

---

# Predicting the Dynamics of Infections with Strong Dose-Dependent Responses

Christopher N. Davis<sup>1</sup>, Emma L. Davis<sup>1</sup>, Cameron P. Lack<sup>1</sup>, Sophie R. Meakin<sup>1</sup>, Ian M. Hall<sup>2</sup>, Matt J. Keeling<sup>3</sup>

**1** EPSRC & MRC Centre for Doctoral Training in Mathematics for Real-World Systems, University of Warwick

**2** Bioterrorism and Emerging Disease Analysis programme, Public Health England

**3** WIDER Centre, Mathematics Institute and School of Life Sciences, University of Warwick

---

## Abstract

We present a multi-scale model for intracellular bacterial infections that links the behaviour of within-host and between-host dynamics. While traditional models of infection consider the state of individuals as binary, infected or not infected, our deterministic within-host model defines the infection state by the current pathogen load. We then define a stochastic mapping from the pathogen load to the transmission potential, which we use to simulate the early spread of infection on ego networks taken from contact survey data. We observe that dose dependence can translate across scales: the average number of secondary cases increases with the initial dose received.

---

## 1 Introduction

In population-level models of disease, the infectiousness of an individual is binary: individuals are either infected or not. Instantaneous switching occurs between the two states at times dependant on fixed rates describing processes such as infection and recovery. However, there are examples of cases where this oversimplification breaks down, such as the heightened infectiousness of an HIV positive individual immediately following infection [1] or the well-documented phenomenon of highly infectious super-spreaders, who generate the majority of secondary infections [2]. The effect of the latter is usually attributed as a significant factor in the 2003 SARS outbreak [3].

For both naturally emerging infections and bioterrorism threats, it is critical to have methods in place to enable prompt assessment of the potential public health impact. Understanding the factors that influence the time-scale of infections, and the impact this can have on transmission potential, is vital for planning and executing outbreak countermeasures, such as vaccination, quarantine and travel restrictions. For example, epidemiological modelling is an important component of the management of future Foot and Mouth Disease outbreaks in Great Britain [4]. Incorporating the effects of dose heterogeneity would allow for increased understanding of individual response and transmission profiles, which would provide greater certainty when planning control and response programs. Whilst recent theoretical studies have considered infectiousness as a

function of time since infection [5], there has been little investigation into the effect of initial received dose on this relationship. A mathematical model describing growth of bacteria prior to activation of the adaptive immune system has been developed for *Francisella tularensis* [6], but there is a need for further development to enable insight into the individual biological dynamics across the entire course of infection.

Numerous experimental studies provide evidence of dose dependent infection dynamics. An experiment by Meynell *et al.* [7] investigates the relationship between initial dose of *Salmonella typhimurium* and time to mortality in a sample of 150 mice. The study shows strong negative correlation between size of initial dose and time to mortality (correlation coefficient  $r = -0.9713$ ; p-value  $p < 0.0001$ ). An experiment on the effect of initial dose on symptom severity for *Vibrio cholerae* infection in humans demonstrates strong positive correlation [8]. The results of an experiment on the effectiveness of different vaccines for *Pasteurella tularensis* infections in humans provides data on the incubation period for a range of initial doses [9]. Initial dose and incubation period are strongly negatively correlated ( $r = -0.7133$ ,  $p = 0.0019$ ). As a result, it may be possible to infer an estimate of initial dose from observed incubation period.

Using a dose-dependent within-host model, such as the model described herein, it would be possible to predict infection dynamics, such as length of infection and timing of infectious peak. For highly infectious diseases, such as cholera and tuberculosis [10], this

---

would improve estimations for quarantine times [11]. During the early stages of an epidemic or the release of a bioterrorist agent it would also be important to identify the source of infection: accurate estimation of initial dose would provide more information when mapping secondary cases back to any source cases [11].

## 1.1 Within-host models

The interactions between the pathogen and the host immune system determine the within-host dynamics of infection. How exactly the pathogen and immune system interact will be highly pathogen-dependent; as a result, within-host models are pathogen-specific or specific to a type of pathogen, such as viruses, bacteria or parasites. For example, due to the unique pathogenesis of HIV, realistic models of HIV dynamics cannot be applied to other viral diseases [12]. Constructing realistic and useful within-host models therefore requires an understanding of the key mechanisms of the immune system and the growth and elimination of the pathogen.

The immune system is a complex system of structures and processes that work to protect the body against disease. It can be split into two subsystems: the innate immune system and the adaptive immune system. The former is an immediate but non-specific response, whilst the latter is a highly specialised but delayed response. At the most basic level, the innate immune response provides physical, mechanical and chemical barriers that prevent pathogens entering the body. For example, the skin is a physical barrier against pathogens; coughing and sneezing are mechanisms that remove pathogens from the respiratory tract; and gastric acid in the stomach is a chemical barrier against ingested pathogens. Once a pathogen has bypassed these barriers, the detection of the pathogen stimulates a cascading chemical response that in turn promotes the cellular response of the innate immune system. The cellular response includes marking pathogens for destruction, the removal of pathogens and the recruitment of additional immunological cells to the site of infection.

The mechanisms by which the pathogen is cleared from the body are of particular interest for a within-host model of infection. As part of the innate immune response, pathogens are removed from the body through a mechanism called phagocytosis. Phagocytosis is carried out by a family of specialised white blood cells, called phagocytes, that includes macrophages. Phagocytes engulf, or phagocytose, the pathogen and trap it inside an intracellular vesicle called a phagosome. The pathogen is destroyed when the phagosome


fuses with the lysosome, another vesicle that contains enzymes and acids that digest the pathogen.

A further function of the innate immune system is to activate the pathogen-specific response of the adaptive immune system. The two main cells of the adaptive immune system are T cells and B cells. Once activated, T cells and B cells eliminate pathogens from the body; the method by which these cells are activated is called antigen presentation. Antigens are molecules found on the surface of the pathogen that identify the pathogen to the body as non-self. After a pathogen has been digested by a macrophage, the macrophage displays the pathogen's antigens on its surface. Antigen presentation stimulates immature T cells to differentiate to become either cytotoxic T cells or T helper cells; the former kill infected cells by releasing cytotoxins that induce apoptosis (programmed cell death). T helper cells aid the activity of other immune cells by releasing cytokines, a cell-signalling molecule: for example, they aid the growth of cytotoxic T cells and recruitment of macrophages to the infection site. Immature B cells are activated when they encounter both their matching antigen and matching T helper cell; the B cell replicates and matures into an antibody-producing plasma cell. Antibodies bind to antigens on the surface of the pathogen; this either neutralises the pathogen or promotes phagocytosis.

One of the major difficulties in constructing within-host models is in attaining a balance between biological realism and mathematical tractability. Current within-host models either present simplified ecological analogies [6, 13–16] or attempt to fully describe the interactions between the pathogen various cellular and chemical components of the immune system, comprising of several ordinary differential equations (ODEs) [17–19]. A limitation of many models is the lack of experimental data to which to fit parameters. This limitation is exacerbated by models with many parameters. In such cases, parameters are determined through a combination of experimental data and, where no data is available, mathematical estimation [18]. Experimental studies may be *in vivo* or *in vitro* and in humans, non-human primates, mice or other mammals. It should be noted that there may be differences in the exact values of parameters taken from human versus non-human experimental studies [20].

A simplification of many models is to treat the host as a single system. Compartmental models can include yet further detail by dividing the body into compartments each with different and interacting infection dynamics. This is particularly relevant for diseases that develop to affect multiple organs, or where specific

---



organs play an important role in the immune response. For example, tularemia infections caused by tick bites or ingestion of contaminated food or water primarily affect the lymphatic system but can spread to the lungs via the blood if left untreated [21]. In tuberculosis infections, the lymphatic system plays an important role in the recruitment of macrophages as part of the immune response; compartmental models of tuberculosis treat the lungs and lymph node separately [18, 22].

In the majority of the literature, heterogeneity in the dynamics comes solely from difference in initial dose. When scaling this up to a population level model, homogeneity in the response across all individuals is unrealistic and undesirable. Heterogeneity can be introduced into deterministic ODE models through the inclusion of stochastic elements, such as by varying the threshold beyond which the adaptive immune system takes effect; this can be done by assuming that the threshold follows a log-normal distribution [23] or a Weibull distribution [24], for example. The single hit model assumes host homogeneity post-infection, but allows the possibility that some infections do not take hold at all [6, 25]. Heppell *et al.* [13] model the probability that an inhaled pathogen is deposited in the lungs as a beta distribution, introducing variation in the initial dose. Wood *et al.* [6] use a stochastic birth-death process to model infections of intracellular bacteria.

In the same way that the immune system has developed complex defences against pathogens, so pathogens have evolved strategies to evade the immune system or resist destruction by phagocytes [26]. Certain bacteria are not only able to resist destruction by phagocytes, but use them as their primary host for growth and reproduction: such bacteria are called intracellular bacteria. By inhabiting the host's own cells, intracellular bacteria are shielded from antibodies and can only be eliminated by the cellular immune response. Such bacteria possess highly specialised mechanisms that allow them to resist destruction and proliferate inside phagocytes. For example, *Mycobacterium tuberculosis*, the bacterium that causes tuberculosis, avoids destruction by interfering with phagosome-lysosome fusion [27]. *Francisella tularensis*, the bacterium that causes tularemia, is able to escape from the phagosome before it fuses with the lysosome [28]. *Coxiella burnetii*, the bacterium that causes Q-fever, does not avoid phagosome-lysosome fusion, but is highly resistant to environmental stresses and is able to replicate even in the presence of the enzymes and acids contained in the host cell's lysosome [29]. After avoiding destruction, intracellular bacteria are then able to replicate

inside the host cell; the resulting intracellular bacterial population are subsequently released when the macrophage ruptures due to internal pressure. This mechanism is included in models for *Mycobacterium tuberculosis* infections [18, 22], *Francisella tularensis* infections [6, 15, 24, 30, 31] and *Coxiella burnetii* infections [13, 32, 33]. Since phagocytosis alone is an ineffective mechanism to clear intracellular bacteria from the body, the role of the adaptive immune response becomes much more important; the most realistic within-host models for intracellular bacteria should therefore include the mechanics of both the innate and adaptive immune system.

## 1.2 Between-host models

Whilst within-host modelling describes the infection dynamics within an individual, between-host modelling describes the transmission between individuals and so informs on spread of disease across a population. The transmission of respiratory diseases occurs when an infectious individual expels microscopic water droplets carrying the pathogen; this occurs during coughing, sneezing, or simply breathing, resulting in continual bacterial output throughout the day. Models of transmission therefore need to account for variation in both the pathogen count on each water droplet and the number of droplets that a susceptible contact inhales in a given time period. Although there is believed to be significant variation in the number of viruses on each droplet [34], bacteria are significantly larger than viruses and thus it is unlikely to find more than one bacterium per droplet [35, 36].

