Mathematical models of *Clostridium difficile* transmission

Angus McLure

A thesis submitted for the degree of
Doctor of Philosophy
of The Australian National University

May, 2019

Declaration

Candidate

This thesis is an account of research undertaken between February 2015 and September 2018 at The Research School of Population Health, College of Health and Medicine, The Australian National University, Canberra, Australia.

I declare that the work contained in this thesis is the result of original research and has not been submitted to any other University or Institution. This thesis by compilation is based on the following papers to which I made significant contribution as first author:

1. Angus McLure, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. Healthcare-associated *Clostridium difficile* infections are sustained by disease from the community. *Bulletin of Mathematical Biology*, 79(10):2242-2257, 2017.

2. Angus McLure, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. *Clostridium difficile* classification overestimates hospital-acquired infections. *Journal of Hospital Infection*, 99(4):453-460, 2018.

3. Angus McLure, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. Diverse sources of *Clostridium difficile* in the community: importance of animals, infants and asymptomatic carriers. *Epidemiology and Infection*, 147:e152, 2019.

4. Angus McLure, Luis Furuya-Kanamori, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. Seasonality and community interventions in a mathematical model of *Clostridium difficile* transmission. *Journal of Hospital Infection*, (In Press) 2019.

5. Angus McLure and Kathryn Glass. Simple rules for estimating reproduction numbers in the presence of reservoir exposure or imported cases. Submitted to *Theoretical Population Biology* (under review), 2018.

The details of my contribution to each of these papers is as follows:

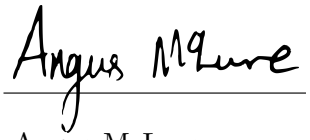
1. I surveyed the literature and integrated this with the input and expertise of co-authors to develop the model framework. I developed the novel method of sensitivity analysis. I wrote MATLAB code that implemented the model, ran simulations, produced the numerical results, and performed the sensitivity analysis. I contributed to the interpretation of the results, made the figures and tables, and drafted the manuscript. I managed the submission process including responses to reviewer feedback.

2. I developed the study concept and the method to assess classification definitions with the model. I wrote MATLAB code that produced the numerical results and performed the sensitivity analysis. I contributed to the interpretation of the results, made the figures, and drafted the manuscript. I managed the submission process including responses to reviewer feedback.

3. I surveyed the literature and integrated this with the input and expertise of co-authors to extend our previous model framework to community and infants. I extended the existing method of assessing classification definitions to the new model. I compiled published data and wrote MATLAB code to estimate model parameters from this data. I conceived the food-driven threshold. I wrote MATLAB code to produce all numerical results and figures. I drafted the manuscript and managed the submission process.

4. I developed the study concept. I wrote MATLAB code that implemented the model, ran simulations, produced numerical results, and performed the sensitivity analysis. I contributed to the interpretation of results and produced the figures and tables. I drafted the manuscript and managed the submission process.

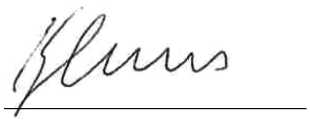
5. I developed the study concept and produced the mathematical results. I applied the results to the two case studies on *Clostridium difficile*. I drafted the manuscript and managed the submission process.



Angus McLure
May, 2019

Collaborating Authors

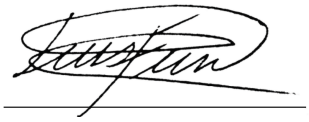
I agree that Angus McLure made the contribution to the authorship and research of the paper(s) on which I am a co-author, as stated in the preceding pages.



Kathryn Glass
May, 2019

Archie C. A. Clements
May, 2019

Martyn Kirk
May, 2019



Luis Furuya-Kanamori
May, 2019

Acknowledgements

First and foremost I would like to thank my primary supervisor Katie Glass. She has been so generous with her time. She has endured at least one draft (and sometimes half a dozen) of pretty much everything I have written over the last three and a half years. These acknowledgements are probably one of the few pieces of writing she hasn't edited. Our weekly meetings were invaluable. One of my faults is a total inability to estimate how long things will take. Even our meetings where I only had 'a couple things' to talk about, seemed to fill the (generous) allotted time. I thank her for all her technical, professional, and academic guidance.

I would also like to thank my others supervisors, Martyn Kirk and Archie Clements. I thank them for their insight, for finding my research interesting even when I was over it, and for encouraging me throughout my candidature.

I thank the Australian Government Research Training Program, the ANU, and the Research School of Population Health (RSPH), which have made this research possible through financial and practical support. I also thank all the other societies and conferences which provided travel funding, particularly PRISM^2.

I would like to thank the wonderful community of students and staff at the RSPH who have helped make the PhD so enjoyable. I thank Luis Furuya-Kanamori and Yingxi (Cimo) Chen for some really insightful conversations about *Clostridium difficile* which informed the basic model underlying Chapters 3 to 6 and helped frame the discussion in Chapter 4. I thank Alexandra Hogan for welcoming me at the RSPH and introducing me to countless people at the school and at mathematics conferences. A special thank you is needed for the members of the Finisher's Club who have been a great source of encouragement and friendship even before the club was founded: Jenny Welsh, Naomi Clarke, Laura Ford, Chatu Yapa, Tehzeeb Zulfiqar, Kefyalew Alene, Ruth Kharis, and Jodie Leu.

I thank my parents for their love and support over many years and all their interest and encouragement in my research. Ezra, my son and the happiest little five-month-old that ever was, also deserves a thank you. Though he has been exceedingly distracting, he has also been a nearly endless source of joy. He has made what could have been a very dreary five months of thesis writing some of the more enjoyable months of my life.

Tamsin, my wife, deserves a whole chapter of acknowledgements. She has supported me through every stage of the PhD. She has shared my excitement when things were going well, counselled me through disappointment and setbacks, encouraged me when I wanted to give up, and celebrated all the little victories along the way. Most importantly, she has helped me keep the right perspective: that a PhD is not an end itself, that the respect

(and financial benefits) that a doctorate confers are not worth very much, but that the real purpose of this work is to, in some little way, carry out God's great work of healing our hurting world. I thank God for providing all the resources, people, and skills that have been essential. I pray for a time when this research is no longer needed.

> *And I heard a loud voice from heaven saying, "Behold, the dwelling of God is with men, and He will live with them, and they shall be His people. God Himself will be with them and be their God. And God will wipe away every tear from their eyes; there shall be no more death, nor sorrow, nor crying. There shall be no more pain, for the former things have passed away." Then He who sat on the throne said, "Behold, I make all things new."*
>
> *(Revelation 21:3-5)*

> *And he showed me a pure river of water of life, clear as crystal, proceeding from the throne of God and of the Lamb. In the middle of its street, and on either side of the river, was the tree of life, which bore twelve fruits, each tree yielding its fruit every month. The leaves of the tree were for the healing of the nations. And there shall be no more curse, but the throne of God and of the Lamb shall be in it, and His servants shall serve Him.*
>
> *(Revelation 22:1-3)*

Abstract

Clostridium difficile infections (CDIs) are some of the most common hospital-acquired infections and the most common cause of antibiotic-associated diarrhoea. CDIs lead to great loss of life, severe health outcomes, and incur very high financial costs through treatment, extended hospital stays, and readmissions. Despite extensive research and many resources committed to the prevention and treatment CDIs in hospitalised patients, hospitals continue to be hotspots for this disease. Meanwhile, there is an emerging awareness of the burden this disease places on the broader community including patients who have not recently been hospitalised. In the community approximately 5% of adults and a higher proportion of infants are asymptomatically colonised. Colonisation is also common in livestock and the pathogen has been isolated from meat and vegetables. However, the various sources of transmission in the community and the consequences for infections within and beyond hospitals are not well understood.

This thesis develops and employs mathematical models of *C. difficile* transmission to explore three themes: improving models to capture the complex epidemiology of *C. difficile*, populations that sustain *C. difficile* transmission, and the classification of CDIs as hospital or community-acquired. Addressing the first theme, I argue that the essential epidemiology of *C. difficile* is captured by modelling the interactions of three key factors: pathogen, immunity, and gut flora. I argue that modelling transmission in an integrated model of adults and infants across hospitals and communities provides insights that hospital-only and adult-only models cannot. By incorporating seasonality into these models, I argue that seasonal variation of antibiotic prescription rates is more likely to be the main driver of CDI seasonality than seasonal transmission.

In the second theme, I argue that most hospitals – though hotspots for transmission – are not disease sustaining populations. Instead, transmission outside hospitals maintains the disease in the hospital and community. I argue that reducing transmission in the hospital cannot eliminate the disease in the broader population, but that reducing transmission from adults or infants in the community could interrupt transmission in the human population. Similarly, I argue that *C. difficile* in the community may be driven by transmission from animal reservoirs if as few as 3.5-26.0% of human infections are acquired from animal or food sources.

In the final theme, I argue that an illusion of hospital-driven disease is in part perpetuated by surveillance definitions that systematically misclassify many community-acquired cases as hospital-acquired. The incubation period for *C. difficile* infections often exceeds the two-day or three-day cut-offs commonly used to classify patients recently admitted to hospital. I argue that many patients who acquire the pathogen prior to admission develop symptoms after the cut-off and are therefore incorrectly classified as having acquired the

infection during their hospital stay. Furthermore, I argue that time since hospital discharge is a poor indicator of whether a CDI is hospital or community-acquired.

Contents

Chapter 1

Introduction

1.1 Background and Research Motivation

Clostridium difficile infection is one of the most common causes of antibiotic-associated diarrhoea and hospital-associated infection globally [1–7]. In the USA alone *C. difficile* is estimated to cause 450,000 infections, 30,000 deaths, and result in 4.8 billion USD in excess accute-care costs annually [8, 9]. Historically, most reported infections were in elderly, hospitalised patients, but there is a growing awareness of infections in the broader community including children [10, 11]. *C. difficile* infection (CDI) is characterised by diarrhoea and nausea, and in more severe cases psuedomembranous colitis, fever, toxic megacolon, and possible renal failure. A recurrence of symptoms occurs in approximately 20% of infections [12]. Transient asymptomatic carriage is also common. The prevalence of asymptomatic carriage is approximately 5% in healthy adults but may be higher amongst hospitalised adults, long-term care facility residents, and infants [13]. Natural gut flora inhibits the proliferation of *C. difficile* in exposed individuals [14]. Most CDIs follow the disruption of the patient's gut flora by the use of broad-spectrum antibiotics or proton-pump inhibitors [15, 16].

C. difficile infections are a major health burden, but much remains unknown. Asymptomatically colonised individuals contribute to transmission, but many never develop symptoms and remain undetected [17, 18]. This makes it difficult to ascertain when and where transmission occurs and the source of the exposure. Furthermore, strains of *C. difficile* that are pathogenic to humans are also common in livestock and pets and have been isolated on retail meat and other produce [19]. Transmission is known to occur from many sources and in many settings but the relative contributions and interactions of these sources and settings remain unclear.

Mathematical models offer a system-level understanding of transmission dynamics by unifying what is known about individual-level factors for disease (e.g. human behaviour, host-pathogen interactions) and population-level epidemiology (e.g. incidence, seasonal trends, differences between settings) [20]. Suitable models can allow investigators to infer the hidden dynamics of transmission and population-level interactions that would not be visible with traditional epidemiological tools. Appropriately calibrated models can be used to assess current practice and possible interventions that could reduce or eliminate the burden of the disease [20–22]. Relatively few mathematical models of *C. difficile* have

been published and these have focused mostly on within-hospital transmission. Much remains to be learnt through mathematical models of *C. difficile* transmission.

1.2 Research themes

The central chapters of this thesis – which consist of five published or submitted research articles and their supplementary materials – are ordered so that each chapter extends ideas or models presented in previous chapters. Three themes run through these chapters:

1. improving mathematical models to capture the complex epidemiology of *C. difficile*
2. identifying populations that sustain *C. difficile* transmission and quantifying interactions between populations, and
3. classifying CDIs as hospital or community-acquired.

1.2.1 Theme 1 – Improving mathematical models of *C. difficile*

I argue that models of *C. difficile* transmission should capture three main factors or dimensions of *C. difficile* epidemiology: pathogen, gut flora, and immunity. Many existing models of *C. difficile* capture only some of these factors or make unrealistic assumptions about these factors. The main models developed and used in this thesis attempt to combine these three factors in a parsimonious, multi-dimensional framework. I build on this model to incorporate other neglected aspects of *C. difficile* epidemiology including seasonality, transmission in the community, animal reservoirs, and the unique interactions of infants with the pathogen.

1.2.2 Theme 2 – Populations that sustain *C. difficile*

In an isolated population, the reproduction number for a pathogen – the average or typical number of secondary cases arising from each primary case in a fully susceptible population – is a measure of the pathogen's epidemic or endemic potential in that population. If the reproduction number exceeds one, upon introduction the pathogen will either quickly die out due to random effects or spread through the population in an epidemic. The disease may then persist, becoming endemic in the population. If the reproduction number is less than one in the population, some onward transmission may occur following the introduction of a small number of infective persons, but the disease will soon die out in the population. However, a disease can persist in a population even if the reproduction number is less than one if the pathogen is continually reintroduced from another human population or from an animal or environmental reservoir. Furthermore, it is possible that two or more populations, which would not sustain the pathogen alone, sustain the disease through mutual exposure or movement between populations. Given the continual admission and discharge of patients to and from hospitals and the possibility of transmission between animals and humans, it is of great interest to understand the interactions of

hospital, communities, and animals. I use mathematical models of transmission in hospitals and the community that account for movement between these settings to determine whether *C. difficile* would persist in either the hospital or community without continual reintroductions from the other setting. I also leverage these models to estimate the threshold amount of transmission from animals that, if exceeded, would imply that *C. difficile* in the human population is driven by transmission from animal reservoirs.

1.2.3 Theme 3 – Classifying *C. difficile* infections

It is important to know where CDIs are acquired in a population: whether in one hospital or another, or in a long-term care facility, or in the broader community. This knowledge can guide the design of interventions to interrupt or reduce transmission in these locations. However, CDIs have a latent period of days or weeks [23, 24], so the location of the onset of symptoms (which is usually known) may not be the same as the location of acquisition (which is usually unknown), especially if the patient has recently been admitted to, discharged from or transferred between healthcare facilities. Common surveillance definitions, first introduced in 2007 and later endorsed by the Infectious Diseases Society of America and the European Centre for Disease Prevention and Control, classify cases as hospital or community-acquired using arbitrary cut-offs for time between onset of symptoms and the most recent hospital admission or discharge [23, 25, 26]. These definitions and their variants are widely used to inform estimates of the proportion of cases that are hospital-acquired, compare incidence between regions and hospitals, define infection control standards, and judge the performance of individual healthcare facilities [6, 8, 27]. However, these definitions have never been tested or verified, so it is not known whether they adequately distinguish hospital and community-acquired cases. I use mathematical models to assess the various elements of the surveillance definitions and suggest possible improvements.

1.3 Thesis structure

This thesis consists of eight chapters including this introduction. Chapter 2 lays out the background information necessary to understand *C. difficile* infections and the mathematical tools used to explore the themes. Chapter 3 presents a new model of *C. difficile* infections in a hospital population that captures the interaction of pathogen, gut flora and immunity (Theme 1). This model is the basis of the analyses in Chapters 3 and 4 and provides the basic framework for the extended models in Chapters 5 and 6. Chapter 3 uses the base model to explore the dynamics of transmission and the role of the admission and discharge of patients colonised with *C. difficile* (Theme 2).

Chapter 4 uses the model presented in Chapter 3 to assess standard surveillance definitions that classify CDIs presenting in hospitals as acquired in the current hospitalisation or acquired prior to admission (Theme 3). Chapter 5 presents an extension of the model in Chapter 3 that captures the hospital, the community, the role of infants, and transmission

from animal reservoirs of *C. difficile* (Theme 1). The expanded model is used to extend the assessment of surveillance definitions begun in Chapter 4 to community-onset infections (Theme 3). The model is used to determine whether hospitals and communities are populations that sustain *C. difficile* and identify the conditions under which the persistent presence of *C. difficile* in the human population would be dependent on continual exposure to animal reservoirs (Theme 2). Chapter 6 uses the model presented in Chapter 5 to explore possible mechanisms for seasonality (Theme 1) and estimate the potential effect of hospital-based and community-based interventions on infections and colonisations in each of these populations (Theme 2).

Chapter 7 extends the idea of the animal-driven or reservoir-driven threshold employed in Chapter 5 to general compartmental models. Due to its generality, this chapter can be read independently from the preceding chapters but corroborates the findings of Chapters 3 and 5 with cases studies focused on *C. difficile* (Theme 2). The thesis concludes with Chapter 8 which discusses the key findings in each theme, placing them in the context of the literature and identifying the key limitations.

Chapter 2

Background

This chapter provides an introduction to the pathobiology, history, and epidemiology of *Clostridium difficile*. It then provides a brief introduction to mathematical modelling of infectious diseases before reviewing the existing literature in the mathematical modelling of *C. difficile* transmission, highlighting some of the gaps in the literature that this thesis addresses.

2.1 *Clostridium difficile*

2.1.1 Pathogen basics

Clostridium difficile is an anaerobic, spore-forming bacteria that colonises the lower intestinal tract of many mammals including humans [19]. The bacteria was first discovered in 1935 in the intestinal flora of healthy newborn infants [28]. However, many strains of *C. difficile* produce toxins (toxin-A, toxin-B and binary toxin) that lead to intestinal damage and potentially life-threatening diarrhoea in some patients [29]. Approximately 5% (but in some populations up to 15%) of adults are asymptomatically colonised with toxigenic and non-toxigenic strains of *C. difficile* [13]. Asymptomatic colonisation is transient, lasting approximately one month [30]. Colonisation with non-toxigenic strains does not lead to symptoms and may prevent colonisation of toxigenic strains [31, 32]. However many (if not most) will remain asymptomatic for the duration of their colonisation, even if the colonising strain is toxigenic. In particular, nearly all infants will be colonised in the first year of life [33] – with prevalence peaking at 40% or higher at age 3-6 months [33–35] – but *C. difficile* rarely causes disease in infants. Children and adults develop specific antibody-mediated immune responses to toxins A and B which protect them from symptomatic disease [36]. One study found that seroprevalence for these antibodies increased with age, reaching 60-70% by adolescence in US participants and 86-97% in Panamanian participants [37]. Though protective against infection, a serological response to these toxins is not protective against asymptomatic carriage [15].

Taking antibiotics kills many of the species that make up the human gut flora, disrupting the balance of the ecosystem for days, weeks, or months and allowing some species to temporarily proliferate above normal levels [38, 39]. For this reason and because various strains of *C. difficile* are resistant to many classes of antibiotics [40], receiving antibiotics

dramatically increases the risk of developing a *C. difficile* infection (CDI) [16, 41]. In hospital settings, where antibiotic prescription rates are high, nearly all cases have a recent history of antibiotic use, though this observation may be due in part to ascertainment bias. A large study of CDI acquired in the community, where antibiotic prescription rates are much lower, found that half of CDI cases had antibiotic exposure in the 45 days prior to infection and that clindamycin and cephalosporins were associated with 32-fold and 15-fold increases in infection risk [42]. Other factors associated with disruption of intestinal flora, such as the consumption of proton-pump inhibitors and irritable bowel disease, are also associated with an increased risk of CDI [43, 44]. The main treatment for CDIs that do not resolve when any existing antibiotic therapy is discontinued is further antibiotic treatment, usually with vancomycin or metronidazole [12]. These treatments leave gut flora damaged and so recurrent CDI is common, especially in those without immune responses to *C. difficile* toxins [12, 45, 46]. Some alternative therapies, such as faecal transplants and probiotics, aim to restore infection resistance by replenishing damaged gut flora. Others, such as *C. difficile*-specific bacteriocin and bacteriophage therapies, attempt to remove *C. difficile* without disrupting the patient's gut flora [47]. Faecal transplants have been shown to prevent recurrent CDI [48] and are now a recommended treatment option for cases with multiple recurrences [49].

2.1.2 The epidemiology of *C. difficile* infections

The emergence of highly virulent 'epidemic' strains of *C. difficile* in the mid 2000's was associated with rapidly increasing incidence and mortality, particularly in North America and Europe [50]. Incidence has continued to rise in many parts of the world including Australia [6] and East Asia [7]. National targets to reduce transmission, improve antimicrobial stewardship, and increase reporting in the UK were followed by reductions in fluoroquinolone prescriptions and near elimination of infections caused by the predominantly fluoroquinolone-resistant epidemic strains [51]. However, the disease continues to lead to great loss of life and impose a large burden on health-systems around the world. Today the estimated annual burden of *C. difficile* in the USA alone is half a million infections, 30 thousand deaths, and direct health-care costs in excess of 4.8-billion USD [8, 9]. Most reported infections affect patients who are currently in hospital or who have recently received some form of healthcare. A study from the USA estimated that 94% of CDIs had either been hospitalised for >72h, resided in a long-term care facility, or had received inpatient or outpatient care in the past 12 weeks at the time of positive test for CDI [8]. Other risk factors for infection include age over 65 years, renal disease, irritable bowel disease, solid-organ transplants, hematopoietic stem cell transplants, and exposure to antibiotics and proton-pump inhibitors [49]. The incidence of CDI exhibits moderate seasonality in the Northern Hemisphere, peaking in late winter [52]. Though one Australian study has identified seasonal differences in colonisation prevalence amongst hospitalised patients [53], the seasonality of colonisation in the community is unknown. Time-series analysis has shown that CDI incidence follows antibiotic prescriptions and

winter respiratory infections such as influenza and respiratory syncytial virus with a lag of one or two months [54, 55]. The reported incidence of CDI also increases during outbreaks of other winter gastroenteritis infections such as norovirus [56, 57]. Others have identified an association between the proportion of positive tests for *C. difficile* and rainfall in tropical and subtropical climates [58].

2.1.3 Transmission of *C. difficile*

C. difficile is transmitted via spores through the faecal-oral route. The spores are very hardy, resisting many hospital-grade disinfectants [59] and can remain viable on surfaces for months [60]. Spores in contaminated meat can survive cooking for two hours at recommended minimum safe food temperatures [61]. Spores can be spread via direct contact with infected patients or via hands of healthcare workers. Unfortunately, the alcohol-based hand-washes commonly used in hospitals for routine hand hygiene are much less effective than soap and water for removing *C. difficile* [62].

Though the very high rates of infections in hospitalised patients suggests extensive transmission within hospitals, it is very difficult to identify actual transmission events. A potentially long period of asymptomatic carriage before the onset of symptoms [63] obscures when an individual acquired the pathogen, and the large proportion of carriers that are asymptomatic [15, 27] obscures the source of infection. A landmark study used whole genome sequencing from all known infections in a defined population to identify closely related isolates and determined that only a quarter of infections could be reasonably attributed to ward-based contact with another individual with CDI [64]. The source of the remaining transmission is highly uncertain. Possible sources include symptomatic carriers in the community and asymptomatic carriers in any setting [17, 63], especially residents of aged-care facilities and infants who have particularly high rates of carriage [13]. Infants are of particular interest not only because of their very high asymptomatic carriage rates but because the density of spores in their stools is comparable to stool from symptomatic adults [35]. Animal carriers are also potential sources of human infections, through direct contact with companion animals [65] and livestock [66] or indirect exposure via contaminated water and food [19]. Phylogenetic analysis of *C. difficile* ribotype 078 isolates from humans and livestock suggest frequent transmission between human and livestock populations [66], while the presence of *C. difficile* spores on retail meat [67] and vegetables [68] provides a vehicle for wide-spread exposure to livestock-derived *C. difficile*.

In the absence of simple means to determine where or from whom a given CDI is acquired, simple surveillance definitions that classify cases as hospital-acquired (or hospital-associated) and community-acquired (or community-associated) have been recommended [23]. The definitions use the time from hospital admission or discharge to onset of symptoms to distinguish hospital and community-acquired cases. These definitions or minor variants are widely used (*e.g.* [6, 8]) but have not been validated with empirical studies. These definitions underpin estimates of the relative contributions of hospital and community-based transmission [8] and are used in government-imposed limits for hospital

acquired infections [27].

2.2 Mathematical models

2.2.1 Introduction to mathematical modelling of infectious diseases

A core task in population health research is to determine and quantify the risk factors for a disease. For infectious and non-infectious diseases, individual characteristics and exposures – such as age, sex, smoking status or environmental exposures – are taken into account when determining the risk or susceptibility of individuals. However, according to McMichael a complete understanding of a disease must also include social-ecologic interactions [69]. This broader perspective is particularly important for infectious diseases where each individual's risk of infection depends not only on themselves but on the prevalence of infection and carriage in the people around them. The characteristics of each individual in a population influences the risk of acquiring the disease for each other individual. Treating infected individuals reduces the prevalence of a disease and thus has benefits for all individuals in the community. Herd immunity – established naturally or through vaccination – protects even those who lack immunity. McMichael also highlights the importance of the dynamic nature of disease and its risk factors. Many infectious diseases cause seasonal epidemics or outbreaks, so the prevalence of infectious persons – and hence each individual's risk of infection – may vary over the course of a season, epidemic or outbreak. For these reasons models of infectious diseases must be *dynamic*, must model whole *populations*, and must account for the *diversity* of individuals. Compartmental mathematical models – which are employed throughout this thesis – have all these characteristics and thus can be useful tools to study infectious diseases. I provide a brief introduction to these models. More detailed treatments of the subject and numerous examples can be found in a growing number of textbooks (*e.g.* [70–74]).

The first step in constructing a compartmental model is to divide the population of interest into groups or compartments relevant to the disease(s) being modelled. Consider a disease where recovered patients are immune to further reinfection in a population with two distinct risk categories: high and low. For this disease and population, a simple compartmental structure might consist of four compartments: high-risk susceptible patients (S_H), low-risk susceptible patients (S_L), infected patients (I) and recovered/immune patients (R).

The next step is to describe how the number of people in each class changes over time by considering the frequency of events or the rate at which disease or demographic processes affect the population in each compartment. These changes might introduce new individuals (*e.g.* the birth of a new susceptible patient), remove existing individuals (*e.g.* the death of an infected patient), or move an individual from one class to another (*e.g.* recovery of infected patient becoming immune). The rates at which these events or processes occur are defined in terms of the number of people in the various classes at that time (*e.g.* S_H, S_L, I or R) and a rate parameter which is to be estimated (typically

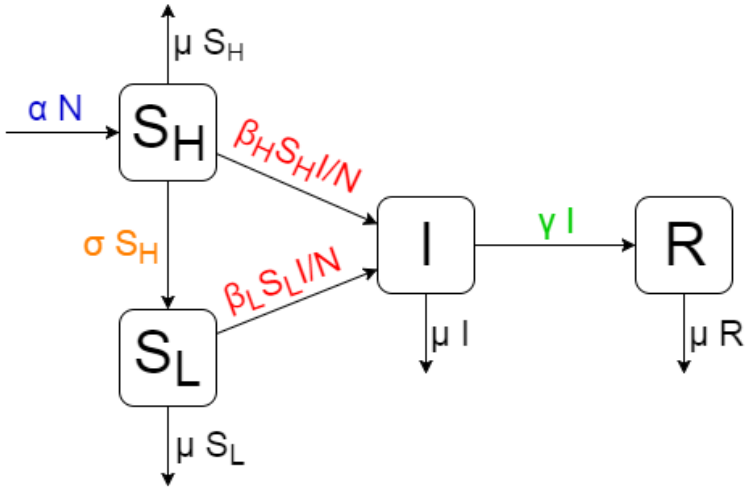


Figure 2.1: Example diagram summarising classes and transitions. The boxes represent the classes: high risk susceptible patients (S_H), low risk susceptible patients (S_L), infected patients (I) and recovered/immune patients (R). An arrow pointing away from a box represents events which remove individuals from that class, while arrows pointing towards a box indicate events which add individuals to that class. The term written by each arrow represents the rate at which these events occur. Greek letters (*e.g.* α, μ, and γ) are rate parameters that determine the rate of movement between classes. Latin upper-case letters indicate the number of people in the class (*e.g.* I is the number of people in the infectious (I) class). N is the total population: $N = S_H + S_L + I + R$. The colour coding indicates the kind of transition: blue for births, black for deaths, red for infections, green for recovery from infection, and orange for some process which reduces susceptibility, *e.g.* prophylactic treatment. The transmission parameter for transmission to high-risk individuals (β_H) is higher than for low-risk individuals (β_L) so high-risk individuals are infected more rapidly.

represented with a Greek letter). For instance, the number of people in the population that recover from an infection in unit time may be $I\gamma$, where I is the number of people infected and $1/\gamma$ is the average duration of the infection. Many transitions or events in compartmental models will be like this example, depending only on the characteristics of individuals that are affected. Such events will occur at a rate proportional to the number people in the class that is being left, but independent of the number of people in other classes. However, the rate at which people go from susceptible to infected classes (*i.e.* the incidence rate) is proportional to the number susceptible (S_H, S_L) and the proportion of population that is infected (I/N, where $N = S_H + S_L + I + R$). The compartmental structure, the possible transitions between these compartments, and the rate at which these transitions happen are often summarised with a diagram like Figure 2.1.

This kind of compartmental structure can be translated into many mathematical frameworks or equations, each with their own sets of assumptions, strengths, and weaknesses. Here I will briefly discuss the types that I and others have used to model *C. difficile* infections. The analytically simplest approach is to use *ordinary differential equations* (ODEs). This framework assumes that the numbers of people in each compartment are continuous,

deterministic (non-random) functions of time. The rate of change of the number of people in each compartment (*i.e.* the derivatives with respect to time) are expressed in terms of the rate of transitions in and out of each compartment. For the model described by Figure 2.1 the equations are:

$$\begin{aligned} S_H' &= \alpha N - \sigma S_H - \beta_H S_H I/N - \mu S_H \\ S_L' &= \sigma S_H - \beta_L S_L I/N - \mu S_L \\ I' &= \beta_H S_H I/N + \beta_L S_L I/N - \gamma I - \mu I \\ R' &= \gamma I - \mu R. \end{aligned} \tag{2.1}$$

This can be written in matrix form

$$\mathbf{X}' = A(\mathbf{X})\mathbf{X}, \tag{2.2}$$

where

$$\mathbf{X} := \begin{bmatrix} S_H \\ S_L \\ I \\ R \end{bmatrix} \tag{2.3}$$

and

$$A(\mathbf{X}) := \begin{bmatrix} -(\sigma + \beta_H I/N + \mu) + \alpha & \alpha & \alpha & \alpha \\ \sigma & -(\beta_L I/N + \mu) & 0 & 0 \\ \beta_H I/N & \beta_L I/N & -(\gamma + \mu) & 0 \\ 0 & 0 & \gamma & -\mu \end{bmatrix}. \tag{2.4}$$

Though there are typically no neat general solutions to these equations, given the number of people in each compartment at time 0, the number of people in each compartment over time can be calculated using standard, rapid computational methods (*e.g.* [75]). The solutions of many ODE systems will converge to an equilibrium point ($\mathbf{X}^*$), where the *rate* that people move in and out of each compartment balances, so the *number* of people in each compartment does not change over time ($\mathbf{X}' = A(\mathbf{X}^*)\mathbf{X}^* = 0$). Many other ODE systems converge to stable, cyclic patterns where the number of people in each compartment changes over time but repeats the same pattern at fixed time intervals.

