UNIVERSITY OF CALIFORNIA,
IRVINE

Mathematical models for assessment of human health risk of pathogens in the environment

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematical, Computational and Systems Biology

by

Srikiran Chandrasekaran

Dissertation Committee:
Professor Sunny Jiang, Chair
Associate Professor Jun Allard
Professor Scott Bartell

2019

DEDICATION

To the lives lost
(some human, many not)
in the pursuit
of science

TABLE OF CONTENTS

Page

LIST OF FIGURES **v**

LIST OF TABLES **vi**

ACKNOWLEDGMENTS **vii**

CURRICULUM VITAE **ix**

ABSTRACT OF THE DISSERTATION **xi**

1 Introduction **1**
1.1 Background 1
1.2 Hypothesis 3
1.3 Objectives 4

2 A transport model for quantifying norovirus internalization in lettuce **6**
2.1 Background 6
2.2 Methods 10
2.2.1 Model structure 10
2.2.2 Model parameters to estimate viral transport in lettuce 11
2.2.3 Fitting model of viral transport in hydroponic grown lettuce 15
2.2.4 Estimating risk of consumption of lettuce 17
2.2.5 Sensitivity analysis 18
2.3 Results 19
2.3.1 Model fitting and parameter prediction 19
2.3.2 Health risks from lettuce consumption 22
2.4 Discussion 23
2.4.1 Fitting model to data 24
2.4.2 Model parameter estimates 25
2.4.3 Risk estimates 25
2.4.4 Contribution and future directions 27
2.5 Conclusion 29

3 Quantifying the infection risk of antibiotic-resistant bacteria 30
3.1 Background . . . 30
3.2 Results . . . 35
3.2.1 Dose-response expression . . . 35
3.2.2 Relationship with the existing DRMs . . . 36
3.2.3 Accounting for residual AB concentration . . . 37
3.2.4 Accounting for fraction of ARB . . . 38
3.2.5 Possible health outcomes . . . 39
3.2.6 Effect of ARB fraction f_r and residual AB concentration C . . . 41
3.3 Discussion . . . 44
3.4 Methods . . . 49
3.4.1 Fitting the model to data . . . 49
3.4.2 Sensitivity Analysis . . . 50
3.4.3 Data availability . . . 51

4 Role of quorum sensing in microbial pathogensis 52
4.1 Background . . . 52
4.2 Results . . . 54
4.2.1 Deterministic model of kinetics . . . 54
4.2.2 Stochastic model of dose-response . . . 56
4.2.3 Parameterizing the model . . . 58
4.2.4 Assessing the fit . . . 59
4.2.5 Parameters in absence of alcohol pre-treatment . . . 62
4.2.6 Parameters for MRSA . . . 64
4.2.7 Case study . . . 64
4.3 Discussion . . . 65
4.4 Methods . . . 69
4.4.1 Fitting kinetic data . . . 69
4.4.2 Simulating the stochastic model . . . 70
4.4.3 Fitting dose-response data . . . 70
4.4.4 Case study . . . 70

5 Discussion 72

Bibliography 75

Appendix A A transport model for quantifying norovirus internalization in lettuce 88

Appendix B Quantifying the infection risk of antibiotic-resistant bacteria 97

Appendix C Role of quorum sensing in microbial pathogensis 104

LIST OF FIGURES

Page

2.1 Overview of model components and information flow 10
2.2 Fit of model to data for hydroponic lettuce 19
2.3 Estimated parameter distributions 20
2.4 Annual risk and disease burdens of norovirus 21
2.5 Sensitivity analysis of transport models 22

3.1 Overview of simple death process 33
3.2 Effect of varying f_r and C on illness outcomes 40
3.3 Model fits for DS1 and DS2 43
3.4 Sensitivity analysis for exponential and beta-Poisson DRMs 44

4.1 Overview of 2C model 55
4.2 2C model fit and predictions 61
4.3 2C model in the absence of alcohol pre-treatment 63
4.4 Outcomes from a case study for 2C model 65

LIST OF TABLES

Page

2.1 Equations used in this study 9
2.2 Summary of symbols and parameter values 14
2.3 Summary of predicted parameters 16

3.1 Parameters used in this study. 34
3.2 Possible human health outcomes 39
3.3 Summary of model fits 42

4.1 Dose-response data used in this study 58
4.2 Summary of model fits 59

ACKNOWLEDGMENTS

I want to start by thanking my adviser, Prof. Sunny Jiang. From the time I joined her lab in Spring 2016, she has been a great source of support and inspiration. Under her guidance, I had the freedom to explore knowing that she would steer me from unproductive routes should I get stuck. She was also extremely approachable and provided face-time when I wanted to talk through ideas. I would not have improved my writing without her patience in reading the many attempts I made at telling a story. By trusting my computational decisions, she provided me with great freedom and the great responsibility that accompanies it. She also taught me to look for the big picture without getting lost in the details, something that I continue to work on. I am greatly indebted to her for providing me with opportunities to teach classes and attend international conferences, which contributed to my professional development. Perhaps most importantly, I want to thank her for having my back these last few years. I hope I met her expectations occasionally.

I would also like to extend my thanks to my committee members Prof. Jun Allard and Prof. Scott Bartell. Their insights helped instill greater rigor in my arguments. They were extremely approachable and lent support at various stages. Prof. Allard also provided me with an opportunity to design and instruct a short lecture, from which I learned to appreciate the effort that goes into teaching. Prof. Bartell's inputs on making my research appeal to a broader audience helped me improve my communication. I would also like to thank my advancement committee members Prof. Jasper Vrugt and Prof. Elizabeth Read, for their contributions. I also learned the art of fitting models to data in the course of Prof. Jasper Vrugt, which was very helpful in Chapters 2 and 4. For this and other valuable philosophical conversations, I would like to thank him. I must also thank Prof. Frederic Wan, whose courses taught me the analysis of differential equations and stochastic processes. These tools helped me think about ideas that matured into Chapters 3 and 4 of this dissertation.

I would not have secured admission to graduate school without the guidance of my undergraduate mentor Prof. Karthik Raman. He provided me with opportunities for undergraduate research, sowing the seeds for the technical approaches that I would adopt through graduate school. I would also like to thank Prof. Sridhar Rajaram for motivating me to apply for Ph.D. programs.

I am indebted to my labmates (Marisa Chattman Nielsen, Hamsa Gowda, Muyue Han, Yen Hsiang Huang, Hunter Quon) for their feedback on my research over the years and their support. Dr. Keah-Ying Lim helped me get started in the lab and provided data that I used in Chapter 2. Dr. Derek Manheim's work ethic helped me sustain momentum, and I found the discussions with him on parameter estimation helpful in solidifying my understanding.

I exhaust my vocabulary quickly in thanking these individuals, but there are many more to whom I would like to extend my gratitude: The administrators at CCBS who were proactive in creating a welcoming and supportive culture; Karen Martin for meeting me on a Sunday to ensure that I was prepared for the first day of class on Monday; Naomi Carreon for ensuring that I (and others) had something to eat at CCBS organized events; Cely Dean for

making many administrative tasks a breeze; Tina Rimal for helping me secure a venue for my defense; and Harry Mangalam, Joseph Farran, and others maintaining the HPC facility at UC Irvine, who treated me with more kindness than I deserved when I submitted a billion jobs.

I would not have made it this far without the support of my family and friends. They stayed with me every step of the way despite my preoccupation with puns, among other things.

I would like to thank Elsevier for permission to include Chapter 2 of this dissertation, which was originally published in Science of the Total Environment. I would like to thank Springer Nature for permission to include Chapter 3 of this dissertation, which was originally published in Scientific Reports. This work was supported in part by the Bill and Melinda Gates Foundation, and the National Science Foundation.

CURRICULUM VITAE

Srikiran Chandrasekaran

EDUCATION

Doctor of Philosophy in Mathematical, Computational and Systems Biology — **2019**
University of California, Irvine — *Irvine, California*

Bachelor of Technology in Biotechnology — **2015**
Indian Institute of Technology - Madras — *Chennai, Tamil Nadu*

RESEARCH EXPERIENCE

Graduate Research Assistant — **2015–2019**
University of California, Irvine — *Irvine, California*

TEACHING EXPERIENCE

Instructor : Intro to Python and Jupyter notebooks — **2019**
Teaching Assistant : Generalized linear models for differential gene expression analysis — **2018, 2019**
Teaching Assistant : Intro to MATLAB — **2016, 2018**
Teaching Assistant : Environmental Processes — **2016, 2017**

REFEREED JOURNAL PUBLICATIONS

A dose response model for quantifying the infection risk of antibiotic-resistant bacteria — **2019**
Scientific Reports

A dynamic transport model for quantification of norovirus internalization in lettuce from irrigation water and associated health risk — **2018**
Science of the Total Environment

REFEREED CONFERENCE ABSTRACTS

A dose-response model for assessing risks from antibiotic-resistant *Escherichia coli* in agricultural water reuse **2019**
SETAC Europe, Helsinki

Infection risk of antibiotic-resistant bacteria **2018**
q-bio, Houston

Parameter estimation in biological systems – a comparison of stochastic search algorithms **2018**
Asia Pacific Bioinformatics Conference, San Francisco

ABSTRACT OF THE DISSERTATION

Mathematical models for assessment of human health risk of pathogens in the environment

By

Srikiran Chandrasekaran

Doctor of Philosophy in Mathematical, Computational and Systems Biology

University of California, Irvine, 2019

Professor Sunny Jiang, Chair

Microbial pathogens in the environment present a growing threat to human health. They are found in waters used for recreation, irrigation etc., which present a multitude of pathways in which a person may be exposed to them. Moreover, the increasing interest in reusing treated wastewater raises questions of water quality and associated public health risks. The overarching goal of my dissertation is to develop quantitative tools to improve the accuracy of methods used to assess this risk and provide insights into disease development. I hypothesize that mathematical models rooted in well-grounded theory and data can augment the overall understanding of microbial risk and steer experiments in the most productive direction. My specific aims are: 1) To quantify the health risk posed by norovirus-contaminated water used for lettuce irrigation, by means of a dynamic transport model 2) To develop a dose-response framework applicable for antibiotic resistant bacteria, specifically the human enteric pathogen, *Escherichia coli* 3) To develop a modeling framework to probe the importance of cooperativity in helping *Staphylococcus aureus* establish skin infections. In my research, I have constructed a transport model using ordinary differential equations to predict the norovirus load in lettuce at harvest given the load in the irrigation water. By fitting this model to published experimental data, I found that attachment of the virus to the growth medium strongly influences the amount of virus in lettuce at harvest. Towards the second aim, I have used stochastic processes to develop an analytical expression for *E.*

coli dose-response. I then fitted this to clinical data and extended the model to predict, for the first time, the risk posed by a mixture of antibiotic sensitive and antibiotic resistant strains. Towards the third aim, I have developed a two-compartment stochastic model with cooperativity between cells to predict *S. aureus* dose-response. Using experimental data to reject the hypothesis of absence of cooperativity, I show the possible role of quorum sensing in *S. aureus* establishing skin infections. The outcomes of this research will enable better understanding of microbial risk associated with environmental exposure and improve human health protection.

Chapter 1

Introduction

1.1 Background

With the growing population and limited freshwater resources, there is increased interest in water conservation practices like using recycled wastewater and hydroponic agriculture. The presence of pathogens in the associated environmental compartments exposes a large fraction of the general populace to infection risks. Therefore, a need of the hour is ensuring that our infrastructure meets the safety requirements designed to protect human health. Proper disposal and treatment of wastes generated at hospitals, industries and residences help meet this goal by reducing the pathogen loads in the environment. However, complete elimination of pathogens is not an option [33]. Therefore, a framework to quantify the threat to human health is desired. The popularly adopted framework is called Quantitative Microbial Risk Assessment or QMRA [57].

With the growing population and limited freshwater resources, there is increased interest in water conservation practices like using recycled wastewater and hydroponic agriculture. The presence of pathogens in the associated environmental compartments exposes a significant

fraction of the general populace to infection risks. Therefore, a need of the hour is ensuring that our infrastructure meets the safety requirements designed to protect human health. Proper disposal and treatment of wastes generated at hospitals, industries, and residences help achieve this goal by reducing the pathogen loads in the environment. However, complete elimination of pathogens is not an option [33]. Therefore, a framework to quantify the threat to human health is desired. The popularly adopted framework is called Quantitative Microbial Risk Assessment or QMRA [57].

The main tenets of QMRA are as follows [57]: 1) hazard identification; 2) exposure assessment; 3) dose-response modeling; 4) risk characterization, and 5) risk management. Hazard identification constitutes deciding on the system of interest and listing out the pathogens present/expected in that system. After identifying the hazard, the interaction of people with the system are modeled to quantify exposure to the pathogen(s). Suppose the system of interest is a particular lake used for recreation, and the hazard identified is E. coli. Exposure assessment would entail enumerating the *E. coli* finally ingested by the person (or the *dose*). These processes have a lot of associated variability and uncertainty. Therefore, quantities are stratified by groups (age of swimmers, seasonal changes in pathogen concentration) or represented by distributions rather than point estimates. Estimating the risk while accounting for these variabilities and uncertainties is done by Monte Carlo sampling.

Dose-response models (DRMs) relate the number of the pathogen (dose) to the probability of a person falling ill (response or *risk*). They are constructed with data from clinical trials in which a predetermined dose of pathogens is administered to a cohort of subjects and the number falling ill counted. The latter is then divided by the total number of subjects to reflect the probability of a single person falling ill. This process is repeated for different pathogen doses to generate data for the models. While these clinical trials may use animals, datasets generated from human trials are preferred since they better reflect the human situation. Popular DRMs are the exponential and beta-Poisson models [57]. DRMs

for different pathogens may share the same functional form but differ in the numerical values of model parameters as a consequence of the biological differences between the pathogens.

Risk characterization involves calculating the risk posed by the hazard by integrating the output of the exposure assessment (dose) with the DRM of choice for that pathogen. One then compares these estimates with guidelines established by the U.S. Environmental Protection Agency (EPA) or the World Health Organization (WHO). Based on these comparisons, risk management measures can be investigated in an iterative process by computing the risk posed by the intervention measures.

1.2 Hypothesis

The larger topic of my interest is in building mathematical models to augment our understanding of microbial risk assessment. The system encompassing pathogens in the environment, and their interaction with humans is complex. Important variables like the weather, the habits/demographics of the people under consideration, and the pathogen's health effects must be accounted for when making decisions to safeguard human health. Experiments spanning this entire scale are not feasible on accounts of the size of the system and the variability at each step. Mathematical modeling becomes indispensable in deriving insights and is the cornerstone of microbial risk analyses.

Nevertheless, I believe that the models in this field abstract away important details and favor simplified computational approaches for their parsimony and ease of use. Examples include the DRMs, such as exponential or beta-Poisson DRMs discussed above. I hypothesize that mathematical models that account for these details and leverage sophisticated computational techniques open up new avenues of scientific inquiry and bring existing paradigms into question. Besides, they can increase the accuracy of risk estimates and help generate

experimentally testable hypotheses that aid risk management.

1.3 Objectives

I aim to investigate this hypothesis by attempting to answer the following three specific questions: 1) if the viral load in the water used to irrigate agricultural produce is known, can the safety of consuming the resulting produce be determined? 2) can we use existing data on dose-response to develop a predictive understanding of dose-response of antibiotic-resistant bacteria (ARB)? and, 3) how do we mathematically investigate the presence or absence of cooperativity between individual bacteria in establishing infection? I motivate these questions below.

Limited freshwater supplies coupled with the increasing demand for it exert stress on the water security of the populace. Using recycled wastewater for irrigating agricultural produce is a sustainable way forward that has received much attention. However, this alternative is not without risk. Pathogens present in the irrigation water can enter the food supply, causing outbreaks of different diseases when contaminated produce is consumed uncooked. When pathogens adhere to the surfaces of the plant, they can be washed off or treated with UV radiation. However, viruses sometimes find their way into the vasculature of the produce where they are shielded from treatment and can thus cause disease. The question then is to find the relationship between viral loads in the input irrigation water (which is easily measured) and the viral loads in the produce to understand its safety for consumption. Using a simple output to input ratio may suffice to predict the risk but does not provide any room for investigation of the factors influencing this ratio and evaluating the effectiveness of risk management measures. Hence there is a need for a first principle driven transport model of viruses (and potentially other pathogens) to understand their internalization in the vasculature of plants.

Understanding the risk posed by ARB has been stymied by the absence of DRMs parameterized for ARB. This difficulty arises from the clinical trials used to parameterize current DRMs, which were performed using antibiotic sensitive bacteria (ASB). While we have *in-vitro* kinetic information relating ARB to ASB, the biophysical/kinetic interpretation of the parameters of the popular exponential and beta-Poisson DRMs is not straightforward. Moreover, the dose-response outcome is potentially complicated by the other processes at play, such as horizontal gene transfer and the differential influence of antibiotics on ASB and ARB death rates. The resulting illness may or may not respond to antibiotic treatment if the ARB subpopulation persists. These challenges require a mathematical framework capable of handling the underlying processes, which can then be used to perform risk assessments of ARB and determine the best course of action.

A point of longstanding debate in QMRA, and broadly the topic of disease progression, is the hypothesis of independent action[57, 117, 93]. It proposes that pathogens act independently of one another, and each has a probability p of initiating infection. The alternative hypothesis is one of cooperation where infection is expected when more than one organism survives to overwhelm the host's defenses collectively. DRMs assuming independent action (also called single hit models) have wider acceptance than DRMs which assume cooperativity (also called multi-hit) [57]. However, DRMs with cooperativity consider the cumulative effects of bacteria but not the potential synergistic interactions between bacterial cells or quorum sensing. I believe that incorporating cell-cell interaction in dose-response is an essential step to developing a better understanding of the development of disease and its treatment.

Chapter 2

A transport model for quantifying norovirus internalization in lettuce

The contents of this chapter appear in the journal Science of the Total Environment [25].

2.1 Background

The growth of the human population places an ever-increasing demand on freshwater resources and food supply. The nexus of water and food is now well recognized. One promising strategy to sustain food production in the face of competing water demands is to increase the reuse of treated human wastewater. Municipal wastewater reuse for food production has been successfully adopted in some regions of the world. For example, Israel uses 84% treated wastewater in agriculture production [113]. However, Southern California, a region that suffers from a similar degree of water shortage, currently uses less than 3% of municipal wastewater in agriculture, while discharging 1.5 million acre-feet effluent per year into the Pacific Ocean [59, 60]. Secondary municipal wastewater effluent for ocean discharge is often

sufficient to support both the nutrient and water needs for food production. Water reuse in agriculture can bring municipal water reclamation effluent to nearby farms within the city limit, thus promoting local agriculture and also reducing the rate of farmland loss to urban development.

While the use of reclaimed water in agriculture offers a multitude of societal and agronomical benefits, broader adoption faces great challenges. One of the important challenges is ensuring the safety of food products in light of a plethora of human pathogens that may be present in recycled wastewater. Past studies have identified risks associated with irrigating food with recycled wastewater through the retention of the irrigation water on edible plant surfaces during overhead irrigation [11, 58, 84, 89, 96, 119]. With the emphasis on water conservation and reduction of transevaporation, subsurface drip irrigation is gaining popularity [132]. Since there is lesser contact between water and the plant surface, the chance of surface contamination of pathogens is reduced. However, this new practice presents risk of uptake of microbial pathogens into plants. Such internalized pathogens are of greater concerns as washing, even with disinfectants, may not affect pathogens sheltered in the vasculature. Although pathogen transport through root uptake and subsequent internalization into the plant has been a growing research area, results vary due to differences in experimental design, systems tested, and pathogens and crops examined [22, 35, 37, 62, 71, 111, 134, 141, 150].

Among the array of pathogens causing foodborne illness that may be carried by treated wastewater, viruses are of the greatest concern but least studied. According to the CDC, 60% of U.S. foodborne outbreaks associated with eating leafy greens were caused by noroviruses (NoV), while Salmonella and E. coli only accounted for 10% of the outbreaks (`http://www.cdc.gov/features/norovirus/`). Estimates of global foodborne illness prevalence associated with NoV (124 million) surpass all other pathogens considered [61]. Viruses are also of concern because they persist in secondary wastewater effluents in high concentrations [32, 42]. They do not settle well in sedimentation basins and are also more resistant to

degradation than bacteria [145]. Therefore, in the absence of solid scientific understanding of the risks involved, the public are likely less receptive to adopting treated wastewater for agricultural irrigation.

NoV internalization in hydroponic systems has been quantified by [37]. Internalization in crops grown in soil is considered lesser [141] but nevertheless occurs. However, the only risk assessment [119] that considered the possibility of NoV internalization in plants assumed a simple ratio of viruses in the feed water over viruses in produce at harvest to account for internalization. The time dependence of viral loads in lettuce was not explored and such an approach did not permit insights into the key factors influencing viral uptake in plants.

In this study, I introduce a viral transport model to predict the viral load in crisphead lettuce at harvest given the viral load in the feed water. It is parameterized for both hydroponic and soil systems. I demonstrate its utility by performing a quantitative microbial risk assessment (QMRA). I explore strategies to reduce risk enabled by such a model, and use a sensitivity analysis to identify influential factors that affect risk. (2.1)

#	Description	Equation	Reference
Rate of change in viral conc (count cm^{-3} day^{-1})			
2.1	Growth medium (first-order decay)	$\frac{dC}{dt} = \frac{\eta_{gr}F(t)C_g(t)}{V_g(t)} - k_{decf}C_g(t) - \frac{C_g(t)}{V_g(t)}\frac{dV_g(t)}{dt}$	This work
2.2	Growth medium (first-order decay with att-det)	$\frac{dC}{dt} = \frac{\eta_{gr}F(t)C_g(t)}{V_g(t)} - \frac{C_g(t)}{V_g(t)}\frac{dV_g(t)}{dt} - (k_{att} + k_{dec})C_g(t) + \frac{k_{det}N_{att}(t)}{V_g(t)}$	"
2.3[a]	Surface attached	$\frac{dN_{att}(t)}{dt} = k_{att}C_g(t)V_g(t) - (k_{dec}+k_{det})N_{att}(t)$	"
2.4	Root	$\frac{dC_{ro}(t)}{dt} = \frac{\eta_{gr}F(t)C_g(t)}{V_{ro}(t)} - \frac{\eta_{rs}F(t)C_{ro}(t)}{V_{ro}(t)} - k_pC_{ro}(t) - \frac{C_{ro}(t)}{V_{ro}(t)}\frac{dV_{ro}(t)}{dt}$	"

2.5	Shoot	$\frac{dC_{\text{sh}}(t)}{dt} = \frac{\eta_{\text{rs}} F(t) C_{\text{ro}}(t)}{V_{\text{sh}}(t)} - k_p C_{\text{sh}}(t) - \frac{C_{\text{sh}}(t)}{V_{\text{sh}}(t)} \frac{dV_{\text{sh}}(t)}{dt}$	
2.6^h	Hydroponic viral transport rate (cm^3 day^{-1})	$F(t) = a_t + b_t \left(\rho_{\text{shoot}} d_{\text{shoot,h}} \frac{dV_{\text{sh}}(t)}{dt} \right)$	[28]
Growth rates (cm^{-3} day^{-1})			
2.7^h	Root, hydroponic	$\frac{dV_{\text{ro}}(t)}{dt} = \frac{\exp(r_1 + r_2 t + r_3 t^2)}{\rho_{\text{root}} d_{\text{root,h}}}$	[17]
2.8^h	Shoot, hydroponic	$\frac{dV_{\text{sh}}(t)}{dt} = \frac{\exp(s_1 + s_2 t + s_3 t^2)}{\rho_{\text{shoot}} d_{\text{shoot,h}}}$	"
2.9^s	Shoot, soil	$\frac{dV_{\text{sh}}(t)}{dt} = r_g V_{\text{sh}}(t) \left(1 - \frac{V_{\text{sh}}(t) \rho_{\text{shoot}}}{w_f} \right)$	[129]
Others			
2.10	Daily consumption	$\lambda_k = B \times L \times C_{\text{sh}}(t)(t_{\text{ht}}) \times \rho_{\text{shoot}}^{-1}$	This work
2.11	Approximate Beta Poisson	$P_{\text{inf,k}} = 1 - \left(1 + \frac{\lambda_k}{\beta_B} \right)^{-\alpha_B}$	[136]
2.12	1F1 Hypergeometric	$P_{\text{inf,k}} = 1 -_1 F_1(\alpha_H, \alpha_H + \beta_H, -\lambda_k)$	"
2.13	Fractional Poisson	$P_{\text{inf,k}} = P_{\text{F}} \left(1 - \exp(-\frac{\lambda_k}{\mu_a}) \right) \right)$	"
2.14	Annual infection Risk	$P_{\text{inf,ann}} = 1 - \prod_{k=1}^{365} (1 - P_{\text{inf,k}})$	[74]
2.15	Annual illness Risk	$P_{\text{ill,ann}} = P_{\text{ill\|inf}} P_{\text{inf,ann}}$	[89]
2.16	Annual disease burden	$D_{\text{annual}} = P_{\text{ill\|inf}} P_{\text{ill,ann}} D_p$	"
2.17	Volume of growth medium (soil)	$V_{\text{g,s}} = V_e \theta$	[29]

Table 2.1: Equations used in this study

[a] Units of (count day^{-1}), [h] Specific to hydroponic model, [s] Specific to soil model

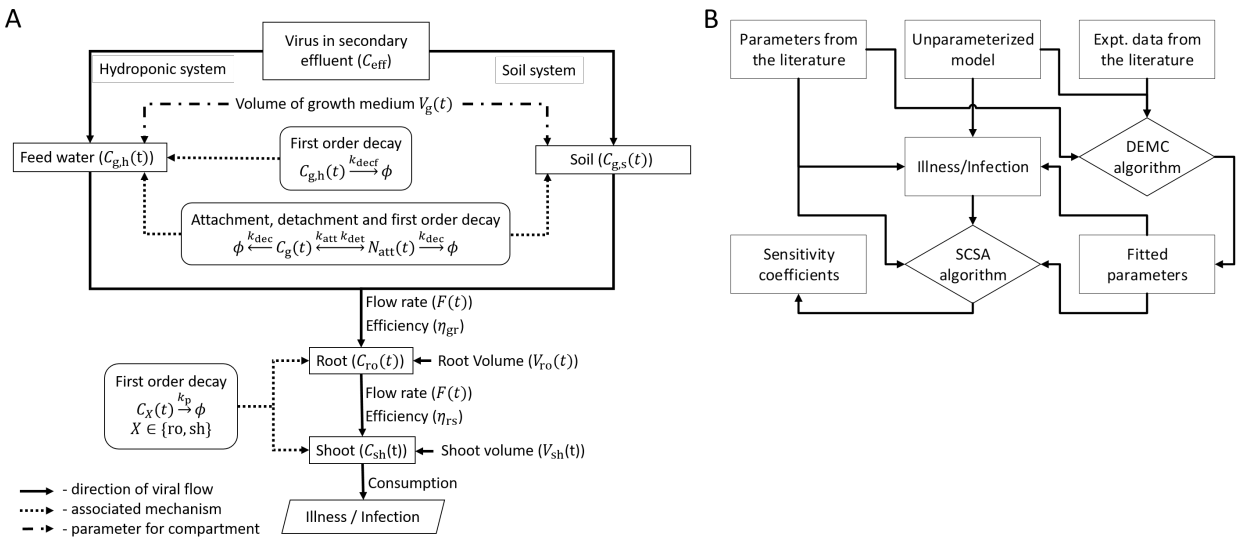

Figure 2.1: Overview of model components (panel A) and information flow (panel B). Here ϕ represents the null species. Symbols are defined in Table 2.2.

