA is assumed to be produced at a constant rate α, and is exported as virions and degraded with per capita rates ρ and μ, respectively.

Microscopic model of a single infected cell

I will now consider the dynamics of a single infected cell from a stochastic point of view, based on the assumptions of the intracellular dynamics from the model in Eq. (5.2.1). In particular, a continuous-time Markov chain, $\mathcal{Z} = \{Z(t) : t \geq 0\}$, can be used to describe the viral dynamics inside a single infected cell, where $Z(t)$ represents the amount of vRNA (copies of the RNA genome) inside a cell at time t after it becomes infected. It is assumed that $Z(t)$ only counts the vRNA copies that are available to be assembled into virions and released from the cell. When the process is in state $Z(t) = 0$, the cell is still infected and able to produce vRNA copies from some background level of replication complexes that are not counted by this Markov chain. As shown in Figure 5.14, the intracellular viral load of a single cell in the infectious phase is modelled by a birth-and-death process, where the birth event represents synthesis of a new copy of intracellular positive vRNA, and the death event represents either the decay of an intracellular vRNA copy, or the release of a single copy of vRNA from the cell, in the form of a virus particle. Following the assumptions of the intracellular virus dynamics in the multi-scale model for HCV by Guedj *et al.* (2013), vRNA is produced within infected cells with constant rate α (day^{-1}), degraded with rate μ (vRNA $\cdot$ day)$^{-1}$, and exported as virions with rate ρ (vRNA $\cdot$ day)$^{-1}$. After an Erlang-distributed time, the process will enter state ϕ, which is an absorbing state representing that viral production no longer occurs from the infected cell, for example if the cell has died due to the action of the immune system. Up until this point, the process describing the amount of intracellular vRNA and the process of cell death are assumed to be independent of each other. When the cell dies, the remaining intracellular vRNA does not get released as virions, since it is assumed that these vRNA copies have not yet been packaged into virus particles. Therefore, the total number of virions released (the burst size) is just as a result of the budding that has occurred throughout the infectious period with rate ρ.

Some models have included multiple viral RNA species (Quintela *et al.* (2018)) and viral proteins (Heldt *et al.* (2012)), so that more complicated feedback mechanisms can be considered. For a model that includes only one viral RNA species, one might assume that as the amount of vRNA inside an infected cell increases,

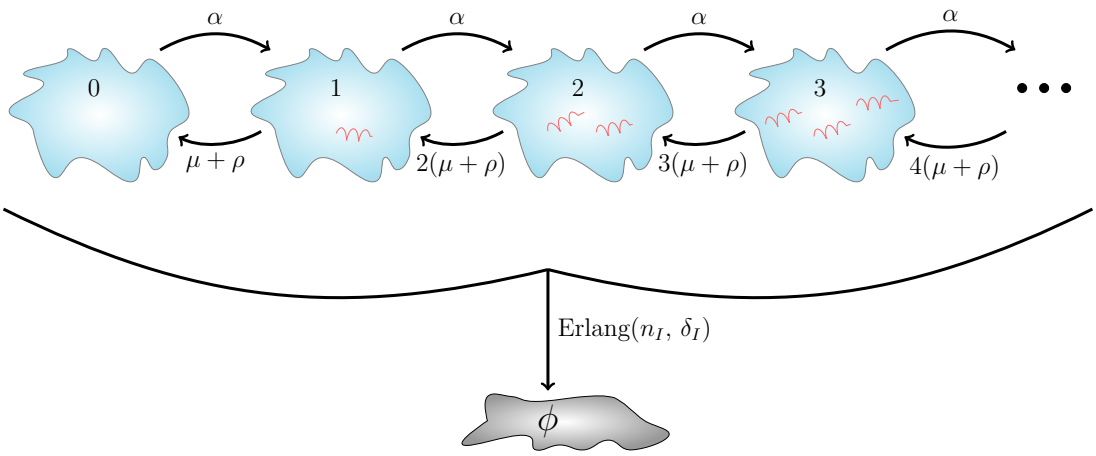

Figure 5.14: A depiction of the continuous-time Markov chain, $\mathcal{Z}$, used to model a single infected cell. State i indicates that the cell contains i vRNA copies that are available to be assembled into virions and released from the cell. When the process is in state 0, the cell is still infected and able to produce vRNA copies from some background level of replication complexes that are not counted by this Markov chain. vRNA is produced within the cell with constant rate α, degraded with rate μ per vRNA, and exported as virions with rate ρ per vRNA. After an Erlang-distributed time, the process will enter state ϕ, which is an absorbing state representing that viral production no longer occurs from the infected cell. For example, the cell might have suffered a virus-induced death, or might have been killed due to the action of the immune system. This model is different to the one in Section 5.1, since here the two competing mechanisms of cell death are not explicitly modelled, but it is assumed that this competition results overall in an Erlang distribution for the time until cell death.

faster replication of the genome will occur. In general, this assumption can be justified early on in the cellular infection, since the copies of vRNA are used as templates for replication. However, a linear replication rate also assumes that there are unlimited resources available for this replication. The process of viral genome replication requires sufficient levels of viral proteins to be present in the cell, and also depends on the cell machinery. Hence these may become limiting factors for the speed of the genome replication. Therefore, since the model in Eq. (5.2.1) is a very simple model that does not explicitly include the dynamics of viral proteins and host factors, it is more realistic to assume a constant rate of viral replication, α, which does not depend on the intracellular viral load. On the other hand, the budding rate is mainly limited by the amount of intracellular vRNA, and hence this rate is assumed to be linear, so that higher intracellular viral loads will lead to increased rates of budding. Therefore, this model assumes that the viral budding rate can change over time, depending on the amount of vRNA inside the cell.

5.2.1 Burst size probability distribution

For the stochastic model of a single infected cell, depicted in Figure 5.14, I will describe how to find the probability distribution for the number of virions released by an infected cell. Once this distribution has been obtained, the distribution for the number of secondary infections produced by a single infected cell in a completely susceptible population can be calculated as described in Section 5.1.2.

Let $\hat{\mu} = \mu + \rho$. Then the process depicted in Figure 5.14, for the amount of vRNA inside an infected cell, is of the form of an M/M/∞ queueing system with arrival rate α and service rate $\hat{\mu}$ (D'Auria (2010)). That is, vRNA is produced or "arrives" in the cell with constant rate α, and each vRNA that is produced stays in the cell for an exponentially distributed amount of time, before either degrading with probability $\mu/\hat{\mu}$ or being exported from the cell by budding with probability $\rho/\hat{\mu}$. Therefore the history of events for each vRNA (e.g time that it is produced, time that it leaves the cell, and whether it is removed by degradation or budding), is independent of any other vRNA.

Assume that the cell begins the infectious phase with zero vRNA copies ready to be packaged into virions, i.e. state $Z(0) = 0$. Let us for the moment ignore the process of cell death, and let $N(t)$ be the total cumulative number of vRNA produced inside the cell by time t. Then $N(t)$ includes the number of vRNA currently inside the cell at time t, denoted $Z(t)$, those that have been removed by budding before time t, denoted $B(t)$, and those removed by degradation, denoted $D(t)$. That is, $N(t) = Z(t) + B(t) + D(t)$. Following ideas from D'Auria (2010), I will now show that at a fixed time $t \geq 0$, $B(t), D(t)$, and $Z(t)$ are independent and Poisson distributed.