The numerical solution and analysis of ODEs is straightforward, but the underlying assumptions have unrealistic consequences. Because ODE models assume that the number of people in each compartment is a continuous function of time, rather than leaving or entering a compartment at a single point in time, people will gradually move in and out of compartments. Therefore, there will almost always be fractional numbers of people in each compartment. This is particularly problematic when diseases are dying out. Rather than being eradicated at a single point in time when the last infected person recovers, the number of people infected will decrease from one through ever smaller fractions of a person, only reaching zero asymptotically. Furthermore, since ODEs are deterministic,

these equations have only one solution and predict only a single epidemic curve for each initial condition. However, people and diseases are not deterministic – the time taken to recover from a disease or the number of onward transmission events cannot be known ahead of time and will be different from person to person, epidemic to epidemic and season to season.

Stochastic compartmental models address the key shortcomings of ODE models at the cost of computational and analytic complexity [72]. In the most common formulation, people move in and out of compartments at random points in time such that the whole population is represented by a continuous time Markov chain (CTMC) [74]. For each kind of event in a CTMC model (*e.g.* births or infections), the number of events or transitions that occur in a given time interval is governed by a time-heterogeneous Poisson process. The rates of these Poisson process are functions of the (fixed) parameters and the (variable) number of people in each compartment with exactly the same form as the equivalent transition rates in the associated ODE model (Equation (2.1)). These compartmental CTMC models can be represented by systems of stochastic integral equations; however, this representation is cumbersome so has typically only been used to explore the general properties of this class of compartmental model (*e.g.* [76, 77]). Instead it is common to represent CTMC models by tabulating all possible model events with the associated state transitions and probabilities (*e.g.* Table 2.1).

Event type	Transition			Probability
Infection	S_H	$\to$	I	$\beta_H S_H I/N\ \Delta t + o(\Delta t)$
	S_L	$\to$	I	$\beta_L S_L I/N\ \Delta t + o(\Delta t)$
Recovery	I	$\to$	R	$\gamma I\ \Delta t + o(\Delta t)$
Susceptibility reduction	S_H	$\to$	S_L	$\sigma S_H\ \Delta t + o(\Delta t)$
Birth		$\to$	S_H	$\alpha N\ \Delta t + o(\Delta t)$
Death	S_H	$\to$		$\mu S_H\ \Delta t + o(\Delta t)$
	S_L	$\to$		$\mu S_L\ \Delta t + o(\Delta t)$
	I	$\to$		$\mu I\ \Delta t + o(\Delta t)$
	R	$\to$		$\mu R\ \Delta t + o(\Delta t)$

Table 2.1: Example CTMC model, listing all possible events with the associated state transitions and the probability that the event will occur between t and $t + \Delta t$ given the number of people in each compartment at time t (*i.e.* $S_H(t)$, $S_L(t)$, $I(t)$ and $R(t)$). An event with state transition $A \to B$ removes exactly one individual from compartment A and adds exactly one individual to compartment B. State transition $A \to$ removes an individual from compartment A and state transition $\to B$ adds an individual to compartment B.

For sufficiently large numbers of people a stochastic compartmental model and a system of ODEs (based on the same model structure, parameters and initial conditions) will predict approximately the same epidemic curves, with the approximation becoming exact in the limit of infinite population size [76]. However, stochastic compartmental models are able to capture the dynamics of a disease even when there are small numbers of people in any of the compartments, such as when a disease is emerging, dying out or simply if the

population being studied is small. As with ODEs, there are rarely ever simple analytic solutions for CTMC models. Because the time of each transition event is random there is not one but an infinite number of possible epidemic curves and rather than reaching an equilibrium point (or stable seasonal pattern), these models reach an equilibrium distribution that fluctuates around the mean number of people in each compartment. If the population size is fixed or bounded it is in theory possible to calculate the full distribution of the stochastic process exactly. Similar methods can be used to approximate the full distribution for models with unbounded population sizes [78]. However, both the exact and the approximate methods involve calculating the exponentials of very large matrices so are often impractical for even relatively simple models and small populations. Instead, it is common to use Monte Carlo simulation techniques such as the Gillespie's exact stochastic simulation algorithm [79] or approximate tau-leaping algorithm [80] to generate samples of possible epidemic curves. Many repeated simulations of the same system are required to sample the range of possibilities, so working with CTMCs can be computationally expensive. However, if the size of the simulated population or the simulated period of time is small, simulations can be very rapid.

However some of the assumptions underlying stochastic compartmental models are unrealistic. Because transitions are modelled with Poisson processes, the amount of time an individual spends in each compartment is exponentially distributed. This is an unrealistic distribution for many processes such as the length of infectious period, length of immunity or length of life. It is possible to subdivide each compartment so that the total time spent in all of the subdivisions has a more general and realistic gamma distribution [81], but this greatly increases the complexity of these models especially if more than one process is given a more general distribution. Moreover both ODE and stochastic compartmental models assume that all people in each compartment are identical and indistinguishable. Therefore to take into account the differences in risks associated with different characteristics and exposures (*e.g.* age, sex, smoking status, underlying conditions, medication history), every characteristic or exposure group must be modelled with its own compartment (*e.g.* S_H *vs.* S_L). This can lead to a model with very many compartments especially if different types of exposures and characteristics interact or compound. Finally, CTMC and ODE models only track the number of people in each compartment and therefore the histories of individuals are lost.

One way to address the shortcomings of stochastic compartmental and ODE models is to use an agent-based model (ABM) [82]. Rather than modelling only the numbers of individuals in each class, an ABM models an ensemble of agents or individuals moving through the compartments in the model. The transition rates between states for each agent can be modelled with any desired distribution. This makes ABMs extremely flexible and capable of handling a large degree of complexity. However ABMs that incorporate this additional complexity require more data to calibrate or estimate model parameters and simulating every individual requires even more computational power than stochastic compartmental models, especially in large populations.

Another way to simulate individual patient histories is to approximate the behaviour of individuals using a simple Markov chain model. This framework or perspective is complementary to the ODE and stochastic compartmental models and therefore shares its other limitations, however is much simpler than an ABM. In the ODE, stochastic compartmental or ABM frameworks, nearly all transitions rates leaving compartments are equal to the number people in the class that is being left multiplied by a fixed rate parameter. Often only the infection rate is different: proportional to the number of people susceptible (*i.e.* the people that could become infected) *and* the proportion of the population that is infected. If the population can be assumed to be at equilibrium, then the proportion of people that are infected at any point in time (I^*/N^*) is fixed and can be treated like the fixed parameters. The equilibrium proportion infected can be approximated by finding the equilibrium points of the ODE model (i.e the solution(s) of $A(\mathbf{X})\mathbf{X} = 0$) or taking an average over the long term behaviour of stochastic compartmental model simulations. Under this approximation *all* transition rates depend only on the state of the individual at the time but not their history or changes affecting other individuals. Therefore the behaviour of each individual can be approximated as a continuous time Markov chain with a state space consisting of the model compartments and an additional state for death, while the whole population can be modelled as an ensemble of independent individuals. The transition rates for these Markov chains are determined entirely by the constant parameters and constant equilibrium proportion infected. For the simple example model, if $\mathbf{p}(t)$ is the vector of probabilities that an individual is in the living states S_H, S_L, I and R at time t, then $\mathbf{p}(t)$ satisfies a set of differential equations very similar to those above:

$$\mathbf{p}'(t) = \begin{bmatrix} -(\sigma + \beta_H I^*/N^* + \mu) & 0 & 0 & 0 \\ \sigma & -(\beta_L I^*/N^* + \mu) & 0 & 0 \\ \beta_H I^*/N^* & \beta_L I^*/N^* & -(\gamma + \mu) & 0 \\ 0 & 0 & \gamma & -\mu \end{bmatrix} \mathbf{p}(t) \tag{2.5}$$

or

$$\mathbf{p}'(t) = Q\mathbf{p}(t), \tag{2.6}$$

The probability that the individual is dead, $m(t)$, satisfies

$$m(t) = 1 - \mathbf{1}^T\mathbf{p}(t) \tag{2.7}$$

and

$$m'(t) = -\mathbf{1}^T Q\mathbf{p}(t). \tag{2.8}$$

Note that because the model individual has already been born, the matrix Q lacks the birth rate terms (α), but is otherwise identical to $A(\mathbf{X}^*)$. Because Q is a constant matrix, unlike other models, there are simple explicit solutions for these equations. Given the probabilities that an individual is in each state at time s, $\mathbf{p}(s)$, the probability that the

individual is in each state at any future time t is

$$\mathbf{p}(t) = e^{Q(t-s)}\mathbf{p}(s). \tag{2.9}$$

This simple solution and other well known results from Markov chain theory allow us to calculate the range of possible behaviour of an individual over its lifetime very quickly without an ABM. I briefly illustrate a few examples using the above example model.

We can calculate the probability that an individual is in a given state at age a. If the individual is born at time 0, then $\mathbf{p}(0) = [1, 0, 0, 0]^T$ since in the example model all people are born into S_H. Then the probability that an individual is in a each state at age a is given by the vector $\mathbf{p}(a)$

$$\mathbf{p}(a) = e^{Qa} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{2.10}$$

We can calculate the probability that someone has died by age a by substituting the above into Equation (2.7) to yield

$$m(a) = 1 - \mathbf{1}^T e^{Qa} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{2.11}$$

One can also calculate the probability that an individual is in a each state at age a *given* that they are still alive (*i.e.* $\frac{\mathbf{p}(a)}{m(a)}$) by combining the above statements. The mean length of time an individual will spend each state before dying is given by

$$\int_0^\infty \mathbf{p}(a)da = \int_0^\infty e^{Qa}\mathbf{p}(0)da = -Q^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{2.12}$$

and consequently the mean life expectancy at birth is

$$\int_0^\infty 1 - m(a)da = \int_0^\infty \mathbf{1}^T\mathbf{p}(a)da = -\mathbf{1}^T Q^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{2.13}$$

It is relatively easy to evaluate the above matrix exponentials and matrix inverses using symbolic computation software, but the results – even for the simple example model – are unwieldy and don't provide clear insights into the model. For this reason, the individual-level outcomes for the models presented in this thesis (which are more complicated than the example model considered above) are calculated and reported in numeric form only.

The methods for calculating the more complex, individual-level outcomes for these models used in this thesis are described as needed in the chapters in which they are used. One can generalise the above examples to calculate the probability of entering a state or set of states and how long an individual spends in each state [83]. This theory is sketched at Chapter 3 in which it is used to calculate recurrence.

ABMs have been used extensively in the *C. difficile* modelling literature (which is reviewed in the next section). However, the models developed for this thesis use a combination of ODE and compartmental stochastic models, relying on the individual Markov chain perspective to explore the properties of individuals.

The *reproduction number*, often denoted by R, is an important concept in infectious disease epidemiology. The reproduction number of a disease in a population is often defined as the average number of secondary cases arising from a typical primary case in the population. Generally speaking, the incidence of a disease is increasing over time if the reproduction number exceeds one, but is decreasing over time if the reproduction number is less than one. In a population at or near equilibrium, the reproduction number is approximately one.

The *basic reproduction number*, often denoted by R_0, is the reproduction number in an almost entirely susceptible population, and describes the growth of a disease that has just been introduced into a population (or in a population with very low prevalence of infection and immunity). In general, if the basic reproduction number is less than one, a newly introduced disease cannot persist in the population without constant reintroductions. If the basic reproduction number is greater than one, introduction of the disease will probably lead to an epidemic and/or eventual endemicity of the disease, unless by chance the first few cases do not infect others. This makes the basic reproduction number a threshold parameter for disease extinction or persistence in a population [84] (though this property can be blurred in some cases by backward bifurcations [86, 87]). The term is sometimes reserved for the reproduction number in situations without any intervention in place to control the disease (*e.g.* [85]). However, since diseases nearly always elicit some response or change in behaviour in the patient and the people around the patient, this thesis is concerned with the reproduction number one would expect in an entirely susceptible population with current or typical responses and interventions. To avoid confusion between the different definitions of the basic reproduction number this thesis generally avoids the term, instead simply using the phrase *the reproduction number*, accompanied by a definition or a method of calculation.

The numerical value of the basic reproduction number very much depends on how the terms *average* and *typical* are interpreted in its definition; under some interpretations and models the basic reproduction number is not guaranteed to be a threshold parameter. The next-generation matrix (NGM) method is a general method for calculating the basic reproduction number in compartmental disease models. The basic reproduction number in the NGM sense is always a threshold parameter for establishment of endemicity following the introduction of a small number of infective individuals and is also the threshold for disease

extinction for isolated populations [84], unless the model displays a backward bifurcation [86, 87]. Because I wish to identify human populations that are able sustain the transmission of *C. difficile* without exposure to or introduction from an external source, this thesis is concerned with reproduction numbers as threshold parameters in a population (or subpopulation). Therefore this thesis uses the NGM definition for the basic reproduction number unless stated otherwise.

For the example model above, the basic reproduction number in the NGM sense is

$$R_0 = \frac{\beta_H \mu + \beta_L \sigma}{\gamma(\mu + \sigma)} \tag{2.14}$$

if we assume that births balance deaths in the population (*i.e.* $\alpha = \mu$). The analytic form of the reproduction number here provides insight into disease dynamics. For instance, R_0 will decreases if σ increases but other parameters stay the same *i.e.* if the process which reduces susceptibility (*e.g.* prophylactic treatment) occurs more rapidly the basic reproduction number will decrease, potentially preventing outbreaks. However, for more complicated models, the analytic form becomes intractable, or too complex to be useful in generating useful insights. For this reason, the reproduction numbers in Chapters 3 to 6 are calculated and presented in numeric form only.

Type reproduction numbers are complementary measures to the basic reproduction number for populations and diseases where there is more than one type of host (*e.g.* adults *vs.* infants or humans *vs.* animals) [88, 89]. The type reproduction number is the number of secondary cases of a type that arises from a typical case of the same type either through direct transmission from the index case or a chain of indirect transmission through other types. The type reproduction number shares the threshold property of the basic reproduction number and provides a way to determine whether a given host type (or group of host types) constitutes a reservoir that sustains the disease [88, 89]. However, the type reproduction number can only be calculated for host types that are explicitly included in the model, so cannot be used if there is insufficient information to explicitly model all hosts (as is the case for *C. difficile* in animals). Furthermore while type reproduction numbers can be used to determine whether different host types constitute reservoirs (*e.g.* patients colonised in hospitals *vs.* patients colonised in the community) they cannot be used to determine whether transmission can be sustained locally in a setting that is connected with other settings by the exchange of host individuals (*e.g.* hospital and the broader community). Therefore this thesis uses and introduces other minor variations of the basic reproduction number that can be used in these scenarios.

2.2.2 Review of *C. difficile* modelling literature

To my knowledge, 18 works on mathematical models of *C. difficile* epidemiology have been published to date (not including those I have authored), most of these in the last five years. The two earliest [90, 91] are short pieces which provide outlines of proposed models and argue the importance of further work. Many of the journal articles published since

then have proposed their own novel mathematical model or extended an existing model to account for greater complexity or a new setting. I review these articles considering the different mathematical modelling frameworks used; the aspects of the disease captured by these models; and the results, insights and recommendations these models have generated. I then detail the gaps remaining in the mathematical modelling literature and set out how this thesis aims to address some of these gaps.

Modelling framework and setting

Almost all existing mathematical models of *C. difficile* consider only the healthcare setting. Durham *et al.* modelled CDI across a population with separate sub-populations for hospitals, aged-care facilities and the broader community [92]. van Kleef *et al.* focused on patients in an intensive care unit but modelled their movements to and from other wards, long-term care facilities, and the community [93]. Only Yakob *et al.* have modelled *C. difficile* transmission in a general setting [94]. Most of the remaining, exclusively hospital-based models consist of a single, well-mixed population; however, some include multiple wards or rooms within wards, with different modes or rates of intra- and inter-ward transmission [95–97]. Since hospital populations are small, most authors have used stochastic compartmental models [92, 94, 95, 98–102] or ABMs [93, 96, 97, 103–105] though at least one article used an ODE model [106].

Transmission routes

Most authors have not explicitly modelled the mechanism of transmission. Instead they have assumed that – whatever the mechanism – the frequency of transmission is proportional to the number of infectious individuals in the population or subpopulation of interest. However, a number of authors have explicitly modelled environmental contamination [104, 105] or incorporated healthcare workers [96, 97, 99, 105] or visitors [97, 105] as transmission vectors. There are few published studies modelling transmission of *C. difficile* via food or from animals. Durham *et al.* estimated the force of colonisation in the community, giving this as an upper bound for transmission from food in the community [92]. Kwon *et al.* used extensive sampling of hospital food and a compartmental model to estimate the number of colonisations per 1,000 hospital admissions due to contaminated food [107], finding limited evidence for food-borne transmission in the study hospital. However, the extent of transmission from animal reservoirs in the community has not been estimated and the consequences for disease control have not been explored. The role of infants in transmission has also not been captured in any of the articles reviewed.

Gut flora and asymptomatic carriage

Authors have captured the interaction of *C. difficile* with gut flora in different ways. Some have assumed that antibiotic-induced gut flora disruption is required to become colonised with *C. difficile* [95, 98, 103, 104, 106, 107]. Others have assumed that all patients are

equally susceptible to colonisation but that patients with disrupted gut flora are more likely to develop symptomatic CDI and do so more rapidly [92, 99–101]. Rubin *et al.* assumed past and present antibiotic exposure increased susceptibility to colonisation, the rate of development of symptoms and the spore shedding rate (infectiousness) [96]. Some have modelled the gradual recovery of disrupted gut flora and the associated restoration of protection against infection [98, 100, 101, 103, 104, 106, 107]. A few authors have not included the effects of antibiotic exposure or disrupted gut flora in their models [94, 97].

Authors have modelled asymptomatic carriage in many ways. Some have assumed asymptomatic colonisation with toxigenic strains of *C. difficile* can only be followed by development of symptoms, discharge from hospital (leaving the scope of these models) or death [95, 96, 99, 100, 102]. In these models, asymptomatic carriage of toxigenic strains is a precursor to symptomatic disease. In other models, asymptomatically colonised patients can either develop symptomatic disease or clear the colonisation returning to a susceptible state [94, 97, 101, 105]. In the original model by Lanzas *et al.* and later variants, patients remain asymptomatic if they have an immune response, but immune patients cannot clear colonisation [98, 103, 104, 106, 107]. Rubin *et al.* allowed for asymptomatic carriage of non-toxigenic *C. difficile* that cannot be cleared or progress to symptomatic disease but prevents infection by toxigenic strains [96].

Immune responses

An early model by Lanzas *et al.* [98] and subsequent adaptations of this model [103, 104, 106, 107] assume all susceptible patients who become infected have a chance of mounting an immune response that prevents the development of symptomatic disease. Patients who recover from CDI in these models return to the susceptible state and thus may be subsequently recolonised, with another chance of mounting an immune response. In these models a susceptible patient's history of prior infection neither protects nor predisposes patients to subsequent reinfection. Codella *et al.* [97] and Barker *et al.* [105] assumed that patients under 65 and a portion of patients who have recovered from CDI are not susceptible to colonisation, implying some form of immunity. The simplifying assumptions in all these models fail to capture certain aspects of the disease. Seropositivity for *C. difficile* toxin antibodies (which implies prior exposure) is protective against future infection [15, 108], so – contrary to the assumptions of some authors – a patient's history of infection influences their risk of future infection. Similarly, an immune response to *C. difficile* toxins does not prevent colonisation [15] and people of all ages can become colonised and infected [13, 109].

Multiple strains

Most authors have not distinguished the many strains of *C. difficile* in their models. Lanzas *et al.* [103] and Yakob *et al.* [94] each considered two groups of strains: 'epidemic' ribotype 027 strain and other, lower virulence strains. Yakob *et al.* modelled different mechanisms of competition between the two groups of strains, exploring which could ac-

count for the dominance of the epidemic strain in some settings. Rubin *et al.* modelled a range of toxigenic and non-toxigenic strains with different antibiotic susceptibilities, capturing strain interactions by assuming that colonisation and infection with one strain prevents colonisation and infection from other strains [96].

Model Outcomes

The most common purpose of the models reviewed here has been to assess or predict the efficacy of hospital-based interventions for preventing *C. difficile* infection and colonisation. Though they did not explicitly model interventions, Starr *et al.* suggested that reducing the susceptibility of patients to colonisation with *C. difficile* (by reducing antibiotic prescription rates) is more effective than comparable reductions in all forms of within-hospital transmission (environment and person to person) [95]. Lanzas *et al.* evaluated screening hospital admissions to detect and isolate asymptomatic carriers using tests with a range of sensitivities and turnaround times. They assumed that identified *C. difficile* carriers were treated with contact precautions that reduced transmission from these patients by 75% and estimated this intervention would reduce the number of hospital acquired colonisations by 40-52% depending on the sensitivity and speed of the screening test [103]. Grigoras *et al.* modelled a similar intervention, predicting smaller reductions in colonisations (36%), but estimated that the addition of an antimicrobial stewardship program to this intervention would further reduce colonisations (total reduction: 56.6%). Lofgren *et al.* assessed the efficacy of routine bolstering of gut flora with faecal transplants, finding that faecal transplants for patients who had received high-risk medications such as antibiotics and PPIs would decrease CDI incidence, while faecal transplants for those recovering from CDI would reduce recurrent cases of CDI [99].

Some authors modelled the effect of bundles of control measures implemented in hospitals [96, 97, 101, 102, 104]. Rubin *et al.* considered the effects of improved hand hygiene amongst healthcare workers, improved adherence to contact precautions for *C. difficile* patients, improved environmental cleaning, and faster testing for CDI [96]. While they found that the combination of all measures had the greatest effect, the single most effective intervention was increased general hand hygiene. Yakob *et al.* considered the effect of four types of interventions: increasing the recovery rate of gut flora in patients who have received antibiotics using probiotics; reduction of antibiotic prescription; reducing person-to-person transmission by improved hygiene; and decreasing the average length of stay [101]. They found that reducing transmission was much more effective than other control measures in reducing CDI incidence, while antimicrobial stewardship and reducing transmission were most effective at reducing new colonisations in hospital. Rather than model the effect of improved intervention strategies, Codella *et al.* estimated the efficacy of four commonly used interventions implemented singly or as a bundle. They found that the combination of all interventions was better than each individually but found that environmental cleaning and treatment of symptomatic patients resulted in the largest improvements when implemented singly [97]. Barker *et al.* expanded on this model to analyse the potential

effectiveness of additional interventions, such as screening for asymptomatic carriage on admission, improved patient hand hygiene, and improved environmental cleaning, finding that interventions that prevented transmission from the large number of patients colonised at admission were better than improved contact precautions for symptomatic inpatients [105].

Vaccination has received little attention in the *C. difficile* modelling literature. van Kleef *et al.* used their model of an intensive care unit to identify the subset of patients that, if vaccinated, would prevent the most number of cases with the fewest number of vaccinations [93]. They found that vaccinating elective surgery patients prevented the most number of cases, but required many vaccinations. Since 70% of infections were acquired outside the ICU in their model, the most efficient strategy was to vaccinate patients with a high risk of colonisation at admission, such as residents of long-term care facilities. Stephenson *et al.* used their model to determine optimal vaccination strategies to reduce infection and colonisation while minimising vaccine-associated costs [106].

Yakob *et al.* used their multi-strain model to understand the rapid emergence and dominance of epidemic strains of *C. difficile*, considering three possible mechanisms: higher infectiousness of epidemic strains, higher likelihood of patients colonised with epidemic strains developing symptomatic disease, and a within-gut competitive advantage of epidemic strains against non-epidemic strains. They concluded that all three mechanisms would allow epidemic strains to replace endemic strains, but that only the first two mechanisms could do so as rapidly as has been observed in North America and Europe [94].

Even though most models have focused on healthcare settings, many have demonstrated the importance of importation of symptomatic and asymptomatic carriers of *C. difficile* from the community [92, 93, 95, 100, 101, 103, 105]. Starr *et al.* found that a simulated 50% reduction of within hospital transmission only reduced CDI cases by 15% suggesting that many (if not most) cases are imported from outside the hospital [95]. Yakob *et al.* quantified the effect of exposed but uncolonised patients entering the hospital, demonstrating their importance for hospital-onset infections [100]. Lanzas *et al.* found that the basic reproduction number for *C. difficile* in hospital settings was less than one for nearly half of all plausible parameter values, indicating that *C. difficile* is often not self-sustaining in hospitals but sustained by importation [98]. The only article to explicitly model transmission in the hospital and broader community found that reducing transmission in the community could lead to significant reductions in both community-onset and hospital-onset CDIs [92].

Gaps in the modelling literature

The *C. difficile* mathematical modelling literature summarised above is relatively limited and leaves important questions unanswered. Key areas for improvement are: better models of immunity and asymptomatic carriage; understanding the interplay between hospital and community; modelling transmission in the community and its effect on hospitals; incorporating transmission from known, but neglected reservoirs of the pathogen – particularly

infants, and animals; and capturing the mechanisms that lead to seasonality.

The existing mathematical modelling literature has simplified models of the interactions of *C. difficile*, gut-flora, and immune responses and therefore does not capture the essential complexity of CDI and asymptomatic carriage. For instance as highlighted above, many models assume that asymptomatic carriage is only a precursor to infection, while others omit the role of immune responses. In Chapter 3 I argue that it is possible to capture the essential complexity of *C. difficile* parsimoniously by modelling the interaction of three factors: *C. difficile* status, gut-flora health, and immune responses to *C. difficile* toxins. The compartmental structure developed in Chapter 3 and used throughout the thesis, models the many possible combinations of these factors or dimensions to produce the phenomena of asymptomatic colonisation, infection and recurrence. This bottom-up, multi-dimensional approach differs from many existing models that only consider some combinations of these factors or consist of a loose organisation of compartments and transitions designed to suit the available data or the specific interventions being studied (*e.g.* [97, 98, 105, 106]). Some authors have used a multi-dimensional model structure, but these have only considered the interaction of *C. difficile* and gut flora, omitting immune responses [92, 100, 101].

Even though many of the mathematical models reviewed above have highlighted the importance of community-acquired colonisations and infections in hospitals, the interaction between hospitals and communities has not been extensively studied. While Lanzas *et al.* found that the reproduction number may often be less than one in hospitals [98], this at first seems to be at odds with other models [100, 101] and empirical studies [13] that report higher prevalence of colonisation among discharged patients than among admitted patients or the general population. In Chapter 3 I demonstrate that both can be (and probably usually are) true of the same hospital.

If it is the case that the reproduction number for within-hospital transmission is less than one and disease in hospitals is sustained by admissions of colonised patients, this raises a question: what sustains the continual inflow of colonised admissions from the community? It could be that patients discharged from hospitals sustain the presence of *C. difficile* in the community with minimal community-based transmission, or it could be that extensive transmission in the community maintains a pool of colonised patients. Though Durham *et al.* found that small reductions of transmission in the community would reduce the incidence of hospital-onset CDIs, they did not calculate reproduction numbers for the hospital or community and did not explore whether larger interventions in one setting could disrupt transmission in one or both [92]. I address these questions in Chapter 5.

The potentially significant role of infants – who have very high colonisation rates and therefore constitute a large portion of colonised individuals outside hospitals – has not been quantified. I also explore this in Chapter 5. Similarly, transmission via food and animals has received very little attention in the modelling literature. In Chapters 5 and 7 I model person-to-person and animal-to-person transmission, providing conditions under which it would be reasonable to believe that transmission from animals sustains the pathogen in

the human population.

I have not identified any mathematical modelling articles that consider the seasonality of *C. difficile*. However, mathematical models are often used to study the seasonality of infectious diseases [110, 111] and could be used to explore which – if any – of the proposed mechanisms could account for the seasonal patterns of *C. difficile* infections.

A number of the models I have reviewed use national, regional or hospital-specific estimates of the incidence of hospital-acquired CDIs and/or community-acquired CDIs to fit or verify their models. These estimates are based on surveillance definitions that use arbitrary and un-validated cut-offs to distinguish hospital and community-acquired infections [23]. Though the authors of one mathematical modelling article have identified this issue [93], I have not found any mathematical models which assess these definitions or account for misclassification that may arise from these definitions.

This thesis begins to fill these key gaps in the modelling literature. I begin by introducing an improved model of *C. difficile* transmission in a hospital setting that is the basis of the next four chapters.

Chapter 3

Healthcare-Associated *Clostridium difficile* Infections are Sustained by Disease from the Community

3.1 Introduction

This chapter consists of an article published in the *Bulletin of Mathematical Biology* and the accompanying supplementary materials. In this article I introduce a novel compartmental model of *C. difficile* transmission that captures the interaction of three factors that lead to infection or asymptomatic carriage: *C. difficile* status, gut-flora health, and immune responses to *C. difficile* toxins. I use the framework to simulate transmission amongst hospitalised patients. I primarily use a stochastic compartmental framework but use the individual Markov chain approximation to calculate individual-level outcomes. The main outcome of this article is that, for most hospitals or hospital wards, there is enough transmission to create a net export of colonised individuals to the broader community but not enough within-hospital transmission to sustain the disease in the hospital without the regular admission of colonised patients. This is, to my knowledge, the first time the two seemingly contradictory halves of this result have appeared together and been reconciled in the modelling literature. The article uses extensive sensitivity analysis to determine the robustness of the main result and demonstrate how different factors differentially affect transmission, infection prevalence and recurrence. The supplementary materials provide additional information on the model structure, parameterisation and sensitivity analysis.

The model presented in this chapter is the basis of the analysis for the article in Chapter 4. The model framework is extended in Chapters 5 and 6 to capture the community, infants, animal reservoirs, and seasonality.

3.2 Article and supplementary material

Angus McLure, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. Healthcare-associated *Clostridium difficile* infections are sustained by disease from the community. *Bulletin of Mathematical Biology*, 79(10):2242-2257, 2017.

Bull Math Biol (2017) 79:2242–2257
DOI 10.1007/s11538-017-0328-8

Society for Mathematical Biology

ORIGINAL ARTICLE

Healthcare-Associated *Clostridium difficile* Infections are Sustained by Disease from the Community

Angus McLure[1] · **Archie C. A. Clements**[1] · **Martyn Kirk**[1] · **Kathryn Glass**[1]

Received: 22 February 2017 / Accepted: 20 July 2017 / Published online: 3 August 2017

Abstract *Clostridium difficile* infections (CDIs) are some of the most common hospital-associated infections worldwide. Approximately 5% of the general population is colonised with the pathogen, but most are protected from disease by normal intestinal flora or immune responses to toxins. We developed a stochastic compartmental model of CDI in hospitals that captures the condition of the host's gut flora and the role of adaptive immune responses. A novel, derivative-based method for sensitivity analysis of individual-level outcomes was developed and applied to the model. The model reproduced the observed incidence and recurrence rates for hospitals with high and moderate incidence of hospital-acquired CDI. In both scenarios, the reproduction number for within-hospital transmission was less than 1 (0.67 and 0.44, respectively), but the proportion colonised with *C. difficile* at discharge (7.3 and 6.1%, respectively) exceeded the proportion colonised at admission (5%). The transmission and prevalence of CDI were most sensitive to the average length of stay and the transmission rate of the pathogen. Recurrent infections were most strongly affected by the treatment success rate and the immune profile of patients. Transmission within hospitals is substantial and leads to a net export of colonised individuals to the broader community. However, within-hospital transmission alone is insufficient to sustain endemic conditions in hospitals without the constant importation of colonised individuals. Improved hygiene practices to reduce transmission from symptomatic and asymptomatic individuals and reduced length of stay are most likely to reduce within-hospital transmission and infections; however, these interventions are likely to

Electronic supplementary material The online version of this article (doi:10.1007/s11538-017-0328-8) contains supplementary material, which is available to authorized users.

Angus McLure
angus.mclure@anu.edu.au

[1] Research School of Population Health, Australian National University, Canberra, Australia

have a smaller effect on the probability of recurrence. Immunising inpatients against the toxins produced by *C. difficile* will reduce the incidence of CDI but may increase transmission.