2.2 Methods

2.2.1 Model structure

To understand viral transport from treated human sewage to lettuce through internalization, and the final viral concentration in the plant tissue at the time of consumption, I developed a conceptual transport model (Fig. 2.1, symbols in Table 2.2).

In this system, I assumed that the treated wastewater used for cultivating lettuce is secondary sewage effluent that contains C_{eff} of NoV/ml. Viral concentrations in the growth medium ($C_{\mathrm{g,h}}(t)$ and $C_{\mathrm{g,s}}(t)$ for hydroponic and soil, respectively) at any given time are related to the volume of the growth medium ($V_{\mathrm{g}}(t)$) and the viral removal. Viral removal in the hydroponic growth medium may be modeled simply as a first-order decay (eq. (2.1)) or also to include attachment-detachment (AD) of viruses onto the walls of the hydroponic tank (tank effects, eq. (2.2-2.3)). Similarly, I considered AD to soil particles as an important process determining the fraction of viruses transferred from the soil to the plant roots. To

avoid making assumptions on the tank geometry, I expressed the attached viral load in viral numbers ($N_{\mathrm{att}}(t)$), which does not depend on the volume or surface area. The viral concentrations in the root ($C_{\mathrm{ro}}(t)$) and shoot ($C_{\mathrm{sh}}(t)$) depend on: 1) the viral transport rate ($F(t)$) from the growth medium to the plant, 2) the volumes of these compartments ($V_{\mathrm{ro}}(t)$ and $V_{\mathrm{sh}}(t)$), which change with time as the plant grows, and 3) the natural decay of viruses in plant tissues. In addition, I included two viral transfer efficiencies (η_{gr} between growth medium and root and η_{rs} between root and shoot) to simulate the potential "barrier" of viral transfer between each compartment (Fig. 2.1).

Consequently, I modeled the viral transport from the growth medium to the root and shoot through internalization as mass transport through three (growth medium, root and shoot) well-mixed reactors ($n-1$,n and $n+1$) in series with changing volume. Thus, the generic equation governing viral concentration ($C_n(t)$) in the reactor n at time t is:

$$V_n(t)\frac{dC_n(t)}{dt} + C_n(t)\frac{dV_n(t)}{dt} = \eta_{n-1,n}F(t)C_{n-1}(t) - \eta_{n,n+1}F(t)C_n(t) - \mathrm{Removal}(C_n(t))$$

where $\eta_{n,n+1}$ is the efficiency of viral transfer between reactors n and $n+1$. The full model equations specific to individual model components are given by eq. (2.1-2.9) in Table 2.1 and used for both soil and hydroponic systems. Finally, I estimated the health risk from the consumption of shoot of lettuce irrigated by recycled wastewater (Fig. 2.1).

2.2.2 Model parameters to estimate viral transport in lettuce

I obtained some parameters to complete the conceptual viral transport model from the literature. I estimated the remaining by fitting the model to published data from experiments using NoV seeded feed water to grow crisphead lettuces in a hydroponic system [37]. I adopted the initial volume of 800 mL for the hydroponic growth medium ($V_{\mathrm{g,h}}(0)$) based on these experiments. I also assumed that the volume reduction over time equivalent to

the plant transpiration rate [28] between refills. For the soil system, the volume of the growth medium ($V_{\mathrm{g,s}}$) equals the volume of water contained in the soil interstitial spaces in an envelope around the roots. I assumed that this envelope is a region (of volume V_e) around the roots that the plant to interact with. $V_{\mathrm{g,s}}$is given by eq. 17, where θ is the volumetric water content obtained from [29]. Estimates for V_e spanned a large range (section A.2.1) and I adopted a middle value of $V_e = 80000$ cm^3 and assumed it to be constant over the lettuce growth period. I also verified that this assumed value wasto have minimal impact on the model outcome (see section 2.2.5 and section 2.3.2).

I adopted the plant transpiration rate as the viral transport rate ($F(t)$) based on: 1) previous reports of passive bacterial transport in plants [13, 138, 146], 2) the significantly smaller size of viruses compared to bacteria, and 3) the lack of known specific interactions between human viruses and plant hosts (section A.1.1). Accordingly, I determined the viral transport rate in hydroponically grown lettuce (eq. (2.6)) from the previously reported transpiration model [28], in which the transpiration rate is proportional to the lettuce growth rate and is influenced by cultivar specific factors (a_t,b_t). I predicted these cultivar specific factors using the hydroponic crisphead lettuce growth experiment carried out by [37] described in section 2.2.3 (and section A.1.4). Since the transpiration rate in soil grown lettuce is significantly higher than that in the hydroponic system, I obtained the viral transport rate in soil grown lettuce from the graphs published by [45] using WebPlotDigitizer [115] section A.2.3 for details).

I estimated the growth rates of lettuce root ($dV_{\mathrm{ro}}(t)/dt$) and shoot ($V_{\mathrm{sh}}(t)/dt$) in hydroponic systems using eq. (2.7-2.8) (parameters in Table 2.2; details and assumptions in section A.1.2 and section A.1.3). For soil grown lettuce, I determined the the shoot growth rate (in terms of fresh volume) using eq. (2.9) (parameters in Table 2.2; details in section A.2.2). In the absence of a published root growth model for lettuce in soil, I used a fixed root volume of 100 cm^3.

Parameter	Symbol	Value/Distribution (Units)	Reference
Common (soil+hydroponic)			
Body weight^{+}	B	[67.0 (10.7, 113.9)] (kg person^{-1})	[73]
Viral load in effluent^{+}	C_{eff}	[4.13 (0.04, 624.32)] (count mL^{-1})	[85]
Concentration of virus in growth medium (root, shoot) *	$C_{\text{g}}(t)$ ($C_{\text{ro}}(t)$, $C_{\text{sh}}(t)$)	$C_{\text{g,s}}(0) = C_{\text{g,h}}(0) = C_{\text{eff}}$ (count mL^{-1})	
DALYs per case of NoV GE	D_p	9×10^{-4} (person^{-1}year^{-1})	[75]
Volumetric flow rate*	$F(t)$	$F_h(t)$ from eq. (2.6), $F_s(t)$ from plot (mL day^{-1})	[16, 45]
Growth medium viral attachment rate*	k_{att}	$k_{\text{att,s}}$(S1,S2)=4.1,0.8; $k_{\text{att,h}}$ [f] (day $^{-1}$)	[121]
Growth medium viral decay rate (with attachment-detachment) *	k_{dec}	$k_{\text{dec,s}}$=0.15; $k_{\text{dec,h}}$ [f] (day $^{-1}$)	[114]
Growth medium viral decay rate (only first-order decay)	$k_{\text{decf}}(t)$	[f] (day^{-1})	
Growth medium viral detachment rate*	k_{det}	$k_{\text{det,s}}(t)$ (S1,S2)= 8.7×10^{-4}, 3.0×10^{-3}; $k_{\text{det,h}}(t)$ [f] (day^{-1})	[121]
Viral decay constant in plant (root, shoot)	k_p	[f] (day^{-1})	
Consumption of lettuce per kg bodyweight^{+}	L	[0.38, (0.04, 2.08)] (g lettuce kg^{-1}person^{-1}day^{-1})	[135]
Number of viruses attached to growth medium	$N_{\text{att}}(t)$	Model intermediate (count)	
Probability of illness if infected	$P_{\text{ill\|inf}}$	0.8	[95]
Last irrigation time*	t_{li}	$t_{\text{li,h}} = 21 t_{\text{li,s}} = 56$ (days)	[16, 45]
Harvest time*	t_{ht}	$t_{\text{ht,h}} = 35, t_{\text{ht,s}} = 70$ (days)	"
Volume of growth medium (root, shoot) *	$V_{\text{g}}(t)$ ($V_{\text{ro}}(t)$, $V_{\text{sh}}(t)$)	$V_{\text{g,h}}(0)$=800 mL $V_{\text{ro,s}}$=100	
Fractional Poisson risk model parameters	$P_{\text{F}}, \mu_{\text{a}}$	0.72, 1	[136]
Beta Poisson risk model parameters	$\alpha_{\text{B}}, \beta_{\text{B}}$	0.104, 32.3	
Hypergeometric risk model parameters	$\alpha_{\text{H}}, \beta_{\text{H}}$	0.46, 1.20	

Shoot density	ρ_{shoot}	0.35 (g cm^{-3})	[72]
Hydroponic specific			
Rate parameters for viral transport by plant	a_t,b_t	[f] (mL day^{-1}, mL)	
Ratio of dry to fresh weight of lettuce root	$d_{\text{root,h}}$	0.057	Estimated from [152]
Ratio of dry to fresh weight of lettuce shoot	$d_{\text{shoot,h}}$	0.045	[16, 28]
Root growth constants	r_1	-8.482	"
	r_2	0.4586 (day^{-1})	"
	r_3	-6.472×10^{-3} (day^{-2})	"
Shoot growth constants	s_1	-7.414	"
	s_2	0.406 (day^{-1})	"
	s_3	-5.579×10^{-3} (day^{-2})	"
Root density	ρ_{root}	0.2 g cm^{-3}	Assumed
Soil specific			
Shoot growth constant	r_g	0.2056 (day^{-1})	[128]
Envelope volume	V_e	80000 (cm^3)	Assumed
Final weight of lettuce	w_f	550 (gm)	[72]
Volumetric water content of soil	θ	0.435	[29]

Table 2.2: Summary of symbols and parameter values

*: These are represented by their subscripts (h: hydroponic, s: soil) where required; +: Empirical distributions, values presented are median (95% interval); [f]: Fitted values, presented in Table 2.3

Additionally, I used a viral transfer efficiency (η) to account for the potential "barrier" between each compartment (i.e. root and shoot). The existence of such a "barrier" is evident from field experiments where some microbial pathogens were internalized in the root but not in the shoot of plants [102]. In addition, viral transfer efficiencies (ranging from 0 – 1) also account for differing observations in pathogen internalization due to the type of pathogen or lettuce. For example, [37] reported the internalization of NoV into lettuce, while [134] did not detect any NoV in another type of lettuce grown in feed water seeded with viruses. I

determined the values of η_{gr} and η_{rs} by fitting the model to experimental data reported by [37] (details in section 2.2.3). I assumed that the values of η_{gr} and η_{rs} for the soil case was the same as predicted for the hydroponic lettuce.

I included both die-off and AD for viral removal in the growth medium, but only natural die-off in the lettuce root and shoot. I obtained the AD kinetic constants (k_{att}, k_{det}) as well as the growth medium viral decay constant (k_{dec}) in the hydroponic case by fitting the model to the data from DiCaprio et al. (2012). Viral AD in soil has been investigated in both lab scale soil columns and field studies (Schijven and Hassanizadeh, 2000). In my model, I obtained viral AD constants ($k_{\mathrm{att,s}}$, $k_{\mathrm{det,s}}$) in soil from the experiments of [121], who investigated MS2 phage kinetics in sandy soil in field experiments. As the MS2 phage was transported with the water in soil, the AD rates changed with the distance from the source of viruses. To capture the range of AD rates, I investigated two scenarios of viral behavior in soils. Scenario 1 (Sc1) used the AD rates estimated at the site closest to the viral source (well 1), while scenario 2 (Sc2) used data from the farthest site (well 6). In contrast to lab scale soil column studies, field studies provide more realistic viral removal rates [120]. I used the surrogate phage MS2 for NoV as it provided conservative risk estimates, since MS2 attached to a lesser extent than NoV in several soil types [91]. I adopted the viral decay rate in the soil ($k_{\mathrm{dec,s}}(t)$) determined by [114] because the experimental temperature (20-30 ^{0}C) and soil type (clay loam) are more relevant to lettuce growing conditions compared to the other decay study [121]. I used the decay rates in the root and shoot from the hydroponic system predictions.

2.2.3 Fitting model of viral transport in hydroponic grown lettuce

I fitted the transport model to $\log_{10}$ viral concentration data from [37], extracted from graphs therein using WebPlotDigitizer [115]. In these experiments, NoV of a known concentration

Parameter	Units	Search bounds	Median (95% Credible Interval)
a_t	$cm^3\ day^{-1}$	(0,100)	19.82 (0.71,39.92)
b_t	$cm^3\ g^{-1}$	(0,300)	40.10 (1.19,96.96)
η_{gr}	-	(0,1)	0.48 (0.07,0.97)
η_{rs}	-	(0,1)	0.74 (0.24,0.99)
k_{att}	day^{-1}	(0,20)	10.66 (0.62,19.55)
k_{det}	day^{-1}	(0,10)	5.19 (0.65,9.76)
k_{dec}	day^{-1}	(0,100)	0.25 (0.03,0.54)
k_p	day^{-1}	(0,20)	0.54 (0.02,6.29)

Table 2.3: Parameter values predicted from fitting model to data from hydroponic experiments

was spiked in the feed water (growth medium) of hydroponic lettuce and was monitored in the feed water, the root and shoot over time. While fitting the model, I adopted an initial feed volume of 800 mL (as used in the experiments) and rejected parameters producing final volumes of <200 mL.

To fit the model while accounting for uncertainty in the data, I used a Bayesian approach to maximize the likelihood of the data given the parameters. I obtained a posterior distribution of the parameters by the differential evolution Markov chain (DE-MC) [130] algorithm, which can be parallelized and can handle multimodality of the posteriors distribution without fine tuning the jumping distribution. I carried out the computation on the High Performance Computing facility at UC Irvine, using MATLAB R2016a (Mathworks) and its ParComp-Tool.

Table 2.3 lists the parameters estimated by model fitting and their search bounds. Fitting data from [37] without including viral AD to the tank walls was attempted but the results were not used in the risk estimates due to the poor fit of model to the data. The rationale behind the model fitting procedure and diagnostics are discussed in section A.2.4.

2.2.4 Estimating risk of consumption of lettuce

I used an empirical distribution for NoV in activated sludge treated secondary effluent [85] to obtain the viral concentration (C_{eff}) in the irrigation water. As justified by [85], I used the sum of the concentrations of two genotypes known to cause illness to construct the distribution. I then estimated the NoV concentrations in lettuce shoot at typical harvest times: $t_{\text{ht,h}}(t)$=35 days in the hydroponic system and $t_{\text{ht,s}}(t)$=70 days in soil. I also assumed that the last irrigation with recycled water occurred on $t_{\text{li,h}}(t)$=21 days for hydroponic and $t_{\text{li,s}}(t)$=56 days for soil grown lettuce. I used these values together with parameters in Table 2.2 and Table 2.3 in eq. (2.1-2.9) (Table 2.1) to generate the probability distribution of NoV concentration for hydroponic or soil grown lettuce.

To estimate the risk from consumption of lettuce, I computed the daily viral dose using eq. (2.10) (Table 2.1) for the k$^{\text{th}}$day. I sampled the body weight (B) from an empirical distribution for all ages and genders in the United States, which I constructed from a report of the percentile data of body weight. I sampled the lettuce consumption rate (L) from an empirical distribution which I constructed from data reported by the Continuing Survey of Food Intakes by Individuals (CSFII). I used the 'consumer only' data for all ages and genders, and hence the reported risk is only for those who consume lettuce. It is important to note that the daily viral dose was computed in (count g^{-1}) from the model output (in count mL^{-1}) using the shoot density ρ_{shoot}(eq. (2.10)) to be consistent with the consumption rate reported in CSFII.

Several different NoV dose-response models have been proposed based on limited clinical data. The validity of these models is a matter of much debate [122, 136], which is beyond the scope of this study. These models differ in their assumptions resulting in large variability of predicted risk outcomes. To cover the range of potential outcomes of human exposure to NoV, I estimated and compared risk outcomes using three models: 1) Approximate Beta-

Poisson (BP) [131, 136]; 2) Hypergeometric ($_1F_1$) [4, 10, 131] ; and 3) Fractional Poisson (FP) [92]. In the risk estimation, I considered NoV in the lettuce tissue exists as individual viral particles (disaggregated form) and used the disaggregated NoV models. The model equations are given by eq. (2.11-2.13), Table 2.1. I calculated ten thousand samples of the daily infectious risks from BP and FP models using MATLAB R2016a. Wolfram Mathematica 11.1 (Wolfram Research) was used for the ($_1F_1$) model estimation as it was faster.

Using a random set of 365 daily risk estimates of ($P_{\text{inf,k}}$ for day k), I calculated the annual infection risk ($P_{\text{inf,ann}}$) according to the Gold Standard estimator [74] using eq. (2.14), Table 2.1. While there appears to be some dose dependence for illness resulting from infection $P_{\text{ill|inf}}$[10, 131], this has not been clearly elucidated for the different dose-response models. Hence, I adopted the procedure used in [89] and calculated annual illness risk with eq. (2.15). The annual disease burden in terms of DALY (disability adjusted life years) lost per case (D_p) was set to 9×10^{-4} pppy for each case of NoV disease [75]. The annual disease burden (D_{annual}) is given by eq. 16 in Table 2.1. As part of the risk characterization process, I compared risk outcomes of this study to the acceptable risk benchmarks established by the U.S. EPA and WHO and the estimates by [119].

2.2.5 Sensitivity analysis

I determined the global sensitivity of the $\log_{10}(P_{\text{inf}})$ (daily risk) to input parameters using the SCSA method [83] since it accounts for correlation in input parameters (not handled by Sobol's method [125]). This method produces three sensitivity indices for each parameter, the structural (S_{struct}), correlative (S_{corr}) and total (S_{tot}) sensitivities. Fitted parameters were used as is, maintaining the observed correlation structures. Parameters drawn from distributions were varied within their 95% credible intervals while other parameters spanned ranges obtained from literature (section A.2.5). The MATLAB implementation of SCSA by

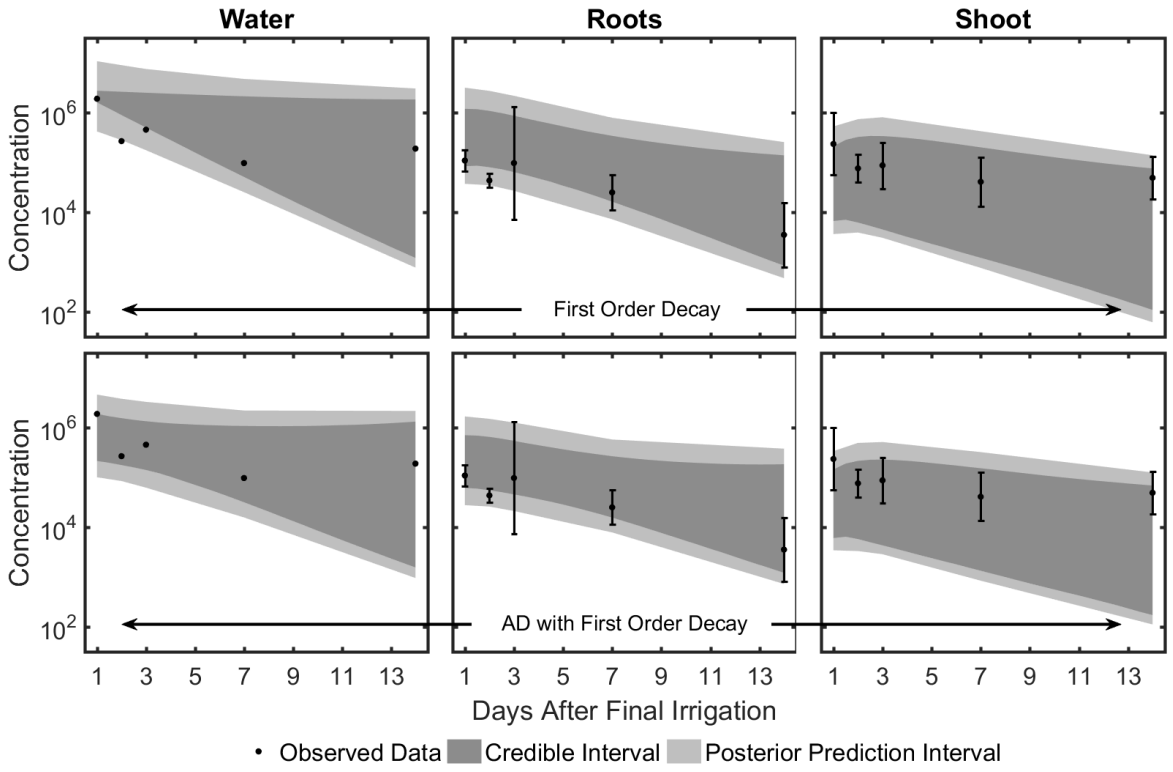

Figure 2.2: Fit of the model to data from [37] for lettuce grown in hydroponic system. Top panel shows model fitting using first-order viral decay only; bottom panel shows model fitting using first-order viral decay and viral attachment/detachment (AD) to cultivation tank wall in growth medium (water). Error bars indicate standard deviations of 3 samples.

Sahin and Vrugt [118] was used.

2.3 Results

2.3.1 Model fitting and parameter prediction

Fig. 2.4 shows a summary of the model fitting exercise for viral transport in hydroponic grown lettuce. Under the assumption of first-order viral decay, NoV loads in water (growth medium) at two time-points did not fall in the credible region of model predictions, indicating that mere first-order decay was unsuitable to capture the observed viral concentration data.

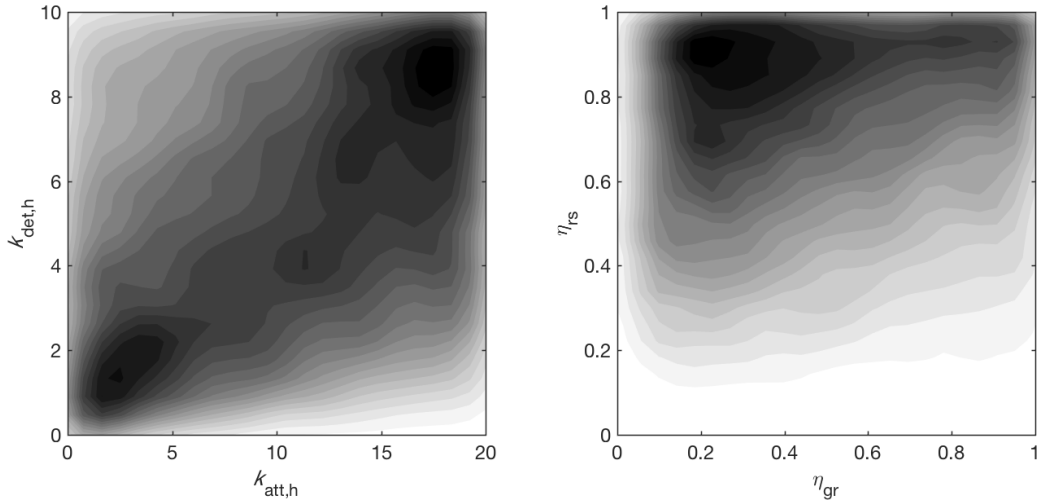

Figure 2.3: Illustration of joint distributions of posterior samples from fitting the model with AD of viruses to hydroponic tank walls. The shaded gradients (light to dark) indicate the localization of parameters in sub-regions of the initial search space, illustrating coupling between the parameters.

The addition of the AD factor into the model addressed this inadequacy and importantly supported the curvature observed in the experimental data. This result indicates the AD of viruses to hydroponic tank wall is an important factor to include in predicting viral concentration in all three compartments (water, root, shoot). Credible and prediction intervals in the shoot at harvest were similar for both models.

The credible intervals of the predicted parameters also show the adequacy of fit for the model with AD (Table 2.3). Four of the predicted parameters: a_t,b_t,$k_{\mathrm{dec,s}}$ and k_p, were restricted to a smaller subset of the search bounds, indicating that they were identifiable. In contrast, the viral transfer efficiency η and the kinetic parameters (k_{att},k_{det}) spanned the entirety of their search space and were poorly identifiable. However, this does not suggest that each parameter can independently take any value in its range because the joint distributions of the parameters (Fig. 2.3) indicate how fixing one parameter (e.g. η_{gr}=0.9) influences the likelihood of another parameter (η_{rs} most likely to be closer to 1). Hence, despite the large range of an individual parameter, the coordination between the parameters constrained the model predictions to produce reliable outcomes (section 2.4.1, Fig. A.5). Hence, I consider the

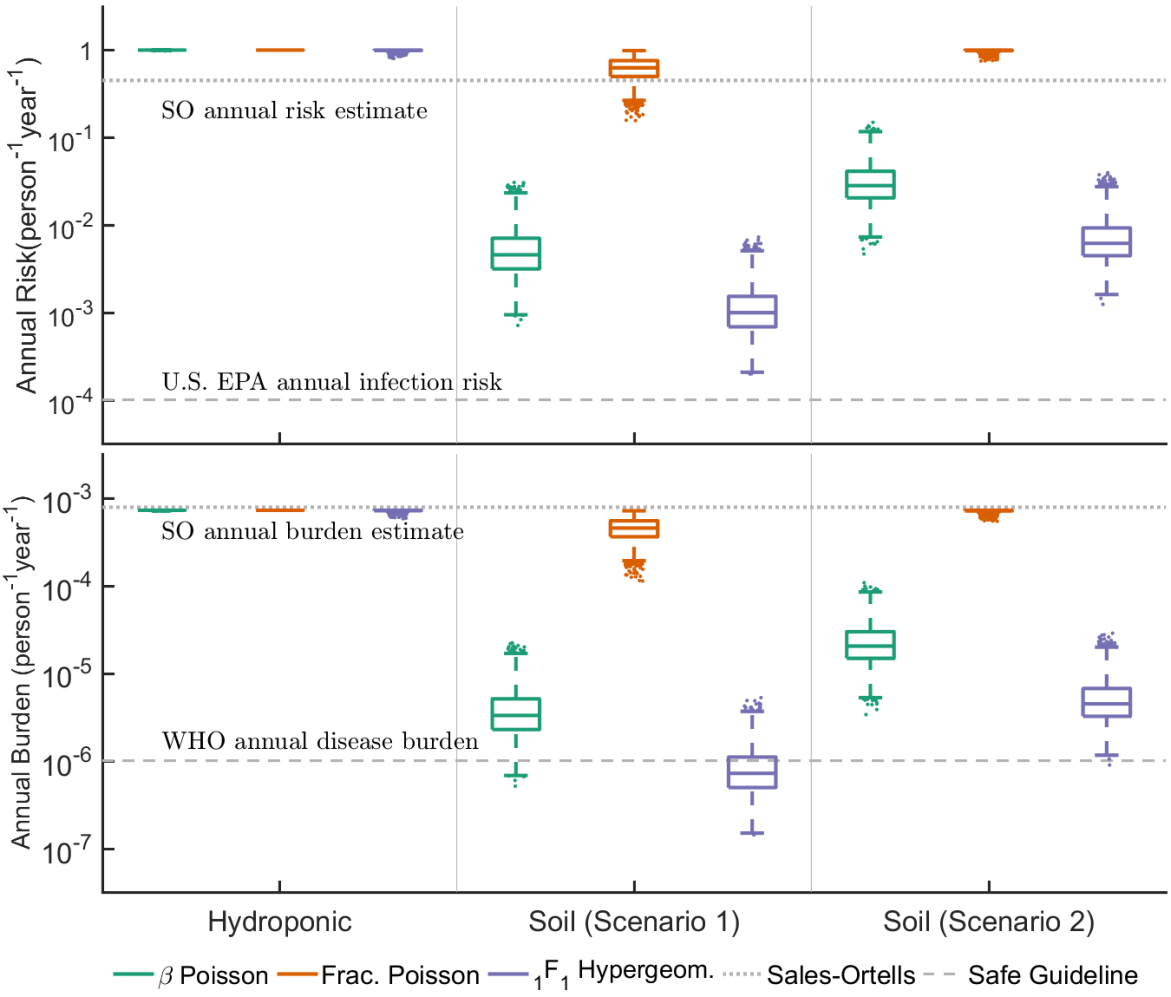

Figure 2.4: Annual risk (top panel) and disease burden (bottom panel) of norovirus infection through consumption of lettuce grown hydroponically or in soil (scenario 1 and 2 explained in section 2.2.2) using treated sewage effluent. The dashed lines indicate existing risk benchmarks or the mean plotted from a previous study by [119].