If it is known that a vRNA has been produced at some random time u that is smaller than a fixed time t, then since the production of vRNA follows a Poisson process, the random variable u is uniformly distributed on the interval $(0, t)$, as shown by D'Auria (2012). Also, the time between a vRNA being produced and removed from the cell is exponentially distributed with parameter $\hat{\mu}$. Hence, given that a vRNA copy has been produced before time t, the probability that it has been removed before time t is,

$$\begin{aligned} p(t) &= \int_0^t (1 - e^{-\hat{\mu}(t-u)}) \frac{1}{t} \, du \\ &= 1 + \frac{1}{t\hat{\mu}}(e^{-\hat{\mu}t} - 1), \end{aligned}$$

and the probability that it is still in the cell at time t is $q(t) = 1 - p(t)$. Let

$$p_b(t) = \frac{\rho}{\hat{\mu}} p(t), \qquad p_d(t) = \frac{\mu}{\hat{\mu}} p(t),$$

be the probabilities that it has been exported or degraded before time t, respectively. Then we have,

$$\begin{aligned} &\mathbb{P}(B(t) = n_b, D(t) = n_d, Z(t) = n_i) \\ &= \mathbb{P}(B(t) = n_b, D(t) = n_d, Z(t) = n_i \mid N(t) = n_b + n_d + n_i) \mathbb{P}(N(t) = n_b + n_d + n_i), \end{aligned}$$

$$= \frac{(n_b + n_d + n_i)!}{n_b! n_d! n_i!} p_b(t)^{n_b} p_d(t)^{n_d} q(t)^{n_i} \frac{(\alpha t)^{n_b + n_d + n_i} e^{-\alpha t}}{(n_b + n_d + n_i)!},$$

$$= \left(\frac{(\alpha t p_b(t))^{n_b} e^{-\alpha t p_b(t)}}{n_b!} \right) \left(\frac{(\alpha t p_d(t))^{n_d} e^{-\alpha t p_d(t)}}{n_d!} \right) \left(\frac{(\alpha t q(t))^{n_i} e^{-\alpha t q(t)}}{n_i!} \right).$$

Thus, if a cell begins the infectious phase with zero vRNA copies ready to be packaged into virions, the number of vRNA copies inside the cell at time t, denoted by $Z(t)$, is Poisson distributed with mean,

$$\alpha t q(t) = \frac{\alpha}{\hat{\mu}}(1 - e^{-\hat{\mu}t}).$$

This distribution converges to a Poisson distribution with mean $\alpha/\hat{\mu}$ as $t \to \infty$, which is the stationary distribution of the Markov chain describing the intracellular vRNA, if cell death is ignored.

Similarly, $B(t)$, the random variable for the number of virions released by budding before time t, is Poisson distributed with mean,

$$\alpha t p_b(t) = \frac{\alpha\rho}{\hat{\mu}}\left(t + \frac{1}{\hat{\mu}}(e^{-\hat{\mu}t} - 1)\right).$$

That is, the probability that b virions have been released by time t, given that the Markov chain $\mathcal{Z}$ began in state $Z(0) = 0$, is,

$$\mathbb{P}(B(t) = b \mid Z(0) = 0) = \exp\left(-\alpha t p_b(t)\right) \frac{\left(\alpha t p_b(t)\right)^b}{b!}. \tag{5.2.2}$$

If we assumed that each newly produced copy of vRNA immediately decays or is assembled and released as a virion, the mean of the Poisson distribution in Eq. (5.2.2) would become $\frac{\alpha\rho}{\hat{\mu}}t$, and we would recover the model with a constant viral production rate of $\frac{\alpha\rho}{\hat{\mu}}$.

For the model shown in Eq. (5.2.1), infected cells are assumed to begin producing and releasing virus as soon as they become infected. However, in reality, when a target cell becomes infected, there is usually a period of time during which the virus is being replicated but no virus is being released, known as the eclipse phase. The dynamics during the eclipse phase may need to be taken into account because the dynamics of viral RNA replication during the eclipse phase may affect the initial state distribution of a cell entering the infectious phase, and in turn will have an effect on the rate that the cell begins to produce virus. For example, in a model for HCV by Quintela *et al.* (2018), the authors included an eclipse phase in the form of a fixed time of 12 hours, during which the viral RNA inside the cell is involved in the processes of replication and translation but is not yet being assembled into virions and released. To incorporate this into

the Markov chain in Figure 5.14, one could assume that prior to the infectious stage, the cell has already begun replicating the viral genome, so that the Markov chain $\mathcal{Z}$ begins with $i \in \mathbb{N}$ copies of vRNA at time $t = 0$ (representing the beginning of the infectious phase), where this initial state i is sampled from some distribution corresponding to the assumptions of the eclipse phase dynamics. In the Markov chain $\mathcal{Z}$, each vRNA is independent of any other vRNA. Hence if we assume that the cell begins the infectious phase with i copies of vRNA, we can think of the number of virions released from the cell as the sum of those released from two independent processes: the ones generated by the initial i copies of vRNA that are within the cell at the beginning of the infectious phase, and the ones generated by new vRNA produced during the infectious phase. If the initial state of the Markov chain is $Z(0) = i$, then the number of these initial copies of vRNA that are released by budding before time t will follow a Binomial$(i, \widetilde{p}(t))$ distribution, with

$$\widetilde{p}(t) = \frac{\rho}{\hat{\mu}}(1 - e^{-\hat{\mu}t}).$$

Therefore, to find the probability that b virions are released in total by time t, we need to sum over the possible values of k, where k out of the original vRNA copies (the ones present in the cell at time 0) are released by budding before t, and $b - k$ virions are released that were produced inside the cell after time 0. Hence, the probability that b virions are released in total by time t is given by,

$$\mathbb{P}(B(t) = b \mid Z(0) = i) = \sum_{k=0}^{\min(i,b)} \binom{i}{k} \widetilde{p}(t)^k (1 - \widetilde{p}(t))^{i-k} \mathbb{P}(B(t) = b - k \mid Z(0) = 0).$$

Since the time that the cell spends in the infectious phase is Erlang(n_I, δ_I) distributed and independent of the intracellular dynamics, the probability that b virions are released over the entire infectious period is

$$\mathbb{P}(B = b \mid Z(0) = i) = \int_0^\infty \mathbb{P}(B(t) = b \mid Z(0) = i) \frac{\delta_I^{n_I} t^{n_I - 1} e^{-\delta_I t}}{(n_I - 1)!}\, dt. \tag{5.2.3}$$

The expected burst size is given by

$$\begin{aligned} \mathbb{E}[B \mid Z(0) = i] &= \sum_{b=0}^{\infty} b\mathbb{P}(B = b \mid Z(0) = i) \\ &= \frac{\rho\alpha}{\hat{\mu}} \left[\frac{n_I}{\delta_I} + \left(\frac{i}{\alpha} - \frac{1}{\hat{\mu}} \right) \left(1 - \left(\frac{\delta_I}{\hat{\mu} + \delta_I} \right)^{n_I} \right) \right], \end{aligned} \tag{5.2.4}$$

and the variance of the burst size is given by

$$\begin{aligned}\mathrm{Var}[B \mid Z(0)=i] = \mathbb{E}[B] + \frac{\rho^2\alpha^2}{\hat{\mu}^2}\Bigg[2\left(\frac{i}{\alpha^2} - \frac{n_I}{\delta_I(\hat{\mu}+\delta_I)}\right)\left(\frac{\delta_I}{\hat{\mu}+\delta_I}\right)^{n_I} \\ + \left(\frac{i(i-1)}{\alpha^2} + \frac{1}{\hat{\mu}^2}\right)\left(\frac{\delta_I}{2\hat{\mu}+\delta_I}\right)^{n_I} \\ - \left(\frac{i^2}{\alpha^2} + \frac{1}{\hat{\mu}^2}\right)\left(\frac{\delta_I}{\hat{\mu}+\delta_I}\right)^{2n_I} - \frac{i}{\alpha^2}\Bigg].\end{aligned} \tag{5.2.5}$$

5.2.2 Reproduction number probability distribution

Once the distribution for the number of virions released has been obtained, using Eq. (5.2.3), the distribution of the number of secondary infections caused by an initial infected cell can be calculated using the methods described in Section 5.1.2.

If the number of target cells is limiting, the method from Eq. (5.1.5) should be used,

$$\mathbb{P}\Big(R=r\Big) = \sum_{b=0}^{\infty} p_r(b)\mathbb{P}(B=b),$$

with the $p_r(b)$ defined recursively in Section 5.1.2 and $\mathbb{P}(B=b)$ defined in Eq. (5.2.3) (with $Z(0)$ chosen according to some initial state distribution that depends on the assumptions of the eclipse phase dynamics). Otherwise, if there is assumed to be a constant number of T_0 target cells, then each virion released from an infected cell will infect a new cell with probability $\theta = \frac{\beta T_0}{c+\beta T_0}$, and will be cleared with probability $1-\theta$. Since only the number of secondary infections produced is being considered and not the timings of these infections, the model in Figure 5.14 can be used to count the number of secondary infections produced rather than the number of virions released, if the rates are adjusted slightly to reflect this. To count the number of secondary infections produced, the rate of budding and subsequent infection is set to $\theta\rho$, and the rate of intracellular decay or budding and extracellular clearance is set to $\mu+(1-\theta)\rho$. Then the reproduction number distribution can be calculated using the same methods as the burst size distribution from Eq. (5.2.3), but with these adjusted rates.