Keywords *Clostridium difficile* · Mathematical model · Sensitivity analysis · Nosocomial infection

1 Introduction

Clostridium difficile is a spore-forming anaerobic bacteria that colonises the large intestine causing mild to severe diarrhoea and inflammation of the colon. Since the emergence of more virulent strains in the early 2000s, the burden of *C. difficile* has been increasing [6,23,38,49]. Although there are no global estimates, the CDC estimate that in 2011 there were 453,000 *C. difficile* infections (CDIs) and 29,300 associated deaths in the USA alone [33]. The primary healthcare costs of treating *C. difficile* in the USA are estimated to be up to USD 4.5 billion per annum [12] or USD 10,000 per infection [39]. The majority of reported CDIs are among hospitalised patients admitted for unrelated conditions [33]. Historically, most hospital-onset CDI has been attributed to transmission from other CDI patients in hospitals. More recent studies suggest that many hospital-onset infections are acquired prior to admission and that much of the transmission within hospitals is from asymptomatic carriers [35,48,52]. Up to 15% of healthy adults in the community and 4–29% of inpatients are colonised in developed countries [19], but most carriers are asymptomatic, making the pattern of transmission difficult to discern. Mathematical models of transmission attempt to reveal this unseen pattern by capturing the key biological mechanisms of the disease and reproducing the incidence observed in real communities. Specific biological mechanisms that play an important role in transmission of CDI are gut flora and immunity as they can prevent or delay the development of symptoms in asymptomatic carriers.

Natural gut flora compete with the pathogen preventing infection [5] but not colonisation [36]. Disruption of the gut flora allows the proliferation and overgrowth of vegetative *C. difficile* and increases spore shedding [9,17,18,30]. In one study, mice receiving antibiotics shed up to a million-fold more *C. difficile* spores than asymptomatic carriers who did not receive antibiotics [30]. Treatment of acute CDI is often with broad-spectrum antibiotics, which suppress *C. difficile* but also damage other gut flora, leaving patients susceptible to recurrent disease from the recrudescence of the original strain or a new infection [32]. Twenty to thirty percent of patients recovering from a primary CDI will have a recurrent infection [32,36].

The symptoms of CDI are not caused by colonisation directly but by the A, B [51] and binary [20] toxins produced by many strains of *C. difficile*. Even in patients with disrupted gut flora, adaptive immune responses to these toxins also protect against CDI. High levels of serum antibodies for toxins A and B are protective against initial [25,27] and recurrent [31] CDI, while the risk of infection increases with advancing age [36]. Passive immunisation with antibodies has been shown to prevent initial [3] and recurrent [37] CDI. Phase I and II trials for a *C. difficile* toxoid vaccine have been completed successfully [10,16], and phase III trials are under way.

A number of mathematical models of *C. difficile* transmission have been reported [4,7,13,28,29,34,46,50,54,55], but only Lanzas et al. [29] have incorporated immune responses in their models, and these immune responses were not adaptive. We present a mathematical model of *C. difficile* that incorporates adaptive immune responses to toxins and the role of antibiotic consumption in triggering spore shedding. It is crucial to understand the effect of these factors on asymptomatic carriage, recurrence and transmission. This requires a more complex model with more parameters. Systematic sensitivity analysis of stochastic disease models is rarely performed but is especially important when there are many parameters. We translate Anderson's coupled finite differences approach [1] to the epidemiological setting and propose a novel method to estimate the sensitivity of individual-level outcomes such as the probability of disease recurrence.

The main aim of this study was to make a mathematical model of *C. difficile* transmission that parsimoniously captures the role of adaptive immune responses, healthy and disturbed gut flora and asymptomatic colonisation. We aimed to reproduce the range and relative proportions of initial and recurrent CDIs and the incidence of hospital and community-acquired asymptomatic colonisation observed in developed countries. We then used the model to determine the key drivers of CDI in hospitals.

2 Methods

2.1 Model Structure

Individuals are distributed between compartments based on three attributes: immune status, *C. difficile* status and commensal gut flora status. There are three immune statuses: able to mount an effective immune response to *C. difficile* toxins conferring resistance or immunity to symptoms but not colonisation (R); naive to *C. difficile* toxins but with a healthy immune system (H); and unable to mount an effective immune response to *C. difficile* toxins because of a suppressed, locally dysfunctional or unhealthy immune system (U).

There are two possible commensal gut flora statuses: disrupted (denoted with a subscript 'a') and not disrupted (without additional subscripts). There are three possible *C. difficile* statuses: *C. difficile* overgrowth (subscript 'o'), colonised (subscript 'c'), and free of *C. difficile* (without additional subscripts). An individual may have almost any combination of these attributes; however, we assumed that *C. difficile* overgrowth can only occur in individuals with disturbed gut flora (Fig. 1).

If *C. difficile* overgrowth is accompanied by a robust immune response to toxins, the individual will not exhibit symptoms (R_{ao}), but otherwise individuals with overgrowth exhibit symptoms (H_{ao} and U_{ao}). All individuals with overgrowth shed significant numbers of spores and so are infectious. Spore shedding has been observed to increase before toxin production [17,18], but be subsequently reduced (but not eliminated) during *C. difficile* treatment [48]. Therefore, individuals with *C. difficile* colonisation and disrupted gut flora who are currently asymptomatic (H_{ac}, U_{ac} and R_{ac}) and individuals with overgrowth (H_{ao}, U_{ao} and R_{ao}) are equally infectious in our model. All other individuals are neither infectious nor symptomatic for CDI.

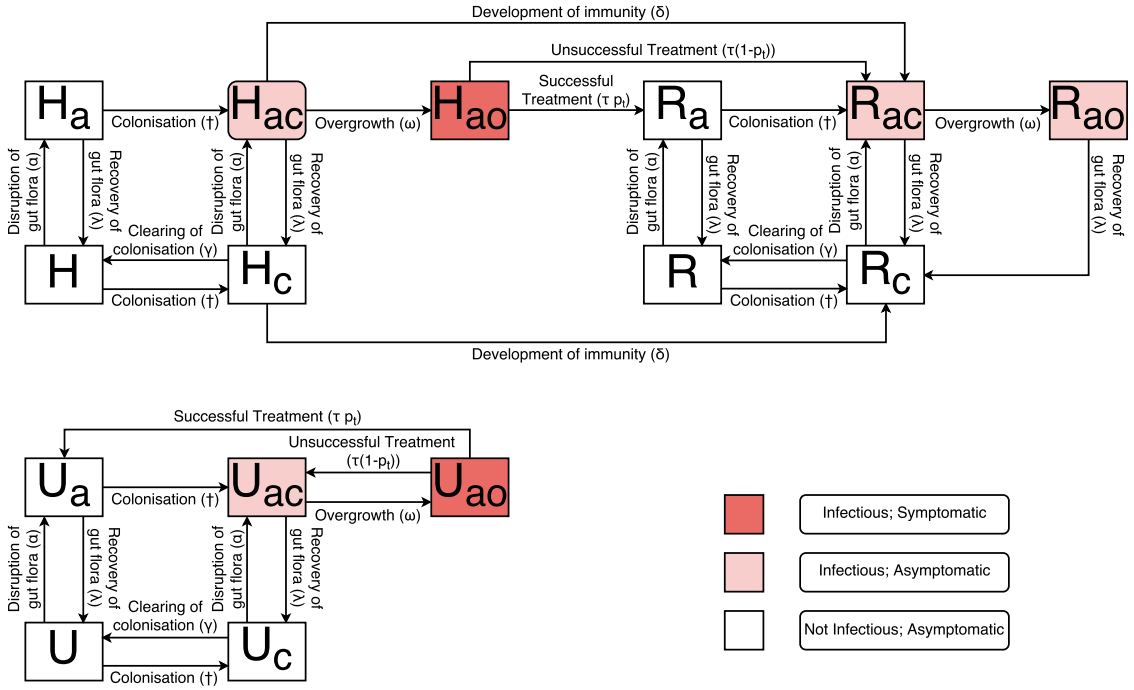


Fig. 1 Model diagram with boxes representing compartments. Arrows indicate possible transitions between compartments with rate parameter given by the associated Greek letter in parentheses. Admissions and discharges are not pictured. †The rate of colonisation—the force of colonisation—depends on the number of infectious individuals, the size of the hospital (N_{bed}), the transmission rate parameter (β) and the efficacy and coverage of contact precautions (q) as: $[(H_{ao} + U_{ao})q + R_{ao} + H_{ac} + U_{ac} + R_{ac}]\,\beta/N_{\text{bed}}$ (Color figure online)

2.2 Model Parameterisation

Parameters were chosen to reflect hospitals based on the available literature (Table 1). Loo et al. [36] reported an exceptionally complete picture of *C. difficile* in a hospital setting: they measured serum levels to toxin antibodies and colonisation status on admission and used extensive stool sampling to estimate the cumulative risk of asymptomatic colonisation and CDI post-admission. We fitted a number of key parameters to these detailed observations; however, as the study represents a high-incidence setting, we chose two parameterisations: a high hospital-acquired CDI incidence hospital, emulating the conditions reported by Loo et al. [36], and a moderate hospital-acquired CDI incidence setting. The moderate-incidence setting differs only in the parameters for which there is significant variation between Loo et al. and the broader literature (length of stay; proportion of patients immunocompromised; and proportion of patients positive for *C. difficile* toxin antibodies at admission).

2.3 Reproduction Number

The reproduction number ($\mathcal{R}$) was calculated using the next-generation method [11]. The two parameterisations of the model we used have no disease-free equilibria because new colonised patients are admitted constantly. Therefore, the reproduction number was calculated for each parameterisation assuming all admitted patients are

Table 1 Parameter definitions, values and references for these values. All rates are in units of per day

Parameter	Description	High incidence	Moderate incidence	References
Hospital protocol and environment				
β	Transmission rate coefficient	0.09[a]	0.09[a]	[36]
$1/\kappa$	Mean length of stay for patients without CDI (days)	10	6	[36,42]
α	Rate at which patients begin new antimicrobial treatment (proxy for rate of gut flora disruption)	0.24	0.24	[36,44]
q	Efficacy and coverage of special contact precautions (0: total prevention of transmission; 1: no reduction in transmission)	0.45	0.45	[23,40]
N_{bed}	Hospital capacity	100	100	
Treatment				
τ	Rate at which the treatment of patients with active CDI progresses towards resolution	0.1	0.1	[26,32]
p_t	Probability that CDI treatment will remove all *C. difficile*	0.55[a]	0.55[a]	[36]
μ	All cause mortality rate for patients with active CDI	0.0075	0.0075	[21,35]
Patient profile				
p_c	Proportion of admissions colonised with a toxigenic strain of *C. difficile*	0.05	0.05	[27,35,36]
p_a	Proportion of admissions with disrupted gut flora (due to recent antibiotic exposure)	0.14	0.14	[43]
p_o	Proportion of admissions with *C. difficile* overgrowth	0.006[a]	0.006[a]	[36]
p_u	Proportion of admissions unable to mount a sufficient immune response	0.55[a]	0.46	[2,36,41]
p_{AB}	Proportion of admissions positive for *C. difficile* toxin antibodies	0.2	0.7	[25,36]
Intestinal processes				
λ	Rate at which damaged commensal gut flora recovers	0.03	0.03	[8,45]
γ	Rate at which *C. difficile* is cleared by competition with gut commensal flora	0.023	0.023	[24]
ω	Rate at which *C. difficile* overgrowth occurs in colonised individuals with disrupted gut flora	0.2	0.2	[17,18]
δ	Rate of development of immunity	0.1	0.1	[10,16]

[a] These parameters are estimated by fitting the model to data reported by Loo et al. [36]

C. difficile negative (p_c, $p_o = 0$). As the model only captures transmission within the hospital, the reproduction number does not include transmission arising from primary cases after they have been discharged to the community.

For comparison, we calculated the reproduction number assuming that colonised patients are not discharged. This model has the same disease-free equilibrium as the main model, but the average infectious period in hospital is longer.

2.4 Patients Discharged with Asymptomatic Colonisation

To complement the reproduction numbers, we calculated the proportion of patients that are colonised at discharge for the high and moderate-incidence settings. We found that this proportion was sensitive to the average length of stay ($1/\kappa$) and the transmission parameter (β) and so also calculated this proportion for average lengths of stay between 4 and 30 days and transmission parameter values between 0 and 0.18 (double the estimated value).

2.5 Sensitivity Analysis Using Coupled Finite Differences

Anderson's coupled finite differences approach to sensitivity analysis [1] uses coupled pairs of model simulations where the parameter of interest differs by a small quantity. The mean difference in model outcomes between the paired simulations for many (psuedo)randomly generated pairs of simulations is a Monte Carlo estimate of the derivative of the model outcome with respect to the varied parameter. By coupling the pairs of simulations to move 'in step', it is possible to reduce the variance without introducing further bias (see [1] for further details).

Parameter sensitivity was estimated using the derivative of each outcome with respect to each parameter divided by the outcome multiplied by the parameter. For small changes in parameter value, the sensitivity was approximately the proportion change in outcome per proportion change in the parameter, allowing comparison of the sensitivities for different outcomes and parameters. We used this approach to estimate the sensitivity of hospital-level outcomes (CDI prevalence, prevalence of all infectious individuals and the force of colonisation).

2.6 Sensitivity Analysis for Individual-Level Outcomes

Many outcomes of interest, such as the probability of recurrence, depend on the histories of individuals. We developed a method for estimating individual-level outcomes and thus the sensitivity of these outcomes to parameter values. If the force of colonisation (FOC) is constant, then each individual in the hospital acts as an independent continuous-time Markov chain with transition rates described in Fig. 1. This is true in general for many SIR-type compartmental models where the only modelled interaction between individuals is transmission. Although the FOC is a non-constant stochastic process, for endemic diseases it fluctuates around an equilibrium value and the behaviour of each individual in the model can be approximated by an independent

Table 2 Comparison of reported (Loo et al. 2011) and simulated values of key outcomes in a high-incidence and moderate-incidence setting.

Outcome	High incidence Reported [95% CI]	Simulated	Moderate incidence Simulated
Incidence of HA colonisation[b]	57.6 [50.5, 65.4]	60.4	42.6
Symptomatic (CDI)[b]	28.1 [23.2, 33.7]	27.1	9.9
Asymptomatic[b]	29.5 [24.5, 35.2]	33.3	32.7
CDI < 72h after admission[b]	14.4 [11.0, 18.5]	11.5	11.9
Force of colonisation[a]	0.007 [0.004, 0.011]	0.007	0.005
Prob of ≥ 1 CDI recurrence	0.248 [0.170, 0.326]	0.199	0.180
Prob of ≥ 2 CDI recurrences	0.068 [0.022, 0.114]	0.058	0.041

[a] Incidence is per 10,000 patient-days
[b] Units of per day

Markov chain. Assuming a constant FOC, we used the theory of Markov chain sojourn times and hitting probabilities [47] to calculate the probability and the mean time to onset of recurrent CDI. We used a simple central difference scheme to estimate the parameter sensitivity of individual-level outcomes and the reproduction number.

3 Results

The model reproduced the incidence of hospital-acquired colonisation, hospital-acquired CDI and recurrence rates observed in both high- and moderate-incidence settings (Table 2). The moderate-incidence parameterisation—which had shorter lengths of stay, more patients with immunity to *C. difficile* toxins and fewer immunocompromised patients—had much lower CDI incidence (11.8 per 10,000 patient-days compared with 28.7 per 10,000 patient-days). Of this reduction in incidence about half was attributable to reduced length of stay and half to the changed immune profile of admitted patients: simulations where only the immune profile was altered reduced incidence to 19 per 10,000 patient-days. The incidence of asymptomatic colonisation and the probability of recurrence was similar in the high- and moderate-incidence settings.

3.1 Reproduction Number and Proportion Colonised at Discharge

The reproduction number was 0.67 in the high-incidence setting compared to 0.44 in the moderate-incidence setting. The lower value in the moderate-incidence setting was due to the difference in the length of stay; using the moderate-incidence parameterisation with mean length of stay equal to that of the high-incidence setting, the reproduction number was 0.69—slightly *larger* than the high-incidence parameterisation. Similarly, though the mean force of colonisation was lower in the moderate-incidence setting ($0.0047\,\text{day}^{-1}$) than in the high-incidence setting ($0.0070\,\text{day}^{-1}$), it was highest in a setting with long mean length of stay but many

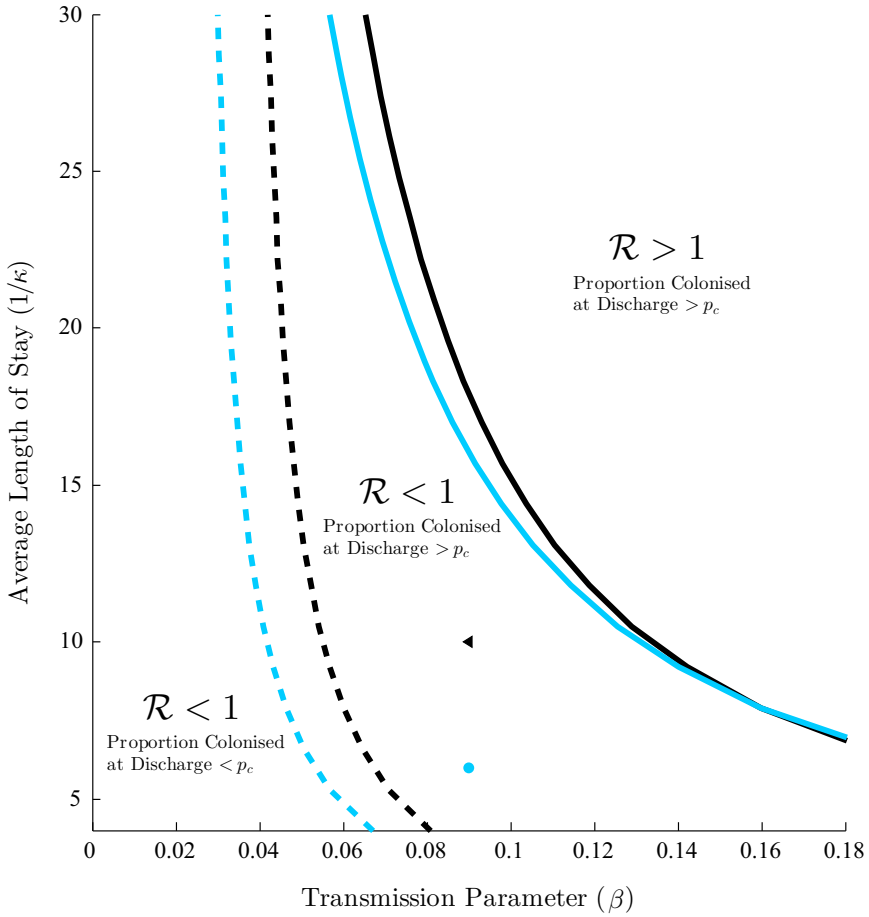


Fig. 2 Critical values of the transmission parameter (β) and mean LOS ($1/\kappa$) such that the reproduction number is 1 (*solid curves*) and such that that proportion colonised at admission (p_c) and at discharge is equal (*dashed curves*), where *black* shows the high-incidence setting, and *light blue* shows the moderate-incidence setting. The *triangle* and *circle* show the values of β and $1/\kappa$ for the high-incidence setting ($\mathcal{R} = 0.67$; proportion colonised at discharge $= 7.3\%$) and the moderate-incidence setting ($\mathcal{R} = 0.44$; proportion colonised at discharge $= 6.1\%$), respectively. In both settings, transmission is insufficient to sustain endemic disease in the absence of importation ($\mathcal{R} < 1$) but does result in a net 'export' of *C. difficile* colonised individuals (proportion colonised at discharge $> p_c = 5.0\%$) (Color figure online)

immune patients ($0.0073\,\text{day}^{-1}$). The reproduction numbers calculated assuming colonised patients were not discharged were 11.7 and 16.7 for the high- and moderate-incidence settings.

The reproduction number was most sensitive to the transmission parameter (β) and the average length of stay ($1/\kappa$). Figure 2 shows the combinations of these two parameters for which $\mathcal{R}$ is equal to one for the moderate- (light blue curves) and high- (black curves) incidence settings. For $\mathcal{R}$ to be greater than 1 under the moderate-incidence setting, the mean length of stay would have to be more than 15 days (double the OECD average [42]) or the transmission parameter would have to be more than double the estimated value. The reproduction number was not sensitive to the proportion of individuals with toxin antibodies or the proportion unable to mount an immune response (Fig. 3). Moreover, the difference in immune profile between the two parameterisations only had a small effect on the threshold values of transmission and length of stay (Fig. 2).

Simulations of the model found that 7.3 and 6.1% of discharged patients are colonised in the high- and moderate-incidence settings, respectively, compared to only 5.0% at admission in both settings. The reduction in the proportion colonised at discharge in the moderate-incidence setting was entirely due to the shorter mean length of stay. In a setting with long lengths of stay like the high-incidence setting, but more immune patients like the moderate-incidence setting, 8.4% of patients were colonised at discharge. The proportion colonised at discharge was found to be sensitive to the length of stay and the transmission parameter. Figure 2 presents threshold values of these quantities above which the proportion colonised at discharge exceeds the proportion colonised at admission. To reduce the proportion colonised at discharge to less than the proportion colonised at admission, within-hospital transmission would have to be reduced by $\geq$39% in the high-incidence setting or by $\geq$41% in the moderate-incidence setting. In a setting with long mean length of stay like the high-incidence setting but an improved immune profile like the moderate-incidence setting, the required reduction would be $\geq$54%.

3.2 Sensitivity Analysis

Figures 3, A.1 and A.2 present the sensitivity of model outcomes to parameter values using the high-incidence setting. Model parameters were divided into four groups: those related to hospital protocol and environment; those describing treatment; those describing the patients; and those describing intestinal processes, as defined in Table 1. The prevalence of CDI, the prevalence of asymptomatic carriage and the reproduction number were all most sensitive to the hospital protocol parameters (Fig. 3). Notably, all of these outcomes were much more sensitive to rate of transmission (β) than to the effectiveness and coverage of special contact precautions for symptomatic patients (q).

CDI prevalence was sensitive to treatment length (τ), but asymptomatic carriage and the reproduction number were not. This suggests that the decreased incidence associated with increased rate of treatment was due primarily to the reduction in the mean duration of infection rather than reduced transmission and incidence. The force of colonisation and $\mathcal{R}$ had a very similar pattern of sensitivity, but the force of colonisation was consistently more sensitive (Fig. A.2).

The probability of recurrence was sensitive to the probability that treatment will remove all *C. difficile* (p_t), average length of stay ($1/\kappa$), the proportion of admitted patients that are immunocompromised (p_u) and was somewhat sensitive to the rate at which *C. difficile* overgrowth develops (ω) and the transmission rate (β) (Fig. A.1). The model only captures recurrences that occur within a single hospitalisation, and thus, reducing the mean length of stay or the rate at which overgrowth develops increases the chance that patients are discharged before they have a recurrence. Increasing the rate of patient discharge also reduces the opportunity for reinfection in hospital. However, this effect is small because in our model most (93%—high incidence; 97%—moderate incidence) recurrent cases are due to the recrudescence of the initial infection not reinfection. Similarly recurrence was much less sensitive than prevalence or the reproduction number to the factors that affect transmission (β and

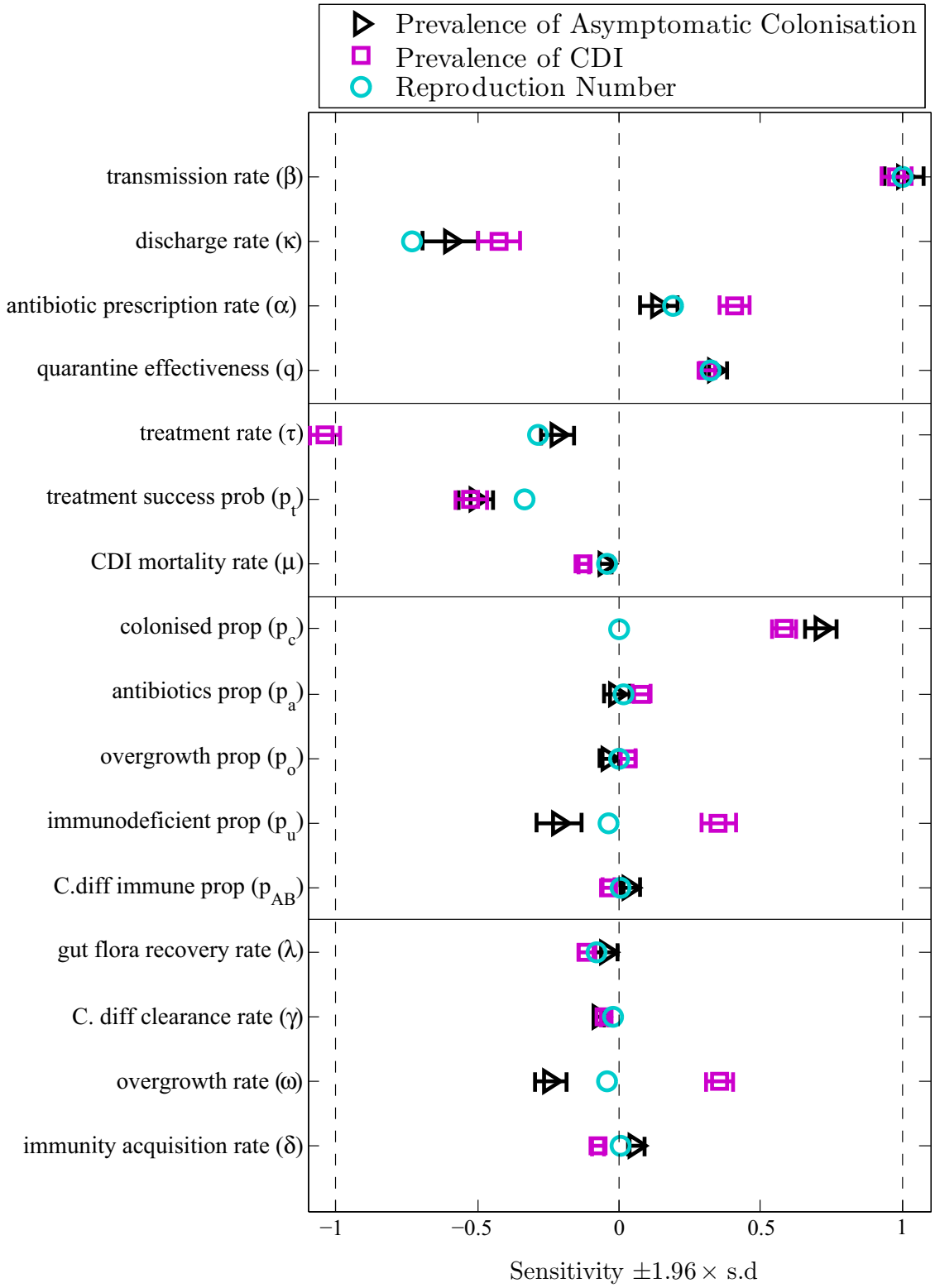


Fig. 3 Sensitivity of the prevalence of asymptomatic colonisation, the prevalence of CDI and the reproduction number model parameters. The bars are the approximate 95% confidence intervals for the Anderson's coupled finite differences estimates of sensitivity. The sensitivity of the reproduction number was calculated using the next-generation method and finite differences. Neither of these are Monte Carlo methods, and so there are no confidence intervals for these estimates

q), but was more sensitive to the probability that all *C. difficile* is removed by treatment (p_t).

All model outcomes were robust parameters for which there are limited data (the rate of acquisition of immunity (δ); the rate at which *C. difficile* colonisation is cleared (γ); and the rate of gut flora recovery (λ)). Although we acknowledge that the sensitivity analysis is local, it is reassuring that these parameters were not highly influential in our analysis.

4 Discussion

The reproduction number for within-hospital transmission is less than one for both the moderate- and high-incidence settings in our model. Since the high-incidence setting was calibrated to reports from the Quebec epidemic period, this suggests that even in epidemic conditions, within-hospital transmission is not sufficient to ensure CDI persists in the hospital without the admission of colonised patients from the broader community. The reproduction number is low because the typical length of stay is short compared to the typical duration of colonisation and infectiousness. The average length of stay would need to be over 15 days or the transmission rate more than doubled to increase the reproduction number above one. Nevertheless, we found that a greater proportion of patients are asymptomatically colonised at discharge than at admission. This finding is in agreement with studies that indicate that colonisation and infection are associated with recent hospitalisation [8,36] and highlights the fact that the reproduction number is calculated for transmission within the hospital only. Some patients colonised in hospital continue to transmit in the community, and so the total reproduction number is higher. If colonised patients spent the duration of their infectious period in hospital, the within-hospital reproduction number would be much larger, suggesting that management and prevention of outbreaks should include a focus on minimising average length of stay of hospital patients.

The immune responses in our model prevent disease but not colonisation or onward transmission. Therefore, colonised and immune individuals are asymptomatic and so are not treated with additional precautions to prevent transmission. Our model suggests that increasing only the proportion of patients that are immune would increase within-hospital transmission, increase the proportion of patients colonised at discharge and require greater reductions in transmission if fewer patients are to be colonised at discharge than at admission. Therefore, our model suggests that antitoxin vaccination programmes for hospitalised patients would reduce the incidence of CDI but increase transmission. This is at odds with a mathematical modelling study by Durham et al. that found that vaccination would reduce the incidence of CDI *and* transmission [13]. This contradictory result arose because, in contrast to our model, they did not explicitly model immunity, but instead assumed that all asymptomatic individuals—whether protected by immune responses to toxins or protected by commensal gut flora—are less infectious than those with symptoms.

Our novel method of sensitivity analysis allows the assessment of individual-level outcomes, such as recurrence, within the framework of compartmental models, rather than the framework of complex and computationally intensive individual-based mod-

els. However, by performing all calculations assuming the force of colonisation is fixed at its mean value, our method does not represent the full stochasticity of the model. More of the stochasticity could be accounted for by calculating individual-level outcomes for many values of the force of colonisation sampled from the equilibrium distribution of the force of colonisation.

The sensitivity analysis revealed that CDI recurrence and CDI prevalence are influenced by different factors. The probability of recurrent CDI was most sensitive to the probability that treatment successfully removes all *C. difficile* (the key factor for recrudescence) but much less sensitive to the rate of within-hospital transmission (the key factor for reinfection), because most recurrences in our model are due to recrudescence not reinfection, in agreement with recent studies [14,15]. The prevalence of CDI is most sensitive to hospital protocols, which suggests differences in hygiene and patient interactions, contact precautions for symptomatic individuals, mean length of stay and antibiotic stewardship may account for much of the difference in asymptomatic carriage and CDI between hospitals. All model outcomes are more sensitive to the transmission rate than to quarantine effectiveness. Therefore, interventions that reduce transmission from all carriers—such as increased use of soap and water [22] or extending special contact precautions to those with a recent history of CDI and other patients likely to be asymptomatic carriers [35]—are likely to be more effective than improving only the existing contact precautions for symptomatic patients.

We have assumed that all patients with CDI commence treatment and special contact precautions as soon as they start exhibiting symptoms. A delay in response could be captured by increasing the average time taken for patients to commence treatment. Our sensitivity analysis for this parameter suggests short delays in identification would not greatly increase total transmission.