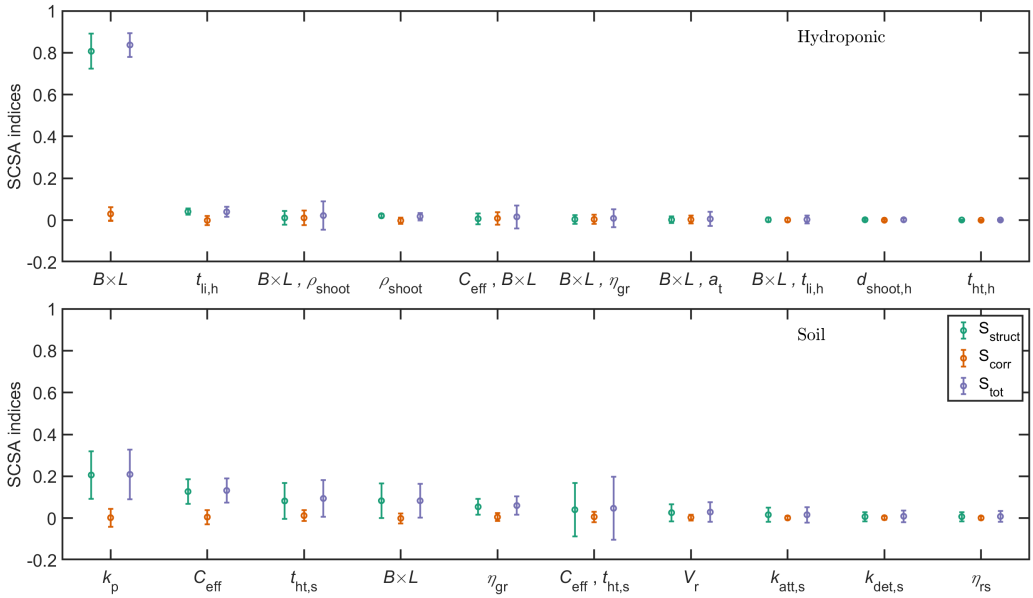

Figure 2.5: Top 10 most significant SCSA sensitivity indices (mean±s.d., 100 bootstrapped samples) for hydroponic (top panel) and soil (bottom panel) grown lettuce. S_{struct}, S_{corr} and S_{tot} correspond to S_a, S_b and S in [83]. A comma denotes second order sensitivity terms for pairs of parameters.

performance of the model with AD adequate for estimating parameters used it for predicting risk.

2.3.2 Health risks from lettuce consumption

Risk estimates for lettuce grown in the hydroponic tank or soil are presented in Fig. 2.4. Across these systems, the FP model predicted the highest risk while the 1F1 model predicted the lowest risk. For a given risk model, higher risk was predicted in the hydroponic system than in the soil. This is a consequence of the very low detachment rates in soil compared to the attachment rates. Comparison of results from Sc1 and Sc2 (section 2.2.2) of soil grown lettuce indicated lower risks and disease burdens under Sc1 (Fig. 2.4).

Comparing with the safety guidelines, the lowest risk predicted in the hydroponic system is higher than the U.S. EPA defined acceptable annual drinking water risk of 10^{-4} (Environ-

mental Protection Agency, 2010) for each risk model. The annual burdens are also above the 10^{-6} benchmark recommended by the WHO [143]. In the case of soil grown lettuce, neither Sc1 nor Sc2 met the U.S. EPA safety benchmark. Two risk models predicted borderline disease burden according to the WHO benchmark, for soil grown lettuce in Sc1, but under Sc2 the risk still did not meet the safety guideline. I found that neither increasing holding time of the lettuce to two days after harvesting nor using bigger tanks significantly altered the predicted risk (Fig. A.1). In comparison, the risk estimates of [119] are higher than range of soil grown lettuce outcomes presented here (Fig. 2.4) for 2 of 3 models.

The SCSA sensitivity indices are presented in Fig. 2.5. For hydroponically grown lettuce, the top 3 factors (by S_{tot}) influencing daily risk are amount of lettuce consumed, time since last irrigation and the term involving consumption and ρ_{shoot}. Also, the risk estimates are robust to the fitted parameters (Table 2.3) despite low identifiability of some model parameters (a_t, b_t, $k_{\text{att,h}}$ and $k_{\text{det,h}}$). For soil grown lettuce, k_p appears to be the major influential parameter, followed by the input viral concentration in irrigation water and the lettuce harvest time. S_{corr} is near zero, suggesting lesser influence of correlation in the input parameters.

2.4 Discussion

In this study, I modeled the internalization and transport of NoV from irrigation water to the lettuce using ordinary differential equations to capture the dynamic processes of viral transport in lettuce. This first attempt is aimed at underscoring the importance of the effect of time in determining the final risk outcome. The modeling approach from this study may be customized for other scenarios for the management of water reuse practices and for developing new guidelines for food safety. Moreover, this study identifies critical gaps in the current knowledge of pathogen transport in plants and calls for further lab and field studies to better understand risk of water reuse.

2.4.1 Fitting model to data

To predict viral transport in plant tissue, it is necessary to couple mathematical assumptions with an understanding of the underlying biogeochemical processes governing virus removal, plant growth, growth conditions and virus-plant interactions. For example, although a simple transport model without AD could predict the viral load in the lettuce at harvest, it failed to capture the initial curvature in the viral load in the growth medium (water). An alternative to the AD hypothesis that could capture this curvature is the existence of two populations of viruses as used in [104], one decaying slower than the other. However, I did not adopt this approach as the double exponential model is not time invariant. This means that the time taken to decay from a concentration C_1 to C_2 is not unique and depends on the history of the events that occurred (Fig. A.2). Other viral models, such as the ones used in [103] faced the same issues. Incorporating AD made the model time invariant and always provided the same time for decay between two given concentrations. This model fitting experience showcases how mathematics can guide the understanding of biological mechanisms. The hypothesis of two different NoV populations is less plausible than that of viral attachment and detachment to the hydroponic tank. While it appears that incorporating the AD mechanism does not significantly improve the accuracy of viral load predictions in the lettuce shoot at harvest, this is a consequence of force fitting the model to data under the given conditions. Changing the conditions, for example, by reducing viral attachment rate to the tank wall, underestimates the viral load in the lettuce shoot in the absence of AD (Fig. A.3). I believe that the model can be improved significantly with new insights on virus-plant interactions.

A potential cause for concern in the model fit is the wide credible and prediction intervals. However, there is significant uncertainty in the data as well (e.g. root day 3, shoot day 1, Fig. 2.4) suggesting that the transport process itself is noise prone. Moreover, from the perspective of risk assessment, the variability between dose-response models is higher than the within dose-response model variability (Fig. A.5). Since within dose-response model variabil-

ity stems from uncertainty in viral loads at harvest among other factors, the wide intervals do not exert a bigger effect than the discordance from different dose-response models.

2.4.2 Model parameter estimates

Some parameters (i.e., $k_{\mathrm{dec,h}}$, k_p) are identifiable to a reasonable degree through model fitting, but there is a large degree of uncertainty in the viral transport efficiencies and the AD kinetic parameters. While this could be a consequence of fitting a limited number of data points with several parameters, the viral load at harvest and risk estimates were well constrained. This large variation in parameters and 'usefully tight quantitative predictions' (Fig. A.5) is termed the sloppiness of parameter sensitivities, and has been observed in physics and systems biology [55, 139]. Well-designed experiments may simultaneously reduce uncertainty in the parameters as well as predictions [6, 24] and therefore increasing confidence in predictions. One possible experiment to reduce parameter uncertainty is recording the transpiration and growth rate to fit eq. (2.6) independently to get a_t and b_t.

2.4.3 Risk estimates

An interesting outcome of my analysis is the strong association of risk with plant growth conditions. The health risks from consuming lettuce irrigated with recycled wastewater are highest in hydroponic grown lettuce, followed by soil grown lettuce under Sc2 and the least in soil grown lettuce under Sc1 (Fig. 2.4 and section 2.2.2). This difference in risk estimates stems to a large degree from the difference in AD kinetic constants ($k_{\mathrm{att,s}}$, $k_{\mathrm{det,s}}$) (Fig. A.4). Increasing $k_{\mathrm{att,s}}$ (holding $k_{\mathrm{det,s}}$ constant) will decrease risk as more viruses will get attached to the growth medium, while increasing $k_{\mathrm{det,s}}$ (holding $k_{\mathrm{att,s}}$ constant) will have the opposite effect (Fig. A.4), as more detached viruses are available for uptake by the plant. The combined effect of the AD parameters depends on their magnitudes and is portrayed

in Fig. A.4. This result indicates that a better understanding of the virus interaction with the growth environment can lead to an improved understanding of risk. More importantly, this outcome indicates that soil plays a vital role in the removal of viruses from irrigation water through the adsorption of viral particles. An investigation focused on understanding the influence of soil composition on viral attachment will help refine the transport model.

The risk predicted by this dynamic transport model is higher than the EPA annual infection risk as well as the WHO annual disease burden benchmark. The reasons for this outcome are many-fold. First, there is a significant variability in the reported internalization of viruses in plants. In research of data for modeling NoV transport in plant, I filtered the existing data using the following criteria: 1) human NoV used as the seed agent, 2) presence of quantitative viral results in the growth medium and different locations of the plant. Based on these criteria, the data from [37] represent the best available data on viral internalization and transport in lettuce. However, it is also important to note that a similar study by [134] did not observe human NoV internalization in lettuce. This discrepancy could be due to the specific subspecies of the plant and growth conditions used in the studies. Besides, minor changes such as damages in roots or decrease in humidity of the growing environment can promote pathogen internalization [62, 137]. Alternatively, tracking viral transport through the growth medium and the plant is challenging, which may yield false results due to reaction inhibitions in genome amplification and inferior detection limit.

The risk outcome of this study is conservative because it assumes an individual consumes the wastewater irrigated lettuce daily for an entire year. This assumption and the corresponding higher risk estimates are only applicable to a small portion of consumers, while most consumers in the U.S. are likely to have a more diverse diet. While the model outcomes presented here represent the best attempt given the available data, it is also possible that the internalization observed by [37] is an extreme case and and typically internalization occurs to a lesser extent.

As previously discussed by others ([122, 136]), risk estimates by different NoV dose-response models differed by orders of magnitude. This study primarily aims to introduce a viral transport model without advocating any one dose-response model. The future refinement of pathogen dose-response models will reduce variability in risk estimates.

The risk of consuming lettuce grown in soil as predicted by [119] is higher than my predictions, although I used the results of [37] in both studies. This is a consequence of considering the greater adsorption capability of soil, which is not reflected when assuming a simple input:output ratio. Using different inoculating concentrations of NoV, body weight and consumption rate distributions also contributed to the difference in the outcomes but to a lesser extent.

I obtained parameters for crisphead lettuce from several different sources, each possibly using a different sub-variety of crisphead. Nevertheless, global sensitivity analysis showed insensitivity of risk estimates to several assumed and fitted parameters (a_t, b_t, $d_{\mathrm{root,h}}$,V_e), lending confidence to the approaches taken to parametrize the model. The sensitivity to $t_{\mathrm{li,h}}$ in hydroponic and $t_{\mathrm{ht,s}}$ in soil cases underscores the importance of considering the dynamics of viral transport. This suggests that given no change in lettuce consumption, changes in irrigation schedule can affect risk outcomes. Such arguments are not possible with the approach of [119]. In soil-grown lettuce, the high sensitivity to k_p indicates the role of plant-specific processes in mediating risk outcomes.

2.4.4 Contribution and future directions

In addition to a transport model predicting the NoV load in lettuce, I explored strategies to reduce the risk of NoV gastroenteritis (Fig. A.1) by increasing holding time of the produce after harvesting or using larger hydroponic culture volumes. Although neither strategy could significantly alleviate the risks, the process highlights two strengths of modeling: 1)

It provides analytical support for arguments that would otherwise be less convincing; 2) It predicts outcomes of experiments without the physical resources required to perform them. For instance, the model can be used to explore alternate irrigation schedules to reduce the NoV internalization risk.

Modeling also helps encapsulate our understanding of the system and generate hypotheses. For example, simple first-order decay did not produce the trend observed in the water, which suggests that additional mechanisms are at play. I postulated the attachment of virus particles on the walls of the hydroponic system as one possible mechanism and examined the fit of the model. Although viral attachment to glass or other materials has been observed before [15], here it stands as a hypothesis that can be tested. In addition to generating and testing hypotheses, some of my model assumptions raise broader questions for future research. For example, I assumed that viruses are transported at the transpiration rate from the growth medium to the roots. However, not much is known regarding the role of roots in the internalization of viruses. Investigating the defense mechanisms of plants' roots to passive viral transport, i.e., through rhizosphere microbiome interactions, may shed light on the broad understanding of plant and microbe interactions.

The question of extending this model to other pathogen and plant systems draws attention to the dearth of data in enabling such efforts. While modeling another virus may not require changes to the model, understanding transport in other plants can be challenging. Data required includes models for growth rate and transpiration, plant growth characteristics including density, water content, as well as internalization studies to determine transport efficiencies. However, from the perspective of risk management, lettuce may be used as the worst-case scenario estimate of risk in water reuse owing to its high consumption with minimal pathogen inactivation by cooking. This worst-case scenario can be used to set water quality standards for irrigation water for the production of fresh produce eaten raw. The models can also be extended to include pathogen transport to the plant tissue from

manure/biosolids that are used as organic fertilizer.

2.5 Conclusion

I developed a dynamic viral transport model to predict the viral load in the lettuce at harvest, given the viral load in the recycled water used for irrigation. Integrating the viral load with the exposure model and NoV dose-response models, I estimated the annual infection risk and disease burden through the consumption of lettuce irrigated with recycled wastewater. My conclusions are:

1. Viral transport in the plant depends on its interaction with the growth medium and the plant tissue to a large extent.

2. The experimental data reported in literature are best explained by the incorporation of attachment and detachment of the viruses to the cultivation tank.

3. Kinetic rates for attachment and detachment, as well as transport efficiencies between plant compartments, were loosely constrained in their search bounds (low identifiability). However, the joint distributions of the parameters and the final risk predictions were well constrained, highlighting the sloppy parameter sensitivities.

4. The overall risk estimates from the model are higher than the commonly accepted infection risk benchmark and annual disease burden. However, there are large uncertainties in the experimental data of viral transport through plants.

5. The model provides a foundation to incorporate new data on pathogen transport and plant-microbe interactions to develop a holistic understanding of pathogen internalization.

Chapter 3

Quantifying the infection risk of antibiotic-resistant bacteria

The contents of this chapter appear in the journal Scientific Reports [26].

3.1 Background

The rise of antibiotic resistance in bacteria, coupled with the slowdown of drug discovery, presents a growing threat to public health [144]. The Centre for Disease Control and Prevention (CDC) attributes at least 2 million illnesses and 23,000 deaths a year in the US to antimicrobial resistance. The economic burden in the US is estimated to be on the order of US $21 to $34 billion [23].

Quantifying the human health risk associated with antibiotic resistance presents several challenges [8, 14, 43, 65, 108, 101, 88]. The human health outcome (e.g., illness not responding to a specific antibiotic or 'resistant' illness) is influenced by antibiotic-resistant bacteria (ARB), antibiotics (ABs), antibiotic-resistant genes (ARG) and their carriers. ARB cause

antibiotic-resistant illness and present the most direct threat. Antibiotics and other selective pressures (e.g., heavy metals) in the environment promote enrichment of ARB and induce *de-novo* resistance mutations in antibiotic susceptible bacteria (ASB) or the uptake of ARG which is known as horizontal gene transfer (HGT). The ARG can come from direct contact with bacteria harboring ARG (conjugation), from phages harboring ARG (transduction) or from free-floating mobile genetic elements (MGEs) in the environment (transformation). ABs in the environment can potentially modulate HGT. Thus the human health outcome from ingesting ARB can be influenced by a complex network of factors.

Studies have found ARB in various sources across the globe, including wastewater treatment plant effluent, recreational water, drinking water (see [101] for a list), and even lettuce at harvest [64]. Nevertheless, microbial risk assessments involving ARB are rare [47, 124], which is due to the lack of a dose-response model (DRM) and uncertainties at each step in a traditional bottom-up risk assessment approach. Few past studies (e.g., [41, 148]) investigated the burden of ARB using a top-down approach by identifying the contribution of veterinary AB use to the overall number of AB resistant disease instances [124]. However, this top-down framework cannot be used to compute the risk posed by an exposure event (such as swimming in the recreational waters discussed in [82]), nor can it be used to set regulatory guidelines for acceptable levels of ARB or residual ABs in the environment.

Attempts to relate ARB concentrations to ARB caused diseases (bottom-up approach) were made in several studies [5, 30, 68, 107]. Since DRMs tailored to ARB do not exist, these studies draw on epidemiological data (e.g., annual illness cases where some AB fails) to predict human health effects. However, there is a large variability of the estimates depending on the scope and geographic region of the investigation. These past studies are useful to draw inferences on the region that the data are based on but may not apply to other regions e.g., resource-limited countries where epidemiological data are not available. Moreover, even if data are available, avoiding confounders to pinpoint the true cause of an antibiotic-resistant

illness is non-trivial. Also, these approaches cannot account for the fraction (f_r) of ARB in a bacteria-contaminated sample or the effects of residual ABs in the environment. Hence, there is a need for DRMs that can account for ARB [82, 101], their fraction (f_r) in the dose and residual ABs.

Obtaining data to develop a DRM for human illness involves infecting a cohort of people with a known pathogen, including AB resistant ones. However, since inoculating people with ARB would result in untreatable illnesses, such data are not available. An alternative approach is to use the existing data collected from human studies involving ASB and assume that in the absence of AB, the same DRM used for the ASB applies to ARB. This assumption is conservative since AB resistance often confers a fitness cost [153]. Some studies have identified mutations that lead to increased fitness have [12, 54]. While data from animal models present a second alternative, usage of human data, where available, is preferred.

The development of DRM for ARB is further complicated when the person under consideration is exposed to residual/sub-inhibitory levels of ABs. This exposure can stem from medication for a previous illness, prophylactic use for surgeries, receiving AB-releasing stents, AB residues in food of animal origin, or even the environmental sources listed earlier. The source of ARB could also contain AB, causing the host to get exposed to sub-inhibitory levels of AB. In this case, the health outcome (not ill, AB treatable illness, or AB untreatable illness) will depend on not only the initial dose (d) and fraction of ARB (f_r), but also the amount of residual AB. This is because the AB will affect the growth rates of the susceptible and resistant subpopulations differently in a concentration-dependent manner. Popular/classical microbial pathogen DRMs, such as the exponential or beta-Poisson models, are not readily amenable to investigation of such changes in growth rates or conversion from ASB to ARB due to mutation or HGT. Hence there is a need for DRMs based on growth processes, which can set the stage for developing a holistic understanding of the dose-response of ARB.

The goal of this research is to quantify the risk posed by ARB and the effect of selection

pressure exerted by sub-inhibitory concentrations of AB. Specifically, I introduce a new DRM based on Simple Death processes. This DRM begins with the following assumptions: 1) a portion of ingested dose of d bacteria may die off (solid lines, Fig. 3.1a) when they encounter the host's defenses. This includes immune factors, gastric acids, and other factors. 2) At greater d, this death rate may not be enough to kill off all the bacteria. In some cases, enough bacteria survive to initiate an infection (Trial 2, Fig. 3.1a), resulting in growth (dotted line, Fig. 3.1a) of bacteria in human body (infection). I model these two assumptions with continuous-time Markov chains (CTMCs) to capture stochastic bacterial kinetics. This differs from existing approaches which derive kinetics from classical dose-response assumptions[67] or which use deterministic kinetics to inform dose-response [90, 112]. My approach is similar to the approaches of [109, 147, 49] and other studies that use CTMCs, but uses analytical solutions instead of simulations. Another study [18] explores an analytical approach specific to anthrax and uses *in vivo* animal data to fit parameters. However, it is not clear if the parameters found in animals apply to humans. Here, I show that my approach integrates with classical DRMs and can hence extend to all pathogens analyzed with the classical DRM framework.

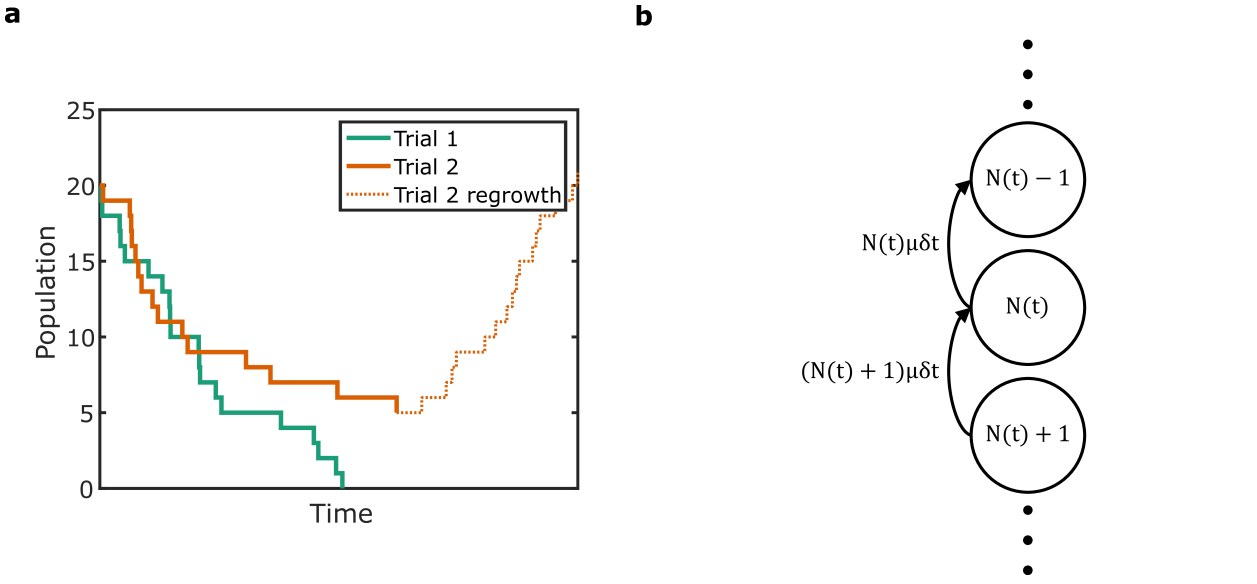

Figure 3.1: Overview of simple death process. (**a**) Plots of two trials of a simple death process. (**b**) The Markov chain of a simple death process.

This study is organized as follows. I introduce the new DRM and fit it to published dose-response data of AB susceptible *E. coli* that causes diarrhea. I then identify a relationship between the kinetic constants in the proposed model and the existing DRMs. I use this relationship to predict the risk of illness and the illness subtype (AB treatable vs. AB untreatable) for mixed doses of ARB and ASB in the presence of residual AB. Finally, I highlight the limitations of this approach and the need for improved data collection complementing this approach. Parameters and their symbols presented in this chapter are listed in Table 3.1.

Symbol	Units	Description
C	mg L^{-1}	AB concentration
d	CFU	Dose of bacteria
$E_{\max,}$ EC_{50}	day^{-1}, mg L^{-1}	AB-bacteria interaction parameters
f_r	-	Fraction of ARB in initial dose
n_{ill}	persons	Number of ill subjects
n_{tot}	persons	Total number of subjects
$P(d,t)$	-	Response probability for initial bacterial load of d at t
$P_{\text{ext}}(d,t)$	-	Probability of extinction of initial bacterial load of d at t
$P(d)$	-	Response probability for initial bacterial load of d
r	CFU^{-1}	Exponential DRM parameter
t	days	Time
t_{fs}	days	The latest time at which a first symptom is observed
α, β	-, CFU	beta-Poisson DRM parameters
μ	day^{-1}	Death rate of bacteria
$\mu_{s,\text{AB}}(C)$	day^{-1}	Death rate of ASB in AB of concentration C
$\mu_{r,\text{AB}}(C)$	day^{-1}	Death rate of ARB in AB of concentration C

Table 3.1: Parameters used in this study.

3.2 Results

3.2.1 Dose-response expression

I model the initial die-off of the bacteria after they enter the host as a stochastic *death process*, which is a kind of CTMC. In such a process, a single death is assumed to occur at a random point in time (solid lines, Fig. 3.1a). The key assumption behind the *Simple* Death (SD) process (Fig. 3.1b) is that in the short time interval $(t, t + \delta t)$, each bacterium dies with probability $\mu\delta t$. Hence, the probability of a death in the time interval $(t,\ t + \delta t)$ is $N(t)\mu\delta t$, where $N(t)$ is the population size at time t. As time goes by, $N(t)$ will reduce as some bacteria begin to die, thus decreasing the probability of a death in the time interval $(t,\ t + \delta t)$. Assuming that the probability of observing a response (here illness) equals the probability of not observing an extinction, I express the relation with eq. 3.1.

$$P(d, t) = 1 - P_{\text{ext}}(d, t) \tag{3.1}$$

where $P(d, t)$ is the probability of observing a response and $P_{\text{ext}}(d, t)$ is the probability of extinction for initial dose d and time t. The expression for $P_{\text{ext}}(d, t)$ [3] is given by

$$P_{\text{ext}}(d, t) = (1 - \exp(-\mu t))^d \tag{3.2}$$

Intuitively, $1-\exp(-\mu t)$ is the probability of the death of one bacterium [3]. The probability of d bacteria dying is the product of the probability of each bacterium dying, resulting in eq. 3.2. Therefore, I express the time-dependent dose-response relationship or the time-dependent SD DRM as:

$$P(d, t) = 1 - (1 - \exp(-\mu t))^d \tag{3.3}$$

I set t in the model (eq. 3.3) to the latest time that the first symptom is observed among all subjects (t_{fs}) in the clinical feeding study. This is because at $t < t_{\text{fs}}$, the n_{ill} used to fit the model is higher than the number of people ill at that time. At $t > t_{\text{fs}}$, the die-off assumption will not hold as the bacterial population will enter the growth phase in the human body, which results in illness. Hence the time-independent dose-response relationship is given by eq. 3.4.

$$P(d) = 1 - (1 - \exp(-\mu t_{\text{fs}}))^d \tag{3.4}$$

Here, d is the ingested dose of bacteria that, $P(d)$ is the response probability, μ is the rate of death, and t_{fs} is the latest time at which the first symptom is observed among all subjects. I refer to this model as the SD DRM through the remainder of this chapter unless specified otherwise.