Parameter	Units	Value
α	d^{-1}	40
ρ	$(\mathrm{vRNA} \cdot \mathrm{d})^{-1}$	8.18
μ	$(\mathrm{vRNA} \cdot \mathrm{d})^{-1}$	1.46
τ_I	d	$1/0.14 \approx 7.14$
n_I	-	10
c	$(\mathrm{vRNA} \cdot \mathrm{d})^{-1}$	22.3

Table 5.5: Parameter values that have been used to obtain numerical results for the HCV model. These are the estimates given for the parameters in the deterministic model by Guedj *et al.* (2013). The only difference is that the model by Guedj *et al.* (2013) assumes an exponentially distributed infectious period ($n_I = 1$), but this may not be realistic, as discussed previously. Therefore a larger value of n_I is used here for illustration.

5.2.3 Numerical results

The results in this section have been calculated using parameter values estimated by Guedj *et al.* (2013) for their model of HCV infection. Guedj *et al.* (2013) estimated the degradation rate μ after the effect of the antiviral daclatasvir, since they fit their model to data from patients treated with this antiviral. However they mention that this treatment is likely to only have a small effect on enhancing vRNA degradation, so the same value of the degradation rate is used in the calculations here. The parameter values used are presented in Table 5.5.

Burst size distribution

B is the random variable representing the amount of virus released by a cell whilst in the infectious phase. The probability distribution for B is shown in Figure 5.15, for varying values of α and n_I. The plot on the left of Figure 5.15 shows how the distribution changes if the intracellular replication rate changes but all other parameters remain equal to their values in Table 5.5. It can be seen that there is quite a lot of variability in the distribution for the number of virions released from an individual infected cell, and this variability increases as α increases. The plot on the right of Figure 5.15 shows how the distribution of B changes for

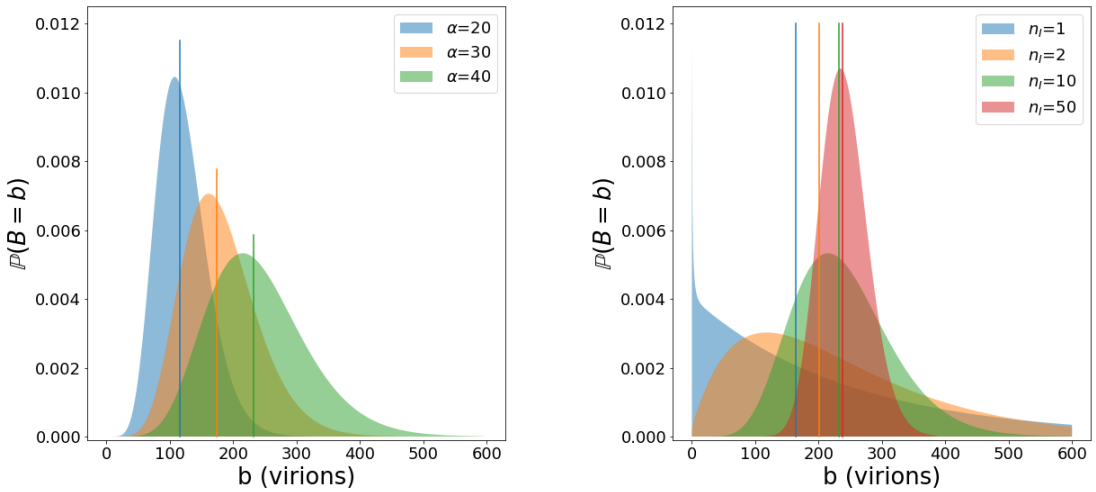

Figure 5.15: Plots showing the probability distribution for B, the random variable for the amount of virus released from a single cell whilst in the infectious phase, for the model in Figure 5.14. In the plot on the left, α is varied and all other parameters are fixed to their values in Table 5.5. In the right hand side plot, n_I is varied, but the average time spent in the infectious phase, $\tau_I = n_I/\delta_I$, remains constant. The medians of each distribution are indicated on the plots.

different infectious period distributions. For each value of n_I, the average time spent in the infectious phase remains constant at the value of τ_I in Table 5.5, and $\delta_I = n_I/\tau_I$. When $n_I = 1$ the infectious period is exponentially distributed and there is a fairly wide distribution for the number of virions released from an individual infected cell, with the mode at zero. As n_I increases, the distribution becomes narrower and peaks much closer to the mean. This is because as n_I increases, the Erlang distribution of time spent in the infectious phase becomes similar to a normal distribution.

The expected values and medians of B when using the different values of n_I are shown in Table 5.6. Here I have used the analytical formula for the expected value of B in Eq. (5.2.4), with the assumption $Z(0) = 0$. The mean of B decreases slightly as n_I increases, even though the average infectious period is kept the same. However the median of the distribution is much more sensitive to changes in n_I, as can be seen in the plot on the right of Figure 5.15, and the values in Table 5.6.

n_I	Mean	Median
1	238.972	164
2	238.924	200
10	238.921	231
50	238.921	237

Table 5.6: Values for the mean and median of the random variable B, when the shape parameter of the Erlang infectious period distribution is varied. For each value of n_I, $\delta_I = n_I/\tau_I$, where τ_I is the average length of the infectious period.

Reproduction number distribution

R is the random variable representing the number of secondary infections caused by a single infected cell in a completely susceptible population. Guedj *et al.* (2013) assumed that no new infections occurred after the initiation of treatment, so they set the infection rate to $\beta = 0$. However estimates of the infection rate from other models of HCV infection are on the order of 10^{-7} and 10^{-8} (Clausznitzer *et al.* (2016); Neumann *et al.* (1998); Rong *et al.* (2013)). With such a small value of β, the two methods described in Section 5.2.2 for calculating the distribution of secondary infections give very similar results, as discussed in Section 5.1.5. Therefore I have used the case of a constant number of target cells to obtain the results in this section. This distribution assumes that each virion released from the initial infected cell will have the same probability of infecting another cell, given by,

$$\theta = \frac{\beta T_0}{\beta T_0 + c},$$

which depends on the infection rate, the number of susceptible cells that are available to be infected, and the viral decay rate. In general, a larger value of θ corresponds to a larger infection rate, more available target cells, or a smaller viral decay rate. Estimates in the literature for these parameters vary, but in general, the value of θ is quite small. For example, Ribeiro *et al.* (2012) estimated the median value across patients studied to be $\theta = 0.018$.

Probability distributions for the reproduction number are shown in the plots in Figure 5.16, for three different values of θ that are similar to estimates from Rong

et al. (2013), Clausznitzer *et al.* (2016), and Neumann *et al.* (1998), respectively, when assuming a constant population of $T_0 = 10^6$ susceptible cells. The top row of plots shows the distributions when $n_I = 1$ and the bottom row corresponds to the case $n_I = 10$. All three values of θ used predict a fairly small number of secondary infections due to an individual infected cell. However, there are substantial differences in the mean, $\bar{R}$, for the different parameter estimates in these three papers, from around 0.5 secondary infections up to 14.

The distribution of R changes dramatically (particularly for larger values of θ) when the infectious period is assumed to be Erlang distributed with $n_I = 10$ rather than exponentially distributed with the same mean. For the value of $\theta = 0.012$, the model with exponentially distributed infectious period predicts that there is around a 27% chance that an initial infected cell would cause zero secondary infections. However, when the infectious period is assumed to be Erlang distributed with $n_I = 10$, the model predicts only an 8% chance of zero secondary infections, even though the mean number of secondary infections is very similar to the exponential case. Therefore, the choice of distribution for the infectious period can be very influential in model predictions of infection outcome at early stages of infection. For example, if the number of secondary infections is distributed according to the model with exponential infectious period, recovery from infection can be fairly likely if there are only a small number of infected cells at the start of an infection, as shown in Figure 5.17. As the number of stages of the Erlang-distributed infectious period increases, recovery from infection becomes much less likely.