There has been considerable research on both within-host infection dynamics for a given dose and between-host population models for given transmission rates, such as the family of deterministic compartmental models that includes the widely-researched SIR model. Yet there is little research that incorporates both scales [16, 37, 38], despite evidence that infection dynamics are dose dependent [31] and individual pathogen load affects onwards transmission. However, the relationship between pathogen load and transmission potential is complicated: it is possible to become infected from interaction with a single virion [39] and it is unclear how to map pathogen load of an individual to a transmission fitness. Indeed, this formulation of the problem may be too simplistic and thus uninformative [40].

Handel and Rohani [40] argues that it is a combination of the pathogen load, the level of the immune response and the symptoms of the individual that affect host infectiousness. For example, symptoms of

respiratory diseases, such as coughing or sneezing, can increase transmission and even vomiting and diarrhoea can be shown to cause an effect [43, 44]. Using data from studies on viral shedding [45, 46] the problem can be simplified to expressing symptoms as a function of pathogen load by considering the relationship between the shedding and the load [40], but it is not clear how to exactly quantify the relationship. Analyses of ferret infections by influenza exemplifies the problem by demonstrating that influenza strains that cause similar viral loads are able to cause contrasting transmission potential, likely caused by factors related to the symptoms, such as the frequency of sneezing [47].

An additional complication to the pathogen-load-transmission-potential relationship is that the factors of pathogen load, immune response and symptoms also affect the host behaviour, potentially causing a reduction in the level of interaction with other individuals, which will further impact transmission potential. A survey of behaviour on infection with influenza showed that sick individuals had approximately a quarter of the daily contacts compared to healthy individuals [48, 49].

Despite the complex interacting factors surrounding transmission potential, in our model we will focus on how the pathogen load affects transmission potential. We acknowledge that while symptoms and host behaviour are important factors, in the interest of developing a tractable model they will be ignored. This assumption has been used in several transmission studies, where the effect of the area under the pathogen load curve is considered on the transmission [50, 51]. Different studies relate transmission to load with a linear [52, 53] or logarithmic scale [53]. Our analysis considers logarithmic scaling only and, using this area under the curve, we link between the within-host and between-host models.

## 2 Model

### 2.1 Within-host

#### 2.1.1 ODE model

The proposed within-host model is a system of three ODEs characterising the number of extracellular bacteria  $B$ , the number of bacteria-containing phagocytes  $P$  and the level of the adaptive immune response  $R$ . We assume this is the first infection by this bacteria and thus both  $P$  and  $R$  are zero at the beginning of the infection. The initial number of extracellular bacteria is given by the initial dose  $D$ .

This proposed model is an extension of the model of [6, 13], in which phagocytes engulf bacteria at rate  $\gamma$ , and phagocytosed bacteria are successfully destroyed with probability  $1 - \Theta$ . However, with probability  $\Theta$ , phagocytosed bacteria are able to resist destruction; the exact mechanism depends on the bacteria. In this case, the bacterium replicates within the phagocyte; bacteria-containing phagocytes rupture at rate  $\lambda$  and subsequently release the resulting intracellular bacterial population into the extracellular environment. We assume that the number of bacteria released  $G$  is constant. It is also assumed that the number of phagocytes are far greater than the number of bacteria, so that bacterial growth is not limited by the availability of phagocytes. This construction captures the key mechanisms of intracellular bacteria and the innate immune response and gives the system of ODEs:

$$\begin{aligned}\frac{dB}{dt} &= \lambda GP - \gamma B, \\ \frac{dP}{dt} &= \Theta \gamma B - \lambda P,\end{aligned}\tag{1}$$

**Table 1.** Variables and parameters that characterise the model for within-host dynamics described by equations (3) and (4). See Appendix A for further discussion of parameter estimation.

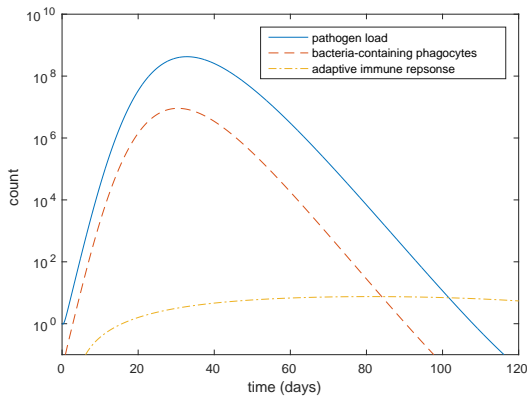
Symbol	Description	Value	Source
$B$	Extracellular bacteria	-	-
$P$	Bacteria-containing phagocytes	-	-
$R$	Adaptive immune response	-	-
$\lambda$	Rupture rate of bacteria-containing phagocytes	0.4	[41]
$G$	Number of bacteria released upon rupturing	50	[41]
$\gamma$	Rate of phagocytosis	0.4	[41]
$\Theta_0$	Baseline probability of bacteria surviving phagocytosis	0.3	[42]
$\rho$	Decay parameter for probability of unsuccessful bacteria kill	0.8	estimated
$\beta$	Growth rate of response	0.025	estimated
$\delta$	Rate of immunological decay	0.01	estimated

with initial conditions

$$\begin{aligned} B(0) &= D, \\ P(0) &= 0. \end{aligned} \quad (2)$$

We adapt the model of Heppell *et al.* [13] with the addition of a third equation for the adaptive immune response,  $R$ . In the same way that the innate immune response is simplified into a single term,  $P$ , we also condense the effect of the adaptive immune response into one term that captures the key mechanisms of the adaptive immune system that affect the within-host bacterial population. The two key mechanisms that we wish to include are the activation of the adaptive immune response by antigen-presenting phagocytes after successful phagocytosis and the positive correlation between the adaptive immune response and the number of bacteria successfully destroyed. Including the effects of the adaptive immune system makes the model more realistic; it is a particularly important mechanism to include for models of intracellular bacteria, since it explains the mechanism that clears the bacteria from the body.

Since the adaptive immune response is triggered by antigen presenting phagocytes during the innate immune response, the response should increase as bacteria are destroyed. We assume that the rate of increase in the response is proportional to the magnitude of the number of bacteria successfully phagocytosed. The adaptive immune system increases the proportion of bacteria that are destroyed, either through the action of cytotoxic T cells or antibodies. To reflect this, we take the probability that a bacterium is successfully destroyed during phagocytosis,  $1 - \Theta$ , to be an increasing function of  $R$ . We choose  $\Theta = \Theta_0 e^{-\rho R}$ . Finally, the response decays at a constant rate  $\delta$ ; this represents



**Fig 1.** The number of extracellular bacteria,  $B(t)$ , bacteria-containing phagocytes,  $P(t)$  and level of adaptive immune response,  $R(t)$ , over time for an initial dose  $D = 1$ .

the loss of immunological memory over time [54]. This construction gives the systems of ODEs:

$$\begin{aligned} \frac{dB}{dt} &= \lambda GP - \gamma B, \\ \frac{dP}{dt} &= \Theta_0 e^{-\rho R} \gamma B - \lambda P, \\ \frac{dR}{dt} &= \beta \log [(1 - \Theta_0 e^{-\rho R}) \gamma B] - \delta R, \end{aligned} \quad (3)$$

with initial conditions

$$\begin{aligned} B(0) &= D, \\ P(0) &= 0, \\ R(0) &= 0. \end{aligned} \quad (4)$$

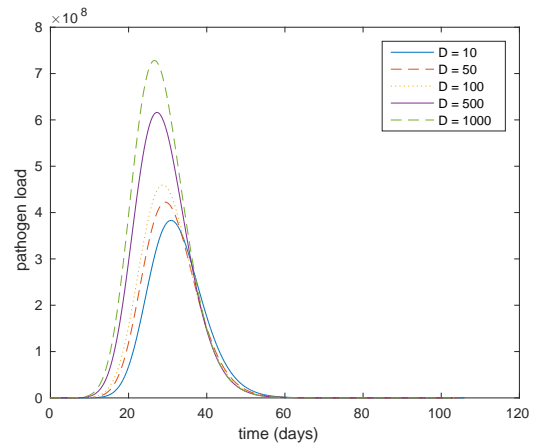
### 2.1.2 Parametrisation

Finally, we parametrise the model given by equations (3) and (4). The values of the seven parameters, along with a brief summary of their meaning, is given in Table 1.

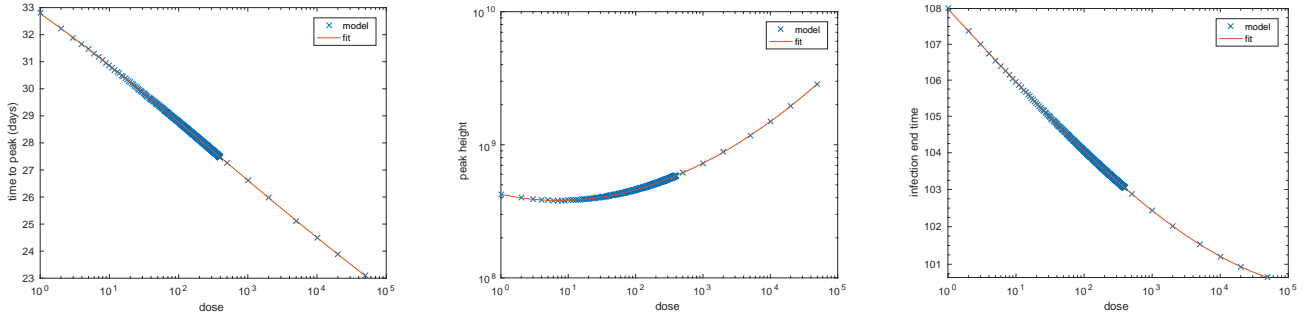
Values for the parameters  $\lambda$ ,  $G$ ,  $\gamma$  and  $\Theta_0$  are taken from two studies modelling *Mycobacterium tuberculosis* infection in the lungs [41, 42]. The remaining three parameters have been estimated based on biological principles; further discussion of our parameter estimation can be found in Appendix A.

### 2.1.3 Results

Figure 1 shows the number of extracellular bacteria (pathogen load)  $B(t)$ , number of bacteria-containing phagocytes  $P(t)$  and level of adaptive immune response  $R(t)$ , as described by equation (3), for initial dose



**Fig 2.** The number of extracellular bacteria,  $B(t)$ , for a range of initial doses  $D$ ; this shows that our within-host model is indeed dose dependent. Peak height increases with dose.



(a) Dose and time to peak are negatively correlated

(b) Dose and peak pathogen load are positively correlated.

(c) Dose and length of infection are negatively correlated.

**Fig 3.** The size of the initial dose  $D$  affects various statistics of the pathogen load curve, such as time to peak pathogen load, size of the peak pathogen load and total length of infection.

$D = 1$ . During the early stages of the infection, the effect of the adaptive immune response is negligible and so the pathogen load increases exponentially. The high pathogen load stimulates an increase in the level of the adaptive immune response, which causes a reduction in the growth rate of  $B(t)$ , resulting in a turning point. Since the decay rate of the adaptive immune response is small, the rate of decay of  $B(t)$  is also exponential.