3.2.2 Relationship with the existing DRMs

The exponential DRM [57] is a widely used and best accepted model for dose-response of pathogenic *E. coli* in humans. It is expressed as

$$P(d) = 1 - \exp(-rd) \tag{3.5}$$

where r is the probability of the pathogen surviving to cause infection once it is ingested. When comparing the SD DRM with the exponential DRM, it is clear that exponential DRM is a special case of the time dependent SD DRM (eq. 3.3) where $t = t_{\text{fs}}$.

$$(1 - \exp(-\mu t_{\text{fs}})) = \exp(-r) \tag{3.6}$$

This result implies that SD DRM will fit any dataset that the exponential DRM fits. In addition, the SD DRM establishes a link between an abstract parameter that is informed by dose-response data (r) and a parameter that has a clear biophysical interpretation (μ). This biophysical parameter becomes instrumental in accounting for the effect of the AB.

Similarly, the above approach can also be extended to the existing beta-Poisson DRM, if a relationship between the beta-Poisson DRM parameters and the death rate (μ) of the SD DRM is established. The beta-Poisson DRM is an approximation of the hypergeometric DRM [57]. In other words, the hypergeometric DRM is a generalized case of the exponential DRM, where the r value is assumed to be beta distributed. The classical beta-Poisson is DRM given by:

$$P(d) = 1 - \left(1 + \left(\frac{d}{\beta}\right)\right)^{-\alpha} \tag{3.7}$$

where α and β are the parameters of beta-Poisson model. An analytical approach to finding the required relationship was not tractable, and instead, I adopted a numerical approach (see section 3.4). With this relationship, the beta-Poisson model can also be investigated under AB loads. Thus, the results demonstrate that the SD approach provides a unified framework to analyze the effect of ABs on dose-response.

3.2.3 Accounting for residual AB concentration

In addition to the dose of ARB (or the f_r, the fraction of resistant bacteria in the dose), the human health outcome also depends on C, the concentration of sub-lethal/residual AB in human body. To account for the effect of AB, one can first adopt models that relate AB concentration and *E. coli* death rate published in literature e.g. [100, 99, 20]). These death rates can then be related to dose-response by using eq. 3.6.

AB causes a significant increase in the death rate of ASB but has less or no effect on ARB. Hence, a simple conservative assumption is that the ARB death rate is not affected by the presence of AB. This assumption is the worst-case scenario because it does not include the fitness cost of AB resistance [153]. I illustrate this relationship with a case study in which I investigate the effect of Gentamicin on a dose of *E. coli*. The increased death rate of the susceptible strain in the presence of AB ($\mu_{s,\mathrm{AB}}$) is given by the sigmoidal model of [99] as

$$\mu_{s,\mathrm{AB}}(C) = \mu + \frac{E_{\max} C}{EC_{50} + C} \tag{3.8}$$

C is the concentration of AB. μ is the death rate of the ASB strain in the absence of AB. It can be obtained from eq. 3.6. $E_{\max}$ (= 1224 day^{-1}, the maximum killing rate) and EC_{50} (= 9.93 mg L^{-1}, AB concentration at half maximum killing rate) are values that determine how C affects $\mu_{s,\mathrm{AB}}(C)$. I obtained numerical values for $E_{\max}$ and EC_{50} from [99], where the effect of Gentamicin on *E. coli* death kinetics was studied. The death rate of the ARB in the presence of AB ($\mu_{\mathrm{r,AB}}$) is μ as per the earlier assumption. The probability of illness of doses consisting of purely ASB or purely ARB can thus be estimated by plugging the estimated $\mu_{s,\mathrm{AB}}(C)$ into eq. 3.4.

3.2.4 Accounting for fraction of ARB

I account for f_r by first assuming, like in existing DRMs, that any two bacteria act independently of each other. Scaling this up implies that the susceptible and resistant subpopulations will also act independent of each other. Therefore, the joint probabilities concerning both subpopulations can be written as the product of the probabilities of each subpopulation. For example, the probability of both subpopulations going extinct ($S_{\mathrm{ext}}R_{\mathrm{ext}}$ in Table 3.2) is given by $P_{\mathrm{ext},s}(d|f_r, C) \times P_{\mathrm{ext},r}(d|f_r, C)$. Therefore the probability of response ($P(d|f_r, C)$) is equal to the complement of the probability of both populations going extinct i.e,

	S_{sur} $(1 - P_{\text{ext},s}(d\|f_r, C))$	S_{ext} $(P_{\text{ext},s}(d\|f_r, C))$
R_{sur} $(1 - P_{\text{ext},r}(d\|f_r, C))$	$(S_{\text{sur}}R_{\text{sur}})$ Illness AB Treatable	$(S_{\text{ext}}R_{\text{sur}})$ Illness AB Untreatable
R_{ext} $(P_{\text{ext},r}(d\|f_r, C))$	$(S_{\text{sur}}R_{\text{ext}})$ Illness AB Treatable	$(S_{\text{ext}}R_{\text{ext}})$ No illness

Table 3.2: Possible outcomes (S: susceptible, R: resistant, $_{\text{sur}}$: survives and $_{\text{ext}}$: goes extinct)

$$P(d|f_r, C) = 1 - P_{\text{ext},s}(d|f_r, C)P_{\text{ext},r}(d|f_r, C) \tag{3.9}$$

where the extinction probability of the susceptible subpopulation is

$$P_{\text{ext},s}(d|f_r, C) = (1 - \exp(-\mu_{s,\text{AB}}(C)t_{\text{fs}}))^{d\times(1-f_r)} \tag{3.10}$$

and the extinction probability of the resistant subpopulation is

$$P_{\text{ext},r}(d|f_r, C) = (1 - \exp(-\mu_{r,\text{AB}}t_{\text{fs}}))^{d\times f_r} \tag{3.11}$$

3.2.5 Possible health outcomes

Two types of health outcomes are possible when a mixed dose of ARB and ASB are involved. When the ASB strain out-competes the ARB strain ($S_{\text{sur}}R_{\text{ext}}$, Table 3.2) to grow and infect the host, the illness would likely be susceptible to the AB treatment under consideration. However, when the ARB subpopulation continues to survive irrespective of what happens to the ASB subpopulation ($S_{\text{sur}}R_{\text{sur}}$ and $S_{\text{ext}}R_{\text{sur}}$), the resulting illness may not be treatable with the AB. I label these health outcomes as 'AB Treatable' and 'AB Untreatable' illness, respectively (Fig. 3.2). To identify the type of health outcome given an illness occurs, I

compare the probabilities of these two events and classify as follows:

$$\begin{cases} \text{AB Treatable}, & (1 - P_{\text{ext},s}(d,t|f_r,C))P_{\text{ext},r}(d,t|f_r,C) > (1 - P_{\text{ext},r}(d,t|f_r,C)) \\ \text{AB Untreatable}, & \text{otherwise} \end{cases}$$

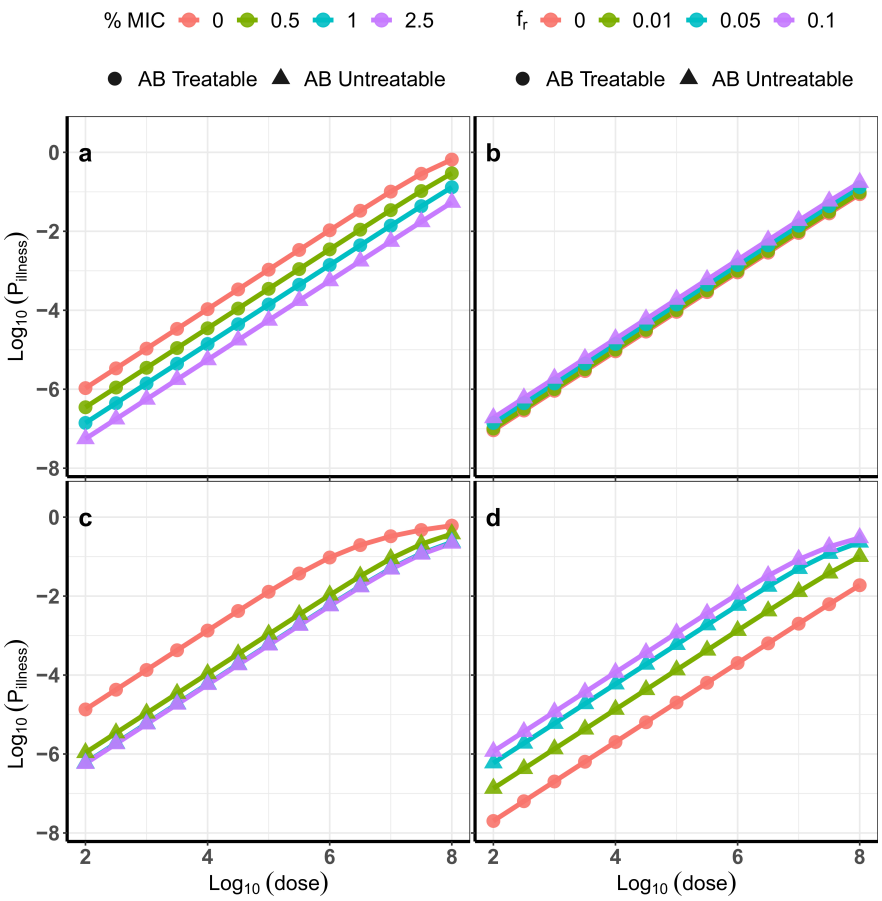

Figure 3.2: Effect of varying f_r and C on illness outcomes. (**a**) Exponential DRM, DS1, $f_r = 0.05$. (**b**) Exponential DRM, DS1, $C = 1\%$ MIC. (**c**) beta-Poisson DRM, DS2, $f_r = 0.05$. (**d**) beta-Poisson DRM, DS2, $C = 1\%$ MIC.

Taken together, the risk of infection due to mixed doses (consisting of ARB and ASB) can be estimated as a function of C. I predict the host response to AB treatment based on the subpopulation of the AB resistant *E. coli* ingested. The implementation of this approach

differs from the currently used exponential and beta-Poisson DRMs.

3.2.6 Effect of ARB fraction f_r and residual AB concentration C

To understand the effect of ARB fraction f_r and residual AB concentration C, the first step is identifying the death rate of the ASB population (μ in eq. 3.8). To this end, I fit the existing exponential and beta-Poisson DRMs to the human clinical datasets (see Table B.1, data obtained from QMRA Wiki [142]) under consideration. Dataset 1 (DS1) uses mild to severe diarrhea as the endpoint to measure positive response, while dataset 2 (DS2) uses diarrhea. Both datasets use oral route of exposure to *E. coli* and exhibit a significant trend at the 0.05 level (one-tailed Cochran-Armitage test [57], n = 6 and P value = 1.91×10^{-4} for DS1, n = 11 and P value = 1.11×10^{-5} for DS2).

Fig. 3.3 presents the DRM fits and Table 3.3 summarizes the DRM choice for each dataset based on the χ^2 test, which is used for model selection for DRMs [57]. The results show that the exponential DRM fits DS1 better, while the beta-Poisson DRM fits DS2 better. The fitted parameters allow one to determine μ, which can then be used to estimate the death rate in the presence of AB ($\mu_{s,\mathrm{AB}}$) using eq. 3.8.

Fig. 3.2 shows the model behavior at different values of f_r and C. For a given dose of bacteria and a given f_r, increasing the concentration of AB decreases the risk (Fig. 3.2a, c). This is expected as the higher AB load kills off more of the bacteria. When $f_r = 0.05$, increasing C from 0 to 2.5% MIC decreases the risk by around 1.5 orders of magnitude. This decrease in risk is more gradual for DS1 (Fig. 3.2a) compared to DS2 (Fig. 3.2c). Not much difference is observed between 1% MIC and 2.5% MIC for DS2, suggesting that the antibiotic effect saturates at small fractions of the MIC. However, increasing C also increases the likelihood of the illness not treatable by the AB (Fig. 3.2a, c, 'AB Untreatable'). This is attributed to the higher survival capability (or lower death rate) of the ARB subpopulation compared

Data	Case	Dev	χ^2	p-value	Conclusion	Model
DS1	Exp	3.19	11.07	0.67	Fail to reject "Exp is a good fit".	Exp
	beta-Poisson	0.95	9.49	0.92	Fail to reject "beta-Poisson is a good fit".	$r = 1.07 \times 10^{-8}$
	Compare	2.24*	3.84	0.13	Fail to reject "Exp fits better than beta-Poisson".	($\mu = 18.35$ day^{-1})
DS2	Exp	57.82	18.31	9.36×10^{-8}	Reject "Exp is a good fit".	beta-Poisson
	beta-Poisson	14.44	16.92	0.11	Fail to reject "beta-Poisson is a good fit".	$\alpha = 0.16$
	Compare	43.38*	3.84	4.51×10^{-11}	Reject "Exp fits better than beta-Poisson".	$\beta = 1.41 \times 10^6$

Table 3.3: Fit of the exponential and beta-Poisson DRMs. 'Dev' is the minimum deviance, except the starred (*) values, which are the differences in minimum deviance between exponential and beta-Poisson RMs. 'Model' represents the preferred DRM based on conclusions. Best fit parameters are also shown.

to the ASB subpopulation. Further difference between DS1 and DS2 is observed, as the concentration of Gentamicin that results in predominantly AB untreatable illness is $> 1\%$ MIC for DS1 but $< 1\%$ for DS2.

For a given dose of bacteria and a given C, increasing f_r also causes an increased risk and greater likelihood of the AB untreatable illness (Fig. 3.2b, d). The higher risk is due to the lower death rate of the ARB subpopulation. The greater likelihood of AB Untreatable illness stems from the greater number of ARB in the initial load, meaning more ARB are likely to survive with time. Specifically, changing f_r appears to have little effect under DS1 when C is fixed to 1% MIC, indicating that antibiotic effect is near saturation. In contrast, increasing f_r from 0 to 0.1 under DS2 increases risk by two orders of magnitude, indicating that antibiotic effect reaches saturation at $> 1\%$ MIC. The switch from AB Treatable occurs

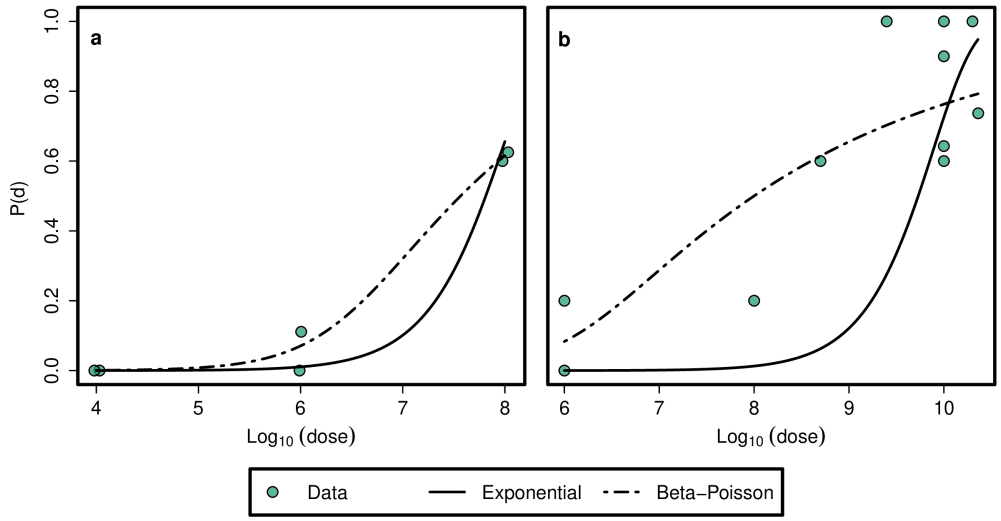

Figure 3.3: Model fits for (**a**) DS 1 and (**b**) DS 2. To avoid overlapping points in the plot, noise is added along the horizontal axis.

at $0.05 < f_r < 0.1$ under DS1 but at $f_r < 0.01$ under DS2.

The magnitude of the impact of f_r and C depends on the DRM and the datasets used to fit the DRM (Fig. 3.2). To better understand this dependence, I performed a sensitivity analysis using the PAWN algorithm [105] (see section 3.4). The results show that the dose of bacteria is the biggest determinant of risk for both the exponential and beta-Poisson DRMs (Fig. 3.4a, b). When one fixes the dose of bacteria, C and f_r play a bigger role in determining risk than the dose-response parameters (Fig. 3.4c, d). The parameters capturing the effect of the AB ($E_{\max}$ and EC_{50}) play a smaller role than the dose-response parameters (Fig. 3.4c, d). Further, I verified that setting $t = t_{\mathrm{fs}}$ has a minimal effect on the final results since sensitivity indices of t_{fs} fall below the threshold value, indicating that t_{fs} is non-influential (Fig. 3.4).

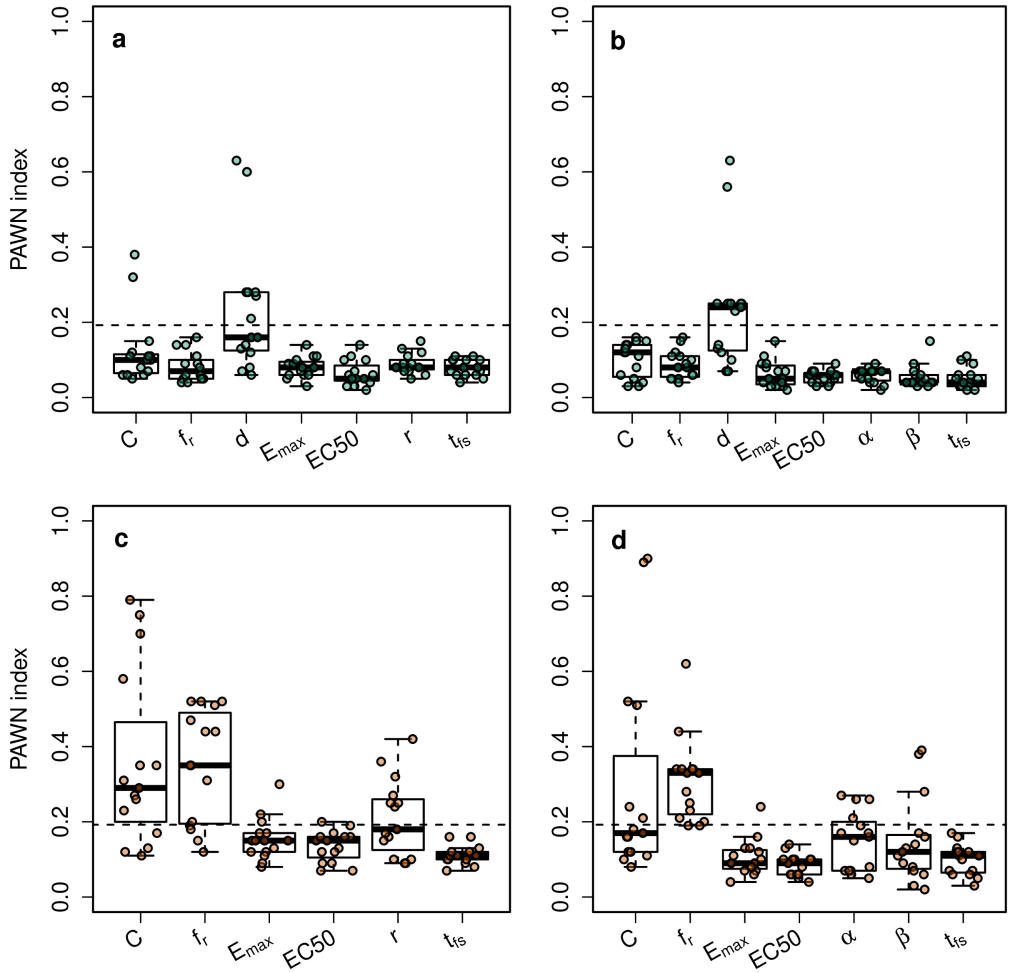

Figure 3.4: PAWN sensitivity index distributions, higher values are more influential parameters. The dashed line is the threshold value at level of significance = 0.05 (see section 3.4). (**a**) Exponential DRM, along with dose of bacteria (d). (**b**) beta-Poisson DRM, along with d.(**c**) Exponential DRM, fixed d.(**d**) beta-Poisson DRM, fixed d.

3.3 Discussion

While AB resistance is recognized as a growing problem, few studies have attempted developing a quantitative understanding of the risk posed by ARB and the associated sub-lethal AB concentration. I propose a new approach to modeling the dose-response of ARB under the Quantitative Microbial Risk Assessment (QMRA) framework. Under this framework, once a pathogen is identified, human exposure to it is quantified. Exposure is coupled with DRMs to estimate the risk of positive response, which is used to design risk management

measures.

The crux of the proposed approach is the stochastic process known as the SD process. I model the kinetics of the bacterial dose under the SD assumption, resulting in an analytical expression for the extinction probability (P_{ext}). Like in existing DRMs, response probability is defined as 1 - P_{ext}. The resulting expression establishes a link between the death rate, which is an experimentally observable parameter, and dose-response parameters (eq. 3.6). Since the relationship between AB concentration and death rates is known (eq. 3.8), we are also able to relate AB load to dose-response. Further, treating the ASB and ARB subpopulations independently, we can compute the odds of the successful treatment with AB in the infected subpopulation.

A significant advantage of the proposed approach is that it meshes well with the existing DRMs. An additional choice of DRM, which could complicate the decision of a practitioner, is not necessary. Moreover, the proposed approach does not change the workflow of traditional QMRAs. Instead, it provides additional capabilities relevant to antibiotic resistance. It can be used to predict the dose-response of ARB if one has information on 1) the dose-response of ASB, and 2) kinetics of ARB and how it relates to ASB (discussed below).

Additionally, if one suspects exposure to antibiotics in the risk assessment, its influence on human health outcomes can be accounted for using eq. 3.8. Since I introduce a novel approach to DRMs, I provide an outline of the steps, along with an example (see section B.4). As shown, the approach works equally well with exponential and beta-Poisson models to compute dose-response for a single exposure. The total risk from multiple exposures can be calculated in the usual way by assuming independence between exposures [57]. For example, the annual risk can be computed from daily risk estimates by assuming that one day's risk does not affect another day's risk.

The conservative assumption enabling this approach is that the ARB and ASB subpopu-

lation have the same death rate in the absence of AB. If there is evidence for significant differences in death rates in the absence of AB (due to the absence of mutations to compensate fitness loss), then this difference can be set to $\mu' - \mu$. Here μ' is the 'base' death rate for the ARB, which will be higher than the death rate of ASB (μ). Then $\mu_{r,\mathrm{AB}} = \mu' + f(C)$, where $f(C)$ is the concentration-dependent effect of the AB on the death rate. Another assumption made is that there are only two kinds of bacteria, ARB and ASB. However, resistance may vary in degree in different subpopulations. If one suspects multiple subpopulations (e.g., with different $E_{\max}$ values), one can assume their independence and adopt the probabilistic framework easily. In fact, the framework presented here can also be used for samples containing a mixture of different pathogens present in an environmental sample to estimate the total risk due to all the pathogens.

As a proof of concept, I investigated the case study of *E. coli* and the AB Gentamicin. I found that the AB concentration C and the fraction of ARB f_r emerged as influential parameters affecting the health outcome from a global sensitivity analysis. A rigorous exposure assessment to quantify C and f_r would thus be necessary to get accurate results in a QMRA. The AB-bacteria interaction parameters ($E_{\max,}$ EC_{50}) seem less influential for Gentamicin. This supports the use of these parameters, that were determined *in-vitro* [99], in a host system. A similar analysis would need to be carried out for different AB-bacteria combinations to draw generalized conclusions.

A critique of this approach is that it fails to account for the details in the biological system. For example, the precise concentration of AB in the vicinity of the bacteria is difficult to estimate and likely varies in time. The bacteria themselves translocate from the point of entry to the site of infection (e.g., travel through the alimentary canal). A more detailed stochastic model for dose-response (with compartments or spatial distributions) may better reflect the underlying system but would be significantly harder to parameterize adequately. Analytical expressions will also be harder to come by, and the approach may require dynamic simulations

for dose-response. Hence, I believe that the proposed approach provides a framework as a first step to solve the problem. For example, the range of variation in the outcome due to the change in C with time can be captured by investigating the outcomes at several fixed concentrations.

When assuming that the probability of observing a response (illness) equals the probability of not observing an extinction (eq. 3.1), I am not accounting for the carrier population. These are the individuals who harbor the pathogen but do not show any visible symptoms (infected but not ill). One could gain insight into the probability of infection in addition to the probability of illness if the dataset had information on the number infected. The proposed approach can be used to compute the probability of observing the response (infection). The probability of illness can then be computed by multiplying this quantity by the constant probability of illness given infection (as is done for norovirus in [89]). However, if evidence suggests that ARB are more virulent than the ASB, this approach will underestimate the illness risk. Here I distinguish between virulence (observed *in-vivo*) and fitness or growth rate, which can be observed *in-vitro*. The latter can be accounted for using the death rates, as discussed above. In summary, although the approach discussed here does not explicitly deal with illness and infection, the framework presented can be applied to this end if the data are available.

HGT and spontaneous mutation can result in the creation of ARB, which can potentially influence the type of health outcome. The SD DRM presented here can be modified to investigate the importance of these processes. For example, conjugation is modeled as a second-order reaction (first order in ASB population ($d(1-f_r)$) and first-order in ARB population (df_r)) [86]. A conservative estimate of rate constant of conjugation for *E. coli* is $r_{\text{conj}} = 2.4 \times 10^{-11}$mL cells^{-1}day^{-1} [86]. Hence the observed rate of conjugation becomes comparable to observed net death rates ($\mu = 18.35$ day^{-1} for DS1, Table 3.3) only if either the ARB or ASB subpopulation number around 10^{10}cells mL^{-1}. This is extremely high

and atypical of what is expected from environmental exposure, suggesting that conjugation does not affect dose-response significantly. Judging the importance of transduction and transformation is non-trivial as the numbers of phages and MGEs must be accounted for, respectively. Nevertheless, if these quantities were known along with the associated rates, the SD framework can be modified to account for them.

The SD framework can be applied to both the exponential and the beta-Poisson DRMs. Hence, this framework can be used to understand the effect of antibiotics on other bacteria for which these models are applicable. However, for a given organism, the best DRM relies on the objective of the QMRA being pursued and the dataset chosen [142]. In this study, the two different datasets for *E. coli* yield two different best fit models. Risk predictions also depend to a large extent on the dataset chosen. Deciding on the 'best' dataset or DRM for *E. coli* is beyond the scope of this chapter and is dicussed in the QMRA Wiki website [142]. Nevertheless, the SD framework accommodates both DRMs and provides the capability to investigate antibiotic resistance.

One limitation of applying this approach is the paucity of experimental data at each step. While *in-vitro* experiments [99] are used to capture the effect of the concentration of AB on bacteria, most studies on AB-bacteria interactions do not model the concentration-rate relationship, and instead report summary metrics of therapeutic importance [100]. While I have worked only with human datasets in this study, animal experiments may provide better quality data. Nevertheless, whether quantitative conclusions can transfer from animals to humans is debatable, as even human datasets (such as DS1 and DS2) show variability. Another area with insufficient data is the enumeration of ARB. Several studies reporting ARB occurrence in the environment report binary results (presence/absence) for each sample. Others report occurrence summaries such as '23% samples tested positive for ARB' (e.g., [87]). Only a few studies (e.g., [46, 56]) report the fraction of resistant bacteria in a single sample. Additional data collection at these stages will enable risk assessment case studies.