5.3 Discussion

The mean value of the basic cellular reproduction number has been calculated from deterministic models for a number of viruses (Baccam *et al.* (2006); González-Parra *et al.* (2018); Iwami *et al.* (2012)) and in some cases these values have been compared between different viral strains as a measure of relative fitness (Iwanami *et al.* (2017)). Although the value of $\bar{R}$ is often used to compare viruses and predict the outcome of infection, there may be complexities that are not revealed by

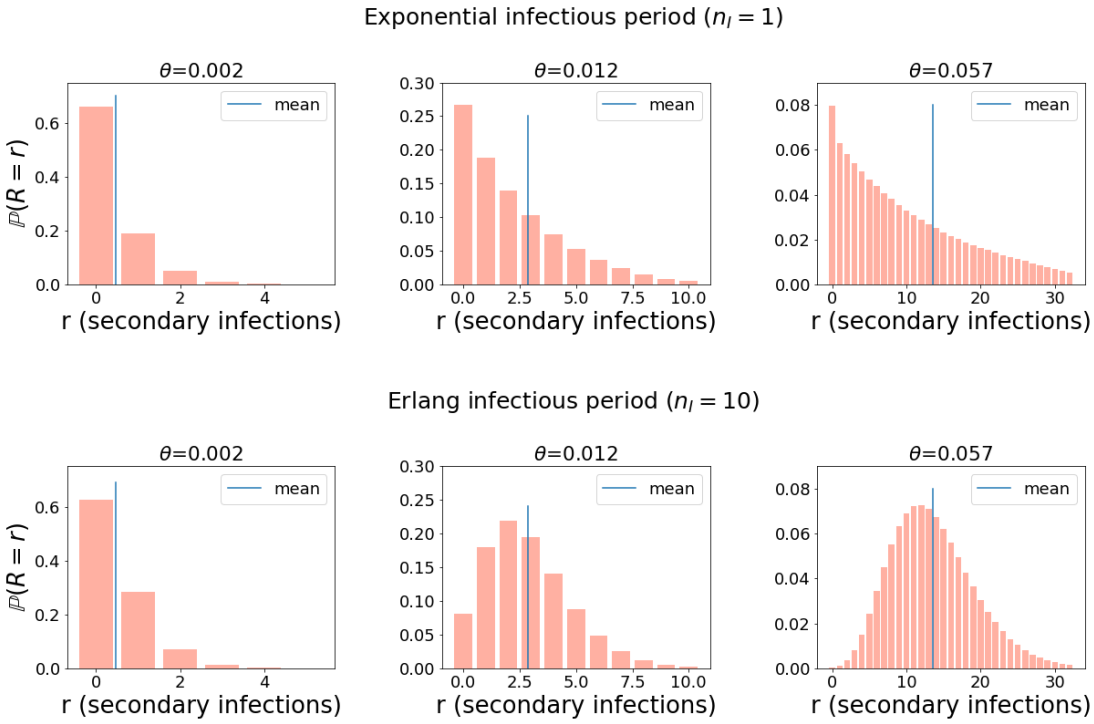

Figure 5.16: Plots showing probability distributions for R, the random variable for the number of secondary infections caused by a single infected cell. The top row corresponds to the case where the distribution of the infectious period is exponential. The bottom row corresponds to the case where the infectious period follows an Erlang distribution with shape parameter $n_I = 10$. The distributions are shown for different values of θ, which is the probability that a progeny virion released from an infected cell will infect a new cell before it decays. These values of θ were obtained by using estimates for the pair of parameters, β and c, from Rong *et al.* (2013), Clausznitzer *et al.* (2016), and Neumann *et al.* (1998), respectively, and assuming a constant population of $T_0 = 10^6$ susceptible cells.

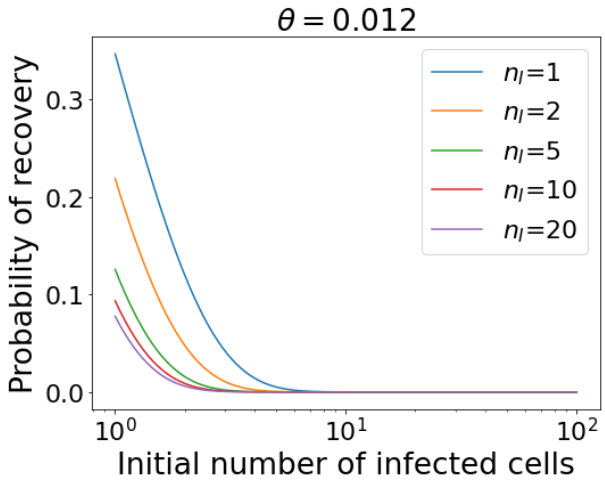

Figure 5.17: Plot showing the probability of viral extinction as a function of the initial number of infected cells, for different values of n_I, which is the number of stages in the Erlang-distributed infectious period.

this single number alone. When stochastic effects are considered, there is heterogeneity in the viral burst size, which is the number of viral progeny generated from an infected cell during its lifetime, and heterogeneity in the reproduction number, which is the number of secondary infections that an infected cell produces. One can capture this heterogeneity by computing the probability distributions of these variables of interest. Indeed, two viruses may have the same $\bar{R}$ but exhibit reproduction number distributions with very different shapes, for example owing to different variability in infected cell lifespans. Here I have shown how to calculate distributions of the random variables for the burst size and the reproduction number, for stochastic versions of two previously used deterministic models of viral dynamics.

I first considered a model in which virus is produced from infected cells at a constant rate. This model, shown in Eq. (5.1.1), includes an eclipse phase to represent the period of time after a virus infects a target cell, in which the virus is replicating but no virus has yet been released. The time until transition of eclipse phase cells into the infectious phase follows an Erlang distribution, as does the time spent in the infectious phase until virus-induced cell death. The time until clearance of infected cells by the immune system is assumed to follow

an Exponential distribution. Thus, the resulting time until infected cell death is the minimum of these two competing mechanisms. In an *in vitro* situation, the effect of the immune system can be removed by setting $\nu = 0$ in the model, as in the models by Liao *et al.* (2020) and Yan *et al.* (2020), from which I have used parameter values to illustrate some numerical examples in Section 5.1.5. In the case where $\nu = 0$, the distribution of the burst size has been shown to follow a negative binomial distribution.

In addition to the model of viral dynamics with constant viral production rate, I also calculated the burst size distribution for an age-structured model of HCV infection in which the rate of virus production is not constant but depends on the intracellular viral dynamics (Guedj *et al.* (2013)). The model considered here only included very simple intracellular dynamics, and it would be more complicated to implement this method for models with more detail around the intracellular viral kinetics. However, it would be fairly straightforward for models in which the production of virus from infected cells is a Poisson process and a functional form of the age-dependent viral production rate can be found. For example, the deterministic model by Guedj *et al.* (2013), presented in Eq. (5.2.1), is equivalent to a model in which the rate of viral production varies continuously over an infected cell's lifespan, described by the function,

$$p(a) = \frac{\alpha\rho}{\hat{\mu}}(1 - e^{-\hat{\mu}a}),$$

where a denotes the age of the infected cell (i.e. the length of time it has been in the infectious phase). The method shown here to calculate the burst size distribution for this particular model of HCV dynamics can be generalised for models with other forms of $p(a)$, and other infectious period distributions. For a general production rate function, the burst size distribution would be calculated as,

$$\mathbb{P}(B = b) = \int_0^\infty \exp\left(-\int_0^t p(a)\ da\right) \frac{(\int_0^t p(a)\ da)^b}{b!} f(t)\ dt,$$

where $f(t)$ is the probability density function of the time a cell spends in the infectious phase.

Given the burst size distribution, I have presented two methods to calculate the reproduction number distribution. One method assumed that the target cell

population is depleted as these cells become infected and the other assumed that the target cell population remains constant. The latter assumption is valid if the system is well mixed and there are a large number of target cells, which could be the case early in infection, for example if a virus is introduced into the blood. However, for some routes of infection, or for particular viruses, the availability of target cells may be limited. Thus, in this case the first method should be used, in which target cell numbers are assumed to decrease due to cells becoming infected.

The reproduction number distribution will have an impact during the very early dynamics of infection, where stochastic effects are not negligible. In particular, the distribution of the reproduction number can be used in order to calculate the probability that disease will become established in an individual, given an initial dose of virus. For example, by comparing a model with a geometric reproduction number distribution and one with a Poisson distribution, Pearson *et al.* (2011) showed that with the greater variability of the geometric distribution, there was a greater chance of viral extinction and a lower probability of successful establishment of disease. This has also been recognised in the case of epidemic models, where increased variability in individual infectiousness has been shown to increase the probability of stochastic extinction in an outbreak beginning with one infected individual (Lloyd-Smith *et al.* (2005)). I have shown that models with the same average burst size and basic reproduction number can produce very different burst size and reproduction number distributions, due to different distributions of infected cell lifetimes for example. The reproduction number distribution determines the probability of viral extinction in a stochastic model, so it is important to consider the range of shapes it may take under different model assumptions.