Our model of *C. difficile* transmission is the first to include adaptive immune responses to toxins and to model the effect of gut flora disruption in triggering spore shedding. The balance between biological realism and parsimony gives the model the flexibility to simulate settings with high and moderate incidence of hospital-acquired CDI, with the moderate-incidence setting differing only in the average length of stay, and the age and immune profile of patients. By modelling the protective effect of intestinal flora, we can capture community-acquired CDI cases who are colonised at admission to hospital and develop CDI after use of antibiotics. By modelling immunity to toxins, we can capture colonised patients who remain asymptomatic despite the disruption of their gut flora. Models lacking one or both of these features have either considered only high-incidence settings [4,53] or used other means to account for the large proportion of colonised individuals that are asymptomatic. Durham et al. assumed that the progression to C. difficile overgrowth was very slow for the majority of colonised patients [13]. Others have limited the susceptible population by assuming disrupted gut flora is required for colonisation to occur [4,29].

Our findings are corroborated by other mathematical models. Yakob et al. found that halving either the length of stay or the transmission parameter halved the incidence of infections, with greater improvements with a combined strategy [54]. They also found that the proportion colonised at discharge exceeded the proportion colonised at admission, but by greater margins than in our study. Lanzas et al. [29] simulated their model for wide range of parameters and found that the reproduction number was as

low as 0.52 but >1 for just over half the parameterisations (median 1.04), with modest differences in transmission and length of stay accounting for 95% of the variation. By contrast, our model is the first to show that the reproduction number is less than 1 under nearly all reasonable conditions, including 'epidemic' conditions. Rubin et al. found that modest improvement in healthcare worker hand hygiene had a greater potential impact than any other intervention including improved quarantine measures for those with CDI or rapid detection and treatment of CDI [46].

Our model does not explicitly model CDI in the community, and so it does not capture readmissions for recurrent infections, or the mechanisms of transmission in the community. We have found that disease importation from the community drives CDI in the hospital. In future work, we will extend our model to the community to determine whether transmission within the hospital sustains endemic conditions in the community, or if the community has self-sustaining endemic disease.

Acknowledgements AM is supported by an Australian Government Research Training Program Scholarship. ACAC is funded by a National Health and Medical Council Research Fellowship (#1058878). We thank Luis Furuya-Kanamori for suggested reading and Thomas Riley for a valuable discussion.

References

1. Anderson DF (2012) An efficient finite difference method for parameter sensitivities of continuous time Markov chains. SIAM J Numer Anal 50(5):2237–2258. doi:10.1137/110849079
2. Australian Institute of Health and Welfare (2015) Admitted patient care 201314: Australian hospital statistics. Technical report, Australian Institute of Health and Welfare, Canberra
3. Babcock GJ, Broering TJ, Hernandez HJ, Mandell RB, Donahue K, Boatright N, Stack AM, Lowy I, Graziano R, Molrine D, Ambrosino DM, Thomas WD (2006) Human monoclonal antibodies directed against toxins A and B prevent *Clostridium difficile*-induced mortality in hamsters. Infect Immun 74(11):6339–6347. doi:10.1128/IAI.00982-06
4. Bintz J, Lenhart S, Lanzas C (2017) Antimicrobial Stewardship and environmental decontamination for the control of *Clostridium difficile* transmission in healthcare settings. Bull Math Biol 79(1):36–62. doi:10.1007/s11538-016-0224-7
5. Britton RA, Young VB (2012) Interaction between the intestinal microbiota and host in *Clostridium difficile* colonization resistance. Trends Microbiol 20(7):313–319. doi:10.1016/j.tim.2012.04.001
6. Burckhardt F, Friedrich A, Beier D, Eckmanns T (2008) *Clostridium difficile* surveillance trends, Saxony, Germany. Emerg Infect Dis 14(4):691–692. doi:10.3201/eid1404.071023
7. Codella J, Safdar N, Heffernan R, Alagoz O (2015) An agent-based simulation model for *Clostridium difficile* infection control. Med Decis Mak 35(2):211–229. doi:10.1177/0272989X14545788
8. Dial S, Kezouh A, Dascal A, Barkun A, Suissa S (2008) Patterns of antibiotic use and risk of hospital admission because of *Clostridium difficile* infection. Can Med Assoc J 179(8):767–772. doi:10.1503/cmaj.071812
9. Diekmann O, Heesterbeek JAP, Roberts MG (2010) The construction of next-generation matrices for compartmental epidemic models. J R Soc Interface 7(47):873–885. doi:10.1098/rsif.2009.0386
10. Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A, Viale A, Ubeda C, Xavier J, Pamer EG (2012) Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. Infect Immun 80(1):62–73. doi:10.1128/IAI.05496-11
11. de Bruyn G, Saleh J, Workman D, Pollak R, Elinoff V, Fraser NJ, Lefebvre G, Martens M, Mills RE, Nathan R, Trevino M, van Cleeff M, Foglia G, Ozol-Godfrey A, Patel DM, Pietrobon PJ, Gesser R (2016) Defining the optimal formulation and schedule of a candidate toxoid vaccine against *Clostridium difficile* infection: a randomized phase 2 clinical trial. Vaccine 34(19):2170–2178. doi:10.1016/j.vaccine.2016.03.028

12. Dubberke ER, Olsen MA (2012) Burden of *Clostridium difficile* on the healthcare system. Clin Infect Dis 55(suppl 2):S88–S92. doi:10.1093/cid/cis335
13. Durham DP, Olsen MA, Dubberke ER, Galvani AP, Townsend JP (2016) Quantifying transmission of *Clostridium difficile* within and outside healthcare settings. Emerg Infect Dis 22(4):608–616. doi:10.3201/eid2204.150455
14. Eyre DW, Walker AS, Griffiths D, Wilcox MH, Wyllie DH, Dingle KE, Crook DW, Peto TEA (2012) *Clostridium difficile* mixed infection and reinfection. J Clin Microbiol 50(1):142–4. doi:10.1128/JCM.05177-11
15. Figueroa I, Johnson S, Sambol SP, Goldstein EJC, Citron DM, Gerding DN (2012) Relapse versus reinfection: recurrent *Clostridium difficile* infection following treatment with fidaxomicin or vancomycin. Clin Infect Dis 55(suppl 2):S104–S109. doi:10.1093/cid/cis357
16. Foglia G, Shah S, Luxemburger C, Pietrobon PJF (2012) *Clostridium difficile*: development of a novel candidate vaccine. Vaccine 30(29):4307–4309. doi:10.1016/j.vaccine.2012.01.056
17. Freeman J (2003) Effects of cefotaxime and desacetylcefotaxime upon *Clostridium difficile* proliferation and toxin production in a triple-stage chemostat model of the human gut. J Antimicrob Chemother 52(1):96–102. doi:10.1093/jac/dkg267
18. Freeman J (2005) Comparison of the efficacy of ramoplanin and vancomycin in both in vitro and in vivo models of clindamycin-induced *Clostridium difficile* infection. J Antimicrob Chemother 56(4):717–725. doi:10.1093/jac/dki321
19. Furuya-Kanamori L, Marquess J, Yakob L, Riley TV, Paterson DL, Foster NF, Huber CA, Clements ACA (2015) Asymptomatic *Clostridium difficile* colonization: epidemiology and clinical implications. BMC Infect Dis 15(1):516. doi:10.1186/s12879-015-1258-4
20. Gerding DN, Johnson S, Rupnik M, Aktories K (2014) *Clostridium difficile* binary toxin CDT. Gut Microbes 5(1):15–27. doi:10.4161/gmic.26854
21. Hensgens MPM, Goorhuis A, Dekkers OM, van Benthem BHB, Kuijper EJ (2013) All-cause and disease-specific mortality in hospitalized patients with *Clostridium difficile* infection: a multicenter Cohort study. Clin Infect Dis 56(8):1108–1116. doi:10.1093/cid/cis1209
22. Jabbar U, Leischner J, Kasper D, Gerber R, Sambol SP, Parada JP, Johnson S, Gerding DN (2010) Effectiveness of alcohol-based hand rubs for removal of *Clostridium difficile* spores from hands. Infect Control Hosp Epidemiol 31(6):565–570. doi:10.1086/652772
23. Jarvis WR, Schlosser J, Jarvis AA, Chinn RY (2009) National point prevalence of *Clostridium difficile* in US health care facility inpatients, 2008. Am J Infect Control 37(4):263–270. doi:10.1016/j.ajic.2009.01.001
24. Johnson S, Homann SR, Bettin KM, Quick JN, Clabots CR, Peterson LR, Gerding DN (1992) Treatment of asymptomatic *Clostridium difficile* carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. Ann Intern Med 117(4):297–302
25. Kelly CP, Kyne L (2011) The host immune response to *Clostridium difficile*. J Med Microbiol 60(8):1070–1079. doi:10.1099/jmm.0.030015-0
26. Khan FY, Abu-Khattab M, Anand D, Baager K, Alaini A, Siddique MA, Mohamed SF, Ali MI, Al Bedawi MM, Naser MS (2012) Epidemiological features of *Clostridium difficile* infection among inpatients at Hamad General Hospital in the state of Qatar, 2006–2009. Travel Med Infect Dis 10(4):179–185. doi:10.1016/j.tmaid.2012.06.004
27. Kyne L, Warny M, Qamar A, Kelly CP (2000) Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. N Engl J Med 342(6):390–397. doi:10.1056/NEJM200002103420604
28. Lanzas C, Dubberke ER (2014) Effectiveness of screening hospital admissions to detect asymptomatic carriers of *Clostridium difficile*: a modeling evaluation. Infect Control 35(08):1043–1050. doi:10.1086/677162
29. Lanzas C, Dubberke ER, Lu Z, Reske KA, Gröhn YT (2011) Epidemiological model for *Clostridium difficile* transmission in healthcare settings. Infect Control Hosp Epidemiol 32(06):553–561. doi:10.1086/660013
30. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P, Raisen C, Mottram L, Fairweather NF, Wren BW, Parkhill J, Dougan G (2009) Antibiotic treatment of Clostridium difficile carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. Infect Immun 77(9):3661–3669. doi:10.1128/IAI.00558-09

31. Leav BA, Blair B, Leney M, Knauber M, Reilly C, Lowy I, Gerding DN, Kelly CP, Katchar K, Baxter R, Ambrosino D, Molrine D (2010) Serum anti-toxin B antibody correlates with protection from recurrent *Clostridium difficile* infection (CDI). Vaccine 28(4):965–969. doi:10.1016/j.vaccine.2009.10.144
32. Leffler DA, Lamont JT (2009) Treatment of *Clostridium difficile*-associated disease. Gastroenterology 136(6):1899–1912. doi:10.1053/j.gastro.2008.12.070
33. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE, Winston LG, Cohen JA, Limbago BM, Fridkin SK, Gerding DN, McDonald LC (2015) Burden of *Clostridium difficile* Infection in the United States. N Engl J Med 372(9):825–834. doi:10.1056/NEJMoa1408913
34. Lofgren ET, Moehring RW, Anderson DJ, Weber DJ, Fefferman NH (2014) A mathematical model to evaluate the routine use of fecal microbiota transplantation to prevent incident and recurrent *Clostridium difficile* infection. Infect Control Hosp Epidemiol 35(01):18–27. doi:10.1086/674394
35. Longtin Y, Paquet-Bolduc B, Gilca R, Garenc C, Fortin E, Longtin J, Trottier S, Gervais P, Roussy JF, Lévesque S, Ben-David D, Cloutier I, Loo VG (2016) Effect of detecting and isolating *Clostridium difficile* carriers at hospital admission on the incidence of *C. difficile* infections. JAMA Intern Med 176(6):796. doi:10.1001/jamainternmed.2016.0177
36. Loo VG, Bourgault AM, Poirier L, Lamothe F, Michaud S, Turgeon N, Toye B, Beaudoin A, Frost EH, Gilca R, Brassard P, Dendukuri N, Béliveau C, Oughton M, Brukner I, Dascal A (2011) Host and pathogen factors for *Clostridium difficile* infection and colonization. N Engl J Med 365(18):1693–1703. doi:10.1056/NEJMoa1012413
37. Lowy I, Molrine DC, Leav BA, Blair BM, Baxter R, Gerding DN, Nichol G, Thomas WD, Leney M, Sloan S, Hay CA, Ambrosino DM (2010) Treatment with monoclonal antibodies against *Clostridium difficile* toxins. N Engl J Med 362(3):197–205. doi:10.1056/NEJMoa0907635
38. McDonald LC, Owings M, Jernigan DB (2006) *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996–2003. Emerg Infect Dis 12(3):409–415. doi:10.3201/eid1205.051064
39. McGlone S, Bailey R, Zimmer S, Popovich M, Tian Y, Ufberg P, Muder R, Lee B (2012) The economic burden of *Clostridium difficile*. Clin Microbiol Infect 18(3):282–289. doi:10.1111/j.1469-0691.2011.03571.x
40. Muto CA, Pokrywka M, Shutt K, Mendelsohn AB, Nouri K, Posey K, Roberts T, Croyle K, Krystofiak S, Patel-Brown S, Pasculle AW, Paterson DL, Saul M, Harrison LH (2005) A large outbreak of *Clostridium difficile*-associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. Infect Control Hosp Epidemiol 26(3):273–280. doi:10.1086/502539
41. National Center for Health Statistics (2014) Health, United States, 2013: with special feature on prescription drugs. Technical reports, National Center for Health Statistics, Hyattsville
42. OECD: Average length of stay in hospitals. In: Health at a Glance 2011: OECD Indicators, Health at a Glance, chap. Average le. OECD Publishing (2011). doi:10.1787/health_glance-2011-33-en
43. Pedersen G, Schonheyder HC, Steffensen FH, Sorenson HT (1999) Risk of resistance related to antibiotic use before admission in patients with community-acquired bacteraemia. J Antimicrob Chemother 43(1):119–126. doi:10.1093/jac/43.1.119
44. Polk RE, Hohmann SF, Medvedev S, Ibrahim O (2011) Benchmarking risk-adjusted adult antibacterial drug use in 70 US academic medical center hospitals. Clin Infect Dis 53(11):1100–1110. doi:10.1093/cid/cir672
45. Rafii F, Sutherland JB, Cerniglia CE (2008) Effects of treatment with antimicrobial agents on the human colonic microflora. Ther Clin Risk Manag 4(6):1343–58
46. Rubin MA, Jones M, Leecaster M, Khader K, Ray W, Huttner A, Huttner B, Toth D, Sablay T, Borotkanics RJ, Gerding DN, Samore MH (2013) A simulation-based assessment of strategies to control *Clostridium difficile* transmission and infection. PLoS ONE 8(11):e80671. doi:10.1371/journal.pone.0080671
47. Rubino G, Sericola B (1989) Sojourn times in finite Markov processes. J Appl Probab 26(4):744. doi:10.2307/3214379
48. Sethi AK, Al-Nassir WN, Nerandzic MM, Bobulsky GS, Donskey CJ (2010) Persistence of skin contamination and environmental shedding of *Clostridium difficile* during and after treatment of *C. difficile* infection. Infect Control Hosp Epidemiol 31(1):21–27. doi:10.1086/649016
49. Slimings C, Armstrong P, Beckingham WD, Bull AL, Hall L, Kennedy KJ, Marquess J, McCann R, Menzies A, Mitchell BG, Richards MJ, Smollen PC, Tracey L, Wilkinson IJ, Wilson FL, Worth LJ,

Riley TV (2014) Increasing incidence of *Clostridium difficile* infection, Australia, 2011–2012. Med J Aust 200(5):272–276. doi:10.5694/mja13.11153

50. Starr J, Campbell A, Renshaw E, Poxton I, Gibson G (2009) Spatio-temporal stochastic modelling of *Clostridium difficile*. J Hosp Infect 71(1):49–56. doi:10.1016/j.jhin.2008.09.013
51. Voth DE, Ballard JD (2005) *Clostridium difficile* toxins: mechanism of action and role in disease. Clin Microbiol Rev 18(2):247–263. doi:10.1128/CMR.18.2.247-263.2005
52. Walker AS, Eyre DW, Wyllie DH, Dingle KE, Harding RM, O'Connor L, Griffiths D, Vaughan A, Finney J, Wilcox MH, Crook DW, Peto TEA (2012) Characterisation of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. PLoS Med 9(2):e1001172. doi:10.1371/journal.pmed.1001172
53. Yakob L, Riley TV, Paterson DL, Clements AC (2013) *Clostridium difficile* exposure as an insidious source of infection in healthcare settings: an epidemiological model. BMC Infect Dis 13(1):376. doi:10.1186/1471-2334-13-376
54. Yakob L, Riley TV, Paterson DL, Marquess J, Clements AC (2014) Assessing control bundles for *Clostridium difficile*: a review and mathematical model. Emerg Microbes Infect 3(6):e43. doi:10.1038/emi.2014.43
55. Yakob, L., Riley, T.V., Paterson, D.L., Marquess, J., Magalhaes, R.J.S., Furuya-Kanamori, L., Clements, A.C. (2015) Mechanisms of hypervirulent Clostridium difficile ribotype 027 displacement of endemic strains: an epidemiological model. Scientific Reports **5**(November 2014):12666. doi:10.1038/srep12666

A Supplementary Materials

A.1 Further details of model structure

A.1.1 Transitions within the Population

In our model the rate of transmission is proportional to the number of infectious individuals and inversely proportional to the size of the hospital N_{bed}. Infectiousness of individuals with *C. difficile* overgrowth is governed by the parameter β. Contact precautions reduce transmission from symptomatic individuals by a factor q. Therefore the rate of transmission to each susceptible person is $[(H_{ao} + U_{ao})q + R_{ao} + H_{ac} + U_{ac} + R_{ac}]\,\beta/N_{bed}$.

Exposure to *C. difficile* spores moves individuals free of *C. difficile* to the associated colonised compartment (e.g. $U_a \to U_{ac}$ or $H \to H_c$). There is no evidence that immune responses to toxins [10,14] or commensal gut flora undisturbed by antibiotics [14] prevents initial colonisation with *C. difficile*. Therefore, in our model all *C. difficile* free individuals are equally susceptible to colonisation.

Almost all CDI cases in hospitals had recently received antimicrobial treatment [14,9,2]. The disruption of commensal gut flora and *C. difficile* susceptibility have been shown to be associated with other chemical agents such as proton pump inhibitors [19,4] and conditions such as irritable bowel disease [15,7,5]. However, we use antimicrobial consumption as a proxy measure of gut flora disruption and associated susceptibility to CDI. Consumption of antimicrobials – administered at rate α – moves individuals with healthy gut flora to the associated compartment with disrupted gut flora (e.g. $R_c \to R_{ac}$ or $U \to U_a$). Only *C. difficile* in colonised individuals with disturbed gut flora can overgrow – at rate ω – moving to the associated overgrowth compartment (e.g. $U_{ac} \to U_{ao}$ or $R_{ac} \to R_{ao}$).

Individuals with symptomatic CDI (H_{ao} and U_{ao}) receive treatment. Treatment occurs at rate τ for all such individuals; however, only a proportion p_t of treatments are fully successful and the remaining proportion $1 - p_t$ are partially successful. Fully successful treatment removes all vegetative *C. difficile* and spores. Partially successful treatment removes sufficient vegetative *C. difficile* to stop symptoms but leaves viable spores that may germinate and overgrow, potentially causing a recurrence of symptoms. A robust immune response to *C. difficile* toxins has been shown to reduce the probability of recurrence [11,12]. Therefore fully and partially successful treatment moves immunocompetent H_{ao} individuals to the immune R_a and R_{ac} compartments respectively but immunocompromised U_{ao} individuals to the U_a and U_{ac} compartments respectively. Individuals in the H_c and H_{ac} classes also acquire immunity to toxins – at rate δ – moving to the R_c and R_{ac} compartments respectively.

The recommended treatment for CDI is with antimicrobials such as vancomycin or metronidazole [13] which prevent the natural recovery of gut flora. Individuals with *C. difficile* overgrowth but no symptoms (R_{ao}) and other asymptomatically colonised individuals do not receive treatment for CDI in our model. Therefore their gut flora recovers – at rate λ – moving them to the associated healthy gut flora compartment (e.g. $R_{ao} \to R_c$, $U_a \to U$, $H_{ac} \to H_c$). The commensal gut flora of individuals colonised without *C. difficile* overgrowth competitively exclude *C. difficile* – at rate γ – moving the individual to the associated *C. difficile* free compartment (e.g. $R_c \to R$ or $H_c \to H$).

A.1.2 Admissions and Discharges

In addition to transitions between compartments, patients are discharged and new patients admitted. All patients without CDI are discharged from hospital at rate κ. Patients with CDI are not discharged but may die at rate μ. To approximately balance discharges, admissions occur at rate κN_{bed}. New admissions are distributed between the 15 compartments according the probabilities of having the associated combination of characteristics. A proportion p_u are immune suppressed and so are admitted in one of the U compartments. Of the remaining $1-p_u$ proportion of admissions, a proportion p_{AB} have serum antibodies for *C. difficile* toxins and are therefore in one of the R compartments. The remainder are in one of the H compartments. A proportion p_a of admissions have recently received antimicrobials or immediately begin treatment with antimicrobials on admission. A proportion p_c of admissions are colonised or have overgrowth on admission. A proportion p_o ($< p_a p_c$) of admissions will have *C. difficile* overgrowth on admission. Table A.1 summarises how these 5 parameters (p_o, p_{AB}, p_u, p_a and p_c) are used to determine the proportion entering each compartment in the model.

A.2 Model Parameterisation

Most of the model parameters have readily measurable, physical interpretations and were chosen to reflect the available literature (Table 1). Parameter choice was less straightforward for some parameters due to limited research in these areas. Little is known about the time to onset of adaptive immune responses for *C. difficile* toxins. Phase I and II vaccine

Immune status			Gut flora and *C. difficile* status				
H_x	U_x	R_x	X	X_a	X_c	X_{ac}	X_{ao}
$(1-p_u)(1-p_{AB})$	p_u	$(1-p_u)p_{AB}$	$(1-p_a)(1-p_c)$	$p_a(1-p_c)$	$(1-p_a)p_c$	$p_a p_c - p_o$	p_o

Table A.1: The proportion of new admissions that arrive with each of the 3 immune statuses and 5 gut flora and *C. difficile* statuses. We assume that immune status is independent of gut flora and *C. difficile* status so the proportion arriving in a particular compartment is the product of the two proportions.

studies [6,3] reported the proportion of study subjects that seroconverted over time. The latest study reported 42% and 56% seroconversion for toxin A and B antibodies 14 days after an initial dose and 100% and 90% seroconversion after 2 further doses. However it is probable that the development of a robust immune response to toxins may happen more rapidly in individuals colonised with *C. difficile* than in vaccine recipients, so we assumed $1/\delta$ is ten days.

The effect of contact precautions on transmission (q) has not been directly measured. Adherence to special contact precautions for CDI patients has been reported to range from 40% to 80% by setting and precaution, corresponding to q in the range $0.2 - 0.6$ as q is a multiplicative parameter. As contact precautions are unlikely to be perfectly effective we have assumed a value for q in the upper half of this range.

The mean length of stay ($1/\kappa$) and the proportion of patients with serum antibodies to *C. difficile* toxins at admission (p_{AB}) reported by Loo et al were both significantly different to the values reported more broadly in the literature. Since the model proved sensitive to the associated parameters, the high hospital-acquired CDI incidence setting used these (smaller) values of the parameters. The parameters β, p_t, p_u and p_o were estimated by fitting to Loo et al (Table 2). We inferred the daily hazard of asymptomatic colonisation or infection – which is approximately equal to the mean force of colonisation – from the reported cumulative risk by assuming that the daily hazard does not change over the duration of a hospital stay. In our model the transmission parameter β was chosen to reproduce this force of colonisation.

For the moderate-incidence setting the mean length of stay and the proportion of patients with serum antibodies to *C. difficile* toxins at admission were set to the typical values found in the literature: $p_{AB} = 0.7$ [8] and $\kappa = 1/6$ day^{-1} [17]. The mean age of patients in the study by Loo et al was 67.4 years. Assuming patients aged over 65 years have weakened immune systems, this was consistent with the fitted value of the proportion of admitted patients that were immunocompromised ($p_u = 0.55$), but significantly more than the mean age from a survey of 130 US hospitals [18] (57.6 years). Therefore, the estimate of the proportion of admissions unable to mount a sufficient immune response (p_u) based on Loo et al. was likely to be an overestimate for a moderate-incidence setting. Therefore p_u is reduced for the moderate-incidence setting to 0.46 – approximately the average proportion of patient-days attributable to persons over 65 years of age (48% in Australia [1]; 44% USA [16]).

A.3 Further Details of Sensitivity Analysis

Cross-covariance plots of long simulations of the model showed that temporal dependence was indistinguishable from noise after 20 days. Therefore when estimating sensitivity using coupled finite differences each simulation was run for 40 days with the prevalence and force of colonisation (FOC) calculated at the end of each simulation.

To calculate individual-level outcomes at a given set of parameters (θ) we first estimated the mean FOC ($\phi(\theta)$) by simulating the full model. Individual-level outcomes (O) estimated using the individual Markov chain approximation depend on the mean FOC and parameter values:

$$O = O(\theta, \phi(\theta)) \,. \tag{1}$$

To estimate sensitivity of individual-level outcomes we used the individual Markov chain approximation and finite differences to calculate the derivative of the individual-level outcomes with respect to the mean FOC ($\frac{\partial O}{\partial \phi}$) and the parameters ($\frac{\partial O}{\partial \theta}$). Using the total derivative rule we combined these with the derivative of the mean FOC obtained using coupled finite differences ($\frac{d\phi}{d\theta}$) to get the derivative of the individual-level outcomes with respect to parameters at the mean FOC:

$$\frac{dO}{d\theta} = \frac{\partial O}{\partial \phi}\frac{d\phi}{d\theta} + \frac{\partial O}{\partial \theta} \,. \tag{2}$$

A.4 Additional Figures

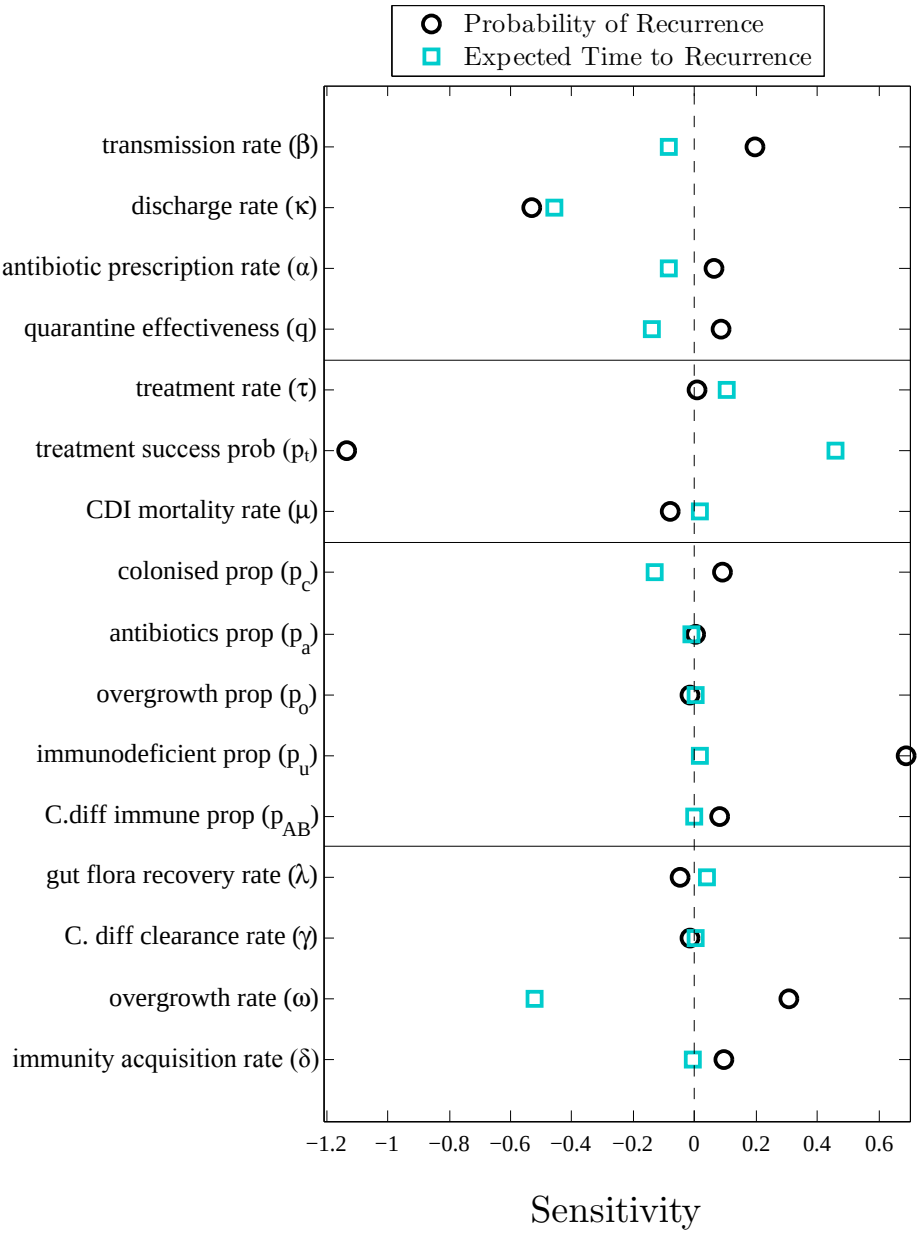


Fig. A.1: The sensitivity of the probability of at least one recurrence of CDI (while hospitalised) given an initial infection and the mean time between the end of the first CDI and the onset of the first CDI recurrence. The sensitivities are calculated using our novel individual Markov chain method at the simulated mean force of colonisation.

Fig. A.2: The sensitivity of the mean force of colonisation and the reproduction number to parameter values. The sensitivity of mean force of colonisation is calculated using Anderson's coupled finite differences, and the sensitivity of the reproduction number is estimated using the next generation method and numerical derivative estimation (second order central difference).