3.4 Methods

3.4.1 Fitting the model to data

I fitted the model to two *E.coli* datasets which are listed in Table B.1 along with the corresponding t_{fs}. I placed a binomial likelihood on the data as follows:

$$n_{\text{ill}} \sim \text{Binomial}(n_{\text{tot}}, P(d, t_{\text{fs}})) \tag{3.12}$$

This approach is commonly used in building DRMs and amounts to minimizing the deviance presented in [57]. I used the differential evolution algorithm [126] from `DEoptim` package [98] in the R[110] programming language to fit the data. The effect of varying ARB f_r at constant AB concentration $C = 1\%$ MIC (minimum inhibitory concentration = 2 µg/mL) and varying C at constant $f_r = 0.05$ are presented in Fig. 3.2. (The effect of antibiotic concentration C is incorporated in the model by increasing the death rate according to eq. 3.8).

The analytical approach described above suffices to fit the exponential DRM. However, this DRM fails to provide a satisfactory fit for some datasets, for which the beta-Poisson DRM provides a better fit and is often used as an alternative. As seen in the case for DS2 (Table 3.3), the beta-Poisson DRM fits the data better. However, this model does not have a ready r value that can be related to the death rate. A numerical simulation is necessary to parameterize the model to include the individual death rate.

The beta-Poisson DRM assumes that the survival probability of the pathogen, r, follows a Beta distribution (with parameters α and β). This is different from the exponential DRM, where r is assumed to be the same for all pathogens. The exact probability of illness can be challenging to compute, and I used the relationship in eq. 3.7. To relate α and β to the death rate of the susceptible strain in the presence of AB ($\mu_{s,\text{AB}}$), I used the following approach:

1. Generate N random values of r $(r_1, r_2, ...r_N)$ from a Beta distribution with parameters α and β.

2. Compute the corresponding μ_i values using eq. 3.6 for each r_i.

3. For known values of C, $E_{\max}$, EC_{50}, and μ_i, compute $\mu_{s,\mathrm{AB},i}$ using eq. 3.8.

4. Compute the corresponding $r_{s,i}$ from $\mu_{s,\mathrm{AB},i}$ using eq. 3.6.

5. Fit a Beta distribution to the N samples of $r_{s,i}$ and obtain the distribution's parameters, α_s and β_s. These can then be plugged into eq. 3.7 to estimate $P(d)$.

The approach outlined above is straightforward, but in practice, numerical issues are encountered due to the extremely small ($\leq 1 \times 10^{-16}$) values of r. Hence I used the `fitdistrplus` package [36] in R. I verified that this procedure produces satisfactory estimates (see Fig. B.1).

3.4.2 Sensitivity Analysis

A global sensitivity analysis was carried out using the PAWN algorithm[105] (algorithm name formed from author names), which measures sensitivity from the entire probability density of the output rather than just the variance of the output. Sensitivity is not characterized by a point estimate, but a list of estimates (PAWN indices) to give a fuller picture, with higher magnitudes representing more influential parameters. The algorithm parameters are $n = 15$, $N_u = N_c = 100$. Both the exponential and beta-Poisson DRMs were considered separately. The analysis was re-run after fixing the dose of bacteria to understand the effect of the remaining parameters (Fig. 3.4). A threshold value was calculated at the 0.05 significance level to identify non-influential parameters (dashed line, Fig. 3.4). Parameters with PAWN indices entirely below the threshold line are non-influential parameters. The parameter ranges in which sensitivity was investigated are described in Table B.2 and Table B.3.

3.4.3 Data availability

Codes reproducing the results in this chapter are available on Github at `https://github.com/JiangLabUCI/AbResistantDoseResponse`.

Chapter 4

Role of quorum sensing in microbial pathogensis

4.1 Background

Pathogens in the environment present a growing threat to human health. Understanding and quantifying the health risk that they pose is critical to make decisions on risk management. To this end, it is essential to quantify the relationship between the level of pathogens a person is exposed to (dose) and the associated probability of an adverse event (response), which is achieved by dose-response models (DRMs). A cornerstone of dose-response is the independent action hypothesis, which assumes that the pathogens in the microbial exposure load do not interact with one another to initiate infection. Making assumptions on distributions of the number of pathogens ingested from the exposure and the number of pathogens that survive to cause disease leads to several different DRMs. Of these, the most common are the exponential and beta-Poisson DRMs [57].

A fundamental assumption of the exponential and beta-Poisson models is the single-hit

hypothesis [94, 57], which postulates that only one organism needs to survive to initiate infection. Meynell and Stocker [94] draw an analogy to bullets fired at a target bottle, of which only one hits the bottle and breaks it. In contrast, the multi-hit hypothesis relaxes this criterion by assuming that at least $k_{\min}(> 1)$ organisms need to survive to initiate infection [57]. In terms of the bullet-bottle analogy, each bullet hits the bottle, but its cumulative effect of bullets that finally breaks the bottle. The multi-hit hypothesis has thus been interpreted as one of cooperation between the organisms in achieving infection, as one organism is incapable of initiating infection [57]. Nevertheless, the single-hit hypothesis is more widely accepted [117], and the corresponding DRMs are more frequently used [57].

The phrase 'multi-hit' has been used interchangeably with 'cooperative' in the literature, even though classical multi-hit models stem from the independent action hypothesis [57]. I resolve this seemingly counterintuitive relationship between cooperativity and independent action by introducing the term 'cooperativity in effect'. I use this term to describe the situations resembling the bullet analogy: when the actions of agents are independent, but the effect of their actions is cumulative. A fired bullet does not interact with any other fired bullet, but the accumulated stress of their impacts breaks the bottle.

In contrast, I also introduce the term 'cooperativity in action', which I define as: when the actions of agents are non-independent. This cooperativity is analogous to soccer (or other team sports), where one individual may score a goal, but it would not have been possible without help from the teammates. In the context of infectious diseases, cooperativity in action occurs when there is quorum sensing, i.e., when signals from one microbe influence the actions of another microbe. The role of quorum sensing in pathogenesis and establishing disease has been observed for several bacteria e.g., *Pseudomonas aeruginosa* [81], *Staphylococcus aureus* [77] and *Streptococcus* [31] (see [34] for other pathogens). Taking the soccer analogy further, winning a match may be attributed to the goals scored by one or more players.

Similarly, infection may develop due to the survival (and subsequent replication) of one or more clones. Classical dose-response does not account for cooperativity in action as the associated models assume the independent action hypothesis. In this work, I introduce a two-compartmental (2C) model of bacterial kinetics for dose-response. A single parameter represents cooperativity in action, and I parameterize the model for the pathogenesis of *S. aureus* on human skin.

Classical dose-response does not address another aspect, the carrier population. These are the individuals who do not show symptoms when colonized, who form a not insignificant proportion of the population in the case of *S. aureus* [19]. Sometimes, this population is estimated when the response data on colonization is available. I show that the response data on symptoms may suffice to predict the carrier population by assuming that response occurs when pathogen load exceeds a threshold (called individual effective dose, [70]). Additionally, being a kinetics based DRM, the 2C model can be used to probe 1) non-instantaneous exposures [109] 2) environment-host transmission dynamics with multiple exposures, without assuming independence between exposures 3) effect of antibiotics and antibiotic-resistant strains [26].

4.2 Results

4.2.1 Deterministic model of kinetics

Studies have observed that pathogens transferred to a host sometimes undergo an initial decay before growing in numbers to cause infection [44, 40, 123]. We hypothesize that this is due to the existence of two distinct states of the pathogen, S1 and S2 (Fig. 4.1A). Bacteria deposited in a new hostile environment (such as the host) are in S1. They can transition to the state S2, which is well adjusted to the environment and exhibits density-dependent

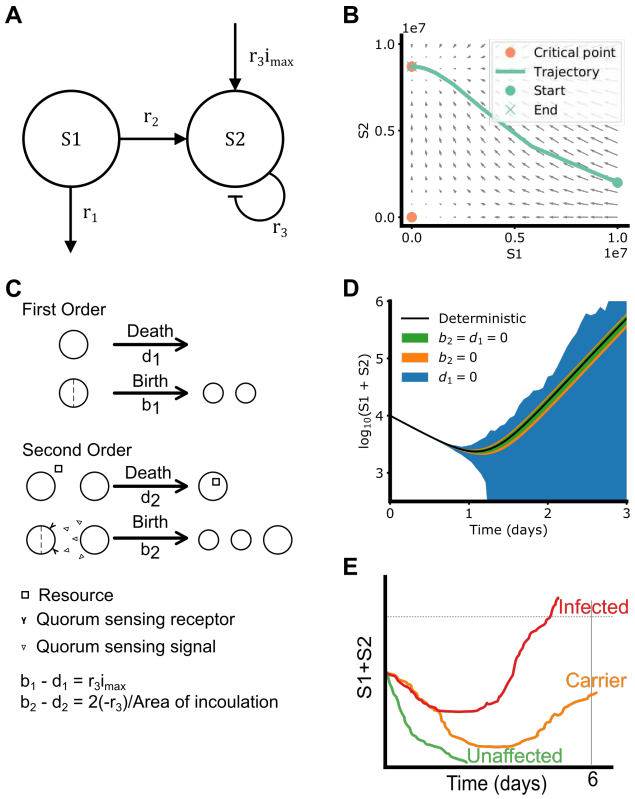

Figure 4.1: A) 2C model schematic B) Phase plot of 2C model C) Logistic process schematic D) Variations in 2C stochastic simulation (each region bounds mean ± SD of 50 simulations) E) Human health outcome classification according to SA dynamics

growth. Suppose $h(t)$ and $i(t)$ represent the density of pathogens in S1 and S2 respectively. The differential equation representing S1-S2 dynamics is given by:

$$\frac{dh(t)}{dt} = -r_1 h(t) - r_2 h(t) \tag{4.1}$$

$$\frac{di(t)}{dt} = r_2 h(t) + r_3 i(t)(i_{\max} - i(t)) \tag{4.2}$$

Here, r_1 is the rate of death of cells in S1. r_2 is the S1-S2 transition rate. r_3 is the logistic growth rate of cells in S2, and $i_{\max}$ is the logistic carrying capacity. I call this model the two-compartment or 2C model (Fig. 4.1A). The model has two critical points at $[0, 0]$ and $[0, i_{\max}]$ (Fig. 4.1B). The former is unstable, whereas the latter is asymptotically stable. In other words, if the system is nudged away from the unstable critical point at $[0, 0]$ (by addition

of bacteria to the system), it will move towards the other critical point at $[0, i_{\max}]$. Note that this system represents the density of pathogens and not their numbers, as differential equations are not restricted to integer values.

4.2.2 Stochastic model of dose-response

The stochastic version of the 2C model is captured with continuous time Markov chains [3], which restrict bacterial numbers to integer values. The first order reactions associated with r_1 and r_2 (eq. 4.1) are easily represented in this framework by a simple death processes with death rates r_1 and r_2 respectively. The r_2 term in eq. 4.2 is represented by a simple birth process with birth rate r_2.

Interestingly, the logistic growth component in eq. 4.2 is more involved and the logistic growth process [3] is used. In essence, the logistic growth process is equivalent to the deterministic logistic equation. It is used when the system has a small number of entities being tracked. It consists of two first and two second-order processes, given by the elementary reactions defined in (Fig. 4.1C). The relationships between the rates of these processes (d_1, b_1, d_2 and b_2) and the deterministic constants (r_3 and $i_{\max}$) are given by [3]

$$b_1 - d_1 = r_3 i_{\max} \tag{4.3}$$

$$b_2 - d_2 = 2(-r_3)/A \tag{4.4}$$

Here, I note that the factor $2/A$ stems from the size of the system (2D system with surface area $= A$), which needs to be accounted for in second-order systems [50]. The first order components (associated with d_1 and b_1) are density-independent and reflect the death and division of cells independent of other cells. The second-order components (associated with d_1 and b_2) are density-dependent and reflect the growth and death of cells as influenced by

the number of cells in their neighborhood. d_2 is the effect of competition of resources used by the cells, creating an upper limit on the population size. I interpret b_2 as the effect of quorum sensing, where signals from the other cells enhance the division rate of a given cell. This effect grows stronger with increasing cell numbers.

To understand the effect of these first and second-order processes on dynamics, I consider the case where the deterministic rates (r_3 and $i_{\max}$) are known. Eq. 4.3 constrains the first and second-order processes, reducing the degrees of freedom to two. Picking values for $d_1 > 0$ and $b_2 > 0$ thus fully determines the behavior of the system (Fig. 4.1B). Setting $b_2 = d_1 = 0$ results in a small variance around the mean, where the mean is determined by the solving the deterministic system (see Fig. 4.1D). Setting b_1 to a nonzero value while holding $b_2 = 0$ increases the variance. Interestingly, giving b_2 a nonzero value and holding $d_1 = 0$ increases the variance dramatically. This increases the odds of the pathogen population going to 0 and causing no symptoms. Therefore, b_2 serves as a knob that controls variance to influence pathogenesis and dose-response outcomes.

I evaluated dose-response outcomes using the concept of individual effective dose or IED [70]. Individuals with a total bacterial load (in states S1 + S2) above IED (represented by i_{thresh}), as done in other studies using CTMCs representing kinetics-based DRMs [147, 109] (see Fig. 4.1E). Individuals in whom bacteria die out completely represent the unaffected. Individuals who fit neither category, in whom the bacterial load takes an intermediate value, represent the carrier population. We have assumed bacteria in the un-adapted S1 state contribute to the threshold, which will not matter much if $r_1 > r_2$ as most of the bacteria in the S1 state will die out before bacteria in S2 establish infection.

Inoculation density (CFU/cm^2)	$\hat{n}_{\text{res}}$	n_{tot}	$\hat{P}_{\text{res}}$
40	4	20	0.20
220	8	20	0.40
2000	13	20	0.65
105000	14	20	0.70
1600000	19	20	0.95
10000000	20	20	1.00

Table 4.1: Dose-response data from [123]. $\hat{n}_{\text{res}}$ and n_{tot} as defined in the text

4.2.3 Parameterizing the model

I parameterized the model with data from the Singh et al. study [123]. They performed a clinical trial in which the participants' hands were cleaned with alcohol before inoculation with a known dose of SA. The area was covered immediately with a patch of polyethylene film to distribute the inoculum underneath it uniformly. Bacterial densities in the covered area was measured over six days (growth data). This was used to identify the parameters r_1, r_2, r_3 and $i_{\max}$ by a procedure outlined in the Methods.

Singh et al. also counted the number of people who developed lesions by day 6 for a given dose (dose-response data, Table 4.1). This data was used to identify the rate parameters (b_2 or d_1) and IED (i_{thresh}). For this, the response probability (P_{res}) was estimated by counting the fraction of stochastic simulations in which $(h(t)+i(t))A \geq i_{\text{thresh}}$. This predicted response probability was used to minimize the deviance given by

$$\sum_j -2\left(\hat{n}_{\text{res},j} \log\left(\frac{P_{\text{res},j}}{\hat{P}_{\text{res},j}}\right) + (n_{\text{tot},j} - \hat{n}_{\text{res},j}) \log\left(\frac{1-P_{\text{res},j}}{1-\hat{P}_{\text{res},j}}\right)\right) \tag{4.5}$$

and identify the unknown parameters. The dose-response data was verified to exhibit a trend according to the Cochran-Armitage test with $Z_{\text{CA}} = 6.29$, P=1.55×10^{-10} [57]. The fitting procedure is outlined in the methods.

Some points of note are: 1) I compute the probabilities of unaffected and carrier outcomes by

Model	Min. dev.	$\chi^2_{\text{degrees},0.05}$	P	Conclusion
2C($d_1 = 0$)	6.34	$\chi^2_{4,0.05} = 9.49$	0.18	Fail to reject
2C($b_2 = 0$)	16.24	$\chi^2_{4,0.05} = 9.49$	2.71×10^{-3}	Reject
Beta-Poisson	6.40	$\chi^2_{4,0.05} = 9.49$	0.17	Fail to reject
RH model	5.50	$\chi^2_{4,0.05} = 11.07$	0.36	Fail to reject

Table 4.2: Summary of model fits. Min. dev. is the minimum deviance from best fit obtained.

their corresponding fractions of simulations i.e, simulations with zero load and simulations with intermediate load ($0 < (h(t) + i(t))A < i_{\text{thresh}}$). 2) the b and d values discussed in this study correspond to stochastic rate constants (c_μ in [50]) and not the standard (deterministic) rate constants. Stochastic rate constants are used for simulating numbers (CFU) with continuous-time Markov chains, whereas standard rate constants are used for simulating densities or concentrations (CFU/cm^2) with differential equations. 3) $i_{\max}$, the carrying capacity, is interpreted as the mean of the SA densities observed in a population of individuals in whom SA is not wiped out. It differs from the IED (i_{thresh}), which is interpreted as the SA load above which an individual from the population will show response.

4.2.4 Assessing the fit

I compared the quality of the dose-response fit with two baselines: 1) the beta-Poisson model and 2) the quasi-mechanistic DRM of Rose and Haas (RH model) [116] with results presented in Table 4.2. The 2C model, with $d_1 = 0$ fits the data as well as the beta-Poisson and RH models at the 0.05 significance level. Setting $b_2 = 0$ fails to fit the data at the 0.05 significance level, and so I rejected this hypothesis.

Biologically speaking, these results mean the following: 1) Under the umbrella of the 2C model and its assumptions (existence of 2 states S1 and S2, response predicted by IED), cooperativity in action is necessary to explain SA dose-response 2) cooperativity in action is an alternate hypothesis to independent action + constant infectivity (RH model) and

independent action + variable infectivity (beta-Poisson model). However, neither the RH model nor beta-Poisson model can explain the carrier population without additional data. Therefore, I favor the 2C model with cooperativity in action ($b_2 = 0, d_1 \neq 0$) and restrict further discussion to its properties and capabilities.

The optimization algorithm used to fit the deterministic model (see Methods) provided a single best-fit parameter set and additional less-optimal parameter sets of varying quality. Using values (r_1, r_2, r_3 and $i_{\max}$) from such sub-optimal solutions to fit eq. 4.5 improved the fit to the dose-response objective function (eq. 4.5). An explanation for this observation is that the best-fit to growth data overfits that data. This two-step optimization procedure sheds light on parameter uncertainties and the trade-off between fitting the growth and dose-response data. The trade-off and the rank one solutions (solutions with no other solution better than them) are presented in Fig. 4.2A and Table C.1 respectively.

I chose two rank 1 solutions to illustrate the fit of the 2C model (colored in Fig. 4.2A). The deterministic model is compared with the model of [116] (labeled RH model) in Fig. 4.2B. The 2C model fits the data better than the RH model, as seen by the sum of squared errors (SSE). The 2C model also predicts a more gradual decay compared to the RH model.

A visual comparison of the fit of the stochastic 2C model with other models is presented in Fig. 4.2C. The better performance of RH and BP models compared to the 2C model is visually less discernible. Compared with Fig. 4.2B, the solution with a lower SSE has a higher deviance, indicating the trade-off between fitting growth and dose-response data.

Fig. 4.2D shows the probabilities of unaffected, response, and carrier outcomes across a range of bacterial doses. The height of each color for a given dose is the probability of the outcome associated with that color. The probability of the response outcome increases with increasing dose, as expected. Carrier probability increases and then decreases with increasing dose.

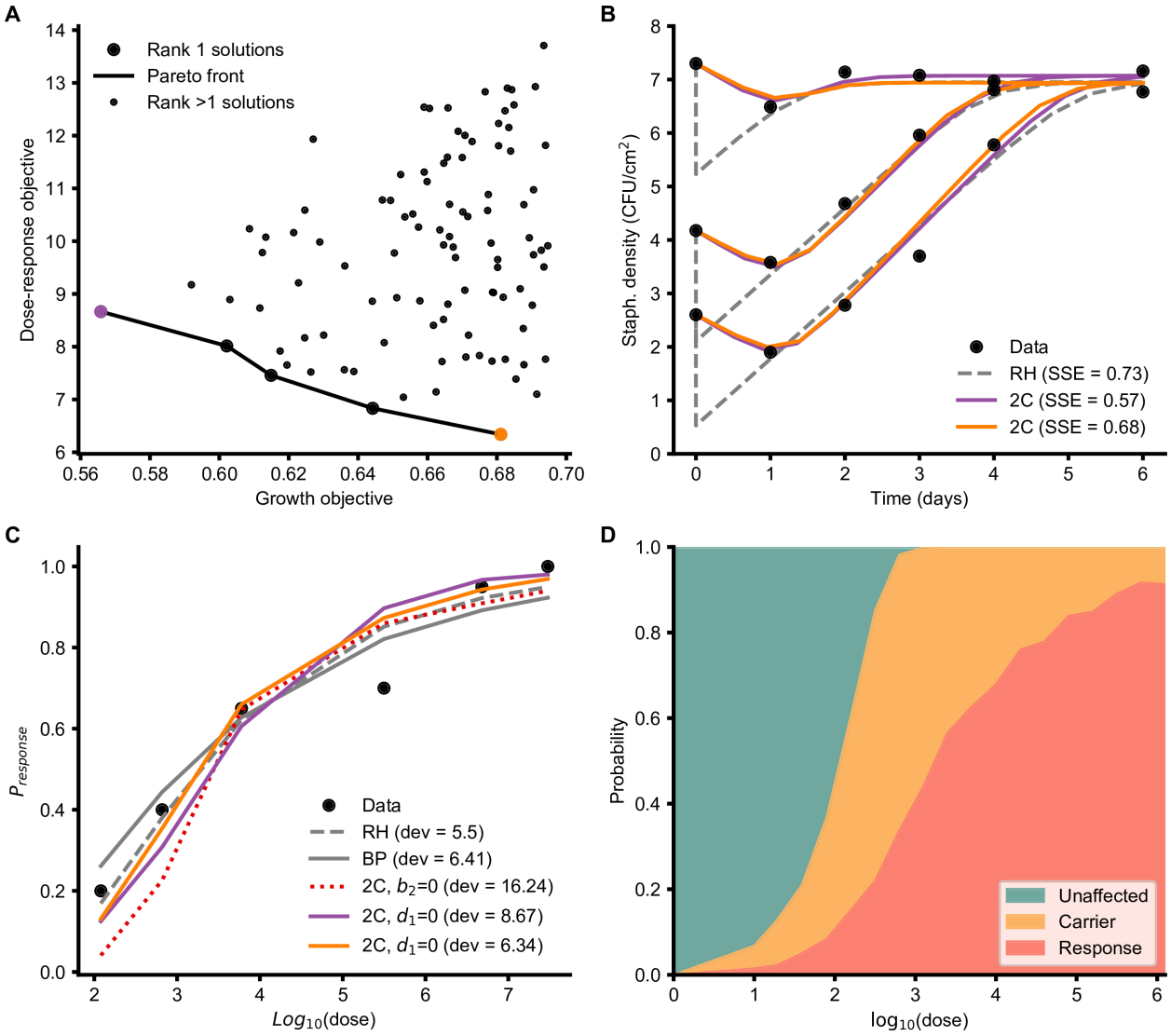

Figure 4.2: 2C model fit and predictions. (A) Plot of the dose-response objective function value (deviance) vs. the growth objective (SSE). Pareto rank 1 solutions (with no solution better than them) are connected by the Pareto front. Solutions at the extremes of the pareto front are colored. (B) Best fits of the 2C and RH models to growth data. (C) Best fits of the 2C and RH models to dose-response data. (D) Plot of the outcome probabilities as a function of the dose. For a given dose, the height of a colored region is the probability of the corresponding outcome, with the sum of the heights of each region totalling 1.

4.2.5 Parameters in absence of alcohol pre-treatment

As mentioned earlier, [123] treated their subjects with alcohol before administering SA, effectively reducing the resident microflora load on the skin. In comparison, [116] report an experiment similar to [123] wherein the subjects were inoculated with SA 24 hours after using regular soap. We assume the skin resident microflora exerts a competitive pressure that acts as a constant first-order death rate on SA. To determine the fit of the 2C model to this more realistic scenario, I explore the model fit under the following two hypotheses:

The first hypothesis (which I call $r1^*$) assumes that only the un-adapted SA are affected by the resident microflora. The model (eq. 4.1 & eq. 4.2) is modified to

$$\frac{dh(t)}{dt} = -r_1^* h(t) - r_2 h(t) \tag{4.6}$$
$$\frac{di(t)}{dt} = r_2 h(t) + r_3 i(t)(i_{\max} - i(t)) \tag{4.7}$$

Here r_1^* reflects the increased r_1, accounting for death by resident microflora. The second hypothesis (which I call r_{mf}) assumes that both un-adapted and adapted SA are affected by resident microflora. The model (eq. 4.1 & eq. 4.2) is modified to

$$\frac{dh(t)}{dt} = -r_1 h(t) - r_2 h(t) - r_{\mathrm{mf}} h(t) \tag{4.8}$$
$$\frac{di(t)}{dt} = r_2 h(t) + r_3 i(t)(i_{\max} - i(t)) - r_{\mathrm{mf}} i(t) \tag{4.9}$$

Other hypotheses may also fit the data from [116]. However, I only consider the hypotheses that can be modeled with one additional parameter. This is because there are only four data points, including the initial condition. Using two parameters to fit the three remaining data points will result in over-fitting. For example, the corrected Akaike Information Criterion (AICc, which corrects for small sample sizes) [69], for such a case (3 data points,

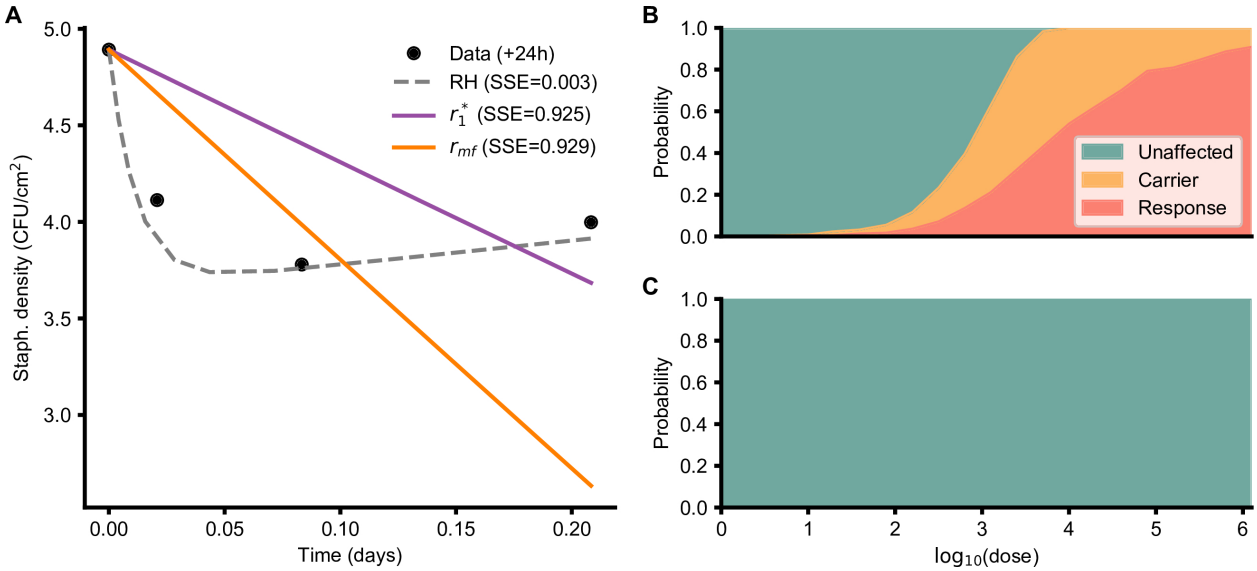

Figure 4.3: 2C model in the absence of alcohol pre-treatment. Two different hypotheses (r_1^* and r_{mf}, described in text) were investigated. (A) Fit of the RH model and the two hypotheses of the 2C model to data obtained in the absence of alcohol pre-treatment. (B and C) Outcome probabilities for r_1^* and r_{mf} hypothesis respectively.