We are most interested in how the pathogen load is affected by the size of the initial dose  $D$ . Figure 2 and Figure 3 shows how the overall shape of the pathogen load varies with the size of the initial dose  $D$ . As  $D$  increases, we observe a higher and earlier peak, followed by an earlier recovery time.

## 2.2 Transmission

We have seen that the dose an individual receives determines the infection dynamics described by equations (3) and (4). We now describe a model that uses the pathogen load  $B_I(t)$  of an infected ( $I$ ) individual to determine the transmission to a susceptible ( $S$ ) individual. Individual  $I$  is infected with dose  $D$  at time  $t = 0$ ; at time  $t = t_1$ , they begin continuously transmitting bacteria to individual  $S$  at a rate proportional to  $\log B_I(t)$ . The logarithmic scale is widely used in literature to map from pathogen load to transmission [53]. We assume that the number of bacteria lost in transmission by individual  $I$  is negligible and does not have any effect on their infection dynamics. Contact ceases at time  $t = t_2$ . This is described by the following

equations:

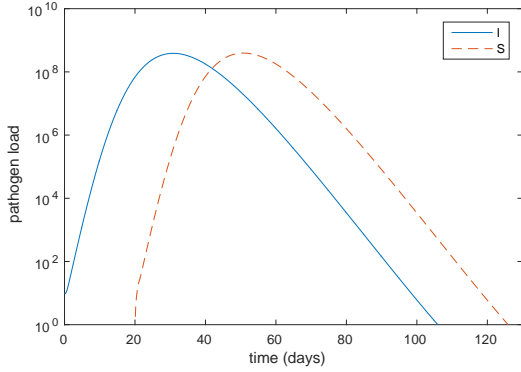
$$\begin{aligned} \frac{dB_I}{dt} &= \lambda GP_I - \gamma B_I \\ \frac{dP_I}{dt} &= \gamma \Theta_0 e^{-\rho R_I} B_I - \lambda P_I \\ \frac{dR_I}{dt} &= \beta \log [(1 - \Theta_0 e^{-\rho R_I}) \gamma B_I] - \delta R_I \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{dB_S}{dt} &= \lambda GP_S - \gamma B_S + \Gamma \log(B_I) \mathbf{1}_{t \in [t_1, t_2]} \\ \frac{dP_S}{dt} &= \gamma \Theta_0 e^{-\rho R_S} B_S - \lambda P_S \\ \frac{dR_S}{dt} &= \beta \log [(1 - \Theta_0 e^{-\rho R_S}) \gamma B_S] - \delta R_S. \end{aligned} \quad (6)$$

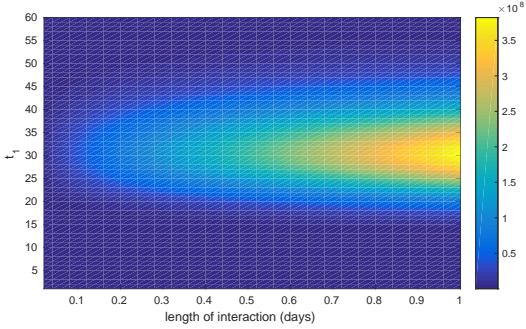
Figure 4 shows the pathogen load curves for individual  $I$  and individual  $S$  based on this model. The effect of the initial dose for individual  $S$  is accelerated by the interaction term, resulting in a higher initial growth rate.

The pathogen load of individual  $I$  at the time of interaction with individual  $S$  affects the subsequent pathogen load of this secondary infection. By considering the total area under the pathogen load curve of individual  $S$  after interaction with individual  $I$  for different intervals  $[t_1, t_2]$ , we show that the area increases with both the length of the interval  $[t_1, t_2]$  and the pathogen load of individual  $I$  in the interval. These results are shown in Figure 5. Therefore, we can assume that the transmission potential is highly dependent on

$$\int_{t_1}^{t_2} \log(B_I(t)) dt. \quad (7)$$



**Fig 4.** The pathogen load curves of an initially infected individual ( $I$ ) and initially susceptible ( $S$ ) individual over time, described by equations (5) and (6) respectively, with  $\Gamma = 1$ . Individual  $I$  is infected with a dose of  $D = 10$  at time  $t = 0$  and has continuous contact with individual  $S$  for  $t \in [20, 21]$ .



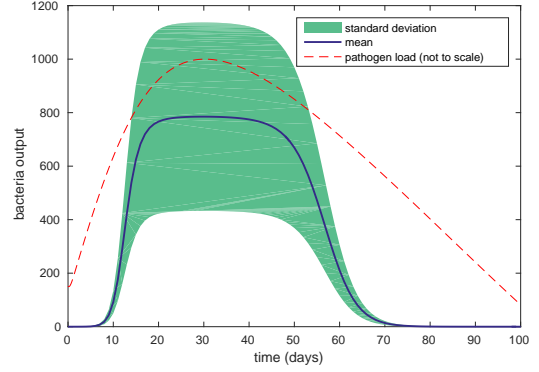
**Fig 5.** The area under the pathogen load curve of the secondary individual increases when the interaction occurs closer to the pathogen load peak of the initial infected individual and also increases when the time interval  $[t_1, t_2]$  is larger. Plot shows  $D = 10$ .

Whilst this model of transmission is useful in determining the relationships between pathogen load and transmission, it also results in individual  $S$  becoming infected even for very short contact times. This is unrealistic, as disease is not transmitted during every susceptible-infected interaction. To prevent this we redefine transmission to be stochastic.

Let  $X \in \mathbb{Z}_+$  be daily bacterial output of an infectious individual. The relationship between pathogen load and bacterial output has been shown experimentally to have a sigmoidal shape [55, 56], where the period of significant increase in output corresponds to the onset of symptoms. Therefore, we let  $\mu_X = \mathbb{E}[X]$  have this sigmoid shape:

$$\mu_X = \frac{A \max_t \log B(t)}{1 + \exp \left[ - \left( \int_t^{t+1} \log B(s) ds - b \right) \right]}. \quad (8)$$

The parameter  $b$  is chosen such that high values for  $\mu_X$



**Fig 6.** The mean and standard deviation of the daily bacterial output is dependent on the pathogen load. Plot shows  $D = 20$

occur around the peak of the infection; we take  $b = 15$ . The parameter  $A$  is chosen such that the range for  $\mu_X$  covers a range of realistic doses; moreover, we choose  $A$  large enough such that infection is possible even for very short contact times. We take  $A = 40$ . We then let  $X \sim \text{NegBin}(r_X, p_X)$ , where  $r_X = 5$  and  $p_X = \mu_X / (r_X + \mu_X)$ . We have chosen to use the negative binomial distribution since it has two parameters, which allows us to control the variance of the distribution through our choice of  $r_X$ ; the variance decreases as  $r_X$  increases. This allows for the possibility of a large range of bacterial outputs. The mean and variation in the daily bacterial output over the course of the infection is shown in Figure 6.

We then let  $Y \in \mathbb{Z}_+$  be the number of bacteria inhaled by a susceptible individual who spends  $c$  hours with an infectious individual on a given day. We assume that the average number of bacteria inhaled by this individual is proportional to the length of time spent with the infectious contact and their bacterial output: so  $\mu_Y = \frac{c}{24} X$ . We then let  $Y \sim \text{NegBin}(r_Y, p_Y)$ , where  $r_Y = 0.1$  and  $p_Y = \mu_Y / (r_Y + \mu_Y)$ . Our choice of  $r_Y$  reflects the significant variation that will occur in the number of bacteria inhaled; in this way it is possible for an individual to inhale a large number of bacteria, even for small duration of contact  $c$ .

Finally, we let  $D \in \mathbb{Z}_+$  be the number of inhaled bacteria that are deposited in the lungs. Many bacteria will be cleared from the respiratory tract before they reach the lungs by the action of cilia in the trachea or by the mechanisms of coughing or sneezing [13]. We make the assumption that bacteria act independently, so  $D \sim \text{Bin}(Y, p_D)$ , where the probability of deposition is  $p_D = 0.124$  as given in [13]. It is still possible that the infection does not take off if all bacteria are successfully phagocytosed. The probability that a single bacterium survives phagocytosis at the point of infec-



tion is  $\Theta_0 = 0.3$ , and thus the probability that the infection takes off is given by the probability that at least one bacteria survives phagocytosis:

$$\mathbb{P}[\text{infection}] = 1 - (1 - \Theta_0)^D. \quad (9)$$

Through these stages we can calculate the deposited dose  $D$  received by a susceptible individual spending  $c$  hours on day  $t$  with an infected individual with pathogen load curve described by  $B(t)$ .

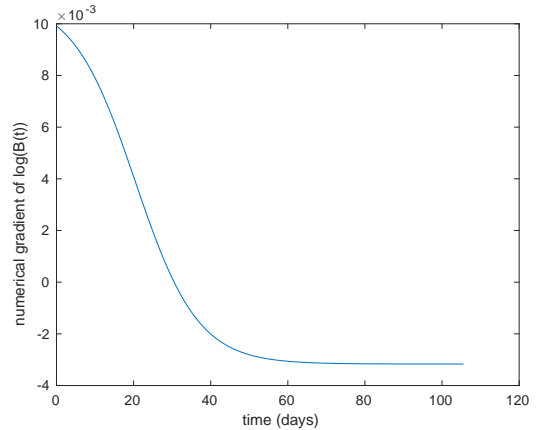
### 2.3 Curve approximations

The final stage of the model considers the effect of the dose-dependent within-host model on the population level dynamics of an epidemic. Our method of transmission requires knowledge of an individual's pathogen load over the entire course of the infection and the exact solution of the system of ODEs is dependent on the initial dose  $D$ . Numerically solving the ODEs for each new infection is computationally expensive and storing the solution over the full range of possible initial conditions is memory expensive. For a population of thousands of individuals both approaches become infeasible. This is the motivation for approximating the pathogen load curve by a function whose parameters are some function of the known initial dose; in this way we are able to generate the pathogen load of an individual from  $D$  alone and thus efficiently simulate epidemics even for large populations.

Due to the qualitative shape of the pathogen load curve, we initially looked at comparing the model output to the probability density functions of the beta and log-normal distributions. This required two scaling parameters, infection end time and total area under the curve, plus two parameters for each distribution; the scaling parameters were taken directly from the model results, whilst the distribution parameters were fitted using the mean and variance of the model results. Whilst the log-normal distribution exhibits a marginally better fit than the beta distribution when compared using a log-likelihood ratio test, the approximation still drastically underestimates the pathogen load during early infection. Detailed fitting methods and outcomes are described in Appendix B.

#### 2.3.1 Sigmoidal approximation

The inaccurate fit of both the beta and log-normal approximations motivate the construction of our third and final approximation, which captures the non-zero initial conditions, exponential growth and decay in the tails and the correct height of the peak that we observe in plots of the pathogen load.