References

1. Australian Institute of Health and Welfare: Admitted patient care 201314: Australian hospital statistics. Tech. rep., Australian Institute of Health and Welfare, Canberra (2015)
2. Britton, R.A., Young, V.B.: Interaction between the intestinal microbiota and host in Clostridium difficile colonization resistance. Trends in Microbiology **20**(7), 313–319 (2012). DOI 10.1016/j.tim.2012.04.001
3. de Bruyn, G., Saleh, J., Workman, D., Pollak, R., Elinoff, V., Fraser, N.J., Lefebvre, G., Martens, M., Mills, R.E., Nathan, R., Trevino, M., van Cleeff, M., Foglia, G., Ozol-Godfrey, A., Patel, D.M., Pietrobon, P.J., Gesser, R.: Defining the optimal formulation and schedule of a candidate toxoid vaccine against Clostridium difficile infection: A randomized Phase 2 clinical trial. Vaccine **34**(19), 2170–2178 (2016). DOI 10.1016/j.vaccine.2016.03.028
4. Deshpande, A., Pant, C., Pasupuleti, V., Rolston, D.D., Jain, A., Deshpande, N., Thota, P., Sferra, T.J., Hernandez, A.V.: Association Between Proton Pump Inhibitor Therapy and Clostridium difficile Infection in a Meta-Analysis. Clinical Gastroenterology and Hepatology **10**(3), 225–233 (2012). DOI 10.1016/j.cgh.2011.09.030
5. Dial, S., Kezouh, A., Dascal, A., Barkun, A., Suissa, S.: Patterns of antibiotic use and risk of hospital admission because of Clostridium difficile infection. Canadian Medical Association Journal **179**(8), 767–772 (2008). DOI 10.1503/cmaj.071812
6. Foglia, G., Shah, S., Luxemburger, C., Pietrobon, P.J.F.: Clostridium difficile: Development of a novel candidate vaccine. Vaccine **30**(29), 4307–4309 (2012). DOI 10.1016/j.vaccine.2012.01.056
7. Furuya-Kanamori, L., Stone, J.C., Clark, J., McKenzie, S.J., Yakob, L., Paterson, D.L., Riley, T.V., Doi, S.a.R., Clements, A.C.: Comorbidities, Exposure to Medications, and the Risk of Community-Acquired Clostridium difficile Infection: A Systematic Review and Meta-analysis. Infection Control & Hospital Epidemiology **36**(02), 132–141 (2015). DOI 10.1017/ice.2014.39
8. Kelly, C.P., Kyne, L.: The host immune response to Clostridium difficile. Journal of Medical Microbiology **60**(8), 1070–1079 (2011). DOI 10.1099/jmm.0.030015-0
9. Khan, F.Y., Abu-Khattab, M., Anand, D., Baager, K., Alaini, A., Siddique, M.a., Mohamed, S.F., Ali, M.I., Al bedawi, M.M., Naser, M.S.: Epidemiological features of Clostridium difficile infection among inpatients at Hamad General Hospital in the state of Qatar, 2006-2009. Travel Medicine and Infectious Disease **10**(4), 179–185 (2012). DOI 10.1016/j.tmaid.2012.06.004
10. Kyne, L., Warny, M., Qamar, A., Kelly, C.P.: Asymptomatic Carriage of Clostridium difficile and Serum Levels of IgG Antibody against Toxin A. New England Journal of Medicine **342**(6), 390–397 (2000). DOI 10.1056/NEJM200002103420604
11. Kyne, L., Warny, M., Qamar, A., Kelly, C.P.: Association between antibody response to toxin A and protection against recurrent Clostridium difficile diarrhoea. The Lancet **357**(9251), 189–193 (2001). DOI 10.1016/S0140-6736(00)03592-3
12. Leav, B.A., Blair, B., Leney, M., Knauber, M., Reilly, C., Lowy, I., Gerding, D.N., Kelly, C.P., Katchar, K., Baxter, R., Ambrosino, D., Molrine, D.: Serum anti-toxin B antibody correlates with protection from recurrent Clostridium difficile infection (CDI). Vaccine **28**(4), 965–969 (2010). DOI 10.1016/j.vaccine.2009.10.144
13. Leffler, D.A., Lamont, J.T.: Treatment of Clostridium difficile-Associated Disease. Gastroenterology **136**(6), 1899–1912 (2009). DOI 10.1053/j.gastro.2008.12.070
14. Loo, V.G., Bourgault, A.M., Poirier, L., Lamothe, F., Michaud, S., Turgeon, N., Toye, B., Beaudoin, A., Frost, E.H., Gilca, R., Brassard, P., Dendukuri, N., Béliveau, C., Oughton, M., Brukner, I., Dascal, A.: Host and Pathogen Factors for Clostridium difficile Infection and Colonization. New England Journal of Medicine **365**(18), 1693–1703 (2011). DOI 10.1056/NEJMoa1012413
15. Matsuoka, K., Kanai, T.: The gut microbiota and inflammatory bowel disease. Seminars in Immunopathology **37**(1), 47–55 (2015). DOI 10.1007/s00281-014-0454-4
16. National Center for Health Statistics: Health, United States, 2013: With Special Feature on Prescription Drugs. Tech. rep., National Center for Health Statistics, Hyattsville (2014)
17. OECD: Health at a Glance 2005. Health at a Glance. OECD Publishing (2005). DOI 10.1787/9789264012639-en
18. Polk, R.E., Fox, C., Mahoney, A., Letcavage, J., MacDougall, C.: Measurement of Adult Antibacterial Drug Use in 130 US Hospitals: Comparison of Defined Daily Dose and Days of Therapy. Clinical Infectious Diseases **44**(5), 664–670 (2007). DOI 10.1086/511640
19. Williams, C., McColl, K.E.L.: Review article: proton pump inhibitors and bacterial overgrowth. Alimentary Pharmacology and Therapeutics **23**(1), 3–10 (2006). DOI 10.1111/j.1365-2036.2006.02707.x

3.3 Model Equations

This section provides an explicit description of the stochastic compartmental model used in this chapter and Chapter 4. Though this section does not provide any information that cannot be inferred from the preceding sections in this chapter, the equations are written out in full for the convenience of the interested reader.Table 3.1 lists the events, transitions and probabilities that define the full CTMC model.

Event type	Transition			Probability
Disruption of gut flora	X	$\to$	X_a	$\alpha X\ \Delta t + o(\Delta t)$
	X_c	$\to$	X_{ac}	$\alpha X_c\ \Delta t + o(\Delta t)$
Recovery of gut flora	X_a	$\to$	X	$\lambda X_a\ \Delta t + o(\Delta t)$
	X_{ac}	$\to$	X_c	$\lambda X_{ac}\ \Delta t + o(\Delta t)$
	R_{ao}	$\to$	R_c	$\lambda R_{ao}\ \Delta t + o(\Delta t)$
Colonisation	X	$\to$	X_c	$fX\ \Delta t + o(\Delta t)$
	X_a	$\to$	X_{ac}	$fX_a\ \Delta t + o(\Delta t)$
Clearing of colonisation	X_c	$\to$	X	$\gamma X_c\ \Delta t + o(\Delta t)$
C. difficile overgrowth	X_{ac}	$\to$	X_{ao}	$\omega X_{ac}\ \Delta t + o(\Delta t)$
Treatment	H_{ao}	$\to$	R_a	$\tau p_t H_{ao}\ \Delta t + o(\Delta t)$
	H_{ao}	$\to$	R_{ac}	$\tau(1-p_t) H_{ao}\ \Delta t + o(\Delta t)$
	U_{ao}	$\to$	U_a	$\tau p_t U_{ao}\ \Delta t + o(\Delta t)$
	U_{ao}	$\to$	U_{ac}	$\tau(1-p_t) U_{ao}\ \Delta t + o(\Delta t)$
Development of immunity	H_c	$\to$	R_c	$\delta H_c\ \Delta t + o(\Delta t)$
	H_{ac}	$\to$	R_{ac}	$\delta H_{ac}\ \Delta t + o(\Delta t)$
Hospital admission		$\to$	X	$1_{\mathrm{B}}\kappa p_X(1-p_a)(1-p_c)N_{\mathrm{bed}}\ \Delta t + o(\Delta t)$
		$\to$	X_a	$1_{\mathrm{B}}\kappa p_X(p_a)(1-p_c)N_{\mathrm{bed}}\ \Delta t + o(\Delta t)$
		$\to$	X_c	$1_{\mathrm{B}}\kappa p_X(1-p_a)(p_c)N_{\mathrm{bed}}\ \Delta t + o(\Delta t)$
		$\to$	X_{ac}	$1_{\mathrm{B}}\kappa p_X(p_a p_c - p_o)N_{\mathrm{bed}}\ \Delta t + o(\Delta t)$
		$\to$	X_{ao}	$1_{\mathrm{B}}\kappa p_X(p_o)N_{\mathrm{bed}}\ \Delta t + o(\Delta t)$
Hospital discharge	X	$\to$		$\kappa X\ \Delta t + o(\Delta t)$
	X_a	$\to$		$\kappa X_a\ \Delta t + o(\Delta t)$
	X_c	$\to$		$\kappa X_c\ \Delta t + o(\Delta t)$
	X_{ac}	$\to$		$\kappa X_{ac}\ \Delta t + o(\Delta t)$
	R_{ao}	$\to$		$\kappa R_{ao}\ \Delta t + o(\Delta t)$
Death of CDI cases	H_{ao}	$\to$		$\mu H_{ao}\ \Delta t + o(\Delta t)$
	U_{ao}	$\to$		$\mu U_{ao}\ \Delta t + o(\Delta t)$

Table 3.1: The CTMC model for *C. difficile* in a hospital population. Parameters and compartment names are as they appear in Figure 1 and Table 1 of the published article. However for brevity X is used instead of H, U, or R when an event occurs for persons of any immune status and does not alter immune status. Similarly, p_X represents the proportion of admissions with each immune status: p_U, $p_R = p_{AB}(1-p_U)$, and $p_H = 1 - p_U - p_R$. Admission only occurred if there were empty beds; 1_B is an indicator variable such that $1_B = 1$ if $N < N_{\mathrm{bed}}$ and 0 otherwise.

Where the individual Markov chain approximation was used for individual-level outcomes, an individual was modelled as a CTMC on state-space

$\{H, H_a, H_c, H_{ac}, H_{ao}, U, U_a, U_c, U_{ac}, U_{ao}, R, R_a, R_c, R_{ac}, R_{ao}\}$ with dynamics given by the Kolmogorov forward equation

$$\mathbf{p}' = \begin{bmatrix} Q_{HH} & 0 & 0 \\ 0 & Q_{UU} & 0 \\ Q_{RH} & 0 & Q_{RR} \end{bmatrix} \mathbf{p} \tag{3.1}$$

where

$$Q_{HH} = \begin{bmatrix} -(\alpha + f^* + \kappa) & \lambda & \gamma & 0 & 0 \\ \alpha & -(\lambda + f^* + \kappa) & 0 & 0 & 0 \\ f^* & 0 & -(\alpha + \gamma + \delta + \kappa) & \lambda & 0 \\ 0 & f^* & \alpha & -(\lambda + \omega + \delta + \kappa) & 0 \\ 0 & 0 & 0 & \omega & -(\tau + \mu) \end{bmatrix}, \tag{3.2}$$

$$Q_{UU} = \begin{bmatrix} -(\alpha + f^* + \kappa) & \lambda & \gamma & 0 & 0 \\ \alpha & -(\lambda + f^* + \kappa) & 0 & 0 & \tau p_t \\ f^* & 0 & -(\alpha + \gamma + \kappa) & \lambda & 0 \\ 0 & f^* & \alpha & -(\lambda + \omega + \kappa) & \tau(1 - p_t) \\ 0 & 0 & 0 & \omega & -(\tau + \mu) \end{bmatrix}, \tag{3.3}$$

$$Q_{RR} = \begin{bmatrix} -(\alpha + f^* + \kappa) & \lambda & \gamma & 0 & 0 \\ \alpha & -(\lambda + f^* + \kappa) & 0 & 0 & 0 \\ f^* & 0 & -(\alpha + \gamma + \kappa) & \lambda & \lambda \\ 0 & f^* & \alpha & -(\lambda + \omega + \kappa) & 0 \\ 0 & 0 & 0 & \omega & -(\lambda + \kappa) \end{bmatrix}, \tag{3.4}$$

$$Q_{RH} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tau p_t \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & \delta & \tau(1 - p_t) \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{3.5}$$

and f^* is the mean force of colonisation at equilibrium, *i.e.*

$$f^* = [(H^*_{ao} + U^*_{ao})q + R^*_{ao} + H^*_{ac} + U^*_{ac} + R^*_{ac}]\,\beta/N_{bed}, \tag{3.6}$$

where H^*_{ao}, U^*_{ao} *etc.* are the means (across the equilibrium distribution) of the number of persons in each infectious class. When calculating individual-level outcomes averaged across all admission types, the initial condition, $\mathbf{p}(0)$, (*i.e.* the probability of being in each

compartment on admission) was

$$\mathbf{p}(0) = \begin{bmatrix} p_H(1-p_a)(1-p_c) \\ p_H p_a(1-p_c) \\ p_H(1-p_a)p_c \\ p_H(p_a p_c - p_o) \\ p_H p_o \\ p_U(1-p_a)(1-p_c) \\ p_U p_a(1-p_c) \\ p_U(1-p_a)p_c \\ p_U(p_a p_c - p_o) \\ p_U p_o \\ p_R(1-p_a)(1-p_c) \\ p_R p_a(1-p_c) \\ p_R(1-p_a)p_c \\ p_R(p_a p_c - p_o) \\ p_R p_o \end{bmatrix} \tag{3.7}$$

where $p_R = p_{AB}(1-p_U)$, and $p_H = 1 - p_U - p_R$.

In the following sections I briefly sketch some of the theory of sojourn times and hitting probabilities as is relevant to calculating the individual-level quantities in this chapter: recurrence probabilities, expected time to recurrence and the probability of developing symptomatic infection within the first 72 hours after admission. These sections assumes some of the basic properties of Markov processes which are discussed in detail in many excellent books (*e.g.* [76, 112]).

3.3.1 Probability of entering a set of states

Let X be a continuous time Markov chain with state space S and rate matrix A. For our model the state space is the set of model compartments plus a seperate compartment for leaving the hospital (death and discharge). The rate matrix A expands the above Q matrix to add the death and discharge transitions:

$$A = \begin{bmatrix} Q & 0 \\ D & 0 \end{bmatrix} \tag{3.8}$$

where D is the set of transitions rates from each state to the discharge/death state:

$$D = \begin{bmatrix} \kappa & \kappa & \kappa & \kappa & \mu & \kappa & \kappa & \kappa & \kappa & \mu & \kappa & \kappa & \kappa & \kappa & \kappa \end{bmatrix}. \tag{3.9}$$

Our first task is to calculate the probability that the Markov chain visits any one of a subset of the states $\Sigma \subset S$ at least once. Let T_1^Σ be the first time X visits any of the states in Σ i.e. $T_1^\Sigma = \inf\{t : X(t) \in \Sigma\}$. Then the probability of ever visiting Σ is $P(T_1^\Sigma < \infty)$.

To calculate this value consider the associated process

$$X^{\Sigma}(t) := \begin{cases} X(t), & t < T_1^{\Sigma} \\ X(T_1^{\Sigma}), & t \geq T_1^{\Sigma} \end{cases} \tag{3.10}$$

which acts as X until it visits one of the states in Σ, after which it remains there. Therefore for each $s \in \Sigma$

$$P(X^{\Sigma}(t) = s) = P(X(T_1^{\Sigma}) = s, T_1^{\Sigma} < t) \tag{3.11}$$

and

$$\lim_{t \to \infty} P(X^{\Sigma}(t) = s) = P(X(T_1^{\Sigma}) = s, T_1^{\Sigma} < \infty). \tag{3.12}$$

Since $X(T_1^{\Sigma})$ must be in Σ

$$P(X^{\Sigma}(t) \in \Sigma) = P(T_1^{\Sigma} < t) \tag{3.13}$$

and

$$\lim_{t \to \infty} P(X^{\Sigma}(t) \in \Sigma) = P(T_1^{\Sigma} < \infty). \tag{3.14}$$

Fortunately X^{Σ} is also a continuous-time Markov chain so this limit is simple to calculate from its rate matrix. X^{Σ} differs from X only in that the states in Σ are absorbing and so its rate matrix A^{Σ} satisfies

$$A^{\Sigma}_{s_2,s_1} = \begin{cases} A_{s_2,s_1}, & s_1 \notin \Sigma \\ 0, & s_1 \in \Sigma \end{cases}. \tag{3.15}$$

We can also write A^{Σ} using sub-matrix notation. For any subsets of states $U, V \subset S$ let $A_{V,U}$ be the sub-matrix of transitions rates from states in U to states in V and for brevity let $A_{U,U} = A_U$, $A_{V,V} = A_V$ etc. If we partition the states into three sets – Σ, transient but not in Σ (which we call U) and absorbing but not in Σ (which we call V) – we can use this notation to write A^{Σ} (up to a permutation of rows and columns so that the states are ordered U, V, Σ) in terms of sub-matrices of A:

$$A^{\Sigma} = \begin{bmatrix} A_U & 0 & 0 \\ A_{V,U} & 0 & 0 \\ A_{\Sigma,U} & 0 & 0 \end{bmatrix}. \tag{3.16}$$

In our case if we are interested in the probability of developing symptoms while in hospital, Σ will be the set of symptomatic states, V a singleton set with the dead/discharged state and U all the set of remaining states. If $\mathbf{p}(t)$ is the vector of probabilities that X will be in a given state in S at time t, the solution to the Kolmogorov forward equation for X is

$$\mathbf{p}(t) = e^{At}\mathbf{p}(0). \tag{3.17}$$

Similarly if $\mathbf{p}^{\Sigma}(t)$ is the vector of probabilities that X^{Σ} will be in a given state in S at

time t, noting that by definition $\mathbf{p}^{\Sigma}(0) = \mathbf{p}(0)$ and using sub-matrix notation we can write the solution to the Kolmogorov forward equation for X^{Σ}

$$\mathbf{p}^{\Sigma}(t) = e^{A^{\Sigma}t}\mathbf{p}(0) \tag{3.18}$$

as

$$\begin{aligned}
\mathbf{p}^{\Sigma}(t) &= \begin{bmatrix} \mathbf{p}_U^{\Sigma}(t) \\ \mathbf{p}_V^{\Sigma}(t) \\ \mathbf{p}_{\Sigma}^{\Sigma}(t) \end{bmatrix} \\
&= \exp\left(t \begin{bmatrix} A_U & 0 & 0 \\ A_{V,U} & 0 & 0 \\ A_{\Sigma,U} & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{p}_U(0) \\ \mathbf{p}_V(0) \\ \mathbf{p}_{\Sigma}(0) \end{bmatrix} \\
&= \left(\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} + \begin{bmatrix} A_U t & 0 & 0 \\ A_{V,U} t & 0 & 0 \\ A_{\Sigma,U} t & 0 & 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} {A_U}^2 t^2 & 0 & 0 \\ A_{V,U} A_U t^2 & 0 & 0 \\ A_{\Sigma,U} A_U t^2 & 0 & 0 \end{bmatrix} + \dots \right) \begin{bmatrix} \mathbf{p}_U(0) \\ \mathbf{p}_V(0) \\ \mathbf{p}_{\Sigma}(0) \end{bmatrix} \\
&= \begin{bmatrix} e^{A_U t} & 0 & 0 \\ A_{V,U} {A_U}^{-1}(e^{A_U t} - I) & I & 0 \\ A_{\Sigma,U} {A_U}^{-1}(e^{A_U t} - I) & 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{p}_U(0) \\ \mathbf{p}_V(0) \\ \mathbf{p}_{\Sigma}(0) \end{bmatrix} \\
&= \begin{bmatrix} e^{A_U t}\mathbf{p}_U(0) \\ A_{V,U} {A_U}^{-1}(e^{A_U t} - I)\mathbf{p}_U(0) + \mathbf{p}_V(0) \\ A_{\Sigma,U} {A_U}^{-1}(e^{A_U t} - I)\mathbf{p}_U(0) + \mathbf{p}_{\Sigma}(0) \end{bmatrix}.
\end{aligned} \tag{3.19}$$

Taking limits we get

$$\begin{aligned}
\lim_{t\to\infty} \mathbf{p}^{\Sigma}(t) &= \lim_{t\to\infty} \begin{bmatrix} \mathbf{p}_U^{\Sigma}(t) \\ \mathbf{p}_V^{\Sigma}(t) \\ \mathbf{p}_{\Sigma}^{\Sigma}(t) \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{0} \\ -A_{V,U} {A_U}^{-1}\mathbf{p}_U(0) + \mathbf{p}_V(0) \\ -A_{\Sigma,U} {A_U}^{-1}\mathbf{p}_U(0) + \mathbf{p}_{\Sigma}(0) \end{bmatrix}.
\end{aligned} \tag{3.20}$$

Therefore the probability that X visits Σ at least once is given by the sum

$$\begin{aligned}
P(T_1^{\Sigma} < \infty) &= \lim_{t\to\infty} P(X^{\Sigma}(t) \in \Sigma) \\
&= \mathbf{1}^T \lim_{t\to\infty} \mathbf{p}_{\Sigma}^{\Sigma}(t) \\
&= \mathbf{1}^T \left(-A_{\Sigma,U} {A_U}^{-1}\mathbf{p}_U(0) + \mathbf{p}_{\Sigma}(0) \right).
\end{aligned} \tag{3.21}$$

Furthermore we can characterise the state of X at T_1^Σ by

$$\begin{aligned}\left[P(X(T_1^\Sigma) = s | T_1^\Sigma < \infty)\right]_{s\in\Sigma} &= \frac{\lim_{t\to\infty} \mathbf{p}_\Sigma^\Sigma(t)}{\mathbf{1}^T \lim_{t\to\infty} \mathbf{p}_\Sigma^\Sigma(t)} \\ &= \frac{A_{\Sigma,U}{A_U}^{-1}\mathbf{p}_U(0) + \mathbf{p}_\Sigma(0)}{\mathbf{1}^T\left(-A_{\Sigma,U}{A_U}^{-1}\mathbf{p}_U(0) + \mathbf{p}_\Sigma(0)\right)}. \end{aligned} \tag{3.22}$$

3.3.2 Recurrence

We also wish to calculate the probability of a first recurrence – the probability that X will enter a state in Σ from a state outside of Σ at least twice given it has done so at least once. Let L_1^Σ be the first time that X leaves Σ; in other words the first time X enters Σ^c after T_1^Σ: $L_1^\Sigma = \inf\{t > T_1^\Sigma : X(t) \in \Sigma^c\}$. Let T_2^Σ be the second time X enters Σ; in other words the first time X enters Σ after L_1^Σ: $T_2^\Sigma = \inf\{t > L_1^\Sigma : X(t) \in \Sigma\}$. Then the probability of a first recurrence is

$$P(T_2^\Sigma < \infty | T_1^\Sigma < \infty) = \frac{P(T_2^\Sigma < \infty)}{P(T_1^\Sigma < \infty)}. \tag{3.23}$$

Because X is a Markov process and $T_1^\Sigma, L_1^\Sigma, T_2^\Sigma, \ldots$ are successive stopping times defined by entry to a set (either Σ or its complement) we can iteratively calculate the probability that they occur (i.e. are finite) by using the above method but with initial condition reset to the conditional distribution of X at the previous stopping time. For example

$$\begin{aligned} P(L_1^\Sigma < \infty) &= \sum_{s\in\Sigma} P(L_1^\Sigma < \infty \mid X(T_1^\Sigma) = s, T_1^\Sigma < \infty) P(X(T_1^\Sigma) = s, T_1^\Sigma < \infty) \\ &= \sum_{s\in\Sigma} P(L_1^\Sigma < \infty \mid X(0) = s) P(X(T_1^\Sigma) = s, T_1^\Sigma < \infty) \\ &= \sum_{s\in\Sigma} P(T_1^{\Sigma^c} < \infty \mid X(0) = s) P(X(T_1^\Sigma) = s, T_1^\Sigma < \infty). \end{aligned} \tag{3.24}$$

3.3.3 Time to enter a set of states

We also want to calculate the distribution of these stopping times and the times between them. For instance we may wish to know the distribution of T_1^Σ or $T_2^\Sigma - T_1^\Sigma$. Because X is Markov if we have a way of characterising T_1^Σ, we can iteratively use the same method to characterise $L_1^\Sigma, T_2^\Sigma, \ldots$ and their differences. We have already calculate enough to define the distribution of T_1^Σ as

$$\begin{aligned} P(T_1^\Sigma = \infty) &= 1 - P(T_1^\Sigma < \infty) \\ &= 1 + \mathbf{1}^T\left(A_{\Sigma,U}{A_U}^{-1}\mathbf{p}_U(0) - \mathbf{p}_\Sigma(0)\right) \end{aligned} \tag{3.25}$$

and

$$P(T_1^\Sigma < t) = \mathbf{1}^T\left(A_{\Sigma,U}{A_U}^{-1}(e^{A_U t} - I)\mathbf{p}_U(0) + \mathbf{p}_\Sigma(0)\right). \tag{3.26}$$

In general $\mathbb{E}\left[T_1^\Sigma\right]$ is not finite since $P(T_1^\Sigma < \infty) \neq 1$. However the conditional expected value is

$$\begin{aligned}
\mathbb{E}\left[T_1^\Sigma \mid T_1^\Sigma < \infty\right] &= \int_0^\infty P(T_1^\Sigma > t \mid T_1^\Sigma < \infty)dt \\
&= \int_0^\infty \frac{1 - P(T_1^\Sigma < t) - P(T_1^\Sigma = \infty)}{P(T_1^\Sigma < \infty)}dt \\
&= \frac{1}{P(T_1^\Sigma < \infty)} \int_0^\infty -\mathbf{1}^T A_{\Sigma,U} {A_U}^{-1} e^{A_U t} \mathbf{p}_U(0)dt \\
&= \frac{\mathbf{1}^T A_{\Sigma,U} {A_U}^{-2} \mathbf{p}_U(0)}{P(T_1^\Sigma < \infty)} \\
&= \frac{\mathbf{1}^T A_{\Sigma,U} {A_U}^{-2} \mathbf{p}_U(0)}{\mathbf{1}^T \left(-A_{\Sigma,U} {A_U}^{-1} \mathbf{p}_U(0) + \mathbf{p}_\Sigma(0)\right)}.
\end{aligned} \tag{3.27}$$

In the special case where all absorbing states are in Σ (i.e. $V = \emptyset$) the unconditioned expectation does exist since $P(T_1^\Sigma < \infty) = 1$. Moreover we can write

$$A^\Sigma = \begin{bmatrix} A_U & 0 \\ A_{\Sigma,U} & 0 \end{bmatrix}. \tag{3.28}$$

Since A^Σ is a rate matrix it satisfies $\mathbf{1}^T A^\Sigma = \mathbf{0}^T$ and therefore $\mathbf{1}^T A_U = -\mathbf{1}^T A_{\Sigma,U}$. Thus eq. (3.27) becomes

$$\begin{aligned}
\mathbb{E}\left[T_1^\Sigma\right] &= \mathbb{E}\left[T_1^\Sigma \mid T_1^\Sigma < \infty\right] \\
&= \frac{\mathbf{1}^T A_{\Sigma,U} {A_U}^{-2} \mathbf{p}_U(0)}{\mathbf{1}^T \left(-A_{\Sigma,U} {A_U}^{-1} \mathbf{p}_U(0) + \mathbf{p}_\Sigma(0)\right)} \\
&= \frac{-\mathbf{1}^T {A_U}^{-1} \mathbf{p}_U(0)}{\mathbf{1}^T \mathbf{p}_U(0) + \mathbf{1}^T \mathbf{p}_\Sigma(0)} \\
&= -\mathbf{1}^T {A_U}^{-1} \mathbf{p}_U(0).
\end{aligned} \tag{3.29}$$

The last equality holds since U and Σ partition S and therefore

$$\mathbf{1}^T \mathbf{p}_U(0) + \mathbf{1}^T \mathbf{p}_\Sigma(0) = P(X(0) \in U) + P(X(0) \in \Sigma) = 1. \tag{3.30}$$

We may be interested not only in how long it takes to arrive in Σ, but how long X spends in each of the transient states prior to entering Σ. For each $s \in U$ this quantity is the expected value

$$\mathbb{E}\left[\int_0^\infty 1_{\{X(t)=s\}} 1_{\{t<T_1^\Sigma\}} dt \,\middle|\, T_1^\Sigma < \infty\right]. \tag{3.31}$$

Because $X^\Sigma(t) = s \in \Sigma$ if and only if $X(t) = s \in \Sigma$ and $t < T_1^\Sigma$ this is the same as

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty 1_{\{X^\Sigma(t)=s\}} dt \,\middle|\, T_1^\Sigma < \infty\right] &= \frac{\mathbb{E}\left[\int_0^\infty 1_{\{X^\Sigma(t)=s\}}\, 1_{\{T_1^\Sigma<\infty\}} dt\right]}{P(T_1^\Sigma < \infty)} \\ &= \frac{\int_0^\infty P(X^\Sigma(t) = s,\, T_1^\Sigma < \infty) dt}{P(T_1^\Sigma < \infty)} \\ &= \frac{\int_0^\infty P(T_1^\Sigma < \infty \mid X^\Sigma(t) = s) P(X^\Sigma(t) = s) dt}{P(T_1^\Sigma < \infty)}. \end{aligned} \tag{3.32}$$

X^Σ is time-homogeneous and memoryless so

$$P\left(T_1^\Sigma < \infty \mid X^\Sigma(t) = s\right) = P\left(T_1^\Sigma < \infty \mid X^\Sigma(0) = s\right) \tag{3.33}$$

and so the expected value is equal to

$$\frac{P(T_1^\Sigma < \infty \mid X^\Sigma(0) = s) \int_0^\infty P(X^\Sigma(t) = s) dt}{P(T_1^\Sigma < \infty)}. \tag{3.34}$$

The vector of these expected values for each $s \in U$ can be written in matrix notation

$$\begin{aligned}&\mathbb{E}\left[\int_0^\infty 1_{\{X(t)=s\}}\, 1_{\{t<T_1^\Sigma\}} dt \,\middle|\, T_1^\Sigma < \infty\right]_{s\in U} \\ &= \frac{1}{P(T_1^\Sigma < \infty)} \left[P(T_1^\Sigma < \infty \mid X^\Sigma(0) = s) \int_0^\infty P(X^\Sigma(t) = s) dt\right]_{s\in U} \\ &= \frac{1}{P(T_1^\Sigma < \infty)} \left[P(T_1^\Sigma < \infty \mid X^\Sigma(0) = s)\right]_{s\in U} \odot \left[\int_0^\infty P(X^\Sigma(t) = s) dt\right]_{s\in U} \\ &= \frac{1}{P(T_1^\Sigma < \infty)} \left[-\mathbf{1}^T A_{\Sigma,U} A_U^{-1} e_s\right]_{s\in U} \odot \int_0^\infty \mathbf{p}_U^\Sigma(t) dt, \end{aligned} \tag{3.35}$$

where $\odot$ is the Hadamard product (element-wise multiplication) and e_s is the fundamental basis vector with a 1 in the s position and 0 elsewhere. Since concatenating e_s for each $s \in U$ is just the identity matrix

$$\left[-\mathbf{1}^T A_{\Sigma,U} A_U^{-1} e_s\right]_{s\in U} = \left(-\mathbf{1}^T A_{\Sigma,U} A_U^{-1}\right)^T. \tag{3.36}$$

Also from eq. (3.19) we have

$$\int_0^\infty \mathbf{p}_U^\Sigma(t) dt = \int_0^\infty e^{A_U t} \mathbf{p}_U(0) dt = A_U^{-1} \mathbf{p}_U(0). \tag{3.37}$$

Putting this all together

$$\mathbb{E}\left[\int_0^\infty 1_{\{X(t)=s\}}\, 1_{\{t<T_1^\Sigma\}} dt \,\middle|\, T_1^\Sigma < \infty\right]_{s\in U} = \frac{(\mathbf{1}^T A_{\Sigma,U} A_U^{-1})^T \odot A_U^{-1} \mathbf{p}_U(0)}{\mathbf{1}^T \left(-A_{\Sigma,U} A_U^{-1} \mathbf{p}_U(0) + \mathbf{p}_\Sigma(0)\right)}. \tag{3.38}$$

In special case where all absorbing states are in Σ (i.e. $V = \emptyset$) the unconditioned expectation exists because $P(T_1^\Sigma < \infty) = 1$. Moreover we can use the two identities

$\mathbf{1}^T A_U = -\mathbf{1}^T A_{\Sigma,U}$ and $\mathbf{1}^T \mathbf{p}_U(0) + \mathbf{1}^T \mathbf{p}_\Sigma(0) = 1$ again to simplify the expected value:

$$\begin{aligned}
\mathbb{E}\left[\int_0^\infty 1_{\{X(t)=s\}}\, 1_{\{t<T_1^\Sigma\}} dt\right]_{s\in U} &= \mathbb{E}\left[\int_0^\infty 1_{\{X(t)=s\}}\, 1_{\{t<T_1^\Sigma\}} dt \,\middle|\, T_1^\Sigma < \infty\right]_{s\in U} \\
&= \frac{-\mathbf{1} \odot A_U^{-1}\mathbf{p}_U(0)}{\mathbf{1}^T\mathbf{p}_U(0) + \mathbf{1}^T\mathbf{p}_\Sigma(0)} \\
&= -A_U^{-1}\mathbf{p}_U(0). \qquad (3.39)
\end{aligned}$$

Clostridium difficile classification overestimates hospital-acquired infections

4.1 Introduction

This chapter consists of an article published in the *Journal of Hospital Infection* and the accompanying supplementary materials. In this article I use the model of *C. difficile* transmission introduced in Chapter 3 to assess commonly used surveillance definitions that are used to classify *C. difficile* infections that present in hospitals. Since the classification system being assessed relies on knowledge of the time from admission to onset of symptoms for individual cases, the individual Markov chain approximation is used in this article. The main outcome of this paper is that commonly used surveillance definitions recommended by numerous infection disease organisations overestimate the proportion of CDIs presenting in hospitals that are acquired during the immediate period of hospitalisation. The article provides definitions that improve the classification of *C. difficile* infections and thus may help correct existing classification biases. The supplementary materials provide a table of model parameters and two additional figures that provide additional detail on classification performance under various scenarios and model assumptions. The article is included as it appears in print. The supplementary materials are as they appear online with the exception of a typographical error that has been corrected in the caption to supplementary figure 2.