Chapter 6

Concluding remarks

In this thesis, a multi-scale mathematical model has been developed for the infection dynamics of inhalational anthrax, which is caused by the bacterium *Bacillus anthracis*. Mathematical models of virus dynamics have also been studied.

The multi-scale inhalational anthrax model features a stochastic model for the dynamics of *B. anthracis* spores and bacteria inside an infected phagocyte, presented in Chapter 3. This model considers the germination rate of spores to be heterogeneous across the spore population, and two hypotheses for the germination rate distribution were compared. Results from the calibration of the model with experimental *in vitro* data suggested that the hypothesis of a Bernoulli germination rate distribution is better supported by the data. Furthermore, this hypothesis agrees with other experimental evidence that has indicated the existence of superdormant spores, which have a much slower germination rate than the majority of the spore population (Setlow (2013)). However, the hypothesis of a Bernoulli distribution for the germination rate is the simplest way to define a bi-modal distribution, and more complex distributions may help to describe the data more accurately. Furthermore, for a given spore, the rate of germination into a newly germinated bacterium, and the rate of maturation into a vegetative bacterium are assumed to be equal. In future, if further data were to become available that allowed one to distinguish between newly germinated and vegetative bacteria, then a separate rate could be incorporated for the maturation step of the germination process. This would also allow for more complicated distributions to account for heterogeneity in the germination and maturation rates.

The dynamics of the intracellular model were analytically studied through the calculation of a number of summary statistics, such as the probabilities that an infected cell will eliminate the infection and recover, or will rupture and release its bacterial content. The distribution of time taken for these fates to occur was also computed. Additionally, an expression was given for the probability distribution of the rupture size, which is the amount of bacteria that are ultimately released from an infected phagocyte. These are key quantities that were used to link the intracellular model to a model at the within-host level, presented in Chapter 4.

A stochastic approach was also used for the within-host model of inhalational anthrax, allowing the model to capture the randomness in the biological processes and describe the probabilities of different outcomes, for example how likely it is that the infection will be cleared by the immune system. This stochasticity is particularly important for modelling the outcome of exposures that result in a relatively small number of spores entering the lungs. However, during the calibration of the model, a deterministic description was used to compare the model with *in vivo* bacterial counts from rabbits and guinea pigs that had been infected with a very high dose of *B. anthracis* spores. At the same time, the probability of infection resulting from the stochastic model was fitted to rabbit and guinea-pig dose-response data sets. This resulted in a single model structure that is able to describe both the *in vivo* dynamics of infection, as well as the probability of response, in two different animal species with slightly different values of the parameters.

In future work, the aim will be to extend the novel multi-scale mathematical model for inhalational anthrax developed in Chapters 3 and 4 to link to human data and the population level scale, in order to estimate infection risk and timescales for symptom onset during an accidental or deliberate release of *B. anthracis* spores. The within-host model could be adapted to be able to describe human inhalational anthrax infection, by leveraging non-human primate data and human outbreak-related epidemiological data (Wilkening (2008)). Then the within-host model could also be linked to Quantitative Microbial Risk Assessment (QMRA) techniques (Sze To & Chao (2010)), making use of similar methodologies previously used by Carruthers *et al.* (2018) for the bacteria *Francisella tularensis*. For example, dispersion models can be used to describe the

airborne spread of *B. anthracis* spores in indoor or outdoor settings, and to predict a distribution of initial doses of spores among an exposed population. This can then be linked to the outputs of the within-host model (e.g. likelihood of symptom onset for a given dose) to predict quantities like the distribution of the number of casualties and fatalities, and the earliest time of symptom onset, in different scenarios such as bioterrorist attacks or accidental releases.

In Chapter 5, stochastic models of virus dynamics were studied in order to obtain probability distributions for the burst size, which is the number of viral progeny generated from an infected cell during its lifetime, and the reproduction number, which is the number of secondary infected cells caused by a single infected cell in a completely susceptible population. It was shown that by considering heterogeneity in the viral burst size and reproduction number, interesting results can be observed, such as the probability of viral extinction becoming an increasing function of the average reproduction number, for some shapes of the reproduction number distribution. The work in this chapter was theoretical and not compared with experimental data. However, some numerical examples were illustrated using parameter values that have been estimated using experimental data in previous modelling work in the literature (Guedj *et al.* (2013); Liao *et al.* (2020); Yan *et al.* (2020)). The models considered here were fairly simple, but future work could focus on implementing these methods for models that include more detail around the intracellular viral kinetics.

References

Akoachere, M., Squires, R.C., Nour, A.M., Angelov, L., Brojatsch, J. & Abel-Santos, E. (2007). Identification of an *in vivo* inhibitor of *Bacillus anthracis* spore germination. *Journal of Biological Chemistry*, **282**, 12112–12118. xi, xxiv, 5, 7, 63, 70, 73, 74, 75, 76, 77, 78, 80, 81

Alahmadi, A.A., Flegg, J.A., Cochrane, D.G., Drovandi, C.C. & Keith, J.M. (2020). A comparison of approximate versus exact techniques for Bayesian parameter inference in nonlinear ordinary differential equation models. *Royal Society open science*, **7**, 191315. 116

Allen, L.J. (2010). *An introduction to stochastic processes with applications to biology*. CRC press. ix, 15, 20, 25, 28

Baccam, P., Beauchemin, C., Macken, C.A., Hayden, F.G. & Perelson, A.S. (2006). Kinetics of influenza A virus infection in humans. *Journal of virology*, **80**, 7590–7599. 200

Bacsik, D.J., Dadonaite, B., Butler, A., Greaney, A.J., Heaton, N.S. & Bloom, J.D. (2022). Influenza virus transcription and progeny production are poorly correlated in single cells. *bioRxiv*. 148

Bai, F., Huff, K.E. & Allen, L.J. (2019). The effect of delay in viral production in within-host models during early infection. *Journal of biological dynamics*, **13**, 47–73. 150, 151

Banks, D.J., Barnajian, M., Maldonado-Arocho, F.J., Sanchez, A.M. & Bradley, K.A. (2005). Anthrax toxin receptor 2 mediates *Bacillus*

anthracis killing of macrophages following spore challenge. *Cellular microbiology*, **7**, 1173–1185. 6, 46, 97

Baral, S., Raja, R., Sen, P. & Dixit, N.M. (2019). Towards multiscale modeling of the cd8+ t cell response to viral infections. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, **11**, e1446. 151

Beauchemin, C.A., Miura, T. & Iwami, S. (2017). Duration of SHIV production by infected cells is not exponentially distributed: Implications for estimates of infection parameters and antiviral efficacy. *Scientific reports*, **7**, 1–13. 149

Bhatnagar, R. & Batra, S. (2001). Anthrax toxin. *Critical reviews in microbiology*, **27**, 167–200. 6

Boyer, A.E., Quinn, C.P., Woolfitt, A.R., Pirkle, J.L., McWilliams, L.G., Stamey, K.L., Bagarozzi, D.A., Hart, J.C. & Barr, J.R. (2007). Detection and quantification of anthrax lethal factor in serum by mass spectrometry. *Analytical chemistry*, **79**, 8463–8470. 9

Boyer, A.E., Quinn, C.P., Hoffmaster, A.R., Kozel, T.R., Saile, E., Marston, C.K., Percival, A., Plikaytis, B.D., Woolfitt, A.R., Gallegos, M. *et al.* (2009). Kinetics of lethal factor and poly-D-glutamic acid antigenemia during inhalation anthrax in rhesus macaques. *Infection and immunity*, **77**, 3432–3441. 9

Boyer, A.E., Gallegos-Candela, M., Quinn, C.P., Woolfitt, A.R., Brumlow, J.O., Isbell, K., Hoffmaster, A.R., Lins, R.C. & Barr, J.R. (2015). High-sensitivity MALDI-TOF MS quantification of anthrax lethal toxin for diagnostics and evaluation of medical countermeasures. *Analytical and Bioanalytical Chemistry*, **407**, 2847–2858. 9

Brittingham, K.C., Ruthel, G., Panchal, R.G., Fuller, C.L., Ribot, W.J., Hoover, T.A., Young, H.A., Anderson, A.O. & Bavari, S. (2005). Dendritic cells endocytose *Bacillus anthracis* spores: implications for anthrax pathogenesis. *The Journal of Immunology*, **174**, 5545–5552. 5

Brookmeyer, R., Johnson, E. & Barry, S. (2005). Modelling the incubation period of anthrax. *Statistics in medicine*, **24**, 531–542. 101, 113