**Fig 7.** Gradient of the log of the pathogen load with initial dose of 15

We plot  $f(t) = \frac{d}{dt} \log(B(t))$  and observe that the shape is approximately sigmoidal:  $f(t)$  starts at a positive constant and slowly switches to a second, negative constant. These three stages correspond to exponential growth at the beginning of the infection, the peak, and exponential decay in the tail of the infection. Figure 7 shows a plot of  $f(t)$  for the pathogen load curve  $B(t)$  with  $D = 15$ .

We construct an approximation  $\tilde{f}(t)$  for  $f(t)$ ; our approximation of  $B(t)$  is then defined by:

$$\frac{d}{dt} \log(\tilde{B}(t)) = \tilde{f}(t) \iff \tilde{B}(t) = Ae^{\tilde{F}(t)}, \quad (10)$$

where  $\tilde{F}(t) = \int \tilde{f}(t) dt$  is an anti-derivative of  $\tilde{f}(t)$ .

Our approximation  $\tilde{f}(t)$  should be constant in the limit as  $t \rightarrow \pm\infty$  and slowly transition between these constants for some specified range of  $t$  at some specified rate. Each of these properties will require parametrisation. In a similar fashion to the construction of the beta and log-normal approximations, we use statistics of the pathogen load curves to estimate these parameters for  $\tilde{f}(t)$ . The following section describes the structure and parametrisation of  $\tilde{f}(t)$ .

We define

$$\tilde{f}(t) = \frac{\alpha - \beta}{1 + e^{\gamma(t-t_p)}} + \beta, \quad (11)$$

where  $\alpha, \beta, \gamma$  and  $t_p$  are parameters to be fitted. Note that  $\tilde{f}(t)$  has the property that  $\lim_{t \rightarrow -\infty} \tilde{f}(t) = \alpha$  and  $\lim_{t \rightarrow +\infty} \tilde{f}(t) = \beta$ ; so  $\alpha$  and  $\beta$  represent the exponential growth and decay respectively. The parameters  $\gamma$  and  $t_p$  characterise the curvature and time of the peak of  $B(t)$  respectively. We find values for  $\alpha$  and  $\beta$  by numerically solving for  $f(t)$  through the points where the gradient is constant;  $\gamma$  and  $t_p$  are then found by minimising the absolute difference in the given load curve, and the approximated load curve.



Finally we define the constant  $A$ ; we do this in terms of the peak pathogen load  $M = \max_t B(t)$ . We define  $T_M$  as the time at which the pathogen load is maximal and so  $M = B(T_M)$ . If  $\tilde{B}(t)$  is to be a good approximation to  $B(t)$  then we must also have  $M = \max_t \tilde{B}(t) = \tilde{B}(T_M)$ . So  $A$  can be written in the following form:

$$M = \max_t \tilde{B}(t) = \tilde{B}(T_M) = Ae^{\tilde{F}(T_M)} \iff A = Me^{-\tilde{F}(T_M)} \quad (12)$$

Therefore we can write  $\tilde{B}(t)$  as follows:

$$\tilde{B}(t) = Me^{\tilde{F}(t) - \tilde{F}(T_M)}. \quad (13)$$

In addition, we can find an expression for  $T_M$  by solving  $\tilde{f}(T_M) = 0$ :

$$\begin{aligned} \tilde{f}(T_M) = 0 &\iff \frac{\alpha - \beta}{1 + e^{\gamma(T_M - t_p)}} + \beta = 0 \\ &\iff T_M = t_p + \frac{1}{\gamma} \log\left(\frac{-\alpha}{\beta}\right) \end{aligned}$$

The sigmoidal approximation to  $B(t)$  is characterised by five parameters:  $\alpha, \beta, \gamma, t_p$  and  $M$ . In the final step of constructing an approximation to  $B(t)$ , we characterise the relationship between the initial dose  $D$  and each of the five parameters. Cubic curves provide a good fit to the model results in all cases; this is shown in Figure 18 in Appendix C. Thus, given only the initial dose  $D$ , we can fully define a sigmoidal approximation to the pathogen load curve  $B(t)$ .

## 2.4 Population level model

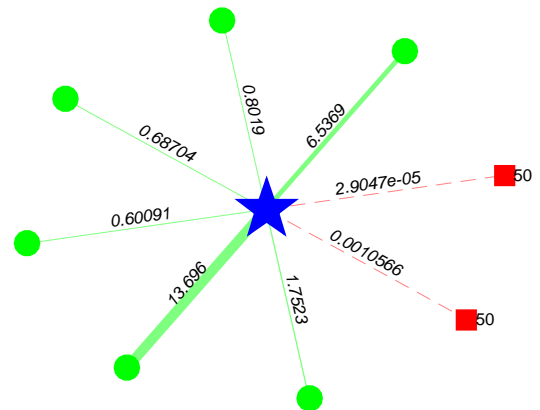
Homogeneous mixing of the population is one of the most unrealistic assumptions of epidemic models, since in reality individuals only meet a subset of the population; moreover, the number of contacts and frequency and duration of contact varies significantly from one individual to the next. The usual solution to this is to place individuals on the nodes of a network, with edges between pairs of individuals who are contacts. However, quantifying and constructing a realistic social contact network is a significant challenge beyond the scope and time-scale of this project. Instead, the population level model is split into two parts: in the first part, we use the results of a large-scale survey of social encounters in order to construct contact networks for single individuals and calculate an estimate for  $R_0$ , the average number of secondary cases generated by an average infected individual in a totally susceptible population. In the second part, we simulate the epidemic on three different network types in order to investigate the effect of initial dose on the final size of the epidemic.

### 2.4.1 Contact survey data and ego networks

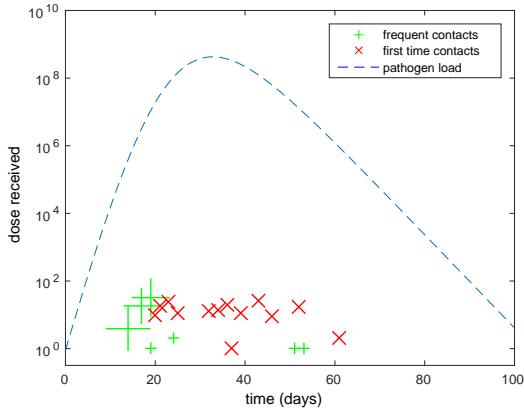
The contact survey data is collected from a large-scale postal and online survey of social encounters conducted in 2009 [57]. In total the survey had 5,388 respondents. Participants were asked to record individuals with whom they had physical or face-to-face conversational contact with over the course of a single day. In order to minimise restrictions on the total number of contacts that can be reported, individuals were able to report groups of contacts who they met in a similar way and for a similar length of time: for example, a shop worker can report the 50 individuals that they meet each for one minute in one group, rather than list each one separately. For each contact or group of contacts, participants are asked to report information about the duration, frequency and location of contact and whether physical contact was made; respondents can also indicate whether they think each pair of contacts have met in the last week, to give information on clustering in the network.

For each respondent we are able to construct a simplified so-called ego network: this network represents an individual's (the ego's) contacts, including the information about the frequency and duration of contact; however edges between contacts are not included. In the survey, participants can report a range of frequency of contacts from every day to first time; we make a further simplification and classify contacts as either frequent or first time. The ego network data is used for two functions: in order to visualise the ego networks and as a way to investigate the effect of initial dose on  $R_0$ .

First, we simply use the data to visualise the networks; such an example is given in Figure 8. The ego is represented by a star at the centre of the network with their contacts spread around it. Contacts that are met



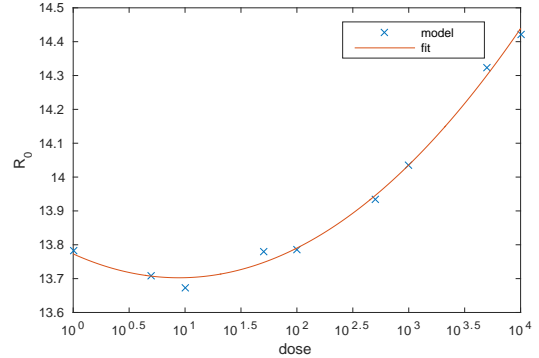
**Fig 8.** An example visualisation of a single ego network.



**Fig 9.** An infected individual generates multiple secondary cases over the course of their infection. Crosses indicate the time and size of each secondary case generated, split into frequent and first time contacts; the size of the cross is proportional to the duration of contact. The pathogen load of the ego is included for reference.

every day are represented by green nodes with a solid edge to the ego; first time contacts are represented by red nodes with a dashed edge to the ego. Individual contacts are represented by circles; groups of contacts are represented by squares, with a number to indicate the group size. Weights on the edges give the duration of contact, given as the number of hours spent with that contact, or each individual in the group, per day. For visual clarity, the thickness of the edge between the ego and the contact is proportional to the duration of contact.

Secondly, we run a simulation to calculate an estimate of  $R_0$  for a range of initial doses; this is one method to determine whether there is dose-dependent behaviour at the population level. The ego in each ego network is infected with the same initial dose  $D$ ; all their contacts are initially susceptible. At each time step, the ego meets their frequent and first time contacts. The ego meets a new set of first time contacts at each time step and thus all first time contacts begin the day as susceptible; so an ego reporting one first time contact for 5 minutes will meet 100 such contacts over the course of a 100 day infection. On the other hand, frequent contacts can be either susceptible or infected due to a previous encounter; since there is little variation in the duration of infection, we assume that such contacts do not recover over the course of the ego's infection and thus cannot become reinfected. At each time step, doses are generated for each susceptible contact and infection occurs according to the model for transmission described in Section 2.2. Over the course of the ego's infection, we count the number of



**Fig 10.** The relationship between  $\log D$  and  $R_0$  is quadratic, that is  $R_0 \propto [\log D]^2$ . For  $D > 10$ , larger doses result in a larger  $R_0$ . The estimates for  $R_0$  are calculated as the average of twenty-five simulations of the model across all the 5276 ego networks with at least one contact, for a range of initial doses.

their contacts that become infected. We repeat this method 25 times for each of the ego networks and take the mean number of secondary cases generated over the 25 repetitions; we calculate an estimate for  $R_0$  by taking the mean over all the ego networks.

For each individual, the number of secondary cases generated over the course of the infection depends on the number of contacts they have and the frequency and duration of contact. Figure 9 is a single realisation of the secondary cases generated by an individual over the course of their infection. In general, frequent contacts are infected earlier than first time contacts, since the duration of contact is longer for the former, and larger doses are received by individuals with a longer duration of contact. However, due to the stochasticity of the model, it is possible for frequent contacts to be infected later in the infection, or for individuals with a short duration of contact to receive a larger dose.

We use the results of the simulation to calculate an estimate for  $R_0$  for a range of doses; results are given in Figure 10. The relationship between  $\log D$  and  $R_0$  is quadratic, that is:

$$R_0 \propto [\log D]^2. \quad (14)$$

The value of  $R_0$  is minimal for  $D \approx 10$ ; this dose coincides with the minimum peak pathogen load (see Figure 3b). The increase in the value for  $R_0$  with increased dose demonstrates that the model is dose-dependent even on the population scale.