Chapter 5 extends the analysis in this chapter, using an expanded model that includes both hospital and community to assess the recommended surveillance definitions of all CDIs (i.e. including those that do not present at hospitals) as hospital or community-acquired.

4.2 Article and supplementary materials

Angus McLure, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. *Clostridium difficile* classification overestimates hospital-acquired infections. *Journal of Hospital Infection*, 99(4):453-460, 2018.

Journal of Hospital Infection 99 (2018) 453–460

Available online at www.sciencedirect.com

Journal of Hospital Infection

journal homepage: www.elsevier.com/locate/jhin

Clostridium difficile classification overestimates hospital-acquired infections

A. McLure*, A.C.A. Clements, M. Kirk, K. Glass

Research School of Population Health, Australian National University, Canberra, Australia

ARTICLE INFO

Article history:
Received 15 October 2017
Accepted 11 December 2017
Available online 16 December 2017

Keywords:
Clostridium difficile
Hospital-acquired infections
Surveillance definitions
Mathematical model

SUMMARY

Background: *Clostridium difficile* infections occur frequently among hospitalized patients, with some infections acquired in hospital and others in the community. International guidelines classify cases as hospital-acquired if symptom onset occurs more than two days after admission. This classification informs surveillance and infection control, but has not been verified by empirical or modelling studies.

Aim: To assess current classification of *C. difficile* acquisition using a simulation model as a reference standard.

Methods: *C. difficile* transmission was simulated in a range of hospital scenarios. The sensitivity, specificity and precision of classifications that use cut-offs ranging from 0.25 h to 40 days were calculated. The optimal cut-off that correctly estimated the proportion of cases that were hospital acquired and the balanced cut-off that had equal sensitivity and specificity were identified.

Findings: The recommended two-day cut-off overestimated the incidence of hospital-acquired cases in all scenarios and by >100% in the base scenario. The two-day cut-off had good sensitivity (96%) but poor specificity (48%) and precision (52%) to identify cases acquired during the current hospitalization. A five-day cut-off was balanced, and a six-day cut-off was optimal in the base scenario. The optimal and balanced cut-offs were more than two days for nearly all scenarios considered (ranges: four to nine days and two to eight days, respectively).

Conclusion: Current guidelines for classifying *C. difficile* infections overestimate the proportion of cases acquired in hospital in all model scenarios. To reduce misclassification bias, an infection should be classified as being acquired prior to admission if symptoms begin within five days of admission.

Introduction

Since *Clostridium difficile* was identified as the causative agent for pseudomembranous colitis in the late 1970s, awareness of the pathogen has grown, as has the burden of disease [1,2]. In 2011, there were an estimated 453,000 *C. difficile* infections (CDIs) and 29,300 deaths in the USA alone [3]. Currently *C. difficile* is implicated as the cause of 71% of hospital-associated gastrointestinal infections [4].

Most CDI cases are observed in healthcare facilities, but there is increasing recognition of community-acquired cases [5]. Symptomatic individuals have mild to severe diarrhoea but patients may also carry the pathogen asymptomatically for weeks or months [6,7]. Because of the potentially long

* Corresponding author. Address: 62 Mills Road, 2601, ACT, Australia. Tel.: +61 423060098.
E-mail address: angus.mclure@anu.edu.au (A. McLure).

https://doi.org/10.1016/j.jhin.2017.12.014
0195-6701/

incubation period, patients displaying symptoms for the first time in a healthcare facility may have acquired the pathogen prior to admission, obscuring the source of transmission [8].

The Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA) have published recommendations for classification of CDIs [9] (Figure 1). They recommend that CDIs with onset of symptoms more than two days after admission to a healthcare facility but prior to discharge be classified as healthcare facility-onset, healthcare facility-associated infections. The recommendation is not evidence-based, but intended to be used as a standard for comparison between healthcare facilities and systems. The classification (or a minor variant) is used to estimate the relative contributions of hospital- and community-based transmission, report temporal changes in incidence, compare the incidence of hospital-acquired cases before and after interventions, and as a case definition for studies comparing hospital-acquired and community-acquired cases [3,10–12]. Therefore, it is important that the classification is fit for purpose, i.e. correctly estimates the number of cases that are hospital- or community-acquired and/or sufficiently discriminates between the two groups. Individuals may be colonized with *C. difficile* for longer than two days before showing symptoms. One study found that the median incubation period was 19 days [13]. Another found that the first quartile and median delays from admission to onset of symptoms were eight days and 17 days, respectively [10]. We hypothesized that increasing the cut-off beyond two days will reduce sensitivity, but greatly improve specificity to identify hospital-onset, healthcare facility-associated CDI.

The aim of this study was to model *C. difficile* transmission in a healthcare setting to simulate the interaction of pathogen and patient from admission through to discharge. The model has been previously shown to reproduce hospital-level outcomes such as the proportions of infections occurring within 72 h of admission [14]. We use this model to assess the current guidance for CDI and to identify potential improvements to the method of classification.

Methods

Mathematical model

A detailed description of the model has been published [14]. Briefly, a stochastic compartmental model of *C. difficile* transmission in a hospital was used. The model divides admitted patients into 15 compartments based on immune responses to *C. difficile* toxins (immune, naïve or immunocompromised), *C. difficile* colonization status (negative, colonized, or overgrowth with substantial toxin load), and the status of commensal gut flora (normal or disturbed). This model simulates the time-course of hospitalized individuals, capturing their state from admission to discharge, including exposure to antimicrobials, colonization with *C. difficile*, onset of CDI, treatment and development of immune responses (Figure 2). Settings where mean incidence of CDI is

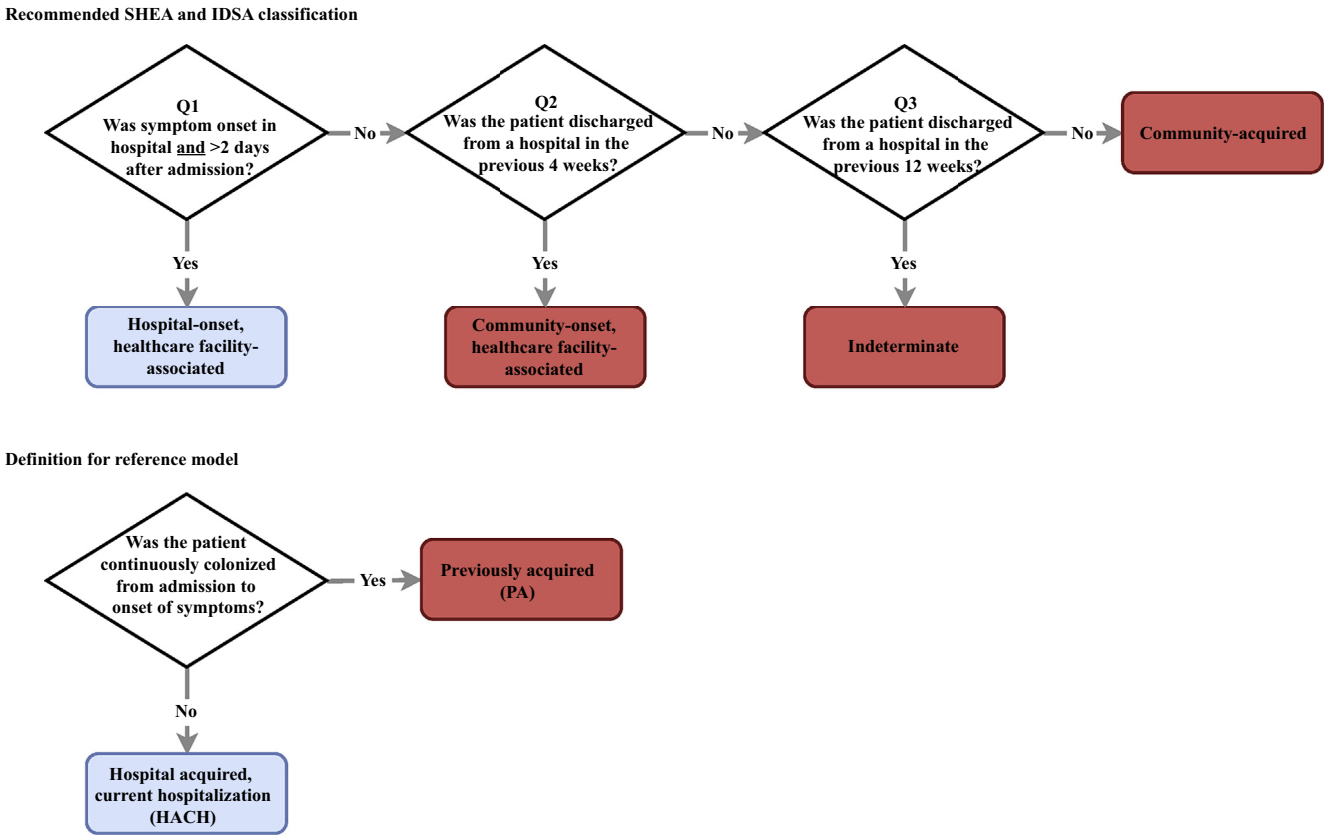


Figure 1. Classification of *Clostridium difficile* infections recommended by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA) compared with the definitions of previously acquired (red boxes) and hospital-acquired in current hospitalization (blue boxes) used in the model.

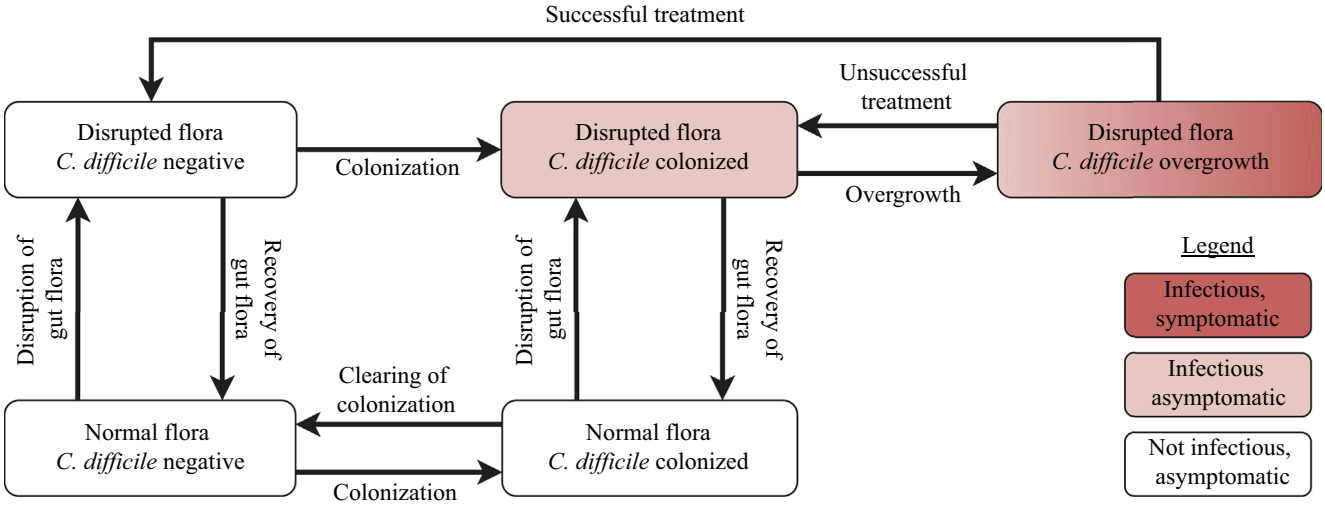


Figure 2. The main states and events simulated in the model. Admitted patients may be in any of the above states, and may change states throughout their hospitalization. Each of the five states is further divided according immunity to *Clostridium difficile* toxins. Only patients with disrupted intestinal flora and *C. difficile* overgrowth but no immunity to toxins are symptomatic for CDI; however, all *C. difficile*-positive patients with disrupted intestinal flora are infectious.

constant were considered and therefore a constant force of colonization was assumed, allowing use of the individual Markov chain approximation described in previous work to classify CDI origin [14]. This enabled estimation of the probability distribution of patient outcomes at the individual level (e.g. probability that a patient colonized at admission did not develop symptoms within two days of admission, but did develop symptoms prior to discharge) without running individual-based simulations.

Case definitions for origin of infection

Though our model simulates recurrent CDI, only the first period of CDI experienced by patients during their hospitalization was considered. All estimates of incidence and classification were performed for this first CDI episode only. The scope of the model was limited to hospitalized patients, so we could not consider any periods of CDI preceding or following the simulated hospitalization. We therefore were unable to assess the classification of patients by history of recent hospitalization and instead focused on events occurring during a single hospitalization. An infection was considered previously acquired (PA-CDI) if the patient was colonized at admission and was continuously colonized until the onset of symptoms, including where patients had symptoms on admission. This definition necessarily included all community-acquired cases, but also included cases where the infection was acquired during a previous hospitalization. All other CDIs were considered 'hospital-acquired in the current hospitalization' (HACH-CDI). This included CDIs where the patient was not colonized at admission and CDIs where the patient cleared the initial colonization and was re-colonized in hospital prior to the onset of symptoms. These definitions and the model were used to calculate the distribution of time between admission and onset of symptoms for HACH-CDI and PA-CDI.

Assessing the classification of origin of infection by time since admission

The classification of cases was assessed using the time between admission and onset of symptoms, emulating the first step in the IDSA and SHEA recommendations (Figure 1). For a two-day cut-off, all CDIs with onset of symptoms before the cut-off were classified as PA-CDI, with all remaining CDIs classified as HACH-CDI. The incidence of CDIs classified as HACH-CDI or PA-CDI was calculated, and the proportions of these that were correctly and incorrectly classified by comparison with the true history of individuals in the model were recorded. To identify potential improvements to the classification, this process was repeated for different cut-off times from 0.01 to 40 days.

It is usual to design binary classification systems so that they balance sensitivity (the proportion of 'positives' correctly classified) and specificity (the proportion of 'negatives' correctly classified). The cut-off that achieves this was identified and called the 'balanced' cut-off. However, such classifications misclassify larger numbers of individuals from the majority class than from the minority class, overestimating the incidence of the latter. Therefore, the 'optimal' cut-off time that balanced sensitivity with precision (the proportion of individuals classified 'positive' that are actually 'positive') was also determined. Using this cut-off there was one incorrectly classified PA-CDI for each incorrectly classified HACH-CDI, and the total numbers of cases classified as either HACH-CDI or PA-CDI were equal to the true numbers of HACH-CDI or PA-CDI cases.

To determine whether PA-CDI and HACH-CDI cases could be differentiated by the time from admission to onset of symptoms, the concordance probability was calculated, namely the probability that the time since admission to onset of symptoms would be greater in a randomly chosen HACH-CDI than a randomly chosen PA-CDI.

Sensitivity analysis

The parameter values for the base scenario, which reflected a moderate CDI incidence setting, were chosen based on previous work [14]. Sensitivity analysis has shown that the two most influential parameters are those governing the person-to-person transmission rate and the mean length of stay for patients admitted overnight. In addition, the proportion colonized at admission and the mean time for *C. difficile* overgrowth to occur in colonized patients with disturbed gut flora were identified as factors likely to have a significant impact on incidence or the time-course of infection and therefore affect the classification of CDIs. The balanced and optimal cut-off times and the concordance probability were calculated, varying each of these parameters independently.

Results

In the base scenario, the recommended two-day cut-off had good sensitivity but poor specificity to identify CDI acquired in the current hospitalization (Figure 3), overestimating the proportion of CDIs acquired in the current hospitalization by nearly 100% (Figure 4). Longer cut-offs decreased the sensitivity but increased specificity to identify CDI acquired in the current hospitalization. A five-day cut-off was balanced, and a six-day cut-off was optimal (Figure 3). Symptom onset for previously acquired cases was generally closer to the time of admission than cases acquired during the current hospitalization. The concordance probability — the probability that the time from admission to onset of symptoms was shorter in a random previously acquired CDI than in a random CDI acquired in the current hospitalization — was 0.842 in the base scenario. This result was insensitive to the assumptions about the hospital setting, falling between 0.83 and 0.88 for all parameter values considered in the sensitivity analysis (Supplementary Figure 1, Appendix A).

The optimal and balanced cut-off times depended on the characteristics of the hospital setting and our assumption about the rate at which *C. difficile* overgrowth occurs in colonized patients (Figure 5). In our sensitivity analysis, the optimal cut-off time was most sensitive to the person-to-person transmission rate, whereas the balanced cut-off time was most sensitive to the mean length of stay. Both the optimal and balanced cut-offs were somewhat sensitive to the mean

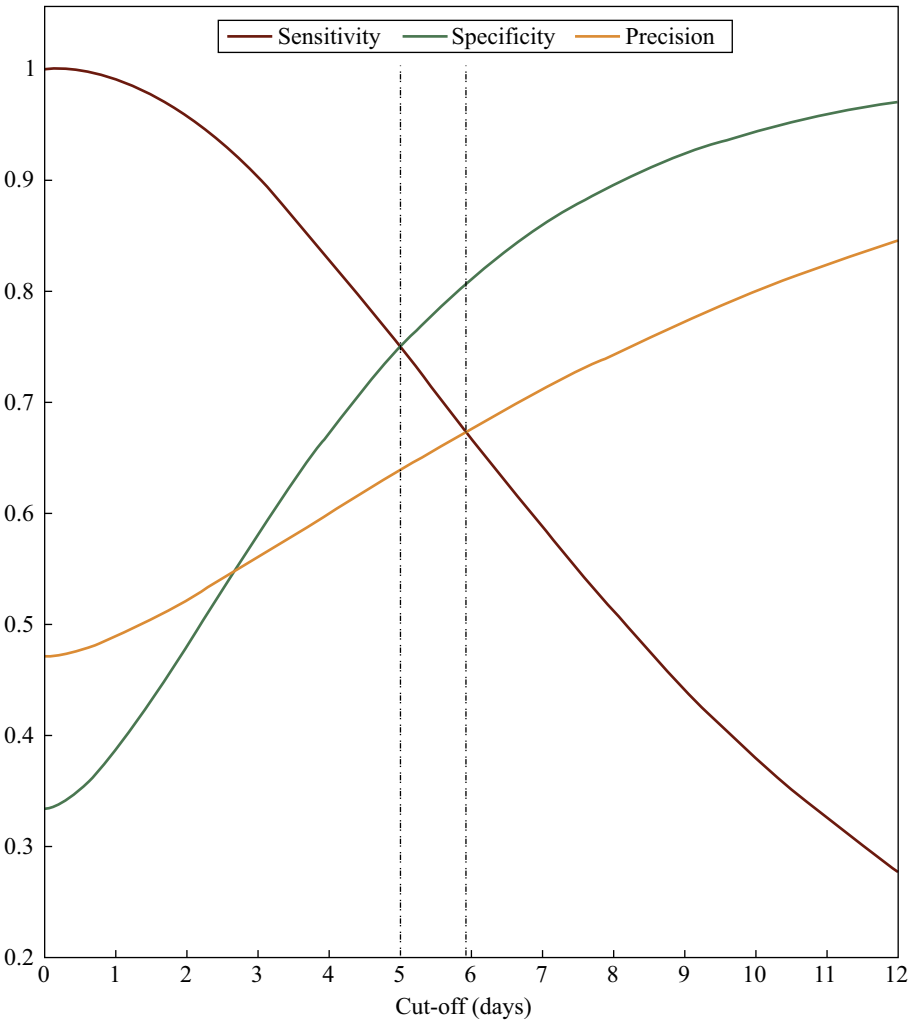


Figure 3. Sensitivity, specificity, and precision of identifying CDIs acquired in the current hospitalization by time since admission. Results are shown for the base scenario only. Sensitivity and specificity are equal with a five-day cut-off. The optimal cut-off (equal sensitivity and precision) is longer at 5.9 days.

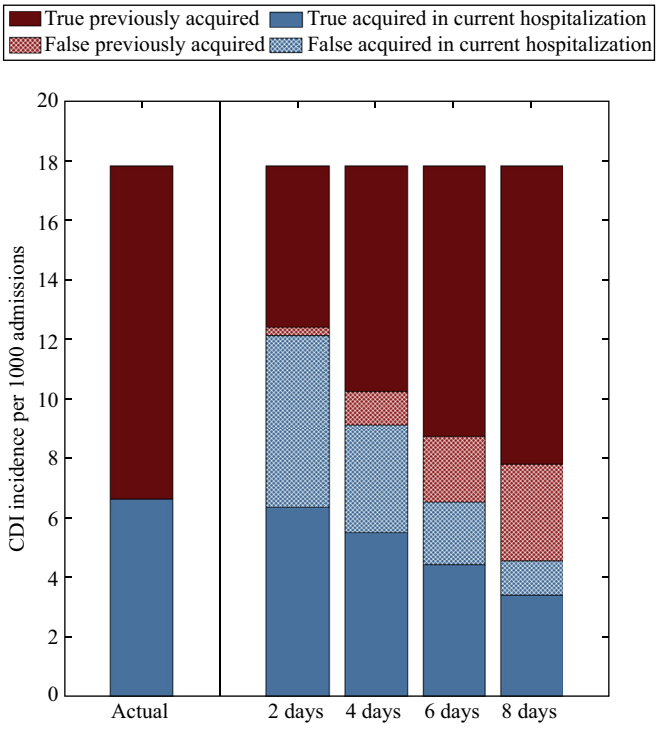


Figure 4. Classification of the origin of *Clostridium difficile* infection (CDI) by time since admission for selected cut-offs, in the base scenario. Shorter cut-offs detect most cases acquired in the current hospitalization but misclassify many previously acquired cases, overestimating the proportion of CDIs that are acquired in the current hospitalization. A cut-off of six days overestimates neither.

time to *C. difficile* overgrowth. The optimal cut-off was longest in settings with low person-to-person transmission, whereas the balanced cut-off was longest in settings with longer mean length of stay. Both the optimal and balanced cut-offs were longer when *C. difficile* overgrowth was assumed to develop more slowly.

A two-day cut-off was not optimal for any of the scenarios considered. No cut-off time was optimal for all scenarios; however, some cut-offs resulted in only moderate over- or underestimation for a wide range of parameters (Figure 5). A cut-off of ~5.5 days did not over- or underestimate incidence of hospital-acquired or previously acquired CDI by more than 20% for a wide range of mean times to *C. difficile* overgrowth (1–9 days), proportion colonized at admission (0.1–15%), mean length of stay (3–16 days), and rate of person-to-person transmission (0.075–0.14). In our sensitivity analysis, the scenario with the shortest optimal cut-off (3.6 days) was the scenario with very high person-to-person transmission (0.18). This extreme scenario had double the person-to-person transmission of the base scenario, and resulted in 45 hospital-acquired CDIs per 10,000 patient-days.

The classification error of a two-day cut-off was much higher in settings with less person-to-person transmission and substantially higher in settings with shorter mean length of stay (Supplementary Figure 2, Appendix A). If transmission was set to 33% of the base rate, a two-day cut-off overestimated the incidence of CDI acquired in the current hospitalization by more than 350%. Even a six-day cut-off overestimated the incidence by 100% in this low-transmission setting. However, the balanced cut-off was only slightly higher in a low-incidence setting (5.0 days) than in a high-incidence setting (4.9 days).

Discussion

Time from admission to onset of symptoms is a reasonable measure for discriminating CDIs acquired in the current hospitalization from previously acquired CDIs. IDSA and SHEA recommend a cut-off of two days for the classification of hospital-onset CDIs as community- or hospital-acquired. This cut-off systematically overestimates the proportion of CDIs that are acquired in the current hospitalization and underestimates the proportion that is acquired prior to admission. Since all community-acquired CDIs observed in healthcare settings must be acquired prior to admission, the current guidelines may also systematically underestimate the proportion of cases that are community-acquired. Moreover, the low specificity of the two-day cut-off for identifying hospital-acquired cases may cause significant misclassification bias in studies that compare hospital and community-acquired cases, reducing apparent differences between the two groups.

Since this model did not differentiate strains of *C. difficile*, our definition of previously acquired CDI does not exclude patients who were colonized at admission but subsequently acquired an additional strain of *C. difficile* prior to the onset of symptoms. This is unlikely to represent a significant portion of previously acquired CDIs and so is unlikely to affect our recommendations [15]. The structure and parameter values in our model synthesize the peer review literature on hospital-associated CDIs, and is not fitted to a single data set. However, our key finding, that the standard classification of CDIs overestimates the proportion of cases acquired in the current hospitalization, is robust to very large variations in all parameters [14].

There is a large variation in mean hospital length-of-stay worldwide. Within the OECD, mean lengths of stay range between 3.9 days (Turkey) and 17.2 days (Japan) [16]. Our recommended optimal cut-off is sensitive to the mean length of stay but a cut-off of five days performs well over this entire range. By contrast, a cut-off of two days consistently overestimates the incidence of CDI acquired in the current hospitalization and overestimates the incidence of CDI acquired in the current hospitalization by more than 100% when the mean length of stay is less than six days.

The rate of person-to-person transmission is difficult to measure directly and is likely to vary significantly between settings due to differences in hygiene protocols and adherence to these protocols. The degree of person-to-person transmission in our base scenario was estimated from hospitals with high incidence of CDI (28.1 cases per 10,000 patient days), and therefore may lie in the upper end of the plausible range [14,17]. The optimal cut-off is longer in settings with less person-to-person transmission. Therefore, in many settings – especially those that have effective infection control programmes – an even longer cut-off may be required to avoid overestimating the incidence of CDI acquired in the current hospitalization. Since it is not usually possible to estimate the

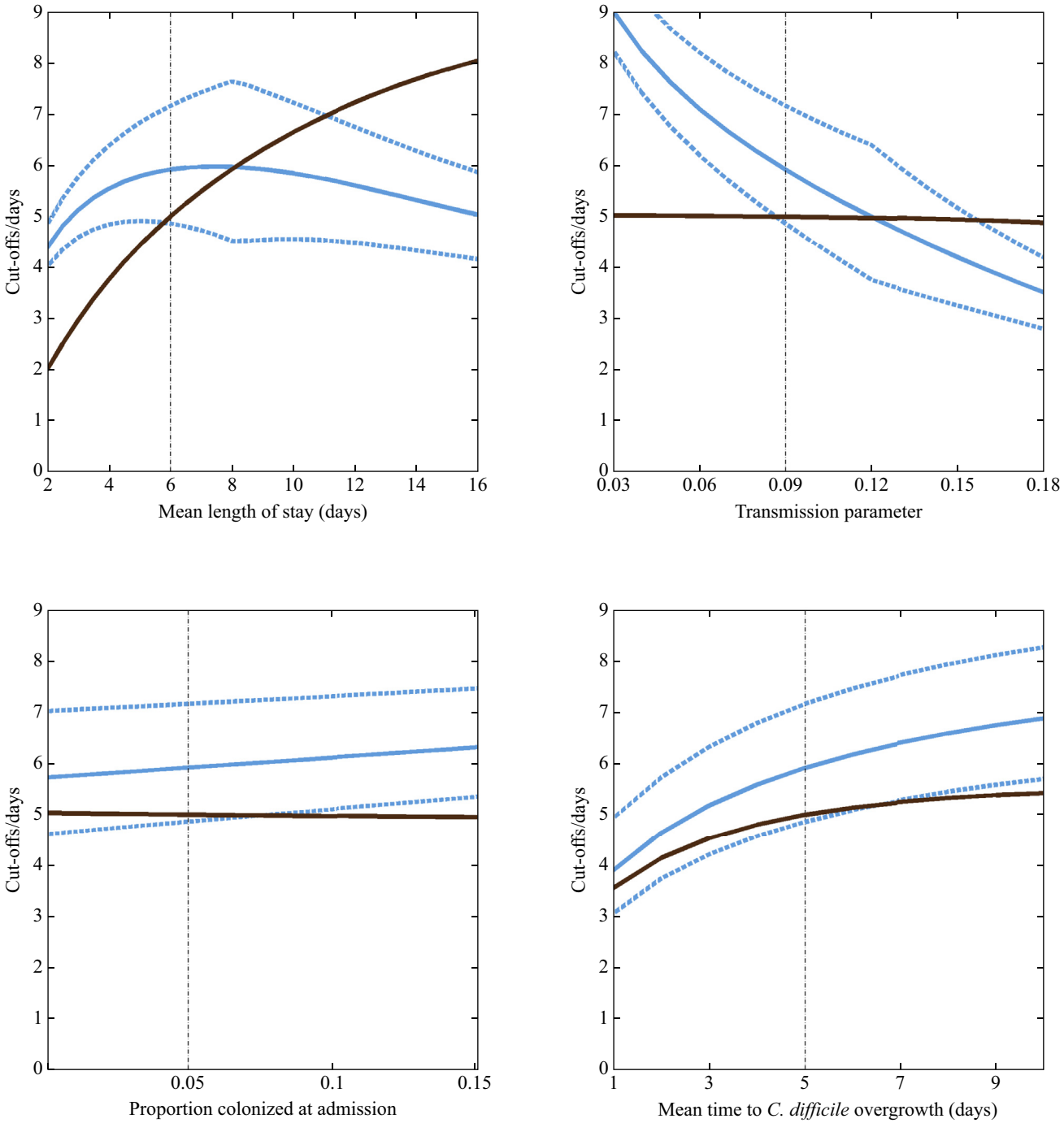


Figure 5. The effect of four parameters on the optimal (light blue: equal sensitivity and precision) and balanced (brown: equal sensitivity and specificity) cut-off times for classifying the origin of *Clostridium difficile* infection (CDI) by time since admission to onset of symptoms. Light blue dashed curves indicate range of cut-offs that over/underestimate the incidence of previously acquired CDIs and CDIs acquired in the current hospitalization by ≤20%. The vertical dashed lines mark the values of the parameters in the base scenario.

rate of person-to-person transmission without knowing the incidence of hospital-acquired infection and colonization, and since most estimates of the incidence of hospital-acquired cases are based on the very classification scheme we are assessing, in practice we cannot calculate the optimal cut-off for a given setting. However, a five-day cut-off balances sensitivity and specificity independent of the rate of person-to-person transmission and is approximately optimal for a range of transmission rates.

IDSA and SHEA also recommend that only those cases arising >84 days (12 weeks) after the most recent hospital discharge should be classified as community acquired [9]. This recommendation may also lead to systematic misclassification, with the extent of misclassification determined by the choice of

cut-off. However, the scope of our model was limited to hospitalized patients, so we could not consider any periods of CDI or asymptomatic colonization preceding or following simulated hospitalizations. Therefore we are unable to make recommendations for the optimal use of a patient's history of hospitalization to classify the origin of CDI for either community- or hospital-onset CDI. This classification should also be assessed with empirical studies or models that simulate patients in communities and hospitals.