2 parameters) will be infinity. The limited data also hindered testing of absolute goodness of fit. Other hypotheses that can be captured by a single parameter change (e.g., resident microbiota affecting r_2) did not yield good fits (data not presented).

The fit of these two hypotheses, along with the approach of [116] are compared in Fig. 4.3A. The RH model fits the data better because of the additional parameter. The r_1^* hypothesis shows a sharper decline than the r_{mf} hypothesis, but the available data does not strongly support either hypothesis in favor of the other as indicated by their similar SSE values.

The difference between the hypotheses is striking when looking at their outcome probabilities with increasing dose. The r_1^* hypothesis predicts a significant P_{res} at higher doses (Fig. 4.3B) whereas the r_{mf} hypothesis predicts almost zero P_{res} and carrier probability (Fig. 4.3C). This is because the adapted SA also dies out from the inhibitory effect of the resident microflora.

4.2.6 Parameters for MRSA

To see if the parameters found for MSSA could be used for MRSA, I looked at studies in which both MSSA and MRSA were grown separately under similar conditions. Since the growth curves for MSSA and MRSA were highly correlated across studies (see Table C.2), I decided to use the MSSA parameters for MRSA.

4.2.7 Case study

To demonstrate the applicability of the model, I use it to predict the outcome probabilities for patients in a hospital. As an example, I look at the bedside rails in a hospital as an environmental reservoir of MRSA and the subsequent risk to a patient using that bed. The patient's exposure to MRSA is calculated as the product of the MRSA density/cm^2 on bed-rails [78], area of hands [1], and the transfer efficiency from bed-rails to hands [2] (see section 4.4). Exposure is assumed to randomly occur at a hand-bedside rail contact frequency of 0.28 contacts/hour [27]. The environmental reservoir is assumed not to be affected by contact events, which is possible if the patient touches different parts of the bedside rails. In this way, I modeled the MRSA load in a patient staying in a hospital with stochastics. I repeated this 1000 times to understand the probabilities of different outcomes.

Sample trajectories of the MRSA load on the patient over time are presented in Fig. 4.4A (r_1^* hypothesis) and Fig. 4.4B (r_{mf} hypothesis). The sudden increases in MRSA loads are indicative of the random contact events. After this increase, the MRSA load encounters a period of decay. In some cases, when the populations cross 10^4 CFU, the contact events are not discernible from the general randomness in the simulation. Some trajectories under the r_1^* hypothesis undergo an explosion while none of those under r_{mf} do. The sharp increase at the ends of the explosive trajectories (Fig. 4.4A, red lines) highlights the numerical difficulties in simulating stochastic systems with a large number of entities.

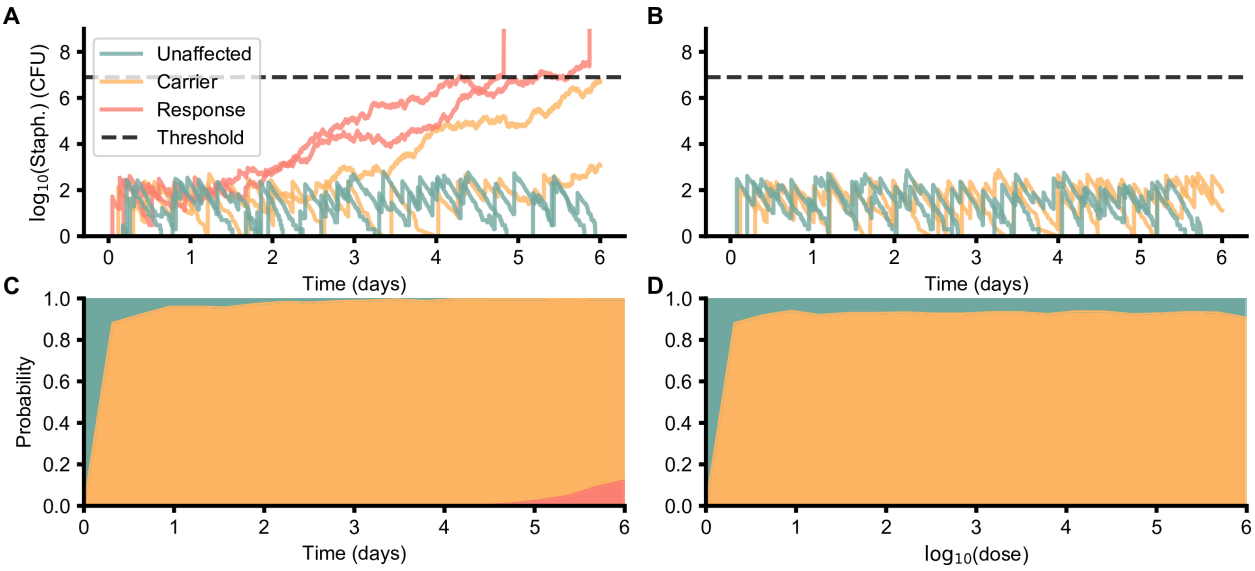

Figure 4.4: Case study (described in text) outcomes for the r_1^* and r_{mf} hypotheses. (A and B) Sample population trajectories of the case study for the two hypothesis. $\log_{10}$ (SA population + 1) is plotted to avoid (negative) infinities. (C and D) Outcome probabilities of the case study for the two hypotheses. Colors are as indicated in (A).

The distribution of outcome probabilities is given in Fig. 4.4C (r_1^*) and Fig. 4.4D (r_{mf}). Initially, both hypotheses predict similar outcomes. At the end of day 1, they start to diverge with r_1^* predicting an increasing likelihood of carrier outcomes and even predicting response outcomes by day 6. On the other hand, r_{mf} predicts a more or less constant probability of carrier status and no response outcomes through day 6 (Fig. 4.4D). The corresponding trajectories (Fig. 4.4B) explain this behavior, as MRSA loads from exposure events are not large enough to overcome the decay and establish response. This is consistent with the trend seen earlier (Fig. 4.2) of r_{mf} predicting less adverse outcomes than r_1^*.

4.3 Discussion

A two compartment model for SA dynamics on the human skin was developed and fitted to data. By assuming that SA transitions from an un-adapted state to an adapted state, the model is grounded in first principles. The stochastic aspect of dose-response emerges

naturally from a stochastic simulation of the growth kinetics. In addition, the model predicts carrier outcomes without additional data.

Armitage et al. [7] interpret results from several studies to posit that pathogens, including bacteria, show an initial exponential increase in all individuals. We argue that this is not inconsistent with the initial decrease assumption for three reasons. Firstly, the exponential increase is observed in organs like the liver or spleen, and not the whole body or site of inoculation (e.g., [66, 53]). This does not refute the possibility of an initial decrease at the inoculation site or the whole body. Secondly, the posited decrease is transient, and samples may not have been collected during this window. Thirdly, the magnitude of decrease is low at higher inocula (e.g., Fig. 4.1B) and consequently less detectable. Further, compared to the initial decrease observed when all bacteria are in the S1 state, one would expect 1) no initial decrease if seeding with bacteria all in the S2 state, and 2) a smaller initial decrease if seeding with a mixture of bacteria in the S1 and S2 state. These trends have been observed when pathogens from *in-vivo* cultures were used for infecting the host [44, 151]. We note that the transition from S1 to S2 is perhaps not instantaneous, and the pathogen population may constitute a continuum of states between S1-S2.

When loads were measured in the whole body, a transient decrease was observed in some cases (see Fig. 6 in Yamamura et al. [151]). Clumping of bacteria was offered as a possible explanation [151], but this does not rule out an actual reduction in viable counts observed in other systems (see [44, 123], Table 2 in [38]).

Armitage et al. [7] also note that non-responders show a subsequent decrease after the initial exponential increase. These were substantiated by measurements from survivors who were killed at later time points [63, 38]. This decrease is probably due to the activation of the adaptive immune response inside the host, which could be incorporated in a within-host variant of the 2C model.

Using the concept of IED to evluate reponse, I am able to explain the data with a single IED. It has been observed that the toxic dose of a chemical can vary between individual subjects or with the season [133]. A similar stochasticity may be expected in IED between individuals which can be attributed to differences in covariates such as body weight, sex, immune history and biological noise. However, assuming this was not necessary to produce an acceptable fit.

The model was fit to data by following a two step optimization procedure. Direct multi-objective optimization was not pursued since the objective functions were very different from each other. The deterministic ODE model was easy to evaluate and a global optimization algorithm was employed to guard against local minima while fitting the growth data. Fitting the dose-response data was computationally challenging for 3 reasons: a non-smooth objective function, stochastic simulations have to be repeated many times, and the number of stochastic entities being modeled is not small. Hence, a simple brute-force optimization was adopted.

The RH model exhibits a sharp initial decline in SA density and predicts values lower than the observed minimum for each initial load (Fig. 4.2B). The 2C model only goes as low as the lowest load observed on the skin. Experiments similar to that of [116] with greater time resolution are necessary to ascertain the time of true minimal SA density. The 2C model stochastic model does not perform as well as the RH model (Fig. 4.2C). However, the 2C model fit to dose-response data improves along the Pareto front (Fig. 4.2A). It is possible that exploring solutions with a higher growth objective may yield a solution that fits as well as, if not better than, the RH fit to the dose-response data. Moreover, the proposed approach offers advantages over the existing approach in that 1) it is fully mechanistic, and hence is more applicable in other scenarios (e.g. using MSSA parameters for MRSA, non-instantaneous exposure, multiple exposure which is discussed here), and 2) in addition to response outcomes, the proposed approach also accounts for carrier outcomes.

Perhaps the most interesting outcome of this study is the incorporation of quorum sensing in dose-response modeling. The rejection of the absence of cooperativity in SA pathogenesis and the adequate fit of cooperativity make a strong case for the cooperativity in action hypothesis. Experimental support for this hypothesis include the well studied Agr system of quorum sensing [77, 80]. In the words of Le. et al, the Agr system "generally enhances pathogenesis by increasing expression of aggressive virulence determinants such as toxins and degradative enzymes" [80]. This system is activated when bacteria reach a certain density, which results in a disease response such as a murine abscess [149]. However, the 2C model posits that quorum sensing enhances bacterial growth rate, for which I propose two possible explanations. The direct explanation is the existence of an as yet undiscovered signaling mechanism responsible for density dependent growth enhancement. A second explanation relates to the events initiating response in a host, which is the interaction of the toxins/enzymes produced by SA with the host tissue. The 2C model captures these dynamics at a higher level of abstraction, with the mathematical variable $i(t)$ (or the S2 state) representing the amount QS signals and toxins. We can interpret b_2 as the rate of enhanced production of these factors. The 2C model does not capture other observed dynamics, such as the down regulation of the Agr system and its role in biofilm formation [80, 77].

The model was extended to more realistic situations where the SA inoculation area was not alcohol treated by means of two competing hypotheses, r_1^* and r_{mf}. These hypotheses predict very different carrier and response probabilities. This is because the former assumes that the adapted state is unaffected by resident microflora while the latter assumes that the adapted and un-adapted states are equally affected. The truth is likely in the middle, i.e., SA in the adapted state are affected by resident microflora to a lesser extent than SA in the un-adapted state. Such an approach was not pursued in the spirit of avoiding over-fitting given the limited data. We note that more data collected within the first day of inoculation will help judge the quality of the hypotheses presented in this study, which need to be evaluated on an absolute scale with a goodness of fit test. Confidence in the predictions of the model

will improve as more data is gathered, either supporting or refuting the hypotheses.

The model's applicability was demonstrated with a simple case study. The mechanistic nature of the model enabled direct simulation of repeated exposures from the environment, without having to assume independence between exposure events. This paves the way for more involved modeling efforts such as accounting for healthcare workers and other hospital surfaces that contain MRSA. Such efforts can be challenging for two reasons: the availability of high quality data to model behaviors and the computational effort in simulating stochastic systems. However, they can supplement our understanding of the environment as a source of MRSA and help devise the most effective control measures in hospitals and the community.

4.4 Methods

4.4.1 Fitting kinetic data

The parameters r_1, r_2, r_3 and $i_{\max}$ were identified by minimizing the following objective function:

$$\sum_c \sum_j \left(\log_{10}\left(h_c(t_j) + i_c(t_j)\right) - \log_{10}\left(\hat{y}_c(t_j)\right) \right)^2 \tag{4.10}$$

Here the subscript c denotes the 3 different initial conditions while t_j is the j^{th} observed time point. $\hat{y}$ is the measured SA density. Since initially none of the bacteria are adapted to the environment, $i(0)$ is set to 0. Hence, $h(0)$ is set to the initial inoculating density i.e., $h(0) = \hat{y}_c(0)$. Using this objective function assumes that both un-adapted and adapted cells are picked up while taking measurements.

A global optimization algorithm, Differential Evolution Markov Chain (DE-MC, [130]), was used to minimize the objective function. The algorithm was run for 40000 iterations and 16

chains were evaluated in parallel. The first half of the solutions were discarded (burn-in). A MATLAB implementation [25] was used.

4.4.2 Simulating the stochastic model

The adaptive τ leaping algorithm [21] specific to the model was implemented in Python. `Numba` was used to speed up the code.

4.4.3 Fitting dose-response data

Running 1000 simulations of the stochastic model is computationally more expensive than evaluating the deviance for a single value of i_{thresh} with eq. 4.11.

$$P_{\text{res}} = \frac{\text{no. of sims with } (h(t) + i(t))A \geq i_{\text{thresh}}}{n} \tag{4.11}$$

Hence, for a guessed value of b_2, simulations were run and the best fitting i_{thresh} was found. When multiple threshold guesses gave the same lowest deviance, i_{thresh} was set to the minimum of these guesses. The best b_2 was found by brute-force search in a hand-tuned interval. Objective functions were evaluated in parallel using Python's built-in `multiprocessing` module on the High Performance Computing facility at UC Irvine.

4.4.4 Case study

Patients were assumed to occupy the same bed for the duration of 6 days. Exposure to MRSA was modeled as a Poisson process and assumed to occur at random with a contact frequency of 0.28 contacts/hour [27]. Hence, the time between two exposure events was sampled from an exponential distribution with rate parameter = 0.28 contacts/hour.

Each exposure event resulted in an increase in the number of SA in S1 (unadjusted state). This inoculation load or increase was computed as the product of 1) MRSA density//cm^2 on bed-rails ([78], sampled from a normal distribution with mean = 159.5 CFU/100 cm^2, std. dev. = 396.4 CFU/100 cm^2, truncated at (0, 1620)) 2) area of hands ([1], women's hand size = 132.42 cm^2) 3) transfer efficiency from bedside rails to hands ([2], sampled from a uniform distribution with bounds [0.22, 0.38]). At each exposure time, the simulation was halted and the number of SA in S1 was increased by the inoculation load.

Chapter 5

Discussion

The broad questions I outlined in my dissertation, and the specific questions that I framed and attempted to answer, concern kinetics of pathogens both outside and within the host. I have developed mathematical models of these processes and parameterized them with data from the literature, often relying on a strongly computational approach. I used these models to make quantitative predictions, generate experimentally verifiable hypotheses, and ask new questions. The data I used has been around for many years, lending support to the idea that it is not always necessary to generate new data; sometimes, it suffices to reflect on what we already know to make progress. Below, I use my work to support my hypothesis - that mathematical models that account for details and leverage sophisticated techniques open up new avenues of scientific inquiry and bring existing paradigms into question.

For my first objective (chapter 2), I focused on the safety of using treated wastewater for agriculture, given that viruses may internalize in the vasculature of the produce. To this end, I introduced a dynamic model for viral transport from the irrigation water to the final produce. Specifically, I tracked the biggest cause of gastrointestinal illness in humans - the human norovirus - in lettuce, which is often eaten uncooked. I modeled lettuce grown in

soil and hydroponic systems from first principles, accounting for lettuce growth, transport of virus within lettuce (internalization), natural viral decay, and the attachment of virus particles to soil particles/walls of hydroponic tanks. I found parameters for the model in the literature spanning several fields where possible. For the remaining parameters, I used published data on norovirus internalization [37] to fit the model. In doing the latter, I greatly benefitted from the computational power of the cluster at UCI, which helped me simulate the system of ODEs repeatedly to identify good parameter sets. I found that using treated wastewater (secondary effluent) for irrigating lettuce did not meet the safety standards of the U.S. EPA nor the WHO. This conclusion was independent of the growth medium (soil or hydroponics), despite the greater attachment of viruses to the soil. I verified that the conclusions were robust to several model assumptions by a global sensitivity analysis – another computationally intensive task. Factors such as the irrigation schedule and time of harvest were deemed influential by this analysis, highlighting the utility of a dynamic model. Changing these factors could reduce the risk of virus internalization, a hypothesis which is experimentally verifiable. The investigation of these interventions would not have been possible without a detailed model of the processes involved.

For my second objective (chapter 3), I focused on antibiotic-resistant bacteria (ARB), which have been isolated in several environmental compartments. To meet this need, I modeled the in-vivo kinetics of bacteria in a host as a simple death process. With this, I was able to establish a link between the popular exponential and beta-Poisson DRMs and observable parameters like death rates – in a framework I call the SD (simple death) framework. This link was analytical for the exponential DRM. I designed an iterative computational procedure for the beta-Poisson DRM, which is easily parallelized to speed up computation. Further, this link enabled inquiries into possibilities that were out of reach of the classical framework. For example, by assuming that ARB act like antibiotic sensitive bacteria (ASB) in the absence of antibiotics, I was able to predict the probability of the resulting illness responding to antibiotics. Additionally, I incorporated the effects of antibiotics on dose-response by using

from *in-vitro* studies. By parameterizing the model for *E. coli* and its Gentamicin resistant variant, I found that a minimal concentration of Gentamicin was sufficient to select for the Gentamicin resistant strain and result in an illness where Gentamicin would not work. Moreover, the SD framework can theoretically account for horizontal gene transfer, although experimental determination of the associated rate constants is a challenge. I also identified other gaps in data required for an accurate QMRA in the context of the SD framework. These include the ARB fraction in environmental compartments and the relationship between antibiotic concentration and death rates for different bacteria.

Towards the third objective (chapter 4), I proposed a two-compartment stochastic model of bacterial kinetics with cooperativity in action captured by a single parameter. This model differed from the existing approaches to dose-response modeling, which either adopted independent action or cooperativity in effect. By developing compiled code that is highly parallelized and using an optimized tau-leaping simulating algorithm [21], I sped up the simulation of the stochastic model to fit it to data on *S. aureus* pathogenesis. Fitting the model to the data would not have been possible without a cluster, as it was a multi-step procedure involving repeated stochastic simulations of large numbers of species. Nevertheless, this huge effort paid off when I found that the hypothesis of cooperativity in action cannot be rejected. I was able to show, for the first time, that synergistic interaction between bacteria (or quorum sensing) can account for their ability to initiate disease symptoms. This differs from the traditional assumption of independent action in dose-response modeling, wherein each bacterium is assumed to act independently.

I believe that the arguments presented above stand in support of my hypothesis.

Bibliography

[1] P. Agarwal and S. Sahu. Determination of hand and palm area as a ratio of body surface area in Indian population. *Indian journal of plastic surgery : official publication of the Association of Plastic Surgeons of India*, 43(1):49–53, 2010.

[2] S. Ali, G. Moore, and A. P. R. Wilson. Effect of surface coating and finish upon the cleanability of bed rails and the spread of Staphylococcus aureus. *Journal of Hospital Infection*, 80(3):192–198, 2012.

[3] L. J. S. Allen. *An Introduction to Stochastic Processes with Biology Applications*. Prentice Hall, 1st edition, 2003.

[4] D. P. Ames, N. W. T. Quinn, A. E. Rizzoli, and G. B. Mcbride. Norovirus dose-response in sewage-related QMRA: The importance of virus aggregation. *Congress on Env. Modelling and Software*, 2014.

[5] S. A. Anderson, R. W. Yeaton Woo, and L. M. Crawford. Risk assessment of the impact on human health of resistant *Campylobacter jejuni* from fluoroquinolone use in beef cattle. *Food Control*, 12(1):13–25, 2001.

[6] J. F. Apgar, D. K. Witmer, F. M. White, and B. Tidor. Sloppy models, parameter uncertainty, and the role of experimental design. *Molecular BioSystems*, 6(10):1890, 2010.

[7] P. Armitage, G. G. Meynell, and T. Williams. Birth-death and other models for microbial infection. *Nature*, 207(4997):570–572, 1965.

[8] N. J. Ashbolt, A. Am?zquita, T. Backhaus, P. Borriello, K. K. Brandt, P. Collignon, A. Coors, R. Finley, W. H. Gaze, T. Heberer, J. R. Lawrence, D. G. J. Larsson, S. A. McEwen, J. J. Ryan, J. Sch?nfeld, P. Silley, J. R. Snape, C. Van den Eede, and E. Topp. Human Health Risk Assessment (HHRA) for Environmental Development and Transfer of Antibiotic Resistance. *Environmental Health Perspectives*, 121(9):993–1001, 2013.

[9] S. Ashraf, U. Chaudhry, A. Raza, D. Ghosh, and X. Zhao. In vitro activity of ivermectin against *Staphylococcus aureus* clinical isolates. *Antimicrobial Resistance and Infection Control*, 7(1):7–12, 2018.

[10] R. L. Atmar, A. R. Opekun, M. A. Gilger, M. K. Estes, S. E. Crawford, F. H. Neill, S. Ramani, H. Hill, J. Ferreira, and D. Y. Graham. Determination of the 50% human

infectious dose for Norwalk virus. *The Journal of Infectious Diseases*, 209(7):1016–22, 2014.

[11] S. F. Barker, J. O'Toole, M. I. Sinclair, K. Leder, M. Malawaraarachchi, and A. J. Hamilton. A probabilistic model of norovirus disease burden associated with grey-water irrigation of home-produced lettuce in Melbourne, Australia. *Water Research*, 47(3):1421–1432, 2013.

[12] A. Beceiro, M. Tomás, and G. Bou. Antimicrobial resistance and virulence: A successful or deleterious association in the bacterial world? *Clinical Microbiology Reviews*, 26(2):185–230, 2013.

[13] C. R. Bell, G. a. Dickie, W. L. G. Harvey, and J. W. Y. F. Chan. Endophytic bacteria in grapevine. *Canadian Journal of Microbiology*, 41(1):46–53, 1995.

[14] T. U. Berendonk, C. M. Manaia, C. Merlin, D. Fatta-Kassinos, E. Cytryn, F. Walsh, H. Bürgmann, H. Sørum, M. Norström, M.-N. Pons, N. Kreuzinger, P. Huovinen, S. Stefani, T. Schwartz, V. Kisand, F. Baquero, and J. L. Martinez. Tackling antibiotic resistance: the environmental framework. *Nature Reviews Microbiology*, 13(5):310–317, 2015.

[15] R. D. Boche and J. J. Quilligan Jr. Adsorption to glass and specific antibody inhibition of iodine-125 labeled influenza virus. *J Immunol*, 97(6):942–950, 1966.

[16] A. Both. Ten years of hydroponic lettuce research. *Knowledgecenter.Illumitex.Com*, 18(5):8, 2003.

[17] A.-J. Both. *Dynamic simulation of supplemental lighting for greenhouse hydroponic lettuce production*. PhD thesis, Cornell University, 1995.

[18] R. Brookmeyer, E. Johnson, and S. Barry. Modelling the incubation period of anthrax. *Statistics in Medicine*, 24(4):531–542, 2005.

[19] F. M. Calia, E. Wolinsky, E. A. Mortimer, J. S. Abrams, and C. H. Rammelkamp. Importance of the carrier state as a source of *Staphylococcus aureus* in wound sepsis. *Journal of Hygiene*, 67(1):49–57, 1969.

[20] J. J. Campion, P. J. McNamara, and M. E. Evans. Pharmacodynamic Modeling of Ciprofloxacin Resistance in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 49(1):209–219, 2005.

[21] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *The Journal of Chemical Physics*, 124(4):044109, 2006.

[22] A. Carducci, E. Caponi, A. Ciurli, and M. Verani. Possible Internalization of an Enterovirus in Hydroponically Grown Lettuce. *International Journal of Environmental Research and Public Health*, 12(7):8214–8227, 2015.

[23] CDC. Antibiotic resistance threats in the United States, 2013. Technical report, 2013.

[24] R. Chachra, M. K. Transtrum, J. P. Sethna, R. N. Gutenkunst, J. J. Waterfall, C. R. Myers, K. S. Brown, R. A. Cerione, and J. P. Sethna. Comment on "Sloppy models, parameter uncertainty, and the role of experimental design". *Molecular BioSystems*, 7(8):2522, 2011.

[25] S. Chandrasekaran and S. C. Jiang. A dynamic transport model for quantification of norovirus internalization in lettuce from irrigation water and associated health risk. *Science of the Total Environment*, 643:751–761, 2018.

[26] S. Chandrasekaran and S. C. Jiang. A dose response model for quantifying the infection risk of antibiotic-resistant bacteria. *Scientific Reports*, 9(1):17093, 2019.

[27] V. C. Cheng, P. H. Chau, W. M. Lee, S. K. Ho, D. W. Lee, S. Y. So, S. C. Wong, J. W. Tai, and K. Y. Yuen. Hand-touch contact assessment of high-touch and mutual-touch surfaces among healthcare workers, patients, and visitors. *Journal of Hospital Infection*, 90(3):220–225, 2015.

[28] D. E. Ciolkosz, L. D. Albright, and A. J. Both. *Characterizing evapotranspiration in a greenhouse lettuce crop*, volume 456. 1998.

[29] R. B. Clapp and G. M. Hornberger. Empirical equations for some soil hydraulic properties. *Water Resources Research*, 14(4):601–604, 1978.

[30] L. A. Cox and D. Popken. A simulation model of human health risks from chicken-borne *Campylobacter jejuni*. *Technology*, 9:55–84, 2002.

[31] D. G. Cvitkovitch, Y. H. Li, and R. P. Ellen. Quorum sensing and biofilm formation in Streptococcal infections. *Journal of Clinical Investigation*, 112(11):1626–1632, 2003.

[32] A. K. da Silva, J.-C. Le Saux, S. Parnaudeau, M. Pommepuy, M. Elimelech, and F. S. Le Guyader. Evaluation of removal of noroviruses during wastewater treatment, using real-time reverse transcription-PCR: different behaviors of genogroups I and II. *Applied and Environmental Microbiology*, 73(24):7891–7, 2007.

[33] A. De Keuckelaere, L. Jacxsens, P. Amoah, G. Medema, P. Mcclure, L. A. Jaykus, and M. Uyttendaele. Zero Risk Does Not Exist: Lessons Learned from Microbial Risk Assessment Related to Use of Water and Safety of Fresh Produce. *Comprehensive Reviews in Food Science and Food Safety*, 14(4):387–410, 2015.

[34] K. V. Deepika and P. V. Bramhachari. *Bacterial Quorum Sensing in Pathogenic Relationships: Relevance to Complex Signalling Networks and Prospective Applications*, pages 67–79. Springer Singapore, Singapore, 2018.