#### 2.4.2 Full network simulation

Another important quantity to calculate in epidemic modelling is the final size distribution, the total number of people infected over the course of the epidemic.

By running our model on three different network structures, lattice, scale-free and small-world, we can look at the effect of the initial dose for the first infected individual on the final size of the epidemic. Whilst none of the networks that we use fully capture the structure social contact network, each demonstrates at least one observable property of real-world networks; where possible, we generate a network with the property matching the contact survey data.

Lattice graphs are a good approximation for social contact networks where contact is determined by social proximity. An  $n \times m$  lattice graph has as its nodes the points in the plane with integer coordinates  $(x, y)$ ,  $1 \leq x \leq n, 1 \leq y \leq m$ ; nodes are connected to their nearest four or nearest eight neighbours.

The second type of network that we consider is a scale-free network; a scale-free network is a network whose degree distribution follows a power law, that is the probability  $p_k$  that a randomly chosen node has degree  $k$  is

$$p_k \propto k^{-\alpha}, \quad 2 < \alpha < 3.$$

Power law distributions have been used to describe human contact networks in both virtual settings, such as the sizes of email address books amongst computer users at a large university [58], and in physical settings, such as the number of sexual contacts across a twelve month period [59]. Whilst the degree distribution of the contact survey data is not fully described by a power law, it is at least a good approximation for the tail of the distribution. We use the Barabási-Albert algorithm to generate a scale-free network whose power-law degree distribution has the same exponent as the tail of the contact survey data degree distribution. For each network generated, we begin with a complete network on  $m_0 = 40$  nodes; each new node added is connected to  $m = 13$  existing nodes by preferential attachment. These values of  $m_0$  and  $m$  are chosen such that the average degree is 26, the average number of contacts given by the contact survey data.

The final type of network that we consider is a small-world network. Small-world networks possess both a high degree of clustering and short average path lengths; graphs that have both of these properties are said to have the small-world property. We use the Watts-Strogatz algorithm to generate small-world networks: we begin with a 26-regular ring lattice and independently rewire each edge with probability  $p = 0.05$ ; this value of  $p$  generates a network with short average path length and a high degree of clustering. The resulting network has mean degree 26, the same as the average degree given in the contact survey data.

For each network type, we generate a social contact

network for 500 individuals and with each edge  $(i, j)$  we associate a weight that corresponds to proportion of a day that individual  $i$  and  $j$  spend together. Individuals can be in one of three states: susceptible, infected or recovered. At the beginning of the simulation, we randomly choose a single individual to be infected with dose  $D \leq 50000$ ; all other individuals are susceptible. At each time step and for each currently infected individual we generate the deposited doses for each of their susceptible neighbours, as described in Section 2.2. Susceptible individuals are able to accumulate dose from multiple infected neighbours. When transmission is complete, the states of individuals are updated. Susceptible individuals with total deposited dose  $d$  become infected with probability  $1 - (1 - 0.3)^d$ ; infected individuals recover when their bacterial load at the end of the time step is less than one and the simulation ends when there are no more infected individuals. The final size of the epidemic is the total number of individuals infected over the simulation; for each network structure and each initial dose  $D$  we average the results over 50 simulations.

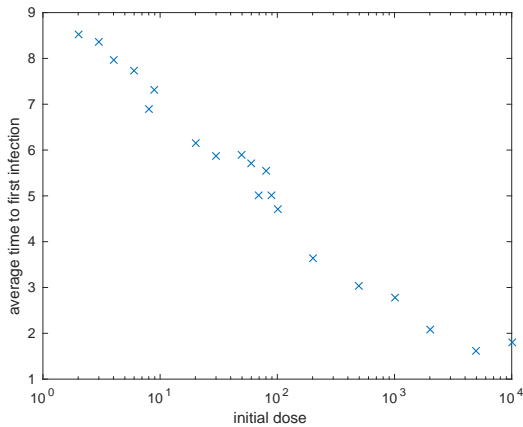
**Table 2.** Correlation coefficients for initial dose and final size of the epidemic by network structure.

Network type	Correlation coefficient
Lattice, 4 neighbours	-0.1280
Lattice, 8 neighbours	0.0383
Scale-free	-0.2455
Small-world	0.3873

For each network, we calculate the correlation coefficient of dose and average final epidemic size; the results are given in Table 2.

For both the 4- and 8-neighbour lattices there is little to no correlation between initial dose and final epidemic size. Although the scale-free network exhibits moderate negative correlation, the final size of the epidemic varies very little: for all doses, we find that the average final epidemic size is at least 499. In this way, the size of the initial dose gives no additional information about the final size of the epidemic. This result is due to the fact that scale-free networks are well connected and thus the infection can easily spread across the entire population. A similar result holds for small-world networks. Whilst there is no correlation between initial dose and final epidemic size for the lattice, we find that the initial dose and the time to the first secondary infection are negatively correlated; this result is shown in Figure 11.

Our network level epidemic model assumes that individuals have a fixed number of contact who they see for the same amount of time every day; this is in



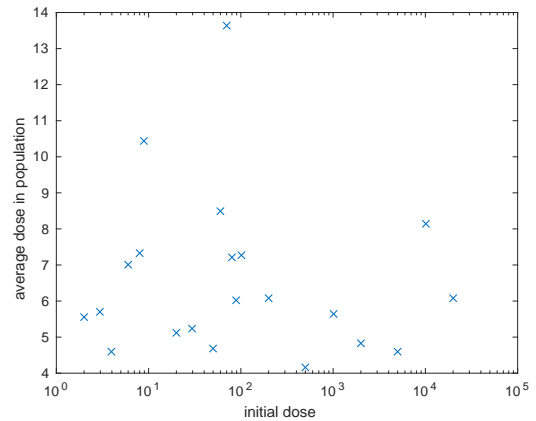
**Fig 11.** Initial dose against time to first secondary infection on an 8-neighbour lattice.

contrast to the contact survey data, where individuals report meeting different contacts at different frequencies. As a result, an individual likely infects their contacts before they reach the peak of their infection, with smaller doses on average than if they were to be infected at the peak. As a result, the initial dose and the average dose received by the secondary cases is uncorrelated; see Figure 12.

### 3 Discussion

In this project we have presented a multi-scale model for dose-dependent infection dynamics. By assuming an individual's bacterial output, and consequently the number of bacteria they transmit to their contacts, depends on their pathogen load [51], we explore the effect of initial dose at the population level. Much current research aims to improve the link between pathogen load and transmission, although methods are still not currently well-developed for bacteria [60]. We contribute to the existing literature by focusing on this link for intracellular bacterial infections.

Results from our within-host model use a set of deterministic ODEs to describe the infection dynamics, which show a clear dependence on the dose, with an increased dose resulting in a higher and earlier peak pathogen load for doses greater than ten (Figure 3b). We use a deterministic model over a stochastic model to give the average pathogen load profile for an individual. We note that this does have the disadvantage of individuals having identical infection dynamics for a specified dose, whereas in fact there will be a complex heterogeneity between the different individuals [61]. However, we can accurately capture the expected effect of dose-dependence without the results being obscured



**Fig 12.** We find no correlation between the initial dose and the average dose received by secondary cases.


by stochastic noise.

We additionally simplify our within-host model to a single count of pathogen load in the entire body of the individual and do not consider migration of the disease around the body. This formulation reduces the number of parameters required to simulate the model, but future work would hope to incorporate a compartmental model. The pathogen load in the lungs and other organs would be modelled separately, since it will be the pathogen load in the lungs that will have the greatest effect on the transmission; the effects of the infection dynamics in other organs will vary [62]. In future analysis, examining the effects of medical countermeasures on the pathogen load could also prove insightful to how they change the course of the infection when administered at different times and hence how this would impact the onward transmission. This would also be of a greater real-world importance in improving the effect and timing of medical intervention and hence public health.

Further considerations would include the effect of interacting infections, where a second infection is superimposed on an earlier one, which has already caused an immune response. This is known as a superinfection. For example, in the case of HIV infections, an increase in the pathogen load can accelerate the progression towards AIDS [63]. It would be interesting to investigate how reinfection with an intracellular bacterial disease, such as tuberculosis, affects the immune response, as well as reactivation of the latent disease [64]. We also could consider pathogen fitness and evolutionary processes [16].

A major difficulty in the development of multi-scale models is that simulations in full mechanistic detail can quickly become computationally intractable when the size of the model becomes large [60]. We incorpo-

---



rate our two scales in the model construction by using curve approximations for the pathogen load curve, as opposed to numerically solving ODEs at every step, since this is our most computationally expensive stage. In other models, realism is compromised in favour of tractability through simplification of the within-host model or using smaller scales [65].

Parameters for our model have been taken predominantly from literature of *Mycobacterium tuberculosis* infections and through mathematical estimation to give biologically relevant results. The lack of experimental parametrisation is particularly notable for the transmission of the infection. This is in part due to the fact that transmission potential is difficult to quantify due to the considerable heterogeneity in the transmission potential between infected hosts [66], but also because there have been few experiments to measure transmission. Handel and Rohani [40] present a potential experimental set-up, whereby an infected animal would on a daily basis be presented with new animals, which would be removed at the end of the day. If measurements were taken for the pathogen load, symptoms, immune response and the behaviour of the host, this data could then be used to define a quantitative function between infection dynamics and infectiousness. However, until such an experiment is carried out theoretical advances in the field will likely be speculative.

In scaling the model up to the population level, we consider the explicit ego networks from participants of a contact survey [57]. Using the data available from the survey provides a realistic contact experience from which we calculate the number of secondary cases from one individual. However, the data does not accurately capture the contact experiences between the other contacts of the ego. Ideally, we would be able to include network statistics such as node degree and clustering coefficient and use this to implement a complete and realistic network for contacts. The exact method by which this would be implemented is an open problem [67], but we have produced some distributions for relevant features such a network should contain; see Appendix D.

The results of calculating values for  $R_0$  with different doses concur that a higher dose induces infection dynamics that initiate higher transmission. However, the range of values is relatively small, with a range of less than one, for doses as different as 1 and 10000. This is perhaps not surprising however, as while dose

has a clear effect on infection dynamics for diseases spread by direct transmission, we would expect little feedback from the infection dynamics on to the infection dynamics of new cases. A disease that is spread by environmental spores, for example, would have far greater reciprocal feedback between the levels of organisation. This is because the infection dynamics will directly affect the number of infective spores in the environment, which will aggregate and thus directly affect dose size [68]. Hence, since we are modelling an intracellular bacterial infection, we do not have such a direct feedback in our model and so we should not expect a strong secondary case dependence on dose. A similar result applies to the full network model.

Finally, we note that our model is effective at allowing frequent contacts not to be infected immediately and indeed sometimes not infected at all, due to the stochastic transmission. Our model can also display cases where very minimal one-time contacts result in infection, while there are much higher probabilities of secondary individuals becoming infected if they meet an infected individual for longer or when their infection is at its peak pathogen load (Figure 9).