It is difficult to determine the source of transmission for CDI cases, and therefore to assess classification by time since admission empirically. Since the same strains of *C. difficile* circulate in hospitals and communities, cases cannot be distinguished by strain type alone [18,19]. Whole genome sequencing can identify transmission events, but the most comprehensive studies have not sequenced isolates from many asymptomatic carriers, and have not been able to identify a transmission source for $\geq$75% of all infections [8,18]. Screening all admissions for asymptomatic colonization, coupled with contact precautions and antimicrobial stewardship for colonized patients, may reduce the incidence of CDI [12,20]. Consequently, studies where screening has occurred may not be representative and may thus be unsuitable for assessing classification of the origin of infections in settings without screening. Therefore, modelling-based approaches may be the best means for assessing the classification of CDIs.

Our findings add to a growing body of evidence suggesting that transmission and reservoirs of *C. difficile* outside hospitals are as least as important as within-hospital transmission. Detailed surveillance has found the same strains circulating in communities and hospitals, demonstrating the interconnectedness of the two populations [18,19]. Further, modelling studies have shown that the reproduction number (the number of secondary colonizations arising from a typical primary colonization in a population of susceptible individuals) is less than one in many healthcare settings, suggesting that CDI is sustained primarily by the admission of colonized individuals, not within-hospital transmission [14,21]. Moreover, only 19% of all CDIs and 25% of CDIs in hospitals can be reasonably attributed to transmission from symptomatic inpatients, with the remainder acquired from asymptomatic carriers or sources in the community [8,18]. Whereas hospital-onset CDIs are carefully monitored and reported, community-onset CDIs are likely to be underreported — especially in patients who have not been hospitalized recently [22,23].

Standardized definitions and reporting of hospital-acquired *C. difficile* infections have value, but the current two-day cut-off is not based on strong evidence and overestimates the proportion of cases acquired during the current hospitalization. Though it may be difficult to change reporting standards, adopting a five- or six-day cut-off will improve the classification of potential sources of infection for *C. difficile*, recognizing the key role of CDI acquired prior to hospital admission, including community-acquired cases.

Conflicts of interest
None declared.

Funding sources
None.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jhin.2017.12.014.

References

[1] Bartlett JG, Chang TW, Moon N, Onderdonk AB. Antibiotic-induced lethal enterocolitis in hamsters: studies with eleven agents and evidence to support the pathogenic role of toxin-producing *Clostridia*. Am J Vet Res 1978;39:1525–30.
[2] Burke KE, Lamont JT. *Clostridium difficile* infection: a worldwide disease. Gut Liver 2014;8:1–6.
[3] Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, et al. Burden of *Clostridium difficile* infection in the United States. N Engl J Med 2015;372:825–34.
[4] Magill SS, Edwards JR, Bamberg W, Beldavs ZG, Dumyati G, Kainer MA, et al. Multistate point-prevalence survey of health care-associated infections. N Engl J Med 2014;370:1198–208.
[5] Hensgens MPM, Keessen EC, Squire MM, Riley TV, Koene MGJ, de Boer E, et al. *Clostridium difficile* infection in the community: a zoonotic disease? Clin Microbiol Infect 2012;18:635–45.
[6] Johnson S, Homann SR, Bettin KM, Quick JN, Clabots CR, Peterson LR, et al. Treatment of asymptomatic *Clostridium difficile* carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. Ann Intern Med 1992;117:297–302.
[7] Otete EH, Ahankari AS, Jones H, Bolton KJ, Jordan CW, Boswell TC, et al. Parameters for the mathematical modelling of *Clostridium difficile* acquisition and transmission: a systematic review. PLoS One 2013;8:e84224.
[8] Walker AS, Eyre DW, Wyllie DH, Dingle KE, Harding RM, O'Connor L, et al. Characterisation of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. PLoS Med 2012;9:e1001172.
[9] Cohen SH, Gerding DN, Johnson S, Kelly CP, Loo VG, McDonald LC, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). Infect Control Hosp Epidemiol 2010;31:431–55.
[10] Lambert PJ, Dyck M, Thompson LH, Hammond GW. Population-based surveillance of *Clostridium difficile* infection in Manitoba, Canada, by using interim surveillance definitions. Infect Control Hosp Epidemiol 2009;30:945–51.
[11] Slimings C, Armstrong P, Beckingham WD, Bull AL, Hall L, Kennedy KJ, et al. Increasing incidence of *Clostridium difficile* infection, Australia, 2011–2012. Med J Aust 2014;200:272–6.
[12] Longtin Y, Paquet-Bolduc B, Gilca R, Garenc C, Fortin E, Longtin J, et al. Effect of detecting and isolating *Clostridium difficile* carriers at hospital admission on the incidence of *C. difficile* infections. JAMA Intern Med 2016;176:796.
[13] Samore MH, Venkataraman L, DeGirolami PC, Arbeit RD, Karchmer AW. Clinical and molecular epidemiology of sporadic and clustered cases of nosocomial *Clostridium difficile* diarrhea. Am J Med 1996;100:32–40.
[14] McLure A, Clements ACA, Kirk M, Glass K. Healthcare-associated *Clostridium difficile* infections are sustained by disease from the community. Bull Math Biol 2017;79:2242–57.
[15] Eyre DW, Walker AS, Griffiths D, Wilcox MH, Wyllie DH, Dingle KE, et al. *Clostridium difficile* mixed infection and reinfection. J Clin Microbiol 2012;50:142–4.
[16] Organisation for Economic Co-operation and Development. Average length of stay in hospitals. Health at a glance 2011: OECD indicators. OECD Publishing; 2011.

[17] Loo VG, Bourgault A-M, Poirier L, Lamothe F, Michaud S, Turgeon N, et al. Host and pathogen factors for *Clostridium difficile* infection and colonization. N Engl J Med 2011;365:1693—703.
[18] Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med 2013;369:1195—205.
[19] Fawley WN, Davies KA, Morris T, Parnell P, Howe R, Wilcox MH, et al. Enhanced surveillance of *Clostridium difficile* infection occurring outside hospital, England, 2011 to 2013. Euro Surveill 2016;21(29).
[20] Zacharioudakis IM, Zervou FN, Pliakos EE, Ziakas PD, Mylonakis E. Colonization with toxinogenic C. difficile upon hospital admission, and risk of infection: a systematic review and meta-analysis. Am J Gastroenterol 2015;110:381—90.
[21] Lanzas C, Dubberke ER, Lu Z, Reske KA, Gröhn YT. Epidemiological model for *Clostridium difficile* transmission in healthcare settings. Infect Control Hosp Epidemiol 2011;32:553—61.
[22] Hensgens MPM, Dekkers OM, Demeulemeester A, Buiting AGM, Bloembergen P, van Benthem BHB, et al. Diarrhoea in general practice: when should a *Clostridium difficile* infection be considered? Results of a nested case—control study. Clin Microbiol Infect 2014;20:O1067—74.
[23] Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, et al. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. BMJ 1999;318:1046—50.

Supplementary Appendix A

Model parameterization

The details of the model we have used to assess the classification of *Clostridium difficile* transmission in hospitals have been published elsewhere [1]. The parameter values and the ranges used for sensitivity analysis in our article, together with references, are shown in Supplementary Table I.

Supplementary Table I

Definitions, values of all parameters used in the model (all rates are in units of per day)

Parameter	Description	Value (range for sensitivity analysis)	References
β	Transmission rate coefficient	0.09 (0.03–0.18)[a]	[1,2][b]
$1/\kappa$	Mean length of stay (days) for patients without CDI	6 (2–16)[c]	[3]
α	Rate at which patients begin new antimicrobial treatment (proxy for rate of gut flora disruption)	0.24	[2,4]
q	Efficacy and coverage of special contact precautions (0: total prevention of transmission; 1: no reduction in transmission)	0.45	[5,6]
N_{bed}	Hospital capacity	100	
τ	Rate at which the treatment of patients with active CDI progresses towards resolution	0.1	[7,8]
p_t	Probability that CDI treatment will remove all *C. difficile*	0.55	[1,2][d]
μ	All-cause mortality rate for patients with CDI	0.0075	[9,10]
p_c	Proportion of admissions positive for *C. difficile*	0.05 (0.001–0.15)[e]	[2,11]
p_a	Proportion of admissions with disrupted gut flora (due to recent antibiotic exposure)	0.14	[12]
p_o	Proportion of admissions with *C. difficile* overgrowth	0.006	[2]
p_u	Proportion of admissions unable to mount a sufficient immune response	0.46	[13,14]
p_{AB}	Proportion of admissions positive for *C. difficile* toxin antibodies	0.7	[15]

λ	Rate at which damaged commensal gut flora recovers	0.03	[16,17]
γ	Rate at which *C. difficile* is cleared in those with intact gut flora	0.023	[18]
$1/\omega$	Mean time (days) until *C. difficile* overgrowth for patients with disrupted gut flora	5 (1–10)[a]	[19,20]
δ	Rate of development of immunity	0.1	[21,22]

CDI, *Clostridium difficile* infection.

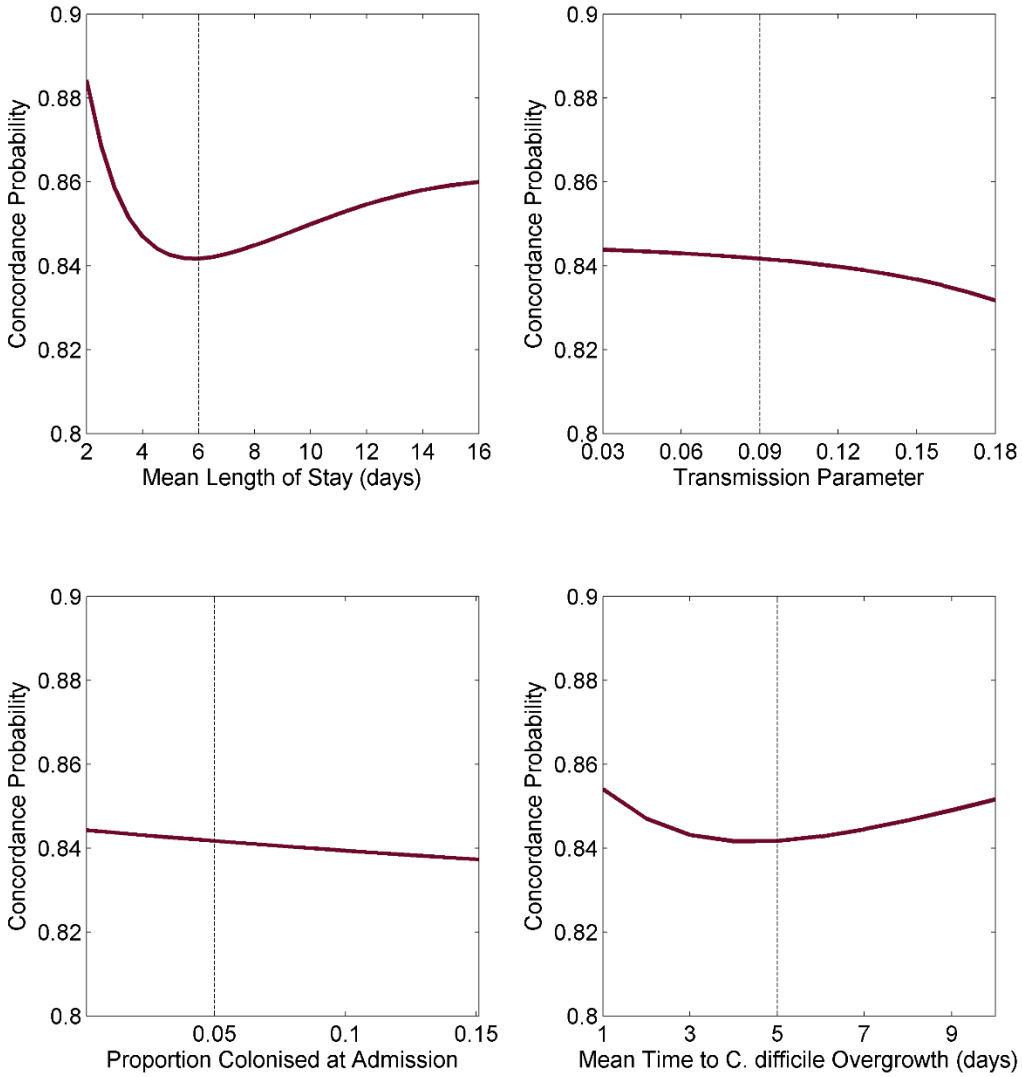
[a]The range is that reported for the general population.

[b]Estimated in [1] from the force of colonization inferred from the risk of colonization and infection as a function of length of stay reported in [2].

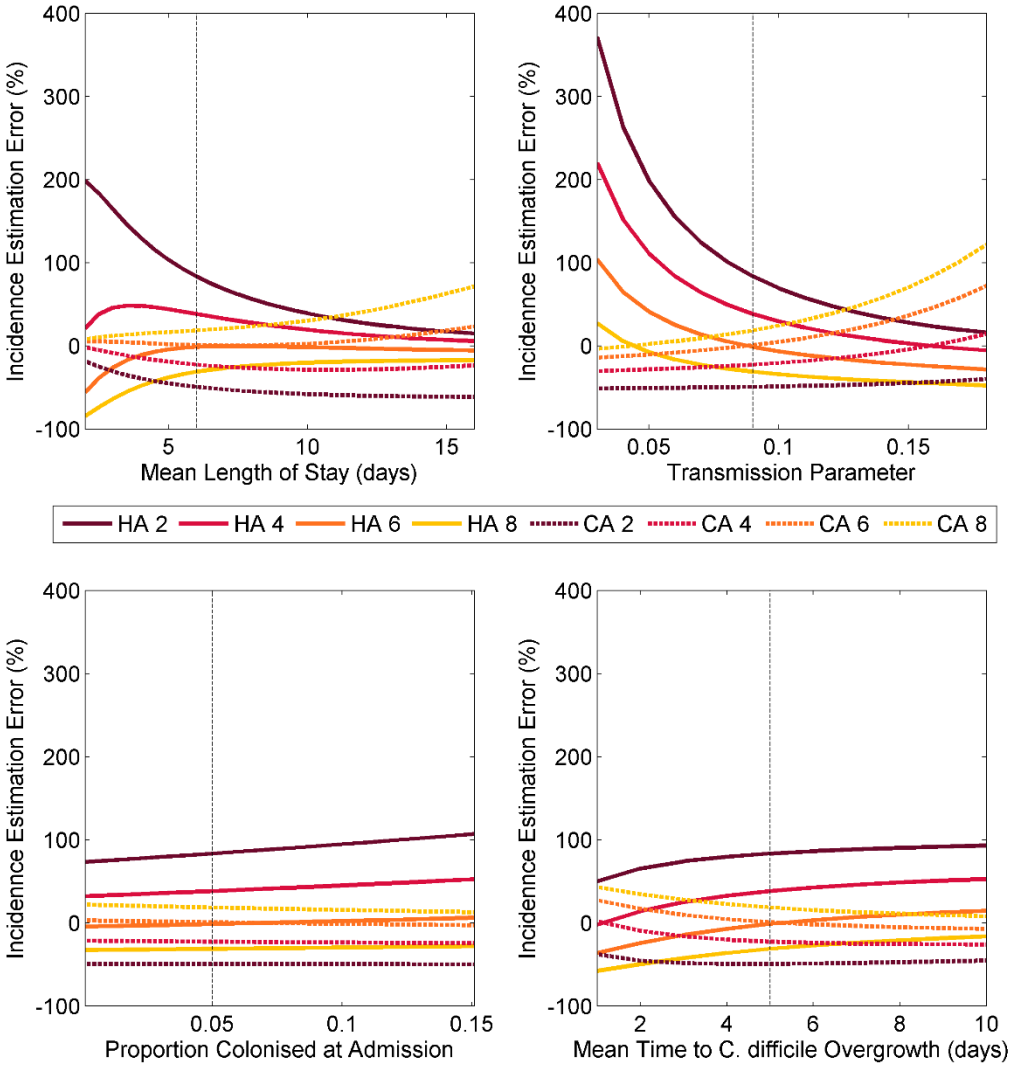
[c]The range of length of stay is from the Organisation for Economic Co-operation and Development (excluding Japan, which used a different definition).

[d]Estimated in [1] from data on the recurrence rate reported in [2].

[e]Limited data are available for this parameter. We have chosen a range we believe to be very wide.



Supplementary Figure 1. The effect of four parameters on the concordance probability for classification of *Clostridium difficile* infection origin by time since admission to onset of symptoms.



Supplementary Figure 2. The effect of four parameters on the classification of *Clostridium difficile* infection (CDI) origin by time since admission to onset of symptoms. The estimation error (as a percentage of actual incidence) for the incidence of previously acquired CDIs and CDIs acquired in the current hospitalization is shown using four different classification cut-offs (two, four, six, and eight days). Positive error is overestimation and negative error is underestimation.

References

[1] McLure A, Clements ACA, Kirk M, Glass K. Healthcare-associated *Clostridium difficile* infections are sustained by disease from the community. Bull Math Biol 2017;79:2242–57.

[2] Loo VG, Bourgault A-M, Poirier L, Lamothe F, Michaud S, Turgeon N, et al. Host and pathogen factors for *Clostridium difficile* infection and colonization. N Engl J Med 2011;365:1693–1703.

[3] Organisation for Economic Co-operation and Development. Average length of stay in hospitals. Health at a glance 2011: OECD indicators. OECD Publishing; 2011.

[4] Polk RE, Hohmann SF, Medvedev S, Ibrahim O. Benchmarking risk-adjusted adult antibacterial drug use in 70 US academic medical center hospitals. Clin Infect Dis 2011;53:1100–10.

[5] Muto CA, Pokrywka M, Shutt K, Mendelsohn AB, Nouri K, Posey K, et al. A large outbreak of *Clostridium difficile*-associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. Infect Control Hosp Epidemiol 2005;26:273–80.

[6] Jarvis WR, Schlosser J, Jarvis AA, Chinn RY. National point prevalence of *Clostridium difficile* in US health care facility inpatients, 2008. Am J Infect Control 2009;37:263–70.

[7] Leffler DA, Lamont JT. Treatment of *Clostridium difficile*-associated disease. Gastroenterology 2009;136:1899–1912.

[8] Khan FY, Abu-Khattab M, Anand D, Baager K, Alaini A, Siddique M, et al. Epidemiological features of *Clostridium difficile* infection among inpatients at Hamad General Hospital in the state of Qatar, 2006–2009. Travel Med Infect Dis 2012;10:179–85.

[9] Longtin Y, Paquet-Bolduc B, Gilca R, Garenc C, Fortin E, Longtin J, et al. Effect of detecting and isolating *Clostridium difficile* carriers at hospital admission on the incidence of *C. difficile* infections. JAMA Intern Med 2016;176:796.

[10] Hensgens MPM, Goorhuis A, Dekkers OM, van Benthem BHB, Kuijper EJ. All-cause and disease-specific mortality in hospitalized patients with *Clostridium difficile* infection: a multicenter cohort study. Clin Infect Dis 2013;56:1108–16.

[11] Furuya-Kanamori L, Marquess J, Yakob L, Riley TV, Paterson DL, Foster NF, et al. Asymptomatic *Clostridium difficile* colonization: epidemiology and clinical implications. BMC Infect Dis 2015;15:516.

[12] Pedersen G, Schonheyder HC, Steffensen FH, Sorenson HT. Risk of resistance related to antibiotic use before admission in patients with community-acquired bacteraemia. J Antimicrob Chemother 1999;43:119–26.

[13] Australian Institute of Health and Welfare. Admitted patient care 2013–14: Australian hospital statistics. Canberra; 2015.

[14] National Center for Health Statistics. Health, United States, 2013: with special feature on prescription drugs. Hyattsville: NCHS; 2014.

[15] Kelly CP, Kyne L. The host immune response to *Clostridium difficile*. J Med Microbiol

2011;60:1070–9.

[16] Rafii F, Sutherland JB, Cerniglia CE. Effects of treatment with antimicrobial agents on the human colonic microflora. Ther Clin Risk Manag 2008;4:1343–58.

[17] Dial S, Kezouh A, Dascal A, Barkun A, Suissa S. Patterns of antibiotic use and risk of hospital admission because of *Clostridium difficile* infection. Can Med Assoc J 2008;179:767–72.

[18] Johnson S, Homann SR, Bettin KM, Quick JN, Clabots CR, Peterson LR, et al. Treatment of asymptomatic *Clostridium difficile* carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. Ann Intern Med 1992;117:297–302.

[19] Freeman J. Comparison of the efficacy of ramoplanin and vancomycin in both in vitro and in vivo models of clindamycin-induced *Clostridium difficile* infection. J Antimicrob Chemother 2005;56:717–25.

[20] Freeman J. Effects of cefotaxime and desacetylcefotaxime upon *Clostridium difficile* proliferation and toxin production in a triple-stage chemostat model of the human gut. J Antimicrob Chemother 2003;52:96–102.

[21] Foglia G, Shah S, Luxemburger C, Pietrobon PJF. *Clostridium difficile*: development of a novel candidate vaccine. Vaccine 2012;30:4307–9.

[22] de Bruyn G, Saleh J, Workman D, Pollak R, Elinoff V, Fraser NJ, et al. Defining the optimal formulation and schedule of a candidate toxoid vaccine against *Clostridium difficile* infection: a randomized Phase 2 clinical trial. Vaccine 2016;34:2170–8.

Chapter 5

Modelling diverse sources of *Clostridium difficile* in the community

5.1 Introduction

This chapter consists of an article published in *Epidemiology and Infection* and the accompanying supplementary materials. In this artcile I extend the model of *C. difficile* transmission in a hospital introduced in Chapter 3 to model transmission in a hospital and the surrounding community. The model also incorporates infants and the possibility of transmission from non-human reservoirs of *C. difficile*. In Chapter 4 I assessed commonly used surveillance definitions that classify CDIs presenting at hospitals as acquired during the current hospitalisation or acquired prior to admission. In this chapter I extend this assessment to the complete surveillance definitions that are used to classify hospital and community-onset infections as hospital or community-acquired. I use an ODE formulation of the model, assuming the population is at endemic equilibrium, and use the individual Markov chain approximation to calculate the findings that require a knowledge of individual patients' histories. The article has four main outcomes. First, transmission in the hospital is not necessary to sustain transmission in the community. Second, even small and plausible amounts of transmission from animal reservoirs are enough to imply that transmission in the community – though primarily person-to-person – requires animal exposure to be sustained. Third, symptomatic carriers account for less than 10% of person-to-person transmission in the community, with infants and asymptomatically colonised adults accounting for the remainder. Fourth, community-onset *C. difficile* infections are under-reported and current surveillance definitions are unable to adequately distinguish community-onset hospital-acquired cases from community-onset community-acquired cases, vastly overestimating the proportion of *C. difficile* infections that are hospital acquired.

The supplementary materials provide a detailed description of the model structure, parameters, fitting and verification. The supplementary materials also provide additional figures showing the impact of model assumptions on the role of infants and asymptomatic

carriers to transmission. The supplementary materials conclude by outlining the method used to simulate individual patient histories when emulating surveillance definitions.

Chapter 6 uses the model presented in this chapter as the basis for an analysis of *C. difficile* seasonality and interventions. The threshold for transmission from animal sources (which if exceeded implies *C. difficile* in the human population is driven by transmission from animals) is extended in Chapter 7 to simple, general models with both local person-to-person transmission and an external source of colonisation or infection.

5.2 Article and Supplementary materials

Angus McLure, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. Diverse sources of *Clostridium difficile* in the community: importance of animals, infants and asymptomatic carriers. *Epidemiology and Infection*, 147:e152, 2019.

Epidemiology and Infection

cambridge.org/hyg

Modelling diverse sources of *Clostridium difficile* in the community: importance of animals, infants and asymptomatic carriers

A. McLure[1], A. C. A. Clements[1,2], M. Kirk[1] and K. Glass[1]

[1]Research School of Population Health, Australian National University, Canberra, Australian Capital Territory, Australia and [2]Faculty of Health Sciences, Curtin University, Perth, Western Australia, Australia

Original Paper

Cite this article: McLure A, Clements ACA, Kirk M, Glass K (2019). Modelling diverse sources of *Clostridium difficile* in the community: importance of animals, infants and asymptomatic carriers. *Epidemiology and Infection* **147**, e152, 1–9. https://doi.org/10.1017/S0950268819000384

Received: 3 September 2018
Revised: 30 November 2018
Accepted: 4 February 2019

Key words:
Clostridium difficile; community-acquired infection; hospital-acquired infection; mathematical disease model; zoonotic infection

Author for correspondence:
A. McLure, E-mail: angus.mclure@anu.edu.au

Abstract

Clostridium difficile infections (CDIs) affect patients in hospitals and in the community, but the relative importance of transmission in each setting is unknown. We developed a mathematical model of *C. difficile* transmission in a hospital and surrounding community that included infants, adults and transmission from animal reservoirs. We assessed the role of these transmission routes in maintaining disease and evaluated the recommended classification system for hospital- and community-acquired CDIs. The reproduction number in the hospital was <1 (range: 0.16–0.46) for all scenarios. Outside the hospital, the reproduction number was >1 for nearly all scenarios without transmission from animal reservoirs (range: 1.0–1.34). However, the reproduction number for the human population was <1 if a minority (>3.5–26.0%) of human exposures originated from animal reservoirs. Symptomatic adults accounted for <10% transmission in the community. Under conservative assumptions, infants accounted for 17% of community transmission. An estimated 33–40% of community-acquired cases were reported but 28–39% of these reported cases were misclassified as hospital-acquired by recommended definitions. Transmission could be plausibly sustained by asymptomatically colonised adults and infants in the community or exposure to animal reservoirs, but not hospital transmission alone. Under-reporting of community-onset cases and systematic misclassification underplays the role of community transmission.

Introduction

Clostridiodes difficile, more commonly known as *Clostridium difficile*, is an emerging pathogen that causes potentially life-threatening diarrhoea and is increasing in burden in many parts of the world [1–3]. In the USA, it caused an estimated 453 000 infections and contributed to 29 300 deaths in 2011 [3]. *C. difficile* infections (CDIs) are common in healthcare facilities where they account for 71% of hospital-associated gastrointestinal infections [4], but there is increasing recognition of community-acquired cases and healthcare-acquired cases with onset of symptoms in the community [3]. It is likely that many CDIs in the community go unreported, either because affected people do not seek treatment [5], do not submit a stool sample when they seek treatment [5] or their stool sample is not tested for *C. difficile* when submitted [6]. However, the extent of under-reporting has not known.

Colonised infants [7–10], contaminated food [11] and animals reservoirs [12] have been identified as possible sources of *C. difficile* outside hospitals, however their contribution to transmission has not been well quantified. Infants under 12 months have much higher prevalence of colonisation than adults [13], can be colonised for over 6 months by a single strain [7] and rarely develop symptoms but shed the same density of spores in their faeces as adults with CDI [8]. However, existing models of *C. difficile* do not capture infant colonisation or their potential role in transmission. Some strains of toxigenic *C. difficile* that cause disease in humans are also isolated from livestock, meat and fresh produce contaminated by animal faeces [11, 12]. However, the proportion of human cases that are acquired from food or animals and the ramifications for disease control are unknown.

The Infectious Disease Society of America (IDSA) and the Society for Healthcare Epidemiology of America (SHEA) recommend that CDI cases be classified as community-acquired or hospital-acquired according to time between onset of symptoms and most recent hospital admission or discharge [14]. Though the recommended system is not evidence-based [14], the system and minor variants are widely used to estimate the incidence of hospital- and community-acquired cases in the USA and many other countries [2, 3, 15, 16]. The recommended classification system has been shown to incorrectly classify many CDIs amongst hospitalised patients, underestimating the proportion of cases acquired prior to hospitalisation [17]. However, there has been no published assessment of the full classification system as applied to hospital-onset and community-onset cases.

CAMBRIDGE UNIVERSITY PRESS

Despite the importance of the community as both a source of new infections and the location of onset for some healthcare-associated infections, there is to date only one published model of *C. difficile* transmission that explicitly models patients outside hospitals [18]. The same model estimates an upper bound to the transmission from food and animals but does not explore the consequences of animal exposure as a source of *C. difficile* transmission. There have been no models that include the potentially important role of infants. We developed a model of *C. difficile* transmission in hospitals and communities to explore the contributions of hospitals, communities, adults, infants, animals and food to the transmission of toxigenic *C. difficile* in human populations. We also estimated the extent of under-reporting in the community and assessed the commonly used definitions of hospital- and community-acquired CDI.

Methods

Model structure

We adapted a compartmental model of toxigenic *C. difficile* transmission in hospitals [19] to model transmission in a hospital and the surrounding community, adding treatment seeking, compartments for infants under 12 months, demographic processes, waning immunity and transmission from animal reservoirs. The model of the non-infant population had the same structure in both the hospital and community, with non-infants distributed amongst different compartments according to their immunity to *C. difficile* toxins, *C. difficile* colonisation state and the state of their gut flora. However, antibiotic prescription rates and treatment-seeking behaviour differed between the hospital and community while infants were only modelled in the community. The structure is summarised in Figure 1 and Figure S1.

Non-infants (whom we call adults from here on) had three immune statuses: able to mount an effective immune response to *C. difficile* toxins conferring resistance or immunity to symptoms but not colonisation; naive to *C. difficile* toxins but with a healthy immune system; and unable to mount an effective immune response to *C. difficile* toxins because of advanced age or a suppressed immune system. Immunity could be conferred to any non-suppressed adult by either extended asymptomatic carriage or recovery from CDI. Any immune person could have their immunity wane when they are not colonised and any non-suppressed individual (including infants) could age to become suppressed.

There were two possible commensal gut flora statuses for adults: disrupted and not disrupted. There were four possible *C. difficile* statuses: free of *C. difficile*, colonised, *C. difficile* overgrowth without treatment and *C. difficile* overgrowth with treatment. As we were concerned primarily with strains that can cause symptomatic disease, we only modelled toxigenic strains of *C. difficile*. An individual could have almost any combination of gut flora and *C. difficile* statuses, but we assumed that *C. difficile* overgrowth could only occur in individuals with disrupted gut flora. Non-immune adults with *C. difficile* overgrowth were considered symptomatic, while all other colonised individuals (infants, immune adults and adults without *C. difficile* overgrowth) were considered asymptomatic. Both symptomatic and asymptomatically colonised individuals shed spores and so were infectious [20]. Spore shedding has been observed to increase before toxin production [21], but decrease during *C. difficile* treatment [22]. Therefore, asymptomatically colonised individuals with disrupted gut flora and individuals with overgrowth were equally infectious, patients receiving treatment had reduced infectiousness determined by the effectiveness and coverage of contact precautions [19] and colonised patients with intact gut flora transmitted at a reduced rate.

Since CDI is only rarely observed in infants under 12 months and antibiotics do not predispose infants to carriage [10], the model for infants was much simpler than for adults, consisting of only three compartments. At birth, infants were not colonised [8, 9] and did not have immunity [23]. As with adults, colonisation conferred immunity, but for simplicity we assumed this occurred immediately so there was no colonised-but-not-immune class for infants. Infants could clear their colonisation [8, 9]. Infants aged by entering the corresponding adult class with intact gut flora that shared the same colonisation and immune states.