Burg, D., Rong, L., Neumann, A.U. & Dahari, H. (2009). Mathematical modeling of viral kinetics under immune control during primary HIV-1 infection. *Journal of Theoretical Biology*, **259**, 751–759. 147

Burnham, K. & Anderson, D. (2002). Model selection and multimodel inference: a practical information-theoretic approach, second edition. *New York, USA: Springer*. 86

Burrell, C.J., Howard, C.R. & Murphy, F.A. (2017). Virion structure and composition. *Fenner and White's Medical Virology*, 27. 1, 10

Canini, L. & Perelson, A.S. (2014). Viral kinetic modeling: state of the art. *Journal of pharmacokinetics and pharmacodynamics*, **41**, 431–443. 152

Carruthers, J., López-García, M., Gillard, J.J., Laws, T.R., Lythe, G. & Molina-París, C. (2018). A novel stochastic multi-scale model of *Francisella tularensis* infection to predict risk of infection in a laboratory. *Frontiers in microbiology*, **9**, 1165. 12, 103, 167, 206

Carruthers, J., Lythe, G., López-García, M., Gillard, J., Laws, T.R., Lukaszewski, R. & Molina-París, C. (2020). Stochastic dynamics of *Francisella tularensis* infection and replication. *PLoS computational biology*, **16**, e1007752. 12, 42, 43, 51, 54, 58

Centers for Disease Control and Prevention (2009). Anthrax Sterne strain (34F2) of *Bacillus anthracis*. 7

Charlton, S., Herbert, M., McGlashan, J., King, A., Jones, P., West, K., Roberts, A., Silman, N., Marks, T., Hudson, M. *et al.* (2007). A study of the physiology of *Bacillus anthracis* Sterne during manufacture of the UK acellular anthrax vaccine. *Journal of applied microbiology*, **103**, 1453–1460. xv, 8, 9, 114, 116, 117, 120, 130, 136, 138

Chung, M.C., Popova, T.G., Millis, B.A., Mukherjee, D.V., Zhou, W., Liotta, L.A., Petricoin, E.F., Chandhoke, V., Bailey, C. & Popov, S.G. (2006). Secreted neutral metalloproteases of bacillus anthracis as candidate pathogenic factors. *Journal of Biological Chemistry*, **281**, 31408–31418. 115

Clausznitzer, D., Harnisch, J. & Kaderali, L. (2016). Multi-scale model for hepatitis C viral load kinetics under treatment with direct acting antivirals. *Virus research*, **218**, 96–101. xxii, 199, 200, 201

Cleret, A., Quesnel-Hellmann, A., Mathieu, J., Vidal, D. & Tournier, J.N. (2006). Resident CD11c+ lung cells are impaired by anthrax toxins after spore infection. *The Journal of infectious diseases*, **194**, 86–94. 5

Cleret, A., Quesnel-Hellmann, A., Vallon-Eberhard, A., Verrier, B., Jung, S., Vidal, D., Mathieu, J. & Tournier, J.N. (2007). Lung dendritic cells rapidly mediate anthrax spore entry through the pulmonary route. *The Journal of Immunology*, **178**, 7994–8001. 5, 97

Cote, C.K., DiMezzo, T.L., Banks, D.J., France, B., Bradley, K.A. & Welkos, S.L. (2008). Early interactions between fully virulent *Bacillus anthracis* and macrophages that influence the balance between spore clearance and development of a lethal infection. *Microbes and infection*, **10**, 613–619. 92, 97

Cote, C.K., Welkos, S.L. & Bozue, J. (2011). Key aspects of the molecular and cellular basis of inhalational anthrax. *Microbes and infection*, **13**, 1146–1155. 3, 97

Cress, B.F., Englaender, J.A., He, W., Kasper, D., Linhardt, R.J. & Koffas, M.A. (2014). Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS microbiology reviews*, **38**, 660–697. 6

Csernica, T. (2015). Extinction in single and multi-type branching processes. 170

Czuppon, P., Débarre, F., Gonçalves, A., Tenaillon, O., Perelson, A.S., Guedj, J. & Blanquart, F. (2021). Success of prophylactic antiviral therapy for SARS-CoV-2: Predicted critical efficacies and impact of different drug-specific mechanisms of action. *PLoS computational biology*, **17**, e1008752. 160

D'Auria, B. (2010). The M/M/∞ queue. *Wiley Encyclopedia of Operations Research and Management Science*. 192, 193

D'Auria, B. (2012). Stochastic processes - notes of February 16, 2012. 193

Day, J., Friedman, A. & Schlesinger, L.S. (2011). Modeling the host response to inhalation anthrax. *Journal of theoretical biology*, **276**, 199–208. 3, 76, 102, 103, 106, 117, 129, 135, 136

Dixon, T.C., Fadl, A.A., Koehler, T.M., Swanson, J.A. & Hanna, P.C. (2000). Early *Bacillus anthracis*–macrophage interactions: intracellular survival and escape. *Cellular microbiology*, **2**, 453–463. 4, 5

European Pharmaceutical Review (2013). GSK signs a multi-year agreement with BARDA to supply the US government with anthrax treatment. ix, 2

Ezzell Jr, J. & Welkos, S. (1999). The capsule of *Bacillus anthracis*, a review. *Journal of applied microbiology*, **87**, 250–250. 98

Foerster, S., Unemo, M., Hathaway, L.J., Low, N. & Althaus, C.L. (2016). Time-kill curve analysis and pharmacodynamic modelling for *in vitro* evaluation of antimicrobials against Neisseria gonorrhoeae. *BMC microbiology*, **16**, 1–11. 145

Gillespie, D.T. (2007). Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, **58**, 35–55. 29

Gimenez, A.P., Wu, Y.Z., Paya, M., Delclaux, C., Touqui, L. & Goossens, P.L. (2004). High bactericidal efficiency of type iia phospholipase A2 against *Bacillus anthracis* and inhibition of its secretion by the lethal toxin. *The Journal of Immunology*, **173**, 521–530. 98

Goel, A.K. (2015). Anthrax: a disease of biowarfare and public health importance. *World Journal of Clinical Cases: WJCC*, **3**, 20. 2, 3, 4

Goggans, P.M., Cao, L. & Henderson, R.W. (2014). Assigning priors for parameters constrained to a simplex region. In *AIP Conference Proceedings*, vol. 1636, 94–99, American Institute of Physics. 74

González-Parra, G., Dobrovolny, H.M., Aranda, D.F., Chen-Charpentier, B. & Rojas, R.A.G. (2018). Quantifying rotavirus kinetics in the REH tumor cell line using *in vitro* data. *Virus research*, **244**, 53–63. 200

Goossens, P.L. (2009). Animal models of human anthrax: the Quest for the Holy Grail. *Molecular aspects of medicine*, **30**, 467–480. 9

Guarner, J., Jernigan, J.A., Shieh, W.J., Tatti, K., Flannagan, L.M., Stephens, D.S., Popovic, T., Ashford, D.A., Perkins, B.A., Zaki, S.R. *et al.* (2003). Pathology and pathogenesis of bioterrorism-related inhalational anthrax. *The American journal of pathology*, **163**, 701–709. 4

Guedj, J., Dahari, H., Rong, L., Sansone, N.D., Nettles, R.E., Cotler, S.J., Layden, T.J., Uprichard, S.L. & Perelson, A.S. (2013). Modeling shows that the NS5A inhibitor daclatasvir has two modes of action and yields a shorter estimate of the hepatitis C virus half-life. *Proceedings of the National Academy of Sciences*, **110**, 3991–3996. xxvi, 11, 149, 150, 188, 190, 197, 199, 203, 207

Guidi-Rontani, C. (2002). The alveolar macrophage: the Trojan horse of *Bacillus anthracis*. *Trends in microbiology*, **10**, 405–409. 3

Guidi-Rontani, C., Weber-Levy, M., Labruyère, E. & Mock, M. (1999). Germination of *Bacillus anthracis* spores within alveolar macrophages. *Molecular microbiology*, **31**, 9–17. 4

Gupta, R.C. (2015). *Handbook of toxicology of chemical warfare agents*. Academic Press. ix, 4

Gutting, B. (2014). Deterministic models of inhalational anthrax in New Zealand white rabbits. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, **12**, 29–41. 102, 105, 121, 123, 128, 129, 135