## 4 Conclusion

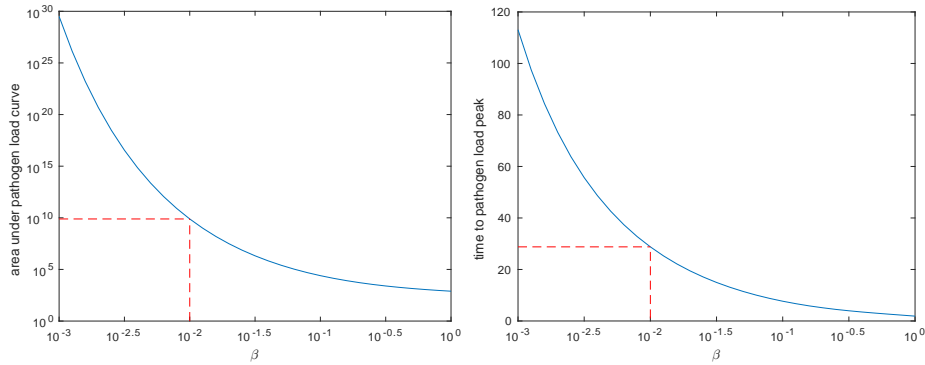
The modelling process undertaken here has served to demonstrate the challenges associated with developing practical multi-scale models of disease, whilst also providing an some insights into the implications of within-host dynamics for outbreaks of highly dose-dependent infections. We have shown that the positive relationship between initial dose and peak size, particularly when combined with varying contact weights, can result in highly heterogeneous transmission. Simulated epidemics where the source cases receive different infecting doses also display differing characteristics, such as higher values of  $R_0$  for larger doses. Although these effects are small relative to changes in dose size, they could still have important ramifications for the design of public health strategy and response; even a small change in  $R_0$  could result in a shift in the number of vaccinations or treatments required for disease control. In providing a general model of intracellular bacterial infections we have highlighted the need for further disease-specific work on this topic, including experimental studies for better parametrisation, to quantify these effects for targeted public health objectives.

---

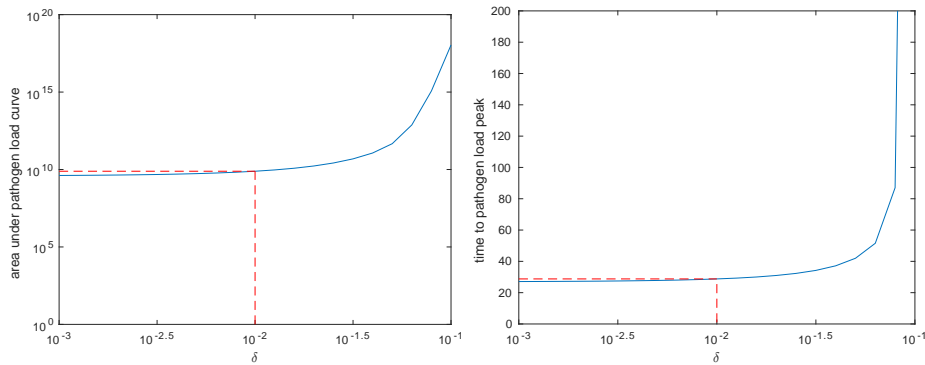
# Appendices

## A ODE parameter estimation

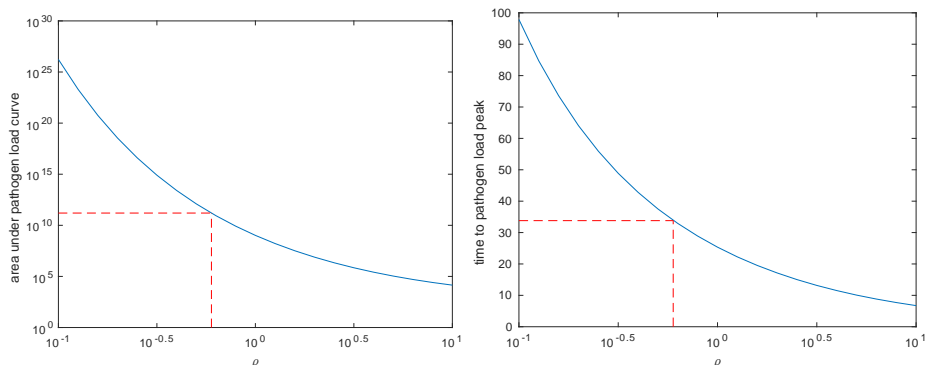
Using parameter values in Table 1, Figures 13, 14 and 15 show how  $\beta$ ,  $\delta$  and  $\rho$  affect the area under the pathogen load curve and the time to reach the pathogen load peak. We expect the area under the pathogen load curve and the time to pathogen load peak to be approximately  $10^{10}$  extracellular bacteria and 30 days respectively and so we choose our parameter values accordingly.



**Fig 13.**  $\beta = 0.01$  is chosen to give biologically realistic output.



**Fig 14.**  $\delta = 0.01$  is chosen to give biologically realistic output.



**Fig 15.**  $\rho = 0.8$  is chosen to give biologically realistic output.

---

## B Beta and log-normal approximations

The linear plot of the pathogen load, given in Figure 2, is reminiscent of the density function of a probability distribution; our first approximations to the pathogen load curves use appropriately scaled beta and log-normal density functions.

Our approximations use four statistics of the pathogen load curves: the mean, variance, time to end of infection and area under the curve over the whole infection. These can be calculated directly from the pathogen load curves. The mean and variance are used to determine the parameters the distribution whilst the time to end of infection and area under the curve are used to scale the distribution. The result of the approximation will be four functions from the initial dose  $D$  to each of the two parameters of the distribution, the time to end of infection and the area under the curve. We now describe our method for the beta approximation.

The beta distribution has support on the unit interval and is parametrised by two positive shape parameters  $\alpha$  and  $\beta$ . The mean and variance of the beta distribution can be written in closed form in terms of  $\alpha$  and  $\beta$ ; if  $X \sim \text{Beta}(\alpha, \beta)$  then:

$$\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta} \quad (15)$$

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (16)$$

The pathogen load curve must be scaled to have support on the unit interval; to do this we scale the x-axis (time) by the time to the end of infection. We also must scale the curve to have an integral of 1 over the support; to do this we scale the y-axis (pathogen load) by the area. We calculate the mean and variance of the scaled pathogen load,  $\bar{B}_D$  and  $s_D^2$ , and use these as approximations for the true mean and variance of the beta distribution. We then solve the following simultaneous equations for  $\alpha$  and  $\beta$ :

$$\begin{aligned} \bar{B}_D &= \frac{\alpha}{\alpha + \beta} \\ s_D^2 &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \end{aligned}$$

to derive the following values for  $\alpha$  and  $\beta$

$$\begin{aligned} \alpha &= \bar{B}_D \left( \frac{\bar{B}_D(1 - \bar{B}_D)}{s_D^2} - 1 \right) \\ \beta &= (1 - \bar{B}_D) \left( \frac{\bar{B}_D(1 - \bar{B}_D)}{s_D^2} - 1 \right). \end{aligned}$$

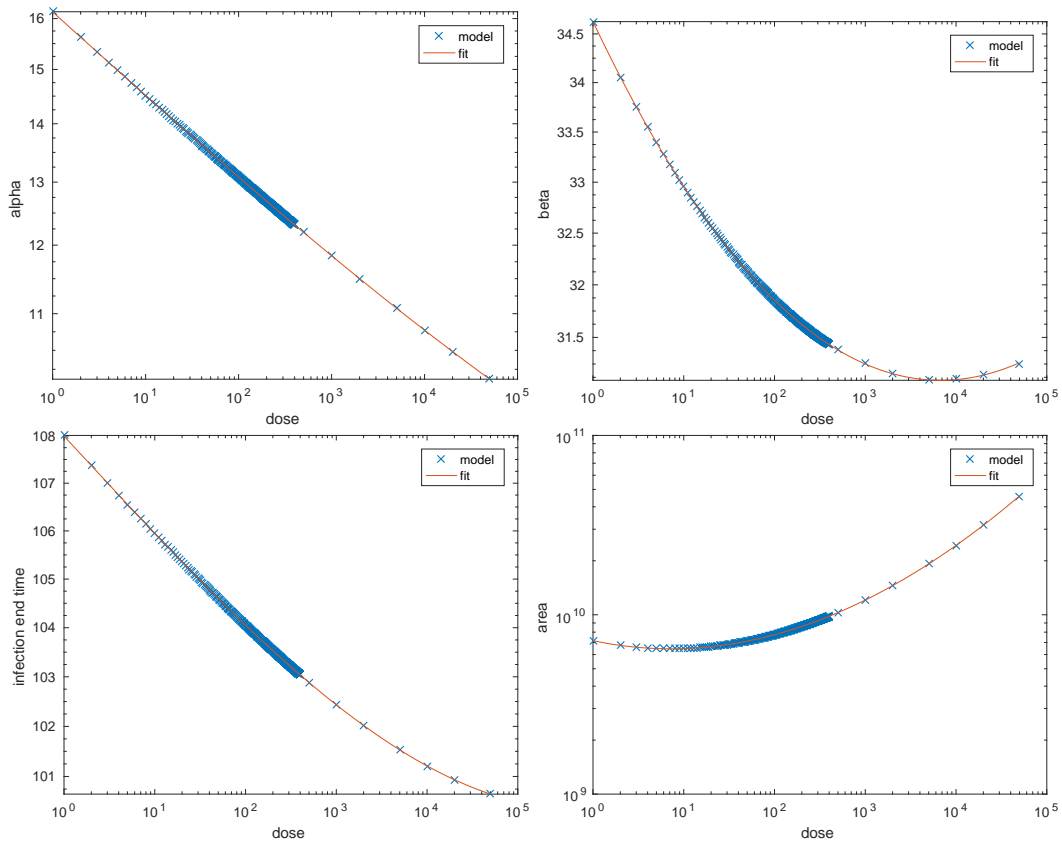
Figure 16 shows, on a log-log scale, the relationship between the initial dose  $D$  and each of  $\alpha$ ,  $\beta$ , time to end of infection and area. Cubic curves are good fit to the points. Thus, given only the initial dose  $D$ , we are able to fully define a beta distribution approximation to the pathogen load curve.

We use an almost identical method to fit the log-normal distribution, but with no scaling on the  $x$ -axis since the log-normal has support on the positive real line. The mean and variance of the log-normal distribution can also be written in closed form in terms of the two parameters  $\mu$  and  $\sigma$ : for  $Y \sim \ln \mathcal{N}(\mu, \sigma^2)$ ,  $\mathbb{E}[Y] = e^{\mu + \sigma^2/2}$  and  $\text{Var}(Y) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$ .

Qualitatively, the beta and log-normal approximations capture the overall shape of the pathogen load; see Figure 17. However, neither approximation accurately represent the time or size of the peak pathogen load or the tails of the infection; this can be seen from Figure 17. The beta approximation underestimates the pathogen load over the majority of the infection; in particular, the pathogen load is slightly underestimated at the peak and more significantly underestimated in the tail of the infection. As a result, the beta approximation severely underestimates the length of the infection by approximately 20 days.