[35] A. J. Deering, L. J. Mauer, and R. E. Pruitt. Internalization of *E. coli* O157:H7 and Salmonella spp. in plants: A review. *Food Research International*, 45(2):567–575, 2012.

[36] M. L. Delignette-Muller and C. Dutang. fitdistrplus : An R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4), 2015.

[37] E. DiCaprio, Y. Ma, A. Purgianto, J. Hughes, and J. Li. Internalization and dissemination of human norovirus and animal caliciviruses in hydroponically grown romaine lettuce. *Applied and Environmental Microbiology*, 78(17):6143–6152, 2012.

[38] J. M. Dolby and A. F. B. Standfast. The intracerebral infection of mice with *Bordetella pertussis*. *Journal of Hygiene*, 59(2):205–216, 1961.

[39] L. Drago, E. De Vecchi, B. Mombelli, L. Nicola, M. Valli, and M. R. Gismondo. Activity of levofloxacin and ciprofloxacin against urinary pathogens. *Journal of Antimicrobial Chemotherapy*, 48(1):37–45, 2001.

[40] H. T. Eigelsbach, J. J. Tulis, M. H. McGavran, and J. D. White. Live Tularemia Vaccine I. : Host-Parasite Relationship in Monkeys Vaccinated Intracutaneously or Aerogenically. *Journal of Bacteriology*, 84(5):1020–107, 1962.

[41] FDA. Human Health Impact of Fluoroquinolone Resistant *Campylobacter* Attributed to the Consumption of Chicken. 2000.

[42] J. Flannery, S. Keaveney, P. Rajko-Nenow, V. O'Flaherty, and W. Doré. Concentration of norovirus during wastewater treatment and its impact on oyster contamination. *Applied and Environmental Microbiology*, 78(9):3400–3406, 2012.

[43] S. Fletcher. Understanding the contribution of environmental factors in the spread of antimicrobial resistance, 2015.

[44] G. M. Fukui, W. D. Lawton, W. A. Janssen, and M. J. Surgalla. Response of Guinea Pig Lungs to In Vivo and In Vitro Cultures of. *Journal of Infectious Diseases*, 100(1):103–107, 1957.

[45] M. Gallardo, R. L. Snyder, K. Schulbach, and L. E. Jackson. Crop Growth and Water Use Model for Lettuce. *Journal of Irrigation and Drainage Engineering*, 122(6):354–359, 1996.

[46] S. Galvin, F. Boyle, P. Hickey, A. Vellinga, D. Morris, and M. Cormican. Enumeration and characterization of antimicrobial-resistant *Escherichia coli* bacteria in effluent from municipal, hospital, and secondary treatment facility sources. *Applied and Environmental Microbiology*, 76(14):4772–4779, 2010.

[47] P. L. Geenen, M. G. J. Koene, H. Blaak, a. H. Havelaar, and a. W. V. D. Giessen. page 121, 2010.

[48] A. Gelman, J. Carlin, H. Stern, D. Dunson, and A. Vehtari. *Bayesian Data Analysis*. Taylor & Francis, 3 edition, 2013.

[49] J. J. Gillard, T. R. Laws, G. Lythe, and C. Molina-ParÃs. Modeling early events in *Francisella tularensis* pathogenesis. *Frontiers in Cellular and Infection Microbiology*, 4(December):1–10, 2014.

[50] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[51] B. Gompertz. On the Nature of the Function Expressive of the Law of Mortality. *Philosophical Transactions*, 27:513–585, 1825.

[52] J. Goudriaan and J. L. Monteithf. A Mathematical Function for Crop Growth Based on Light Interception and Leaf Area Expansion. *Annals of Botany*, 66(6):695–701, 1990.

[53] J. E. Gray, J. R. Wilkins, M. C. Prestrud, and C. T. Nikitas. Further characterization of an experimental staphylococcal infection in mice. *Journal of Infectious Diseases*, 101(2):137–147, 1957.

[54] T. Guillard, S. Pons, D. Roux, G. B. Pier, and D. Skurnik. Antibiotic resistance and virulence: Understanding the link and its consequences for prophylaxis and therapy. *BioEssays*, 38(7):682–693, 2016.

[55] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10):1871–1878, 2007.

[56] S. Guyomard-Rabenirina, C. Dartron, M. Falord, S. Sadikalay, C. Ducat, V. Richard, S. Breurec, O. Gros, and A. Talarmin. Resistance to antimicrobial drugs in different surface waters and wastewaters of Guadeloupe. *PLoS ONE*, 12(3):e0173155, 2017.

[57] C. N. Haas, J. B. Rose, and C. P. Gerba. *Quantitative Microbial Risk Assessment*. John Wiley & Sons, Inc, Hoboken, New Jersey, 2014.

[58] A. J. Hamilton, F. Stagnitti, R. Premier, A.-m. M. Boland, and G. Hale. Quantitative Microbial Risk Assessment Models for Consumption of Raw Vegetables Irrigated with Reclaimed Water. *Applied and Environmental Mcrobiology*, 72(5):3284–3290, 2006.

[59] H. Hauser, D. C. Barilotti, T. Taylor, and S. Barbara. Ocean Wastewater Discharge Inventory for the State of California. Technical Report 805, 2005.

[60] H. Hauser and J. O. Hawkins. State Water Resources Control Board. Draft Staff Report: Proposed Amendments to the Water Quality Control Plan for Ocean Waters of California Addressing Desalination Facility Intakes, Brine Discharges, and the Incorporation of other Nonsubstantive Changes. (805):1–4, 2015.

[61] A. H. Havelaar, M. D. Kirk, P. R. Torgerson, H. J. Gibb, T. Hald, R. J. Lake, N. Praet, and D. C. Bellinger. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. *PLOS Medicine*, 12(12):e1001923, 2015.

[62] K. a. Hirneisen, M. Sharma, and K. E. Kniel. Human Enteric Pathogen Internalization by Root Uptake into Food Crops. *Foodborne Pathogens and Disease*, 9(5):396–405, 2012.

[63] D. Hobson. Chronic bacterial carriage in survivors of experimental mouse typhoid. *The Journal of Pathology and Bacteriology*, 73(2):399–410, 1957.

[64] K. Holvoet, I. Sampers, B. Callens, J. Dewulf, and M. Uyttendaele. Moderate prevalence of antimicrobial resistance in *Escherichia coli* isolates from lettuce, irrigation water, and soil. *Applied and Environmental Microbiology*, 79(21):6677–6683, 2013.

[65] P.-Y. Hong, T. Julian, M.-L. Pype, S. Jiang, K. Nelson, D. Graham, A. Pruden, and C. Manaia. Reusing Treated Wastewater: Consideration of the Safety Aspects Associated with Antibiotic-Resistant Bacteria and Antibiotic Resistance Genes. *Water*, 10(3):244, 2018.

[66] F. L. Horsfall and H. S. Ginsberg. The dependence of the pathological lesion upon the multiplication of pneumonia virus of mice (PVM); kinetic relation between the degree of viral multiplication and the extent of pneumonia. *The Journal of Experimental Medicine*, 93(2):139–150, 1951.

[67] Y. Huang and C. N. Haas. Quantification of the Relationship between Bacterial Kinetics and Host Response for Monkeys Exposed to Aerosolized *Francisella tularensis*. *Applied and Environmental Microbiology*, 77(2):485–490, 2011.

[68] H. S. Hurd, S. Doores, D. Hayes, A. Mathew, J. Maurer, P. Silley, R. S. Singer, and R. N. Jones. Public health consequences of macrolide use in food animals: a deterministic risk assessment. *Journal of Food Protection*, 67(5):980–992, 2004.

[69] C. M. Hurvich and C. L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

[70] J. O. Irwin. Statistical Method Applied to Biological Assays. *Supplement to the Journal of the Royal Statistical Society*, 4(1):1, 1937.

[71] M. Islam, J. Morgan, M. P. Doyle, S. C. Phatak, P. Millner, and X. Jiang. Fate of *Salmonella enterica* Serovar Typhimurium on Carrots and Radishes Grown in Fields Treated with Contaminated Manure Composts or Irrigation Water. *Applied and Environmental Microbiology*, 70(4):2497–2502, 2004.

[72] S. Jenni and G. Bourgeois. Quantifying phenology and maturity in crisphead lettuce. *HortTechnology*, 18(4):553–558, 2008.

[73] H. D. Kahn and K. Stralka. Estimated daily average per capita water ingestion by child and adult age categories based on USDA's 1994-1996 and 1998 continuing survey of food intakes by individuals. *Journal of Exposure Science & Environmental Epidemiology*, 19(4):396–404, 2009.

[74] N. Karavarsamis and A. J. Hamilton. Estimators of annual probability of infection for quantitative microbial risk assessment. *Journal of Water and Health*, 8(2):365–73, 2010.

[75] K. Kemmeren, M. Mangen, V. Duynhoven YTHP, and A. Havelaar. Priority setting of foodborne pathogens: disease burdenand costs of selected enteric pathogens. Technical report, 2006.

[76] D. Knaack, E. A. Idelevich, N. Schleimer, S. Molinaro, A. Kriegeskorte, G. Peters, and K. Becker. Bactericidal activity of bacteriophage endolysin HY-133 against *Staphylococcus aureus* in comparison to other antibiotics as determined by minimum bactericidal concentrations and time-kill analysis. *Diagnostic Microbiology and Infectious Disease*, 93(4):362–368, 2019.

[77] K. F. Kong, C. Vuong, and M. Otto. *Staphylococcus* quorum sensing in biofilm formation and infection. *International Journal of Medical Microbiology*, 296(2-3):133–139, 2006.

[78] E. J. O. Kurashige, S. Oie, and H. Furukawa. Contamination of environmental surfaces by methicillin-resistant *Staphylococcus aureus* (MRSA) in rooms of inpatients with MRSA-positive body sites. *Brazilian Journal of Microbiology*, 47(3):703–705, 2016.

[79] K. L. LaPlante and M. J. Rybak. Impact of high-inoculum *Staphylococcus aureus* on the activities of nafcillin, vancomycin, linezolid, and daptomycin, alone and in combination with gentamicin, in an in vitro pharmacodynamic model. *Antimicrobial Agents and Chemotherapy*, 48(12):4665–4672, 2004.

[80] K. Y. Le and M. Otto. Quorum-sensing regulation in staphylococci-an overview. *Frontiers in Microbiology*, 6(OCT):1–8, 2015.

[81] J. Lee and L. Zhang. The hierarchy quorum sensing network in *Pseudomonas aeruginosa*. *Protein and Cell*, 6(1):26–41, 2014.

[82] A. F. Leonard, L. Zhang, A. J. Balfour, R. Garside, and W. H. Gaze. Human recreational exposure to antibiotic resistant bacteria in coastal bathing waters. *Environment International*, 82:92–100, 2015.

[83] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, and C. E. Kolb. Global Sensitivity Analysis for Systems with Independent and / or Correlated Inputs. *Journal of Chemical Physics*, pages 6022–6032, 2010.

[84] K. Y. Lim and S. C. Jiang. Reevaluation of health risk benchmark for sustainable water practice through risk analysis of rooftop-harvested rainwater. *Water Research*, 47(20):7273–7286, 2013.

[85] K.-Y. Lim, Y. Wu, and S. C. Jiang. Assessment of *Cryptosporidium* and norovirus risk associated with de facto wastewater reuse in Trinity River, Texas. *Microbial Risk Analysis*, 0:1–10, 2016.

[86] A. J. Lopatkin, S. Huang, R. P. Smith, J. K. Srimani, T. A. Sysoeva, S. Bewick, D. K. Karig, and L. You. Antibiotics as a selective driver for conjugation dynamics. *Nature Microbiology*, 1(6):1–8, 2016.

[87] F. Lüddeke, S. Heß, C. Gallert, J. Winter, H. Güde, and H. Löffler. Removal of total and antibiotic resistant bacteria in advanced wastewater treatment by ozonation in combination with different filtering techniques. *Water Research*, 69:243–251, 2015.

[88] C. M. Manaia. Assessing the Risk of Antibiotic Resistance Transmission from the Environment to Humans: Non-Direct Proportionality between Abundance and Risk. *Trends in Microbiology*, 25(3):173–181, 2017.

[89] D. Mara and A. Sleigh. Estimation of norovirus infection risks to consumers of wastewater-irrigated food crops eaten raw. *Journal of Water and Health*, 08(1):39, 2010.

[90] B. T. Mayer, J. S. Koopman, E. L. Ionides, J. M. Pujol, and J. N. S. Eisenberg. A dynamic dose-response model to account for exposure patterns in risk assessment: a case study in inhalation anthrax. *Journal of The Royal Society Interface*, 8(57):506–517, 2011.

[91] J. S. Meschke and M. D. Sobsey. Comparative adsorption of Norwalk virus, poliovirus 1 and F+ RNA coliphage MS2 to soils suspended in treated wastewater. *Water Science and Technology*, 38(12):187–189, 1998.

[92] M. J. Messner, P. Berger, and S. P. Nappier. Fractional poisson - A simple dose-response model for human norovirus. *Risk Analysis*, 34(10):1820–1829, 2014.

[93] G. G. Meynell and E. W. Meynell. The growth of micro-organisms in vivo with particular reference to the relation between dose and latent period. *Journal of Hygiene*, 56(3):323–346, 1958.

[94] G. G. Meynell and B. A. D. Stocker. 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. *Journal of General Microbiology*, 16(1):38–58, 1957.

[95] C. L. Moe. Preventing Norovirus Transmission: How Should We Handle Food Handlers? *Clinical Infectious Diseases*, 48(1):38–40, 2009.

[96] H. F. Mok, S. F. Barker, and A. J. Hamilton. A probabilistic quantitative microbial risk assessment model of norovirus disease burden from wastewater irrigation of vegetables in Shepparton, Australia. *Water Research*, 54(March 2016):347–362, 2014.

[97] H. Motulsky and A. Christopoulos. *Fitting models to biological data using linear and nonlinear regression : a practical guide to curve fitting*. Oxford University Press, 2004.

[98] K. Mullen, D. Ardia, D. Gil, D. Windover, and J. Cline. DEoptim : An R Package for Global Optimization by Differential Evolution. *Journal of Statistical Software*, 40(6), 2011.

[99] E. I. Nielsen, O. Cars, and L. E. Friberg. Pharmacokinetic/Pharmacodynamic (PK/PD) indices of antibiotics predicted by a semimechanistic PKPD model: A step toward model-based dose optimization. *Antimicrobial Agents and Chemotherapy*, 55(10):4619–4630, 2011.

[100] E. I. Nielsen and L. E. Friberg. Pharmacokinetic-pharmacodynamic modeling of antibacterial drugs. *Pharmacological Reviews*, 65(3):1053–90, 2013.

[101] E. O'Flaherty and E. Cummins. Antibiotic resistance in surface water ecosystems: Presence in the aquatic environment, prevention strategies, and risk assessment. *Human and Ecological Risk Assessment*, 23(2):299–322, 2017.

[102] G. Oron, M. Goemans, Y. Manor, and J. Feyen. Poliovirus distribution in the soil-plant system under reuse of secondary wastewater. *Water Research*, 29(4):1069–1078, 1995.

[103] M. Peleg and C. M. Penchina. Modeling microbial survival during exposure to a lethal agent with varying intensity. *Critical Reviews in Food Science and Nutrition*, 40(2):159–72, 2000.

[104] S. R. Petterson, P. F. M. Teunis, and N. J. Ashbolt. Modeling Virus Inactivation on Salad Crops Using Microbial Count Data. *Risk Analysis*, 21(6):1097–1108, 2001.

[105] F. Pianosi and T. Wagener. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67:1–11, 2015.

[106] C. Portas. Development of root systems during the growth of some vegetable crops. *Plant and Soil*, 39(3):507–518, 1973.

[107] P. Presi, K. D. C. Stärk, R. Stephan, E. Breidenbach, J. Frey, and G. Regula. Risk scoring for setting priorities in a monitoring of antimicrobial resistance in meat and meat products. *International Journal of Food Microbiology*, 130(2):94–100, 2009.

[108] F. Prestinaci, P. Pezzotti, and A. Pantosti. Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens and Global Health*, 109(7):309–318, 2015.

[109] J. M. Pujol, J. E. Eisenberg, C. N. Haas, and J. S. Koopman. The Effect of Ongoing Exposure Dynamics in Dose Response Relationships. *PLoS Computational Biology*, 5(6):e1000399, 2009.

[110] R Core Team. R: A Language and Environment for Statistical Computing, 2017.

[111] A. Rababah. Innovative production treatment hydroponic farm for primary municipal sewage utilisation. *Water Research*, 34(3):825–834, 2000.

[112] A. Rahman, D. Munther, A. Fazil, B. Smith, and J. Wu. Advancing risk assessment: Mechanistic dose-response modelling of *Listeria monocytogenes* infection in human populations. *Royal Society Open Science*, 5(8), 2018.

[113] A. Rejwan. The State of Israel: National Water Efficiency Report. Technical report, Planning Department of the Israeli Water Authority, 2011.

[114] B. Roberts, R. Bailey, M. McLaughlin, and J. Brooks. Decay rates of zoonotic pathogens and viral surrogates in soils amended with biosolids and manures and comparison of qPCR and culture derived rates. *Science of The Total Environment*, 573:671–679, 2016.

[115] A. Rohatgi. WebPlotDigitizer, 2015.

[116] J. B. Rose and C. N. Haas. A risk assessment framework for the evaluation of skin infections and the potential impact of antibacterial soap washing. *American Journal of Infection Control*, 27(6):S26–33, 1999.

[117] L. G. Rubin. Bacterial Colonization and Infection Resulting from Multiplication of a Single Organism. *Clinical Infectious Diseases*, 9(3):488–493, 1987.

[118] A. Sahin and J. A. Vrugt. SPACE: Sensitivity of Parameters with Analysis of CovariancE. 2018.

[119] H. Sales-Ortells, X. Fernandez-Cassi, N. Timoneda, W. D??rig, R. Girones, and G. Medema. Health risks derived from consumption of lettuces irrigated with tertiary effluent containing norovirus. *Food Research International*, 68:70–77, 2015.

[120] J. F. Schijven and S. M. Hassanizadeh. Removal of Viruses by Soil Passage: Overview of Modeling, Processes, and Parameters. *Critical Reviews in Environmental Science and Technology*, 30(1):49–127, 2000.

[121] J. F. Schijven, W. Hoogenboezem, M. Hassanizadeh, and J. H. Peters. Modeling removal of bacteriophages MS2 and PRD1 by dune recharge at Castricum, Netherlands. *Water Resources Research*, 35(4):1101–1111, 1999.

[122] P. J. Schmidt. Norovirus Dose-Response: Are Currently Available Data Informative Enough to Determine How Susceptible Humans Are to Infection from a Single Virus? *Risk Analysis*, 35(7):1364–1383, 2015.

[123] G. Singh, R. R. Marples, and A. M. Kligman. Experimental *Staphylococcus aureus* Infections in Humans. *Journal of Investigative Dermatology*, 57(3):149–162, 1971.

[124] E. L. Snary. Antimicrobial resistance: a microbial risk assessment perspective. *Journal of Antimicrobial Chemotherapy*, 53(6):906–917, 2004.

[125] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[126] R. Storn and K. Price. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[127] C. O. Tacket, M. B. Sztein, G. Losonsky, A. Abe, B. B. Finlay, B. P. McNamara, G. T. Fantry, S. P. James, J. P. Nataro, M. M. Levine, and M. S. Donnenberg. Role of EspB in Experimental Human Enteropathogenic *Escherichia coli* Infection. *Infection and Immunity*, 68(6):3689–3695, 2000.

[128] F. Tei, D. P. Aikman, and A. Scaife. Growth of Lettuce, Onion and Red Beet. 2. Growth Modelling. *Annals of Botany*, 78(5):645–652, 1996.

[129] F. Tei, a. Scaife, and D. P. Aikman. Growth of Lettuce, Onion, and Red Beet. 1. Growth Analysis, Light Interception, and Radiation Use Efficiency. *Annals of Botany*, 78(5):633–643, 1996.

[130] C. J. F. Ter Braak. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249, 2006.

[131] P. F. Teunis, C. L. Moe, P. Liu, S. E. Miller, L. Lindesmith, R. S. Baric, J. Le Pendu, and R. L. Calderon. Norwalk virus: How infectious is it? *Journal of Medical Virology*, 80(8):1468–1476, 2008.

[132] G. N. Tindula, M. N. Orang, and R. L. Snyder. Survey of Irrigation Methods in California in 2010. *Journal of Irrigation and Drainage Engineering*, 139(3):233–238, 2013.

[133] J. W. Trevan. The Error of Determination of Toxicity. *Proceedings of the Royal Society B: Biological Sciences*, 101(712):483–514, 1927.

[134] A. Urbanucci, M. Myrmel, I. Berg, C. H. von Bonsdorff, and L. Maunula. Potential internalisation of caliciviruses in lettuce. *International Journal of Food Microbiology*, 135(2):175–178, 2009.

[135] U.S. EPA. CSFII Analysis of Food Intake Distributions. Technical report, U.S. Environmental Protection Agency (EPA), Washington, DC, 2003.

[136] N. Van Abel, M. E. Schoen, J. C. Kissel, and J. S. Meschke. Comparison of Risk Predicted by Multiple Norovirus Dose-Response Models and Implications for Quantitative Microbial Risk Assessment. *Risk Analysis*, 2016.

[137] R. L. Ward and R. J. Mahler. Uptake of bacteriophage f2 through plant roots. *Applied and Environmental Microbiology*, 43(5):1098–1103, 1982.

[138] K. Warriner, F. Ibrahim, M. Dickinson, C. Wright, and W. M. Waites. Internalization of Human Pathogens within Growing Salad Vegetables. *Biotechnology and Genetic Engineering Reviews*, 20(August 2014):117–136, 2003.

[139] J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna. Sloppy-model universality class and the Vandermonde matrix. *Physical Review Letters*, 97(15):150601, 2006.

[140] J. E. Weaver and W. E. Bruner. *Root development of vegetable crops.* McGraw-Hill publications in the agricultural and botanical sciences, E.W. Sinnott, consulting editor. McGraw-Hill Book Company, Incorporated, 1927.

[141] J. Wei, Y. Jin, T. Sims, and K. E. Kniel. Internalization of murine norovirus 1 by *Lactuca sativa* during irrigation. *Applied and Environmental Microbiology*, 77(7):2508–2512, 2011.

[142] M. Weir, J. Mitchell, J. Libarkin, and A. Mraz. Board # 156 : QMRA Wiki: An Educational Tool for Interdisciplinary Teaching of Risk Modeling in Engineering Curricula. In © *2017 ASEE Annual Conference & Exposition*, Columbus, Ohio, 2018. American Society for Engineering Education.

[143] WHO. *Guidelines for Drinking-Water Quality, Fourth edition*, volume 1. World Health Organization, 2011.

[144] WHO. *Antimicrobial resistance: global report on surveillance.* World Health Organization, Geneva, 2014.

[145] K. R. Wigginton, B. M. Pecson, T. Sigstam, F. Bosshard, and T. Kohn. Virus inactivation mechanisms: impact of disinfectants on virus function and structural integrity. *Environmental Science & Technology*, 46(21):12069–78, 2012.

[146] D. Wilson. Endophyte: The Evolution of a Term, and Clarification of Its Use and Definition. *Oikos*, 73(2):274, 1995.

[147] R. M. Wood, J. R. Egan, and I. M. Hall. A dose and time response Markov model for the in-host dynamics of infection with intracellular bacteria following inhalation: with application to *Francisella tularensis*. *Journal of The Royal Society Interface*, 11(95):20140119–20140119, 2014.

[148] M. Wooldridge. Antibiotic Resistance in the European Union Associated with Therapeutic use of Veterinary Medicines Report and Qualitative Risk Assessment by the Committee for Veterinary Medicinal Products Annex IV Qualitative Risk Assessment for Antibiotic Resistance Ca. 1999.

[149] J. S. Wright, R. Jin, and R. P. Novick. Transient interference with staphylococcal quorum sensing blocks abscess formation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5):1691–1696, 2005.

[150] K. M. Wright, S. Chapman, K. McGeachy, S. Humphris, E. Campbell, I. K. Toth, and N. J. Holden. The endophytic lifestyle of *Escherichia coli* O157:H7: quantification and internal localization in roots. *Phytopathology*, 103(4):333–40, 2013.

[151] Y. Yamamura, A. Walter, and H. Bloch. Bacterial populations in experimental murine tuberculosis I. Studies in Normal Mice. *Journal of Infectious Diseases*, 106(2):211–222, 1960.

[152] G. Zhang, S. Shen, M. Takagaki, T. Kozai, and W. Yamori. Supplemental Upward Lighting from Underneath to Obtain Higher Marketable Lettuce (*Lactuca sativa*) Leaf Fresh Weight by Retarding Senescence of Outer Leaves. *Frontiers in Plant Science*, 6(December):1–9, 2015.

[153] P. S. zur Wiesch, J. Engelstädter, and S. Bonhoeffer. Compensation of fitness costs and reversibility of antibiotic resistance mutations. *Antimicrobial Agents and Chemotherapy*, 54(5):2085–2095, 2010.

Appendix A

A transport model for quantifying norovirus internalization in lettuce

A.1 Hydroponic growth medium

A.1.1 Justifying the setting of rate of volume reduction to rate of transpiration

Hydroponic tanks face water loss by two processes, evaporation and transpiration. [28] used generalized linear models to fit the rate of evapotranspiration (ET) in hydroponically grown lettuce. They found that the air temperature and vapor pressure deficit had the lowest significance among the covariates considered. Taken together with the marked increase in water usage as the lettuce grows, the evaporation component in ET is thought to be less important in the hydroponic system. Hence, I made a simplifying assumption that the rate of ET, rather than the rate of transpiration, drives viral transport. I equated the flow rate to the rate of ET for these reasons.

A.1.2 Volume of shoot

The growth model for the lettuce shoot dry weight (given below) was obtained from [28], which was first introduced in [17]. I converted the dry weight to fresh weight using the dry:fresh weight ratio $d_{\text{shoot,h}}$=0.045 [16] and shoot density ρ_{shoot}=0.35 g/cm^3 [72].

$$M'_{\text{shoot}}(t) = \exp(s_1 + s_2 t + s_3 t^2)$$

$M'_{\text{shoot}}(t)$ is the rate of accumulation of dry mass (g dry wt/ mL). s_1, s_2 and s_3 are as defined in Table 2.2. Volumetric growth rate is given by eq. (2.8), Table 2.1.

A.1.3 Volume of roots

I obtained the growth model for the lettuce root dry weight from [16]. I converted this to fresh weight using the dry:fresh weight ratio ($d_{\text{root,h}}$=0.057) estimated from [152] white LED with no supplemental light). I assumed a density ρ_{shoot}=0.2 g/cm^3.

$$M'_{\text{root}}(t) = \exp(r_1 + r_2 t + r_3 t^2)$$

$M'_{\text{root}}(t)$ is rate of accumulation of dry mass (g dry wt/ mL). r_1, r_2 and r_3 are as defined in Table 2.2. Volumetric growth rate is given by eq. (2.7), Table 2.1.

A.1.4 Water volume fitting considerations

While [28] found that the rate of dry mass accumulation had the highest significance, I could not use the coefficients (a_t, b_t) they provided. This is because the cumulative water usage predicted by those coefficients in the monitored period is higher than the tank volume of 800 mL specified by [37]. To account for the lower water usage, I fitted the coefficients while

maintaining the conclusion of the experiment that the rate of ET is a linear function of the dry mass accumulation rate.