Gutting, B., Nichols, T.L., Channel, S.R., Gearhart, J.M., Andrews, G.A., Berger, A.E., Mackie, R.S., Watson, B.J., Taft, S.C., Overheim, K.A. *et al.* (2012). Inhalational anthrax (Ames aerosol) in naive and vaccinated New Zealand rabbits: characterizing the spread of bacteria from lung deposition to bacteremia. *Frontiers in cellular and infection microbiology*, **2**, 87. xxv, 122, 123, 124, 125

Gutting, B.W., Rukhin, A., Mackie, R.S., Marchette, D. & Thran, B. (2015). Evaluation of inhaled versus deposited dose using the exponential dose-response model for inhalational anthrax in nonhuman primate, rabbit, and guinea pig. *Risk Analysis*, **35**, 811–827. xvi, xvii, 119, 121, 122, 132, 133, 136, 137, 139, 141

Heffernan, J.M., Smith, R.J. & Wahl, L.M. (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, **2**, 281–293. 154

Heldt, F.S., Frensing, T. & Reichl, U. (2012). Modeling the intracellular dynamics of influenza virus replication to understand the control of viral RNA synthesis. *Journal of virology*, **86**, 7806–7817. 190

Heldt, F.S., Frensing, T., Pflugmacher, A., Gröpler, R., Peschel, B. & Reichl, U. (2013). Multiscale modeling of influenza A virus infection supports the development of direct-acting antivirals. *PLoS computational biology*, **9**, e1003372. 188

Heppell, C.W., Egan, J.R. & Hall, I. (2017). A human time dose response model for Q fever. *Epidemics*, **21**, 30–38. 144

Holder, B.P. & Beauchemin, C.A. (2011). Exploring the effect of biological delays in kinetic models of influenza within a host or cell culture. *BMC Public Health*, **11**, 1–15. 149

Homma, T. & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, **52**, 1–17. 38

Iwami, S., Holder, B.P., Beauchemin, C.A., Morita, S., Tada, T., Sato, K., Igarashi, T. & Miura, T. (2012). Quantification system for the viral dynamics of a highly pathogenic simian/human immunodeficiency virus based on an *in vitro* experiment and a mathematical model. *Retrovirology*, **9**, 1–12. 200

Iwanami, S., Kakizoe, Y., Morita, S., Miura, T., Nakaoka, S. & Iwami, S. (2017). A highly pathogenic simian/human immunodeficiency virus effectively produces infectious virions compared with a less pathogenic virus in cell culture. *Theoretical Biology and Medical Modelling*, **14**, 1–13. 200

Jang, J., Cho, M., Chun, J.H., Cho, M.H., Park, J., Oh, H.B., Yoo, C.K. & Rhie, G.e. (2011). The poly-γ-D-glutamic acid capsule of *Bacillus anthracis* enhances lethal toxin activity. *Infection and immunity*, **79**, 3846–3854. 6, 98

Jenkins, S.A. & Xu, Y. (2013). Characterization of *Bacillus anthracis* persistence *in vivo*. *PLoS One*, **8**, e66177. 99

Jernigan, D.B., Raghunathan, P.L., Bell, B.P., Brechner, R., Bresnitz, E.A., Butler, J.C., Cetron, M., Cohen, M., Doyle, T., Fischer, M. *et al.* (2002). Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerging infectious diseases*, **8**, 1019. 3

Jones, W.I., Klein, F., Walker, J.S., Mahlandt, B.G., Dobbs, J.P. & Lincoln, R.E. (1967). *In vivo* growth and distribution of anthrax bacilli in resistant, susceptible, and immunized hosts. *Journal of bacteriology*, **94**, 600–608. 94

Kalns, J., Morris, J., Eggers, J. & Kiel, J. (2002). Delayed treatment with doxycycline has limited effect on anthrax infection in BLK57/B6 mice. *Biochemical and biophysical research communications*, **297**, 506–509. 76

Kang, T.J., Fenton, M.J., Weiner, M.A., Hibbs, S., Basu, S., Baillie, L. & Cross, A.S. (2005). Murine macrophages kill the vegetative form of *Bacillus anthracis*. *Infection and immunity*, **73**, 7495–7501. xi, xii, xxiv, 43, 46, 60, 61, 69, 70, 71, 75, 76, 80, 81, 82, 84, 85, 89, 96, 97, 98

Karlin, S. & Tavaré, S. (1982). Linear birth and death processes with killing. *Journal of Applied Probability*, **19**, 477–487. 46, 58

Karlin, S. & Taylor, H.M. (1981). *A Second Course in Stochastic Processes*. Academic Press. 20

Kitagawa, K., Nakaoka, S., Asai, Y., Watashi, K. & Iwami, S. (2018). A PDE multiscale model of hepatitis C virus infection can be transformed to a system of ODEs. *Journal of theoretical biology*, **448**, 80–85. 188

Kulkarni, V.G. (2010). *Modeling and Analysis of Stochastic Systems Second Edition*. Taylor and Francis Group, LLC. 20

Liao, L.E., Carruthers, J., Smither, S.J., Team, C.V., Weller, S.A., Williamson, D., Laws, T.R., García-Dorival, I., Hiscox, J., Holder, B.P. *et al.* (2020). Quantification of Ebola virus replication kinetics *in vitro*. *PLoS computational biology*, **16**, e1008375. xxv, 11, 148, 149, 150, 151, 172, 176, 178, 203, 207

Liu, J.Z., Ali, S.R., Bier, E. & Nizet, V. (2018). Innate immune interactions between *Bacillus anthracis* and host neutrophils. *Frontiers in Cellular and Infection Microbiology*, **8**, 2. 102

Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E. & Getz, W.M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, **438**, 355–359. 165, 204

Lovchik, J.A., Drysdale, M., Koehler, T.M., Hutt, J.A. & Lyons, C.R. (2012). Expression of either lethal toxin or edema toxin by *Bacillus anthracis* is sufficient for virulence in a rabbit model of inhalational anthrax. *Infection and immunity*, **80**, 2414–2425. 9

Loving, C.L., Kennett, M., Lee, G.M., Grippe, V.K. & Merkel, T.J. (2007). Murine aerosol challenge model of anthrax. *Infection and immunity*, **75**, 2689–2698. 145

Mabry, R., Brasky, K., Geiger, R., Carrion Jr, R., Hubbard, G.B., Leppla, S., Patterson, J.L., Georgiou, G. & Iverson, B. (2006). Detection of anthrax toxin in the serum of animals infected with *Bacillus anthracis* by using engineered immunoassays. *Clinical and Vaccine Immunology*, **13**, 671–677. 9

Moayeri, M. & Leppla, S.H. (2009). Cellular and systemic effects of anthrax lethal toxin and edema toxin. *Molecular aspects of medicine*, **30**, 439–455. 9

Moayeri, M., Leppla, S.H., Vrentas, C., Pomerantsev, A.P. & Liu, S. (2015). Anthrax pathogenesis. *Annual review of microbiology*, **69**, 185–208. 2, 3, 4

Nelson, P.W., Gilchrist, M.A., Coombs, D., Hyman, J.M. & Perelson, A.S. (2004). An age-structured model of HIV infection that allows for variations in the production rate of viral particles and the death rate of productively infected cells. *Mathematical Biosciences & Engineering*, **1**, 267. 188

Neumann, A.U., Lam, N.P., Dahari, H., Gretch, D.R., Wiley, T.E., Layden, T.J. & Perelson, A.S. (1998). Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon-α therapy. *Science*, **282**, 103–107. xxii, 199, 200, 201

Nowak, M. & May, R.M. (2000). *Virus dynamics*. Oxford University Press, UK. 147

Oyston, P.C., Sjöstedt, A. & Titball, R.W. (2004). Tularaemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nature Reviews Microbiology*, **2**, 967–978. 92

Pantha, B., Cross, A., Lenhart, S. & Day, J. (2018). Modeling the macrophage-anthrax spore interaction: Implications for early host-pathogen

interactions. *Mathematical biosciences*, **305**, 18–28. x, xii, xxiii, 3, 41, 42, 45, 46, 60, 61, 62, 78, 79, 84, 85, 87, 95, 96, 97, 102

Pearson, J.E., Krapivsky, P. & Perelson, A.S. (2011). Stochastic theory of early viral infection: continuous versus burst production of virions. *PLoS computational biology*, **7**, e1001058. 147, 149, 151, 204

Perelson, A.S. (2002). Modelling viral and immune system dynamics. *Nature reviews immunology*, **2**, 28–36. 149

Perelson, A.S. & Nelson, P.W. (1999). Mathematical analysis of HIV-1 dynamics *in vivo*. *SIAM review*, **41**, 3–44. 152