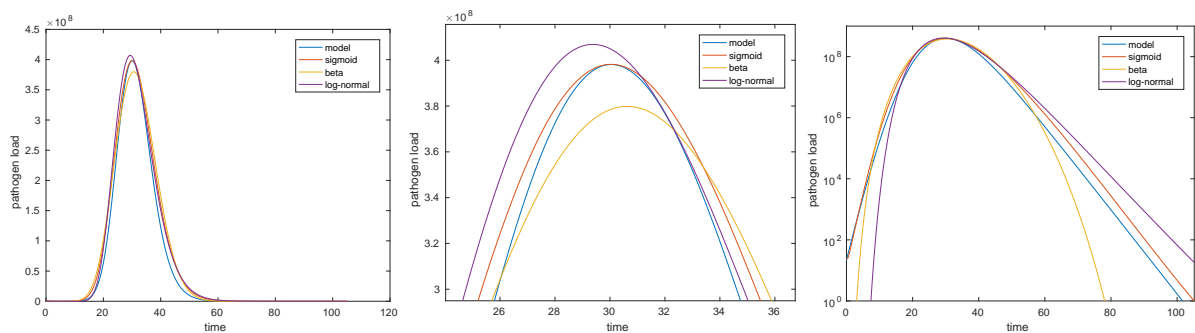
The log-normal approximation is a marginally better fit, at least qualitatively, but underestimates the pathogen load early in the infection and then overestimates both the peak and tail of the infection. As a result, the log-normal approximation overestimates the length of the infection by approximately 20 days.





**Fig 16.** The relationship between the initial dose  $D$  and each of  $\alpha$ ,  $\beta$ , time to end of infection and area for the beta approximation.

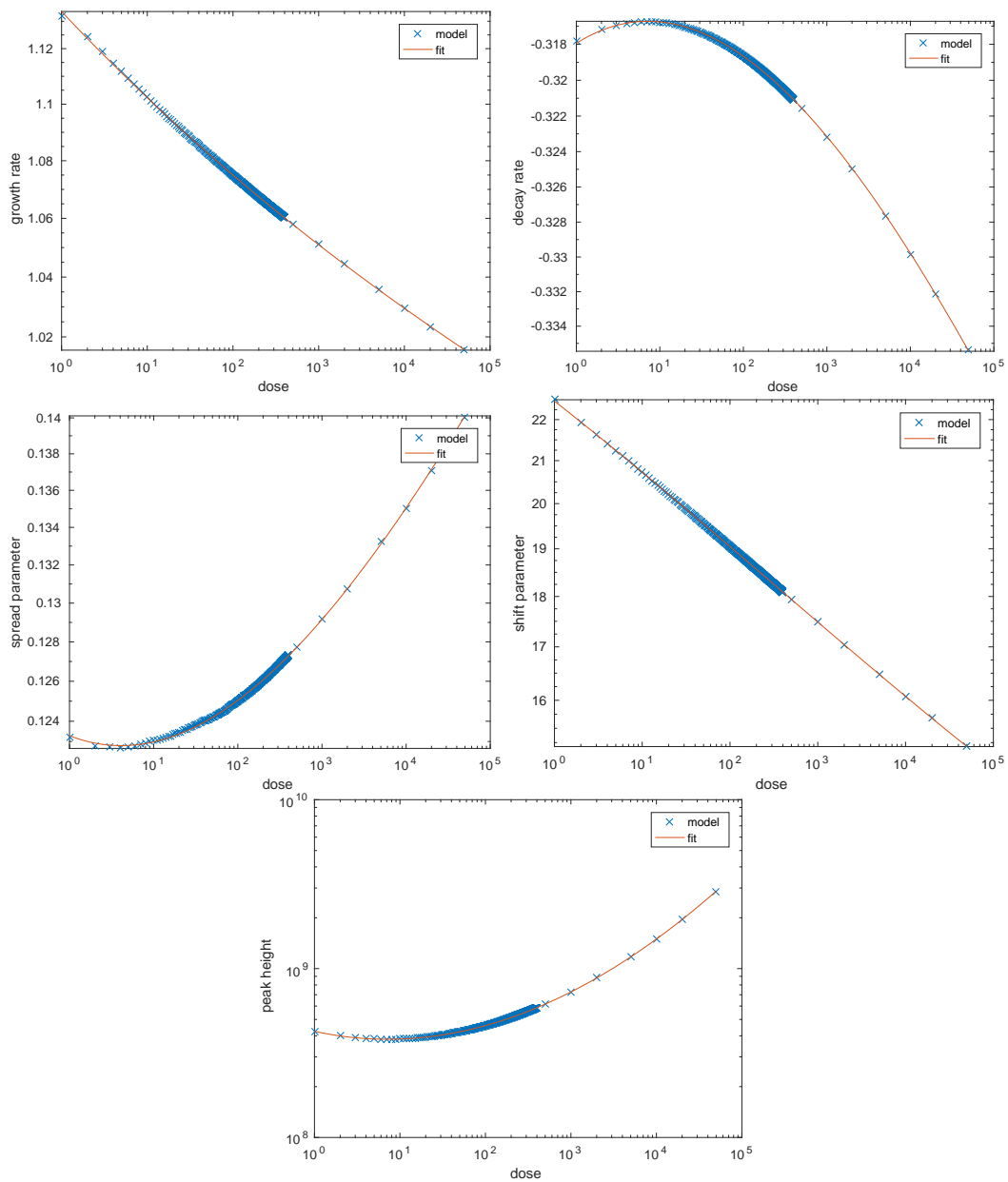
The inaccurate fit of both the beta and log-normal approximations motivate the construction of our third and final approximation, which captures the non-zero initial conditions, exponential growth and decay in the tails and the correct height of the peak.



**Fig 17.** A qualitative comparison of the different approximation function

---

## C Sigmoidal approximation parameters

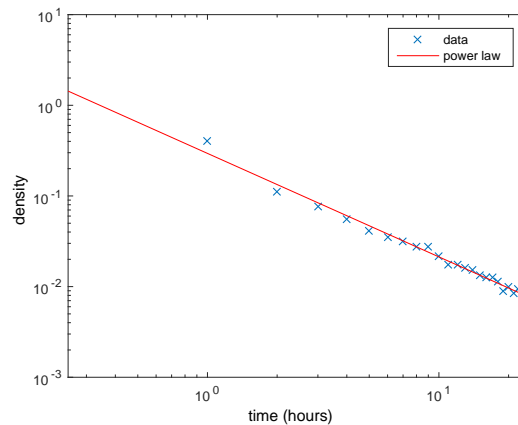


**Fig 18.** Five plots on a log-log scale showing the relationship between the initial dose  $D$  and each of  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $t_p$  and  $M$  for the sigmoidal approximation. Each plot is fitted with a cubic curve that maps the initial dose  $D$  to the parameter.

## D Contact survey

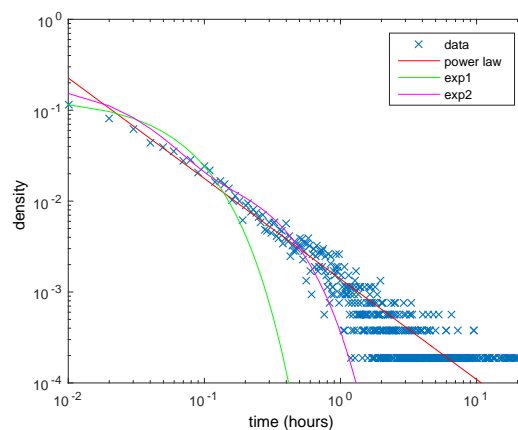
Danon *et al.* [57] fit a double Pareto log-normal (dPIN) distribution to the degree distribution of the contact survey data and quantify the clustering in ego networks. Two additional statistics that may be useful for the construction of a social contact network based on the contact survey data are the proportion of an individual's contacts that are frequent versus first time contacts, and the distribution of contact times for both frequent and first time contacts. We calculate that on average, the proportion of contacts that are frequent contacts is  $p = 0.7852$ , and the proportion that are first time contacts is  $1 - p = 0.2148$ .

The duration of contact for frequent contacts is well described by a power law with exponent  $\alpha = -1.1416$ ; see Figure 20. R-square goodness of fit statistic for this is  $R = 0.9872$ ; this means that over 98% of the variation in the data is explained by the power law fit.



**Fig 19.** A power law with exponent  $\alpha = -1.1416$  is a good fit to the distribution of contact times for frequent contacts.

We fit three distributions to the first time contact duration: a power law with exponent  $\alpha = -1.1054$ , a single exponential model (exp1), described by the equation  $y = Ae^{ax}$  with  $A = 0.1151$  and  $a = -17.36$ , and a double exponential model (exp2), described by the equation  $y = Be^{bx} + Ce^{cx}$  with  $B = 0.1287$ ,  $C = 0.02416$ ,  $b = -37.51$  and  $c = -4.25$ . The R-square statistics, adjusted for the number of parameters in each fit, are  $R = 0.9118$ ,  $R = 0.9416$  and  $R = 0.9888$ . Therefore, according to the R-square statistic, the double exponential model provides the best fit to the data, even when adjusted for the number of parameters.



**Fig 20.** A power law distribution, a single exponential model (exp1) and a double exponential model (exp2) are fit to the distribution of contact times for first time contacts. According to the R-square goodness of fit statistic, the double exponential model is the best fit to this data.