A.2 Soil growth medium

A.2.1 Volume of envelope

The reported soil envelope volume (V_e) in the lettuce Rainha De Maio in [106] for different soil types varied from 3332 – 8568 cm^3 although the growth stage is not reported. [140] used a different cultivar (Early Prize Head) and very different estimates were obtained, especially for the later stages of growth. Fitting a cone of radius 3.46 inches and height 6 inches in the early stage gives a volume of 3706 cm^3. Fitting a cone of radius 1.5 feet and height 5 feet gives a volume of 3.336×10^5 cm^3 for the mature stage. Using these numbers as guidelines, I fixed the soil envelope volume (V_e) to 80000 cm^3 and verified the low sensitivity of the risk estimates to large variations in this parameter (Fig. 2.5).

A.2.2 Volume of shoot

[128] compared 3 lettuce growth models – the logistic, Gompertz [51] and the expolinear [52] model. I chose the logistic model (eq. (2.8), Table 2.1), despite its higher root mean squared error, because there are insufficient data to simulate the expolinear and Gompertz models. This model predicts dry weight/area, and I extended it to fresh volume per unit plant because: 1) The number of plants per unit area (n_p=17.6 plants/m^2) remains constant over the course of that experiment [129]; 2) The ratio of dry weight: fresh weight ($d_{\text{shoot,s}}$) remains constant over the lettuce growth phase [16] ; 3) The density of the shoot (ρ_{shoot}) remains constant over the growth phase. The growth equation in dry weight/area ($W(t)$)

over time is given by:

$$\frac{dW(t)}{dt} = r_g W(t)\left(1 - \frac{W(t)}{W_f}\right)$$

where w_f is the final dry weight per unit area. Substituting $W(t) = V_{sh}(t) n_p d_{\text{shoot,h}} \rho_{\text{shoot}}$ and $W_f = w_f n_p d_{\text{shoot,h}}$ (where w_f is the final fresh weight per plant), eq. (2.8), Table 2.1 is obtained. Hence the same growth rate constant r is applicable for the fresh volume model. I took the final fresh weight (w_f) and density (ρ_{shoot}) from [72] for the Ithaca variety.

A.2.3 Flow rate

[45] used a model to predict the transpiration rate (in mm day^{-1}) of lettuce over its growth phase in soil. To get the volumetric flow rate, I multiplied by an arbitrary area and divided by the number of plants in that area:

$$F(\text{mL plant}^{-1}\text{day}^{-1}) = ROT(\text{mm day}^{-1}) \times \frac{\text{area}}{\text{no. plants in that area}}$$

where F is the flow rate and ROT is the rate of transpiration. Since lettuce was reportedly to occupy three fourths of the total area at harvest time (the rest occupied by soil), I computed the per plant transpiration rate by considering an area of 1m^2 using:

$$F(\text{mL plant}^{-1}\text{day}^{-1}) = \frac{ROT(\text{cm day}^{-1})}{10} \times \frac{1\text{m}^2}{0.75\text{m}^2 \div \text{area of fully grown lettuce}}$$

I obtained the ROT from figure 3 in [45] using WebPlotDigitizer [115] and converted it to flow rate using the method described above.

A.2.4 Model fitting and diagnostics

I fitted $\log_{10}$ of the viral concentration because viral concentrations at different time points differed by orders of magnitude. Fitting concentrations resulted in a good fit for early time points but inaccurate estimates for later timepoints used in estimating the risk. While maximizing the likelihood of the $\log_{10}$ concentration, I did not weight by the standard deviations as these estimates were from small sample numbers [97]. I did not pursue multiobjective approaches to simultaneously maximize the individual likelihoods of water, root, and shoot concentrations due to the similar orders of magnitude of the likelihoods. For the same reason, I did not weight likelihoods differently.

I also investigated convergence diagnostics. I discarded the first half of the 20000 iterations of DE-MC (burn-in). I identified outlier chains by their characteristic higher mean objective function value compared to other chains and discarded them. The $\hat{R}$ values [48] of the remaining chains were close to 1 (data not shown). I randomly subsampled the samples from these chains (after burn-in and outlier removal) and used them for further analysis.

A.2.5 Sensitivity analysis

Parameter	Fixed value	Lower bound	Upper bound	Source
ρ_{shoot}	0.35	0.27	0.35	Other lettuce densities in
$d_{\text{root,h}}(t)$	0.057	0.0444	0.071	95% quantile by assuming normal distributed root weights [152]
$d_{\text{shoot,h}}$	0.045	0.04	0.05	From [16]
$\rho_{\text{root,h}}$	0.2	0.1	1	Assumed to span one order of magnitude.
$t_{\text{ht,h}}$	14	12	16	Assumed two-day difference
$t_{\text{g,h}}$ (0)	800	600	6000	Assumed from experiment, span one order of magnitude
$t_{\text{li,h}}$	21	19	23	Assumed two-day difference
V_e	80000	8000	80000	Assumed to span one order of magnitude

θ	0.435	0.435	0.476	Ranging from sandy loam to clay loam (no loamy sand or sand) from [29]
$k_{\mathrm{dec,s}}$	0.15	0.15	0.199	Surface applied, PBS, 2 soil types from Roberts.et al 2016
$V_{\mathrm{root,s}}(t)$	100	30	300	Assumed to span one order of magnitude
r_{g}	0.2056	0.203	0.2082	From standard error in [128]
w_{f}	550	335	550	From [72]
$k_{\mathrm{att,s}}$	4.1, 0.8	0.8	4.1	From [121]
$k_{\mathrm{det,s}}$	0.00087, 0.003	0.00087	0.003	
$t_{\mathrm{ht,s}}(t)$	14	12	16	Assumed two-day difference

Table A.1: Sensitivity analysis parameters

A.3 Figures

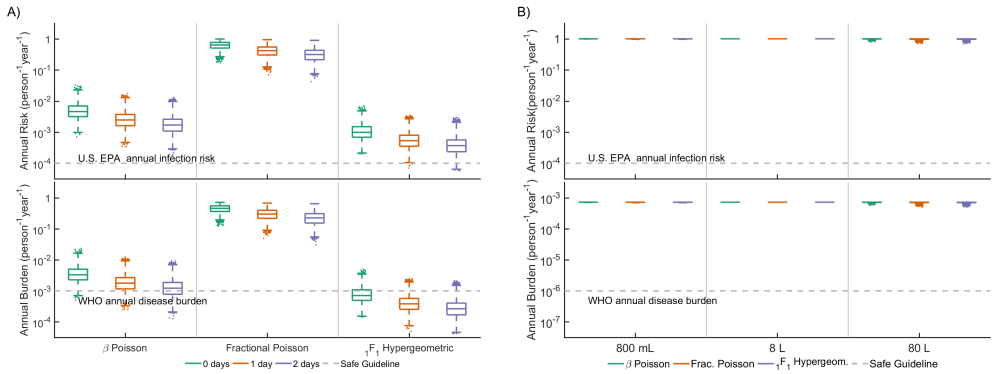

Figure A.1: Comparison of risks by increasing holding time of lettuce after harvesting from soil (A) or increasing tank volume of hydroponic grown lettuce (B). Both strategies failed to reduce risk below the acceptable limits.

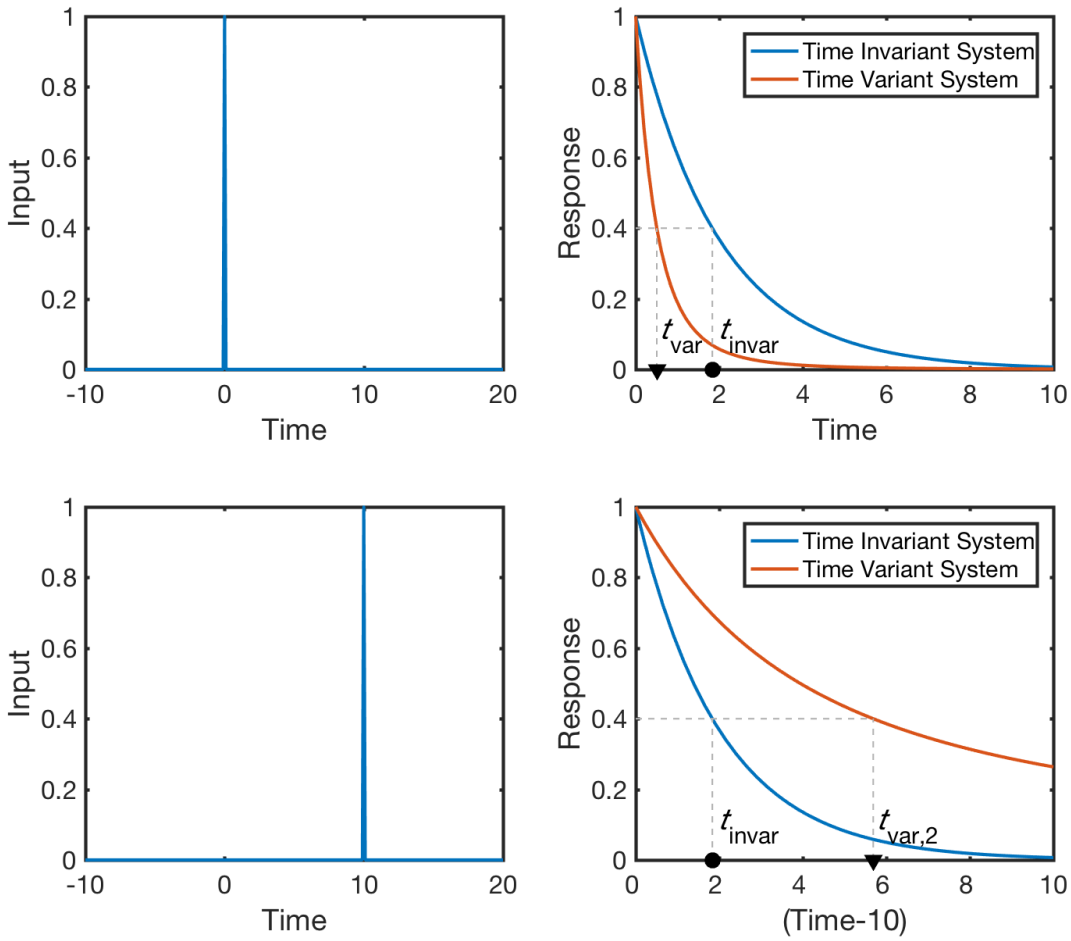

Figure A.2: Illustration of the difference in response to delay input for a time invariant vs. a time variant system. An input pulse at $t = 0$ produces the response in the top right panel for time invariant and time variant systems. Shifting the input pulse to $t = 10$ produces a shifted (but same shape of blue lines) response in the time invariant case but a different characteristic response (different shape of orange lines) in the time variant case. Time to reach a response of 0.4 remains in the same relative position (t_{invar}) for the time invariant system whereas it shifts (t_{var} to $t_{\text{var},2}$) for the time variant system, showing non-unique times for reaching the same response.

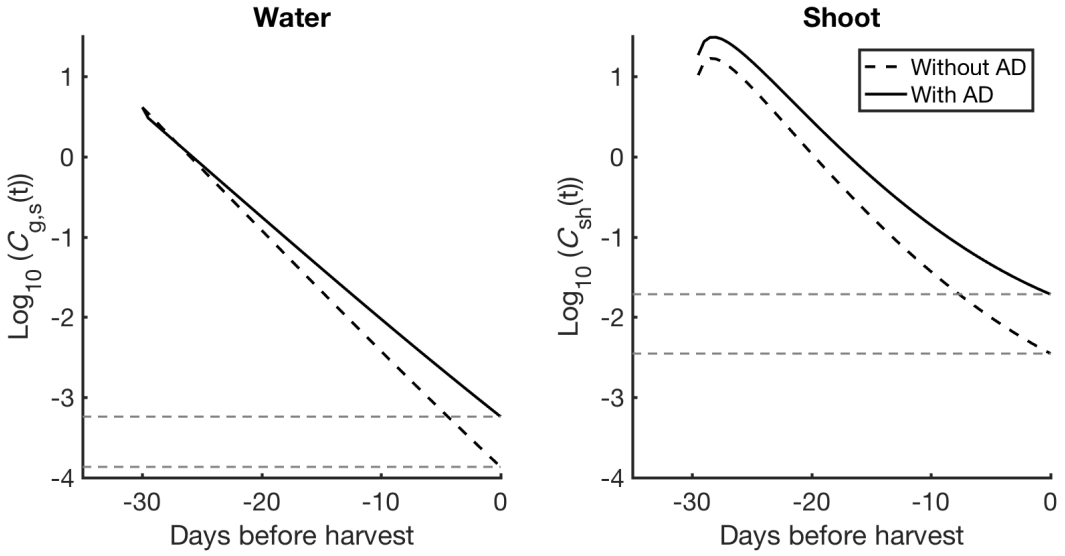

Figure A.3: Comparison of the viral concentration in the growth medium (water) and lettuce shoot using models with and without incorporation of AD of viruses to hydroponic tank walls. In this simulation, a lower $k_{\text{att,s}}$ value (one tenth of the best fit parameter from the) was used. The model without AD underestimates the viral load in the lettuce.

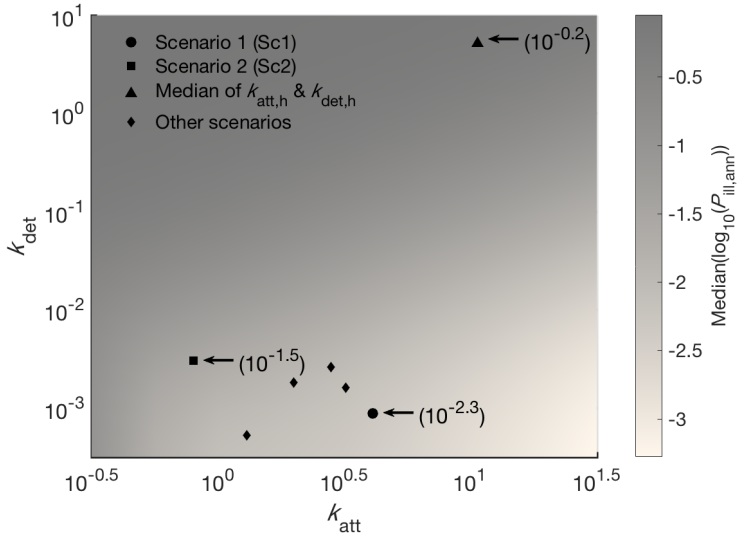

Figure A.4: Illustration of median($\log_{10}$(risk)) as a function of AD kinetic parameters for soil grown lettuce, β Poisson risk model. Simply replacing $k_{\text{att,s}}$ and $k_{\text{det,s}}$ with the medians of $k_{\text{att,h}}$ and $k_{\text{det,h}}$ in the soil model drastically reduces the median($\log_{10}$(risk)). Comparing numerical values of median($\log_{10}$(risk)) listed above (see Table 2.3) with Fig. 2.4 (top panel, median of β Poisson boxes) shows that a change in AD kinetic parameters exerts a major influence in determining the magnitude of risk.

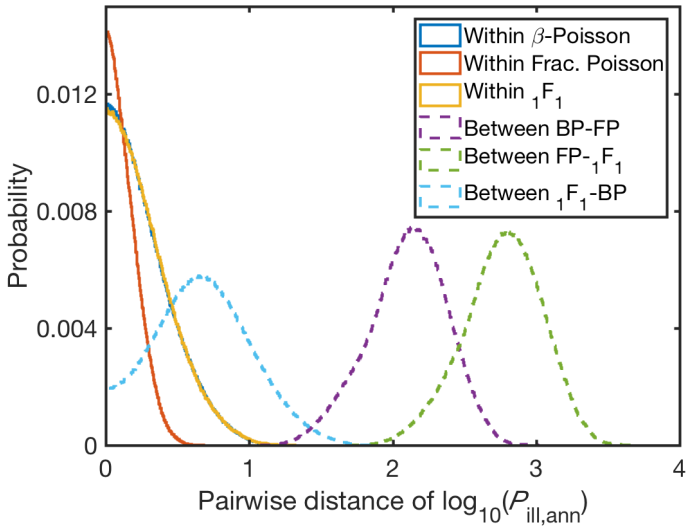

Figure A.5: Histograms of the pairwise distances of annual risk estimates (Sc1), $P_{\text{ill,ann}}$ for the different dose-response models. Pairwise distance is a measure of variability with lower distances indicative of less variability. The within dose-response model variability (solid lines) are lesser than between dose-response model variability (dashed lines).

Appendix B

Quantifying the infection risk of antibiotic-resistant bacteria

B.1 Datasets used

ID	[Ref]	Dose	n_{ill}	n_{tot}	t_{fs}(days)
DS1	[127]	1.00e+04	0	5	1
		1.00e+04	0	5	
		1.00e+06	0	5	
		1.00e+06	1	9	
		1.00e+08	5	8	
		1.00e+08	3	5	
DS2	[127]	1.00e+06	0	4	2.625
		1.00e+06	1	5	
		1.00e+08	1	5	
		5.00e+08	3	5	
		2.50e+09	6	6	
		1.00e+10	9	10	
		1.00e+10	9	14	
		1.00e+10	3	5	
		1.00e+10	5	5	
		2.00e+10	2	2	
		2.30e+10	14	19	

Table B.1: Datasets used in chapter 3.

B.2 Parameter ranges for sensitivity analysis

Parameter	Units	Lower bound	Upper bound
C	mg L^{-1}	0.00	0.05
f_r	-	0.00	0.10
$\log_{10}(d)$	-	1.00	4.00
$E_{\max}$*	day^{-1}	612.00	1836.00
EC_{50}*	mg L^{-1}	4.96	14.89
r*	CFU^{-1}	5.33×10^{-9}	1.59×10^{-8}
t_{fs}	days	1.50	2.50

Table B.2: Parameter ranges for exponential model. Parameters with * are increased and decreased by 50% of the values used in Fig. 3.2.

Parameter	Units	Lower bound	Upper bound
C	mg L^{-1}	0.00	0.05
f_r	-	0.00	0.10
$\log_{10}(d)$	-	1.00	4.00
$E_{\max}$*	day^{-1}	612.00	1836.00
EC_{50}*	mg L^{-1}	4.96	14.89
α*	-	0.08	0.24
β*	-	7.07×10^6	2.12×10^7
t_{fs}	days	2	3

Table B.3: Parameter ranges for beta-Poisson model. Parameters with * are increased and decreased by 50% of the values used in Fig. 3.2.

B.3 beta-Poisson procedure verification

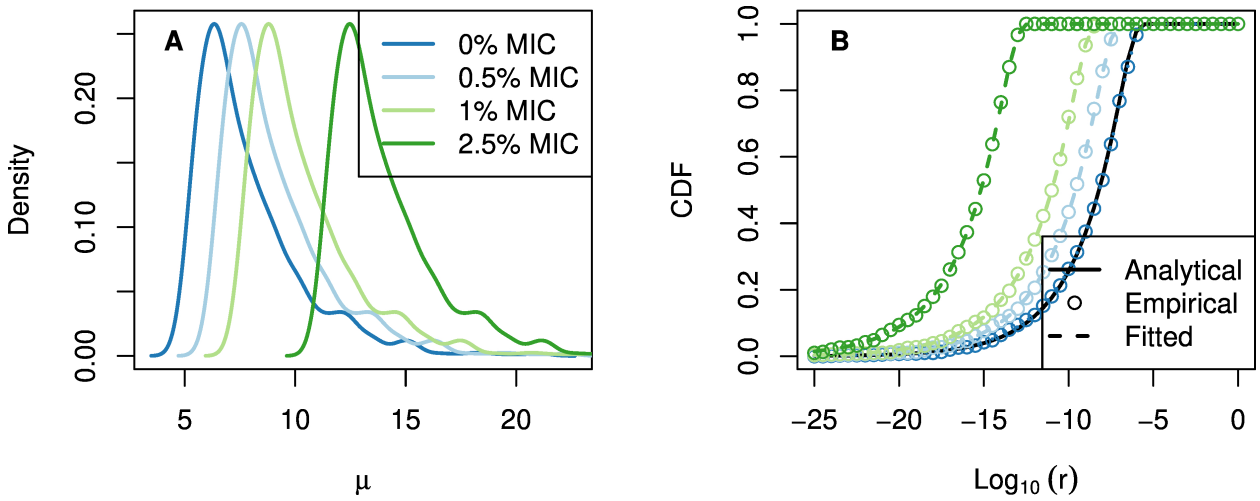

Figure B.1: Verify conversion procedure for beta-Poisson model. (**A**) Plot of μ at various concentration. (**B**) Comparison of empirical CDF of the sampled r values and the CDF from the fitted Beta distributions. The analytical CDF at 0% MIC is also shown for comparison.

B.4 Methods

B.4.1 Using the Simple Death DRM

Suppose one is interested in calculating the response for a pathogen. It is present in an exposure case with $d = 1000$, with 20% of the pathogen being resistant to an antibiotic, and the concentration of antibiotic is $C = 0.025 \times \text{MIC}\,(2\,\mu\text{g mL}^{-1}) = 0.05\,\mu\text{g mL}^{-1}$.

- Identify dose-response data for the pathogen. This can be like DS1 or DS2 listed in Table B.1.
- Identify t_{fs}. This is the latest time at which some subject shows the first symptom. Suppose $t_{\text{fs}} = 1$ day
- Identify $E_{\max}$ and EC_{50} for the antibiotic-pathogen combination of interest. Suppose $E_{\max} = = 1224$ day^{-1}and $\text{EC}_{50} = 9.93$ mg L^{-1}= 9.93 μg mL^{-1}.
- Fit both exponential and beta-Poisson models to this dataset and identify the best fitting model, using methods outlined in [57].
- If best fitting model is exponential, go to section B.4.2. If best fitting model is beta-Poisson, go to section B.4.3.

The sensitivity analyses indicate that getting approximate values for t_{fs} is sufficient to predict response. However, it's value is critical to accurately estimate death rate (μ) if using the exponential DRM.

B.4.2 Using exponential DRM

The exponential model is given by:

$$P(d) = 1 - \exp\left(-rd\right)$$

Suppose the best fit for r is given by $\hat{r} = 1.07 \times 10^{-8}$.

- Compute μ by solving
$$(1 - \exp(-\mu t_{\text{fs}})) = \exp(-\hat{r})$$
to get
$$\mu = -\log(1 - \exp(-\hat{r}))/t_{\text{fs}} = 7.97\,\text{day}^{-1}$$
- Compute $\mu_{s,\text{AB}}(C)$ using
$$\mu_{s,\text{AB}}(C) = \mu + \frac{E_{\max}C}{EC_{50} + C} = 7.97\text{day}^{-1} + \frac{1224\,\text{day}^{-1} \times 0.05\mu\text{g mL}^{-1}}{(9.93 + 0.05)\mu\text{g mL}^{-1}} = 14.10\,\text{day}^{-1}$$
- Set $\mu_{r,\text{AB}} = \mu = 7.97\,\text{day}^{-1}$
- Compute extinction probabilities for the susceptible and resistant subpopulations using
$$P_{\text{ext},s}(d|f_r, C) = (1 - \exp(-\mu_{s,\text{AB}}(C)t_{\text{fs}}))^{d\times(1-f_r)} \approx 0.999400822$$
and
$$P_{\text{ext},r}(d|f_r, C) = (1 - \exp(-\mu_{r,\text{AB}}t_{\text{fs}}))^{d\times f_r} \approx 0.933317727$$
- Compute total response probability with
$$P(d|f_r, C) = 1 - P_{\text{ext},s}(d|f_r, C)P_{\text{ext},r}(d|f_r, C) \approx 0.067241497$$
- If $\big(1-P_{\text{ext},s}(d,t|f_r,C)\big)P_{\text{ext},r}(d,t|f_r,C) > (1-P_{\text{ext},r}(d,t|f_r,C))$, illness is AB treatable. If not, illness is not AB treatable. In this case, this condition evaluates to False and hence the illness is likely not AB treatable.

B.4.3 Using beta-Poisson DRM

The beta-Poisson DRM is given by

$$P(d) = 1 - \left(1 + \left(\frac{d}{\beta}\right)\right)^{-\alpha}$$

Suppose the best fit parameters are $\hat{\alpha} \approx 0.1615058$ and $\hat{\beta} = 1414958$. Computing response probabilities is more involved and requires access to a function that can fit a beta distribution, such as the `fitdistrplus` package in R [36].

- From the values of $\hat{\alpha}$, $\hat{\beta}$, $E_{\max,}$ EC_{50} and C, compute α_s and β_s for the susceptible subpopulation. For this, use the algorithm outlined in the Methods section. We get $\alpha_s = 0.1613020$ and $\beta_s = 1.295420 \times 10^{13}$.

- Set $\alpha_r = \hat{\alpha}$ and $\beta_r = \hat{\beta}$ for the resistant subpopulation.

- Compute extinction probabilities for the susceptible and resistant subpopulations using

$$P_{\mathrm{ext},s}(d|f_r, C) = \left(1 + \left(\frac{d \times (1 - f_r)}{\beta_s}\right)\right)^{-\alpha_s} = 1$$

  and

$$P_{\mathrm{ext},r}(d|f_r, C) = \left(1 + \left(\frac{d \times f_r}{\beta_r}\right)\right)^{-\alpha_r} \approx 0.999977202$$

- Compute total response probability with

$$P(d|f_r, C) = 1 - P_{\mathrm{ext},s}(d|f_r, C) P_{\mathrm{ext},r}(d|f_r, C) \approx 2.28 \times 10^{-5}$$

- If $\big(1 - P_{\mathrm{ext},s}(d, t|f_r, C)\big) P_{\mathrm{ext},r}(d, t|f_r, C) > (1 - P_{\mathrm{ext},r}(d, t|f_r, C))$, illness is AB treatable. If not, illness is not AB treatable. In this case, this condition evaluates to False and hence the illness is likely AB untreatble.

Appendix C

Role of quorum sensing in microbial pathogensis

C.1 Fitted parameter values

r_1 (/day)	r_2 (/day)	r_3 (cm^2/ (CFU day))	i_{max} (CFU/cm^2)	b_2 (/day)	i_{thresh} (1/(CFU day))	F_{gr}	F_{dr}
1.94	1.47e-02	2.71e-07	1.18e+07	1.70	7.34e+06	0.57	8.67
1.90	1.55e-02	3.56e-07	8.93e+06	2.00	4.81e+06	0.60	8.01
2.10	2.68e-02	2.92e-07	1.05e+07	0.70	2.04e+07	0.61	7.46
2.19	1.95e-02	3.53e-07	9.09e+06	1.80	6.88e+06	0.64	6.83
1.69	1.02e-02	3.91e-07	8.70e+06	1.70	7.97e+06	0.68	6.34

Table C.1: Rank 1 solutions in increasing order of F_{gr}, the growth-objective (eq. 4.10). F_{dr} is the dose-response objective (eq. 4.5). Remaining parameters are defined in chapter 4.

C.2 Comparing MRSA and MSSA growth

Study	MSSA strain	MRSA strain	ρ
[79]	1999	494	0.8505
[9]	O9	P22	0.9461
[39]*	1492	211	0.9843
[76]*	29213	43300	0.9995

Table C.2: Pearson correlation coefficients (represented by ρ) for growth curves of various strains of MSSA and MRSA. * represents median of the several correlation coefficients computed