Pflughoeft, K.J., Swick, M.C., Engler, D.A., Yeo, H.J. & Koehler, T.M. (2014). Modulation of the *Bacillus anthracis* secretome by the immune inhibitor A1 protease. *Journal of bacteriology*, **196**, 424–435. 8, 115

Pomerantsev, A.P., Pomerantseva, O.M., Moayeri, M., Fattah, R., Tallant, C. & Leppla, S.H. (2011). A *Bacillus anthracis* strain deleted for six proteases serves as an effective host for production of recombinant proteins. *Protein expression and purification*, **80**, 80–90. 8, 115

Pornillos, O., Garrus, J.E. & Sundquist, W.I. (2002). Mechanisms of enveloped RNA virus budding. *Trends in cell biology*, **12**, 569–579. 11

Powell, J.D., Hutchison, J.R., Hess, B.M. & Straub, T.M. (2015). *Bacillus anthracis* spores germinate extracellularly at air–liquid interface in an *in vitro* lung model under serum-free conditions. *Journal of applied microbiology*, **119**, 711–723. 3

Pratt, A., Bennett, E., Gillard, J., Leach, S. & Hall, I. (2020). Dose–response modeling: Extrapolating from experimental data to real-world populations. *Risk Analysis*. 42, 104

Puziss, M. & Wright, G.G. (1959). STUDIES ON IMMUNITY IN ANTHRAX: VII. Carbohydrate metabolism of *Bacillus anthracis* in relation to elaboration of protective antigen. *Journal of Bacteriology*, **78**, 137–145. 8, 9

Quintela, B.d.M., Conway, J.M., Hyman, J.M., Guedj, J., Dos Santos, R.W., Lobosco, M. & Perelson, A.S. (2018). A new age-structured multiscale model of the hepatitis C virus life-cycle during infection and therapy with direct-acting antiviral agents. *Frontiers in microbiology*, **9**, 601. 188, 190, 194

Ribeiro, R.M., Li, H., Wang, S., Stoddard, M.B., Learn, G.H., Korber, B.T., Bhattacharya, T., Guedj, J., Parrish, E.H., Hahn, B.H. *et al.* (2012). Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the *in vivo* mutation rate. 189, 199

Rong, L., Guedj, J., Dahari, H., Coffield Jr, D.J., Levi, M., Smith, P. & Perelson, A.S. (2013). Analysis of hepatitis C virus decline during treatment with the protease inhibitor danoprevir using a multiscale model. *PLoS computational biology*, **9**, e1002959. xxii, 148, 199, 201

Ross, J.M. *et al.* (1957). The pathogenesis of anthrax following the administration of spores by the respiratory route. *Journal of Pathology and Bacteriology*, **73**, 485–94. 9

Ruthel, G., Ribot, W.J., Bavari, S. & Hoover, T.A. (2004). Time-lapse confocal imaging of development of *Bacillus anthracis* in macrophages. *Journal of Infectious Diseases*, **189**, 1313–1316. 42, 97

Sacha, J.B., Buechler, M.B., Newman, L.P., Reed, J., Wallace, L.T., Loffredo, J.T., Wilson, N.A. & Watkins, D.I. (2010). Simian immunodeficiency virus-specific cd8+ t cells recognize vpr-and rev-derived epitopes early after infection. *Journal of virology*, **84**, 10907–10912. 151

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer physics communications*, **145**, 280–297. 38

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons. 38

Savransky, V., Sanford, D.C., Syar, E., Austin, J.L., Tordoff, K.P., Anderson, M.S., Stark, G.V., Barnewall, R.E., Briscoe, C.M., Lemiale-Biérinx, L. *et al.* (2013). Pathology and pathophysiology of inhalational anthrax in a guinea pig model. *Infection and immunity*, **81**, 1152–1163. xvi, 123, 124, 125, 136

Setlow, P. (2003). Spore germination. *Current opinion in microbiology*, **6**, 550–556. 2, 42, 43, 46, 47

Setlow, P. (2013). Summer meeting 2013–when the sleepers wake: the germination of spores of *Bacillus* species. *Journal of applied microbiology*, **115**, 1251–1268. 42, 43, 47, 48, 95, 99, 205

Setlow, P. (2014). Germination of spores of *Bacillus* species: what we know and do not know. *Journal of bacteriology*, **196**, 1297–1305. 42, 43, 47, 48, 99

Sharma, S., Bhatnagar, R. & Gaur, D. (2020). *Bacillus anthracis* poly-γ-D-glutamate capsule inhibits opsonic phagocytosis by impeding complement activation. *Frontiers in immunology*, **11**, 462. 5, 98

Shetron-Rama, L.M., Herring-Palmer, A.C., Huffnagle, G.B. & Hanna, P. (2010). Transport of *Bacillus anthracis* from the lungs to the draining lymph nodes is a rapid process facilitated by CD11c+ cells. *Microbial pathogenesis*, **49**, 38–46. 97

Sobol, I.M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling Computational Experiments*, **1**, 407–414. 36, 64

Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, **55**, 271–280. 36, 37

Splino, M., Patocka, J., Prymula, R. & Chlibek, R. (2005). Anthrax vaccines. *Annals of Saudi medicine*, **25**, 143–149. 8

Sze To, G.N. & Chao, C.Y.H. (2010). Review and comparison between the Wells–Riley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor air*, **20**, 2–16. 206

Tambe, Y. (2005). Action of anthrax exotoxins with English annotation. ix, 7

Tang, S., Moayeri, M., Chen, Z., Harma, H., Zhao, J., Hu, H., Purcell, R.H., Leppla, S.H. & Hewlett, I.K. (2009). Detection of anthrax toxin by an ultrasensitive immunoassay using europium nanoparticles. *Clinical and vaccine immunology*, **16**, 408–413. 9

Titball, R. & Manchee, R. (1987). Factors affecting the germination of spores of *Bacillus anthracis*. *Journal of applied bacteriology*, **62**, 269–273. 99

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M.P. (2009). Approximate Bayesian Computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, **6**, 187–202. 39, 43, 60, 69, 95

Toth, D.J., Gundlapalli, A.V., Schell, W.A., Bulmahn, K., Walton, T.E., Woods, C.W., Coghill, C., Gallegos, F., Samore, M.H. & Adler, F.R. (2013). Quantitative models of the dose-response and time course of inhalational anthrax in humans. *PLoS pathogens*, **9**, e1003555. 131

Twenhafel, N. (2010). Pathology of inhalational anthrax animal models. *Veterinary pathology*, **47**, 819–830. 7, 98

Weir, M.H. & Haas, C.N. (2011). A model for *in-vivo* delivered dose estimation for inhaled *Bacillus anthracis* spores in humans with interspecies extrapolation. *Environmental science & technology*, **45**, 5828–5833. 104

Welkos, S., Friedlander, A., Weeks, S., Little, S. & Mendelson, I. (2002). *In-vitro* characterisation of the phagocytosis and fate of anthrax spores in macrophages and the effects of anti-PA antibody. *Journal of medical microbiology*, **51**, 821–831. 98

WHO (2008). *Anthrax in humans and animals*. World Health Organization. 98

Wilkening, D.A. (2008). Modeling the incubation period of inhalational anthrax. *Medical Decision Making*, **28**, 593–605. 206

Williams, W.R. & Kendall, L.V. (2015). Blood collection in the guinea pig (Cavia porcellus). *Lab Animal*, **44**, 207–208. 124

Wodarz, D. & Nowak, M.A. (2002). Mathematical models of HIV pathogenesis and treatment. *BioEssays*, **24**, 1178–1187. 147

Yan, A.W., Zhou, J., Beauchemin, C.A., Russell, C.A., Barclay, W.S. & Riley, S. (2020). Quantifying mechanistic traits of influenza viral dynamics using *in vitro* data. *Epidemics*, **33**, 100406. xxv, 11, 147, 149, 150, 151, 181, 182, 183, 203, 207

Yuan, Y. & Allen, L.J. (2011). Stochastic models for virus and immune system dynamics. *Mathematical biosciences*, **234**, 84–94. xxvi, 11, 150, 155, 166, 184, 185

Zai, X., Zhang, J., Liu, J., Liu, J., Li, L., Yin, Y., Fu, L., Xu, J. & Chen, W. (2016). Quantitative determination of lethal toxin proteins in culture supernatant of human live anthrax vaccine *Bacillus anthracis* A16R. *Toxins*, **8**, 56. 8, 9, 115, 119