---

## References

1. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis*. 2007;195(7):951–959.
2. Stein RA. Super-spreaders in infectious diseases. *IJID*. 2011;15(8):510–513.
3. Wang SX, Li YM, Sun BC, Zhang SW, Zhao WH, Wei MT, et al. The SARS outbreak in a general hospital in Tianjin, China – the case of super-spreader. *Epidemiol Infect*. 2006;134(4):786–91.
4. Defra. Foot and mouth disease control strategy for Great Britain. 2011;November.
5. Zelnor JL, Lopman BA, Hall AJ, Ballesteros S, Grenfell BT. Linking time-varying symptomatology and intensity of infectiousness to patterns of norovirus transmission. *PLoS ONE*. 2013;8(7):1–8.
6. Wood RM, Egan JR, Hall IM. A dose and time response Markov model for the in-host dynamics of infection with intracellular bacteria following inhalation: with application to *Francisella tularensis*. *J R Soc Interf*. 2014;11(95):20140119.
7. Meynell GG, Meynell EW. The growth of micro-organisms in vivo with particular reference to the relation between dose and latent period. *J Hyg*. 1958;56(03):323–346.
8. Hornick RB, Music SI, Wenzel R, Cash R, Libonati JP, Snyder MJ, et al. The Broad Street pump revisited: response of volunteers to ingested cholera vibrios. *Bull N Y Acad Med*. 1971;47(10):1181–1191.
9. Saslaw S, Eigelsbach HT, Prior JA, Wilson HE, Carhart S. Tularemia vaccine study II: respiratory challenge. *Arch Intern Med*. 1961;107(5):702–714.
10. U.S. Department of Health & Human Services. What diseases are subject to Federal isolation and quarantine law?; 2009. <http://www.hhs.gov>.
11. Fine PEM. The interval between successive cases of an infectious disease. *Am J Epidemiol*. 2003;158(11):1039–1047.
12. Wodarz D, Nowak MA. Mathematical models of HIV pathogenesis and treatment. *BioEssays*. 2002;24(12):1178–1187.
13. Heppell CW, Egan JR, Hall IM. A Human Time Dose Response Model for Q Fever. In preparation. 2016; p. 1–24.
14. Gilchrist MA, Sasakiz A. Modeling host-parasite coevolution : a nested approach based on mechanistic models. *J Theor Biol*. 2002;218(3):289–308.
15. Gillard JJ, Laws TR, Lythe G, Molina-París C. Modeling early events in *Francisella tularensis* pathogenesis. *Front Cell Infect Microbiol*. 2014;4:169.
16. Read JM, Keeling MJ. Disease evolution across a range of spatio-temporal scales. *Theor Popul Biol*. 2006;70(2):201–213.
17. Day J, Friedman A, Schlesinger LS. Modeling the immune rheostat of macrophages in the lung in response to infection. *P Natl Acad Sci USA*. 2009;106(27):11246–11251.
18. Marino S, Kirschner DE. The human immune response to *Mycobacterium tuberculosis* in lung and lymph node. *J of Theor Biol*. 2004;227(4):463–486.
19. Pawelek KA, Dor D, Salmeron C, Handel A. Within-host models of high and low pathogenic influenza virus infections: The role of macrophages. *PLoS ONE*. 2016;11(2):1–16.
20. Legrand N, Ploss A, Balling R, Becker PD, Borsotti C, Brezillon N, et al. Humanized mice for modeling human infectious disease: challenges, progress, and outlook. *Cell Host Microbe*. 2009;6(1):5–9.
21. Tarnvik A. World Health Organization Guidelines on Tularemia; 2007.
22. Marino S, El-Kebir M, Kirschner D. A hybrid multi-compartment model of granuloma formation and T cell priming in Tuberculosis. *J Theor Biol*. 2011;280(1):50–62.
23. Gaddum JH, Allen P, Pearce SC. Lognormal distributions. *Nature*. 1945;156:463–466.
24. Jones RM, Nicas M, Hubbard A, Sylvester MD, Reingold A. The infectious dose of *Francisella tularensis* (tularemia). *Applied Biosafety*. 2005;10(4):227.
25. Meynell GG, Stocker BAD. Some hypotheses on the aetiology of fatal infections in partially resistant hosts and their application to mice challenged with *Salmonella paratyphi-B* or *Salmonella typhimurium* by intraperitoneal injection. *J Gen Microbiol*. 1957;16(1):38–58.
26. Janeway CA, Travers P, Walport M. Immunobiology: the immune system in health and disease. 5th ed. New York: Garland Science; 2001.
27. Meena LS, Rajni T. Survival mechanisms of pathogenic *Mycobacterium tuberculosis* H37Rv. *FEBS Journal*. 2010;277(11):2416–2427.
28. Clemens DL, Lee BY, Horwitz MA. Virulent and avirulent strains of *Francisella tularensis* prevent acidification and maturation of their phagosomes and escape into the cytoplasm in human macrophages. *Infect Immun*. 2004;72(6):3204–3217.
29. Voth DE, Heinzen RA. Lounging in a lysosome: the intracellular lifestyle of *Coxiella burnetii*. *Cell Microbiol*. 2007;9(4):829–840.
30. Egan JR, Hall IM, Leach S. Modeling inhalational tularemia: deliberate release and public health response. *Biosecur Bioterror*. 2011;9(4):331–343.
31. Huang Y, Haas CN. Quantification of the relationship between bacterial kinetics and host response for monkeys exposed to aerosolized *Francisella tularensis*. *Appl Environ Microbiol*. 2011;77(2):485–490.
32. Brooke RJ, Kretzschmar ME, Muters NT, Teunis PF. Human dose response relation for airborne exposure to *Coxiella burnetii*. *BMC Infect Dis*. 2013;13(1):488.
33. Roman MJ, Coriz PD, Baca OG. A proposed model to explain persistent infection of host cells with *Coxiella burnetii*. *J Gen Microbiol*. 1986;132:1415–1422.
34. Baylor ER, Peters V, Baylor MB. Water-to-air transfer of virus. *Science*. 1977;197(4305):763–764.
35. Xu Z, Shen F, Li X, Wu Y, Chen Q, Jie X, et al. Molecular and microscopic analysis of bacteria and viruses in exhaled breath collected using a simple impaction and condensing method. *PLoS ONE*. 2012;7(7):1–8.
36. Morawska L. Droplet fate in indoor environments, or can we prevent the spread of infection? *Indoor Air*. 2006;16(5):335–347.
37. Antia R, Levin BR, May RM. Within-host population dynamics and the evolution and maintenance of microparasite virulence. *Amer Nat*. 1994; p. 457–472.



38. Lukens S, DePasse J, Rosenfeld R, Ghedin E, Mochan E, Brown ST, et al. A large-scale immuno-epidemiological simulation of influenza A epidemics. *BMC Public Health*. 2014;14:1019.
39. Zwart MP, Hemerik L, Cory JS, de Visser JAG, Bianchi FJ, Van Oers MM, et al. An experimental test of the independent action hypothesis in virus-insect pathosystems. *Proc Biol Sci*. 2009;276(1665):2233–2242.
40. Handel A, Rohani P. Crossing the scale from within-host infection dynamics to between-host transmission fitness: a discussion of current assumptions and knowledge. *Philos Trans R Soc Lond [Biol]*. 2015;370(1675):20140302.
41. Magombedze G, Garira W, Mwenje E. Modelling the human immune response mechanisms to mycobacterium tuberculosis infection in the lungs. *Math Biosci Eng*. 2006;3(3):661–682.
42. Armstrong JA, Hart PD. Phagosome-lysosome interactions in cultured macrophages infected With virulent tubercle bacilli. *J Exp Med*. 1975;142(1):1–16.
43. Marks PJ, Vipond IB, Regan FM, Wedgwood K, Fey RE, Caul EO. A school outbreak of Norwalk-like virus: evidence for airborne transmission. *Epidemiol Infect*. 2003;131(1):727–36.
44. O’Neill PD, Marks PJ. Bayesian model choice and infection route modelling in an outbreak of Norovirus. *Stat Med*. 2005;24(13):2011–2024.
45. Hayden FG, Fritz R, Lobo MC, Alvord W, Strober W, Straus SE. Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. *J Clin Invest*. 1998;101(3):643–649.
46. Fritz RS, Hayden FG, Calfee DP, Cass LM, Peng AW, Alvord WG, et al. Nasal cytokine and chemokine responses in experimental influenza A virus infection: results of a placebo-controlled trial of intravenous zanamivir treatment. *J Infect Dis*. 1999;180(3):586–593.
47. Zaraket H, Baranovich T, Kaplan BS, Carter R, Song MS, Paulson JC, et al. Mammalian adaptation of influenza A(H7N9) virus is limited by a narrow genetic bottleneck. *Nature Commun*. 2015;6:6553.
48. Eames KTD, Tilston NL, White PJ, Adams E, Edmunds WJ. The impact of illness and the impact of school closure on social contact patterns. *Health Technol Assess*. 2010;14(34):267–312.
49. Kerckhove KV, Hens N, Edmunds WJ, Eames KTD. The impact of illness on social networks: implications for transmission and control of influenza. *Am J Epidemiol*. 2013;178(11):1655–1662.
50. Canini L, Carrat F. Population modeling of influenza A/H1N1 virus kinetics and symptom dynamics. *J Virol*. 2011;85(6):2764–70.
51. Handel A, Akin V, Pilyugin SS, Zarnitsyna V, Antia R. How sticky should a virus be? The impact of virus binding and release on transmission fitness using influenza as an example. *J R Soc Interf*. 2014;11(92):20131083.
52. Handel A, Brown J, Stallknecht D, Rohani P. A multi-scale analysis of influenza A virus fitness trade-offs due to temperature-dependent virus persistence. *PLoS Comput Biol*. 2013;9(3).
53. Handel A, Lebarbenchon C, Stallknecht D, Rohani P. Trade-offs between and within scales: environmental persistence and within-host fitness of avian influenza viruses. *Proc R Soc Lond [Biol]*;281(1787):20133051.
54. Mena-Lorcat J, Hethcote HW. Dynamic models of infectious diseases as regulators of population sizes. *J Math Biol*. 1992;30(7):693–716.
55. Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, Alizon S, et al. Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science (NY)*. 2014;343(6177):1243727.
56. Nguyet MN, Duong THK, Trung VT, Nguyen THQ, Tran CNB, Long VT, et al. Host and viral features of human dengue cases shape the population of infected and infectious *Aedes aegypti* mosquitoes. *Proc Natl Acad Sci USA*. 2013;110(22):9072–7.
57. Danon L, House TA, Read JM, Keeling MJ. Social encounter networks: collective properties and disease transmission. *J R Soc Interf*. 2012;9(76):2826–2833.
58. Virkar Y, Clauset A. Power-law distributions in empirical data. *Ann Appl Stat*. 2014;8(1):89–119.
59. Liljeros F, Edling CR, Amaral LA, Stanley HE, Aberg Y. The web of human sexual contacts. *Nature*. 2001;411(6840):907–908.
60. Gog JR, Pellis L, Wood JLN, McLean AR, Arinaminpathy N, Lloyd-Smith JO. Seven challenges in modeling pathogen dynamics within-host and across scales. *Epidemics*. 2014;10:45–48.
61. Cui N, Chen Y, Small DS. Modeling parasite infection dynamics when there is heterogeneity and imperfect detectability. *Biometrics*. 2013;69(3):683–692.
62. Gutting B. Deterministic models of inhalational anthrax in New Zealand white rabbits. *Biosecur Bioterror*. 2014;12(1):29–41.
63. Korenromp EL, Williams BG, Schmid GP, Dye C. Clinical prognostic value of RNA viral load and CD4 cell counts during untreated HIV-1 infection - A quantitative review. *PLoS ONE*. 2009;4(6).
64. Gomes MGM, Aguas R, Lopes JS, Nunes MC, Rebelo C, Rodrigues P, et al. How host heterogeneity governs tuberculosis reinfection? *Proc R Soc Lond [Biol]*. 2012;279(1737):2473–2478.
65. Heffernan JM, Keeling MJ. Implications of vaccination and waning immunity. *Proc R Soc Lond [Biol]*. 2009;276(1664):2071–2080.
66. Woolhouse ME, Dye C, Etard JF, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc Natl Acad Sci USA*. 1997;94(1):338–342.
67. Ball F, Britton T, Sirl D. A network with tunable clustering, degree correlation and degree distribution, and an epidemic thereon. *J Math Biol*. 2013;66(4-5):979–1019.
68. Mideo N, Alizon S, Day T. Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends Ecol Evol*. 2008;23(9):511–517.