Model parameterisation

Many of the parameters used in this model were based on our previous model of *C. difficile* transmission in hospitals [19] and/or drawn from the literature (Table S1). Eight parameters were fitted to data in this study. The likelihood function used to fit the model was composed from data for the prevalence of colonisation [9] and immunity [23] at given ages, longitudinal infant colonisation [8, 9], the proportion of hospital admissions with CDI as the primary diagnosis [24] and the incidence of reported hospital- and community-acquired cases [3]. The reported estimates of the prevalence of toxigenic colonisation in the general adult population have varied considerably between settings and studies. These studies have used different detection methods and often had small sample sizes [13]. Therefore, we considered multiple scenarios with colonisation prevalence from the range 2–10%, with a default of 5%. We determined the values of the eight parameters that (A) ensured that a predetermined proportion (in the range 2–10%) of the general adult population was colonised and (B) maximised the model likelihood. This was repeated for a range of values of the colonisation prevalence in the general adult population. See Supplementary materials for details of all parameters and how they were estimated.

Transmission from infants

Despite their high carriage rates, there has been little research on the contribution of infants to *C. difficile* transmission. Furthermore, the relative infectiousness of infants and adults in the community cannot be determined using our model and available data. It has been shown that the mean density of *C. difficile* per gram of stool is similar for asymptomatically colonised infants and adults with CDI [8]. However, many other unquantified factors (e.g. hygiene practices and the number of social contacts) contribute to infectiousness, so we considered a wide range of assumptions in our sensitivity analysis. In a preliminary analysis, model fit was poor and/or the proportion of transmission from infants implausibly high in scenarios where infant infectiousness exceeded that of symptomatic adults. Therefore, we considered relative infant infectiousness in the range 0–1 for our sensitivity analysis with 0.5 as conservative default assumption.

Accounting for under-reporting and misclassification of CDIs

To fit our model to incidence estimates for CDI [3], we simulated the processes of treatment seeking, reporting and the classification of cases as hospital or community-acquired. We assumed that, as

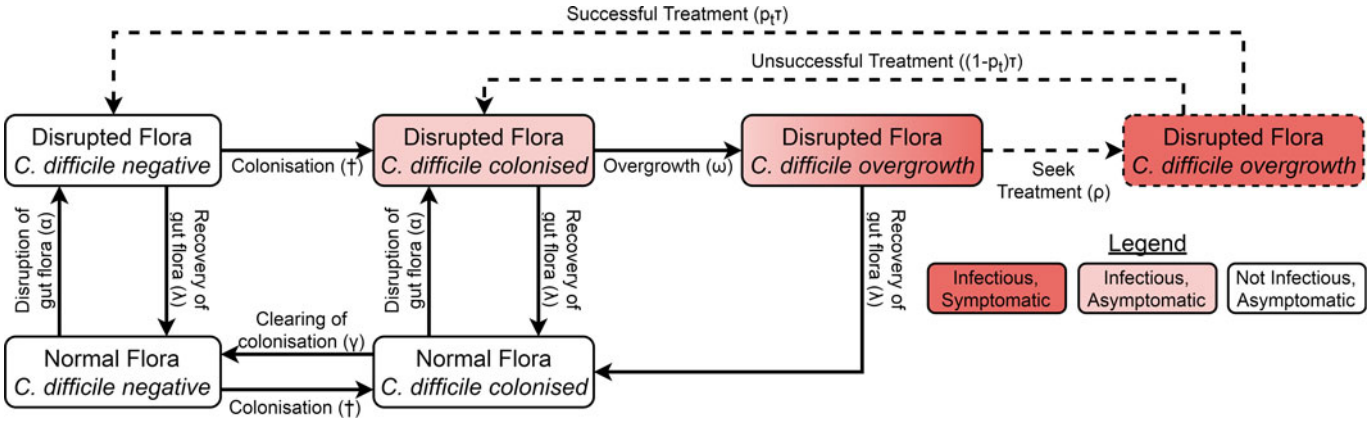


Fig. 1. Model structure showing including colonisation, gut flora status, symptoms and treatment. Adults in the immune classes do not have symptoms and therefore not all individuals with overgrowth seek or receive treatment (dashed arrows and box). The details for infants, immunity, demographics and hospital–community structure are summarised in Figure S1. The definitions and values of the parameters associated with each transition can be found in Table S1. †The force of colonisation depends in the number and type of infectious individuals in the same setting (hospital or community).

with other diarrhoeal diseases, some patients recover from CDI without seeking treatment [25], by modelling treatment seeking in the community and recovery as competing hazards (see Supplementary materials for details). To account for the low testing rate for diarrhoea in general [5, 25] and community-onset CDI in particular [6], we estimated the proportion of cases seeking treatment in the community that were identified, allowing us to compare model outputs to published estimates of disease burden based on notification data [3].

The IDSA and SHEA recommend surveillance definitions that classify where a CDI was acquired by location of onset of symptoms (healthcare facility or community) and by time since the most recent hospital discharge or admission [14] (Fig. 2). Lessa *et al.* [3] employed a variant of these definitions to estimate the incidence of initial (i.e. non-recurrent) hospital- and community-acquired CDIs in the USA (Fig. 2). We therefore emulated this classification system to fit our model to the incidence of hospital- and community-acquired CDIs reported by Lessa *et al.* (see Supplementary materials for further details).

To determine the true origin of an infection in our model, we subdivided each *C. difficile*-positive compartment into hospital-acquired and community-acquired compartments, allowing us to track where infection was acquired even if patients moved between settings once or more between acquisition and onset of symptoms. For simplicity, we assumed that current hospital-acquired colonisation prevented community-acquired colonisation and *vice versa*. This assumption had no effect on overall transmission dynamics. Moreover, coinfection with multiple strains (which may have been acquired from multiple sources) accounts for only approximately 10% of infections [26], so our simplifying assumption was unlikely to substantially affect the classification of infections. For each set of surveillance definitions, we calculated the sensitivity and precision to identify hospital- and community-acquired cases amongst both hospital-onset and reported community-onset cases, using the true origin of infection in our model as a gold standard. We identified cut-offs that improved on the existing definitions amongst reported cases, considering classification systems with a single cut-off for time since hospital admission and a single cut-off for time since most recent hospital discharge (i.e. classifying no cases as indeterminate). The *balanced* pair of cut-offs had equal sensitivity to identify hospital- and community-acquired cases amongst both hospital-onset cases and community-onset cases. The *optimal* pair of cut-offs had equal precision and sensitivity when identifying hospital-acquired cases, amongst both hospital-onset and community-onset cases.

Reproduction number

Since the extent of human exposure to animal reservoirs of *C. difficile* is unknown, we calculated reproduction numbers assuming that all exposure was due to person-to-person transmission – an upper bound for the true reproduction number. We calculated the reproduction number for the whole population. We also calculated reproduction numbers for the community and the hospital separately. The latter calculations were identical to standard next-generation matrix calculations [27], except we only considered the colonised individuals in the setting of interest to be colonised for the purposes of the calculation. The reproduction numbers for hospital and community were the endemic threshold parameters in each setting assuming no external sources of *C. difficile* (movement of patients or animal reservoir).

Food- and animal-driven transmission

The extent of zoonotic or foodborne *C. difficile* exposure is unknown; however, we considered the implications of differing amounts transmission from animal reservoirs. For a given force of colonisation, higher human exposure from food or animals implies less person-to-person transmission and therefore a smaller reproduction number. If a sufficient proportion of exposure originates from food or animals, the reproduction number in the human population is less than one and human disease is sustained by constant exposure to non-human sources of *C. difficile*. In this case, we say that *C. difficile* is *animal-driven*.

For each set of modelling assumptions, we calculated the extent of foodborne exposure that implied *C. difficile* was animal-driven. We expressed this *animal-driven threshold* in terms of exposures leading to colonisation per person per year and as a proportion of all transmission (i.e. foodborne transmission *and* person-to-person transmission).

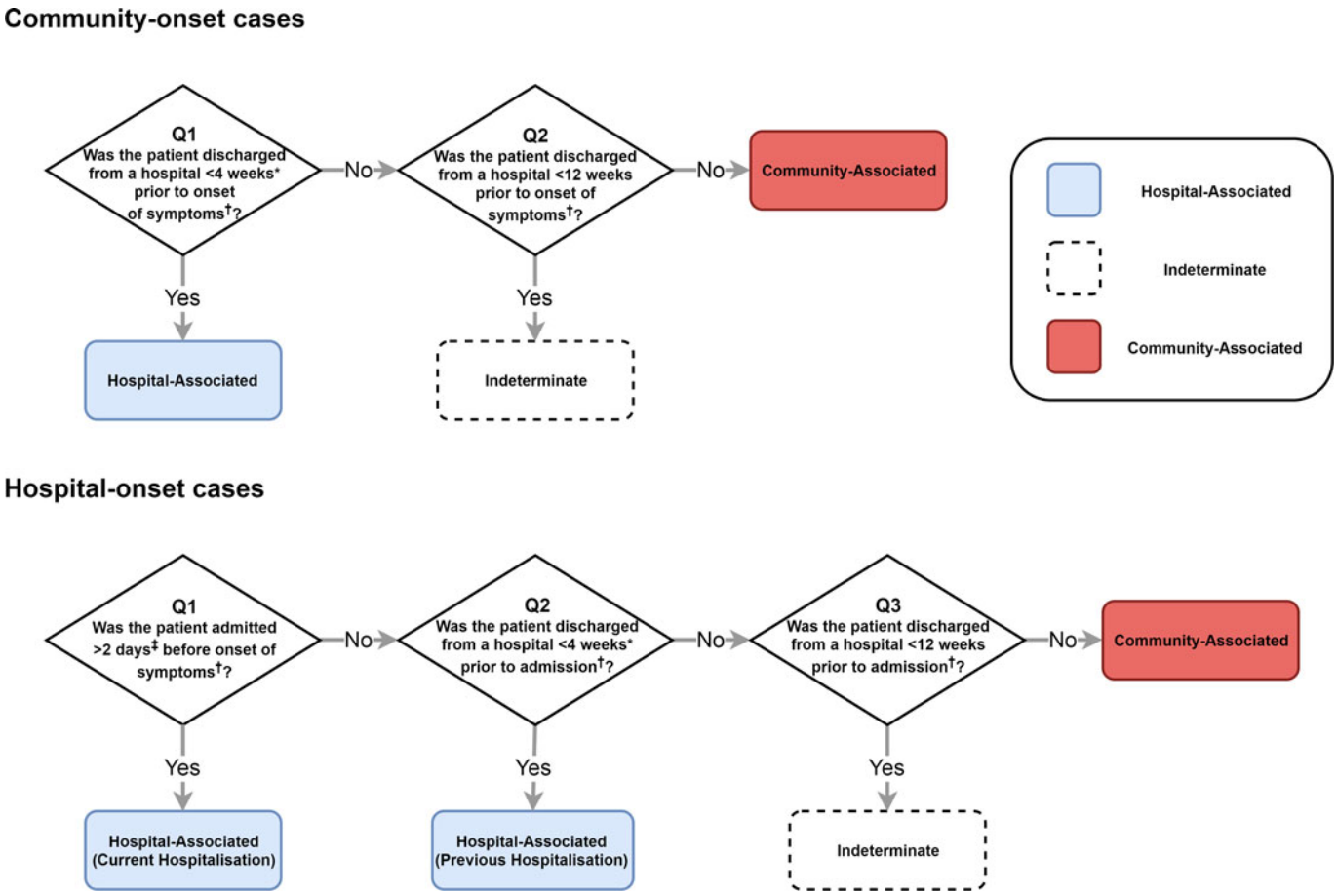


Fig. 2. The classification of CDI cases based on IDSA and SHEA surveillance recommendations that we assessed with our model. Lessa *et al*. used a similar classification scheme to estimate incidence in the USA. *Lessa *et al*. used a 12-week cut-off and therefore do not classify any cases as 'indeterminate'. ‡Lessa *et al*. used a 3-day cut-off. †We used symptom onset or hospital admission as reference points in our simulations. However, the classification system recommended by IDSA and SHEA uses onset of symptoms as the reference point for all cut-offs. Our classification is otherwise identical. Lessa *et al*. used date of positive faecal sample as reference point.

Results

Model fit

The model fitted the data well, reproducing the observed age profile of toxigenic *C. difficile* colonisation, immunity, reported incidence of infection and proportion of admissions for CDI (Figure S2). For most scenarios, infant infectiousness did not affect model fit. However, the model fit was poor for combinations of low colonisation prevalence amongst adults and high infant infectiousness, so these scenarios were not considered further. The model was verified by outcomes not used to fit the model such as recurrence proportion for hospital and community cases, the proportion of hospital-based transmission attributable to symptomatic carriers, the duration of colonisation in infants and the greater proportion of elderly and immune suppressed in hospital-acquired *vs.* community-acquired cases (see Supplementary materials for details). In our model, colonisation prevalence was 17% higher (range: 4–55%) at hospital discharge than in the general adult population, agreeing with the common observation that colonisation is more common amongst those who have been recently discharged from hospital. However, 78% (range: 60–87%) of colonised discharges had acquired the pathogen in the community prior to admission and remained colonised for the duration of their hospital stay. We estimated a mean immune period of 9.4 years (range: 4.0–30.4 years) with the longest immune period when we assumed adult colonisation prevalence was low (2%).

Reproduction number and food-driven threshold

Under the assumption of no foodborne transmission, the reproduction number for the whole population was greater than one for all plausible assumptions (default: 1.11, range: 1.03–1.35) (Fig. 3a). The reproduction number for the hospital was less than one for all plausible assumptions (default: 0.28, range: 0.16–0.46), decreasing with increasing colonisation prevalence of adults in the community and unaffected by assumptions concerning the infectiousness of infants (Fig. 3c). The reproduction number for the community was close to but lower than the reproduction number for the whole population (default: 1.09, range: 0.999–1.34) (Fig. 3b) and increased with increasing infant infectiousness. The reproduction number was less than one in the community only if infants were not infectious and adult colonisation prevalence was 2%.

The animal-driven threshold (the minimum force of colonisation attributable to food and animals that implies the reproduction number in the human population is less than one), was 0.046 exposures per person per year (range: 0.006–0.107) or 10.6% of all transmission in the community (range: 3.5–26.0%) (Fig. 4). This is equivalent to one foodborne or animal exposure

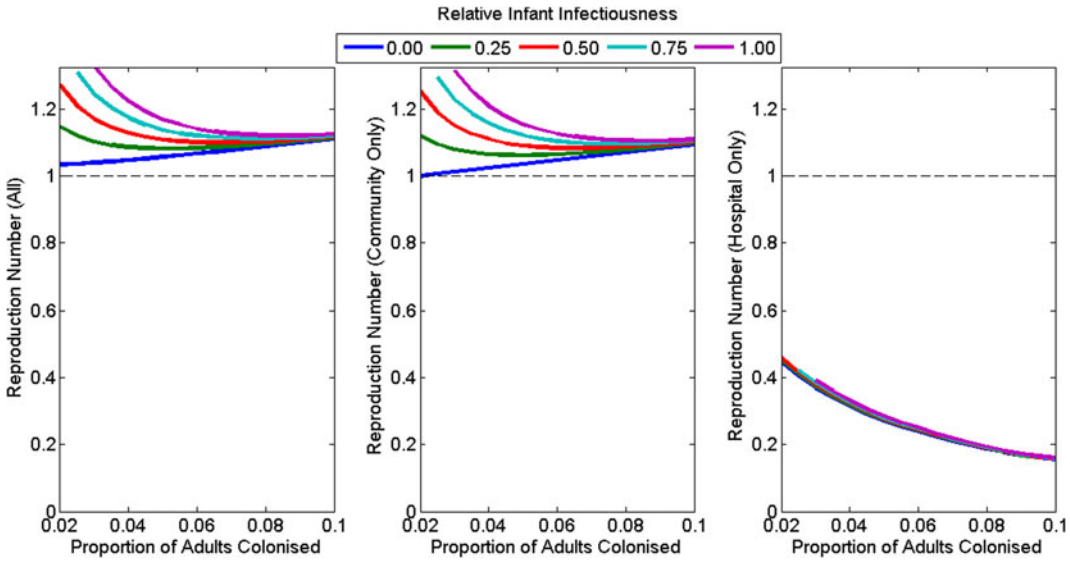


Fig. 3. The reproduction number at the disease-free equilibrium for various plausible assumptions for the colonisation prevalence in adults and relative infectiousness of infants for (a) the whole population, (b) the community only and (c) the hospital only. The model had poorer model fit for the combination of high infant infectiousness and low adult colonisation prevalence, so these combinations are omitted from the figures.

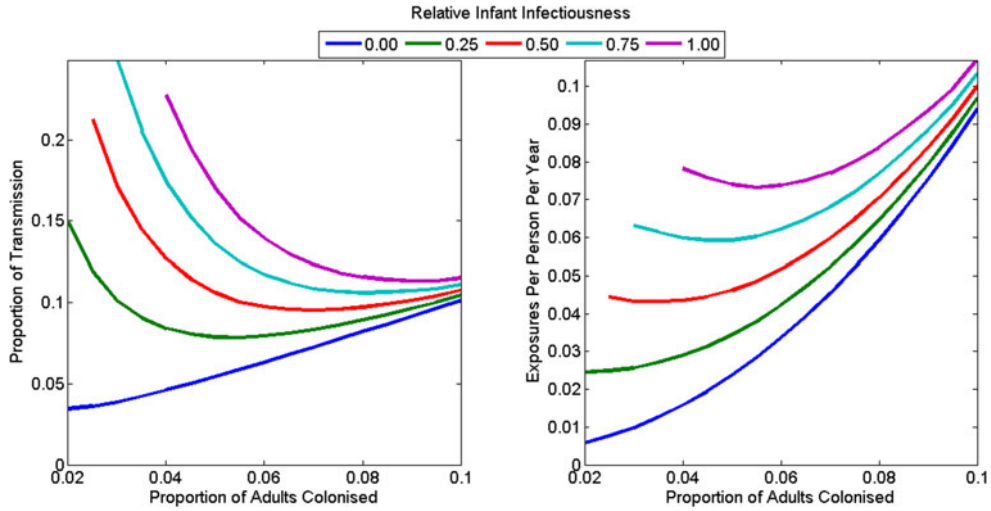


Fig. 4. The animal-driven threshold under various plausible assumptions for the *C. difficile* colonisation prevalence in adults, and the relative infectiousness of infants as (a) a proportion of all transmission in the community and (b) as rate of exposure to adults in the community. The reproduction number is less than one in the community if transmission from animals exceeds the animal-driven threshold. The model had poorer model fit at the animal-driven threshold for the combination of high infant infectiousness and low adult colonisation prevalence, so these combinations are omitted from the figures.

leading to colonisation every 21.7 years per person (range 9.4–175.5 years). The animal-driven threshold was lowest (once every 175.5 years per person) if infants were not infectious and adult colonisation prevalence was low (2%). The animal-driven threshold was highest (once every 9.4 years per person) if infants were as infectious as adults and adult colonisation prevalence was high (10%). The model had poor model fit at the animal-driven threshold when infant infectiousness was high and adult colonisation prevalence was low.

Transmission from infants and asymptomatic adults with intact gut flora

In our main analysis, 13–30% of transmission in hospitals was from patients receiving treatment for CDI, but <10% of all transmission in the community was attributable to symptomatic patients or patients with disrupted gut flora. The remaining transmission was attributable to infants or asymptomatically colonised adults with intact gut flora. The proportion of transmission in the community attributable to infants was 17.4% for our conservative default scenario but was highly sensitive to the relative infectiousness of infants and colonisation prevalence in adults (Figure S3). With infants as infectious as symptomatic adults and adult colonisation prevalence in the community at ⩽5%, ⩾40% of transmission in the community was attributable to infants. The proportion of transmission attributable to asymptomatically colonised individuals with intact gut flora was also highly sensitive to these assumptions (Figure S4). Under default assumptions, this group accounted for 79% of transmission in the community and 25% of transmission in the hospital, but ⩾90% of transmission in

Table 1. Definitions, values and references for eight parameters fitted with the model

Parameter	Description	Value (Sensitivity analysis range)
σ	Rate at which immunity wanes	2.9×10^{-4} ($0.9–6.9 \times 10^{-4}$)
θ	Multiplicative factor for colonisation susceptibility of infants	1.4 (0.6–4.4)
γ_{infant}	Rate at which *C. difficile* is cleared in infants	2.0×10^{-3} ($2.0–2.5 \times 10^{-3}$)
$p_{disrupt}$	Proportion of antibiotics that disrupt gut flora	0.22 (0.12–0.48)
p_{report}	Proportion of all community-treated CDIs that are reported	0.63 (0.57–0.76)
ν_{CDI}	Hospital admission rate for CDI	1.4×10^{-2} ($1.3–1.7 \times 10^{-2}$)
$\beta_{Disrupt}$	Transmission rate coefficient for colonised adults with disrupted gut flora (due to recent antibiotic exposure)	12.8×10^{-2} ($7.1–17.4 \times 10^{-2}$)[a]
β_{Intact}	Transmission rate coefficient for colonised adults with intact gut flora (no recent antibiotic exposure)	1.9×10^{-2} ($0.1–2.6 \times 10^{-2}$)[a]
β_{infant}	Transmission rate coefficient from infants	6.4×10^{-2} ($0–17.4 \times 10^{-2}$)[a]

A full list of parameters can be found in Table S1. All rates are in units of day^{-1}.
[a]Only these parameters were affected by assumptions around infant infectiousness, being estimated under the assumption that $\beta_{infant} = k \times \beta_{Disrupt}$ for k in the range 0–1.

Table 2. Simulated incidence of hospital-acquired (HA) and community-acquired (CA) CDIs, under-reporting of cases and classification errors for two different simulated classification schemes

			IDSA/SHEA recommendations			Lessa *et al.*, classification		
	Proportion reported %	Actual Proportion of all reported cases %	Proportion of all reported cases % [b]	Precision %	Sensitivity %	Proportion of all reported cases % [c]	Precision %	Sensitivity %
Any HA	67 (66–70)	15 (9–23)	38 (38–39)	37 (24–57)	97 (96–97)	41	34 (22–53)	94 (93–95)
Any CA	35 (33–40)	85 (77–91)	57 (56–57)	99 (99–100)	65 (61–72)	59	99 (97–99)	68 (64–74)
HO-HA	100[a]	11 (7–17)	25 (25–25)	44 (29–65)	97 (96–97)	23	46 (31–67)	93 (91–94)
HO-CA	100[a]	17 (11–21)	3 (3–3)	89 (77–94)	15 (14–19)	5	85 (70–92)	27 (25–33)
CO-HA	31 (28–37)	3 (2–6)	13 (13–14)	24 (15–42)	96 (95–97)	18	18 (11–33)	99 (99–99)
CO-CA	30 (27–36)	69 (66–70)	54 (53–54)	>99.8	78 (76–81)	53	>99.8	78 (76–81)

HA, hospital-acquired, CA, community-acquired, HO, hospital-onset, CO, community-onset.
The range in parenthesis is the range across all sensitivity analysis scenarios. Classification sensitivity is amongst reported cases only; thus, multiplying by the reported proportion will return the sensitivity amongst all cases.
[a]We assumed that all hospital-onset infections are reported.
[b]Percentages do not sum to 100% as some cases are classified as 'indeterminate' under this system.
[c]The model was fit to estimates of CDI incidence that used this scheme to classify the location of acquisition. Consequently, these values change very little across the sensitivity analysis scenarios.

the community if colonisation prevalence was 10% amongst adults in the community. Patients with CDI and colonised individuals with disrupted gut flora were 6.6 times more infectious (range: 2.8–131.8) than colonised individuals with intact gut flora, but were much less numerous, especially in the community where the antibiotic prescription rate was low. Infants cleared their colonisation 9.2–11.5 times more slowly than adults with intact gut flora. Under most scenarios, infants were also more exposed or susceptible to colonisation (default: factor of 1.4; range: 0.6–4.4) and more infectious (default: factor of 3.3; range: 0–9.8) than asymptomatic adults with intact gut flora (compare Table 1 and Table S1).

Under-reporting and misclassification of CDIs

Though we estimated that patients with CDI were admitted to hospital at 59 (range: 53–73) times the rate of the general adult population (Table 1), only 48% of adults with community-onset CDIs sought treatment in the community or hospital (Table 2) and only 63% (range 56–76%) of CDIs treated in the community were reported (Table 1). Therefore, while we assume that 100% of symptomatic hospital-onset infections were reported, we estimate that only 30% (range 27–37%) of all community-onset CDIs were reported. Considering both hospital- and community-onset CDIs, only 67% (range 66–70%) of all hospital-acquired CDIs and 35% (range 33–40%) of all community-acquired cases were reported (Table 2).

Standard CDI classification schemes misclassified many of the reported community-acquired cases as hospital-acquired in our model: 63% (range: 43–76%) of cases classified as hospital-acquired with the IDSA/SHEA scheme were actually community-acquired (Table 2). The classification systems were much more precise but less sensitive for community-acquired cases (Table 2). Though total incidence was underestimated due to under-reporting, both classification schemes overestimated the proportion of reported cases that were hospital-acquired (Fig. 5). We estimate that only 40% (range: 26.5–60.6%) of hospital-onset and 4.5% (range: 2.7–8.4%) of reported community-onset infections are hospital-acquired. In contrast, the classification scheme recommended by IDSA and SHEA classified 89.6% (range: 88.9–90.3%)

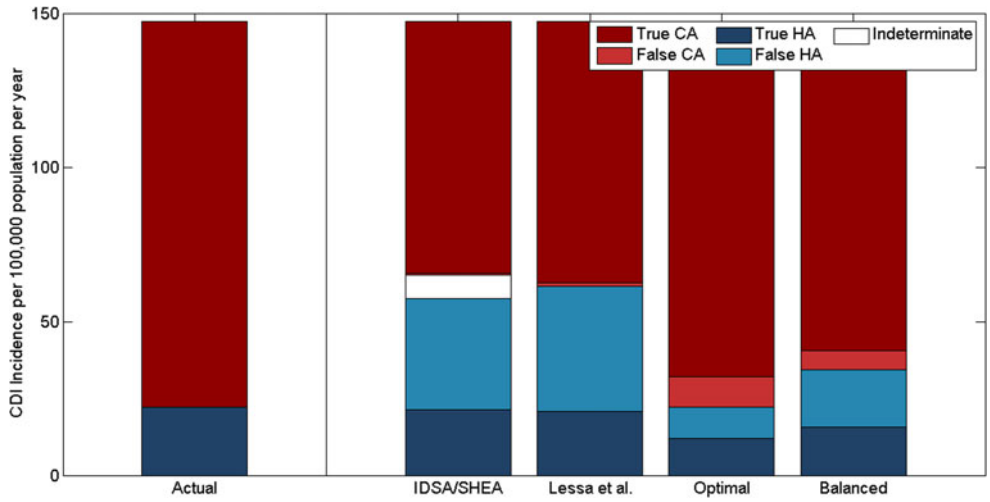


Fig. 5. Classification of the origin of reported CDIs by time since hospital discharge or admission, comparing the actual incidence of reported hospital-acquired (HA) and community-acquired (CA) CDIs *vs*. the classification recommended by IDSA and SHEA and three variants. Lessa *et al*. use a 3-day cut-off for recent hospital admission and a 12-week cut-off for recent hospital discharge. The optimal and balanced classifications we have identified use 7.4- and 6.6-day cut-offs, respectively, for recent hospital admission and 2.1- and 12.5-day cut-offs, respectively, for recent hospital discharge.

of hospital-onset and 19.6% (range: 19.4–20.3%) of reported community-onset infections as hospital-acquired. A 7.4-day cut-off (range: 5.0–9.5) for recent hospital admission (in hospital-onset cases) and a 2.1-day cut-off (range: 1.3–3.9) for prior hospital discharge were the optimal pair of cut-offs. A 6.6-day cut-off (range: 5.8–7.0) for recent hospital admission and a 12.5-day cut-off (range: 11.8–14.5) for prior hospital discharge were the balanced pair of cut-offs. The optimal cut-off correctly estimated the proportion of cases that were hospital- or community-acquired, but had poor precision (≈50%) to identify hospital-acquired cases (Fig. 5).

Discussion

Under all reasonable scenarios and modelling assumptions, transmission between hospitalised adults amplified disease burden (higher force of colonisation and higher colonisation proportion in discharged patients than the general population) but was not the key driver of toxigenic *C. difficile* transmission in the population (hospital reproduction number less than one), in agreement with previous modelling studies [19, 28]. When we simultaneously assumed low prevalence of *C. difficile* colonisation in adults, no infant infectiousness and no transmission from non-human sources, the reproduction number in the community was also less than one. In this unlikely scenario, the movement of colonised individuals between hospital and community was essential for persistence of *C. difficile* in both settings. However, in all other scenarios without transmission from non-human sources, the reproduction number was greater than one in the community, and therefore transmission in the community would persist even in the absence of transmission in hospitals. This is the first time reproduction numbers have been estimated for *C. difficile* in a model including both the hospital and the community.

Symptomatic carriers of *C. difficile* accounted for <10% of transmission in the community in our model. Despite accounting for <2% of the total population, infants under 12 months accounted for 17% of transmission in the community for our conservative default assumptions and ⩾40% of transmission if infants were at least as infectious as symptomatic adults and colonisation prevalence was ⩽5% in the community. However, the exact proportion was highly sensitive to the relative infectiousness of infants (which has not been well quantified) and the colonisation prevalence in adults in the community (which varies considerably between studies and settings [13]). Nevertheless, our results indicate that asymptomatically colonised infants are likely to be a substantial source of transmission in the community. This is in agreement with a number of small studies that found CDI was associated with exposure to infants [29, 30] and a large study that, despite sampling only 1% of infants in Oxfordshire, was able to determine that 2% of all known CDIs in Oxfordshire could be reasonably attributed to recent direct or indirect transmission from these infants [31].

We investigated how transmission from non-human sources affected estimates of the reproduction number for person-to-person transmission. We demonstrated that the reproduction number in the human population was less than one if over 3.5–26.0% of transmission in the community was from non-human sources such as food or water contaminated by livestock animals. If current transmission from animals is above this threshold, *C. difficile* could not persist in the human population without these non-human exposures. This animal-driven threshold in terms of *C. difficile* exposures per person per year was remarkably low: equivalent to one exposure leading to colonisation per adult every 21.7 years under our default assumptions. For comparison, it has been estimated that Australians have an episode of food-borne gastroenteritis (i.e. not counting asymptomatic exposure) on average once every 5 years [32]. Given the overlap of strains between animals and humans, the presence of *C. difficile* spores on raw meats and fresh vegetables [11], and the high survival rate of *C. difficile* spores following cooking at recommended 'safe' temperatures [33], it is plausible that exposure exceeds this low threshold. Though our model has not accounted for multiple strains of *C. difficile*, one could apply the animal-driven threshold to individual strains or types of *C. difficile*. For instance, it is not plausible that ribotype 001, which accounts for a substantial proportion of human cases but is not common in food animals [34], exceeds the animal-driven threshold. On the other hand, it is plausible that ribotypes 078, 027 and other ribotypes that both cause human infection and are commonly isolated from animals [34], exceed the threshold. This is especially true for ribotype 078, as isolates from humans and food animals appear to be closely related [12].

We estimate that approximately 70% of community-onset cases are not reported, either because the patient does not seek treatment, or the pathogen remains unidentified. This is in agreement with the low treatment-seeking rates reported generally for diarrhoea [5, 25] and low testing rates for *C. difficile* in primary care [6]. The simulated proportion of community-acquired cases was higher amongst community-onset cases than hospital-onset cases. Consequently, we estimate that while two-thirds of hospital-acquired infections are reported, only one-third of community-acquired infections are reported. Though we only simulated the under-reporting of community-onset cases, our findings complement an empirical study that found that missed cases of CDI in hospital settings are disproportionately likely to be community-acquired [35]. Existing classification schemes attempt to account for cases that may acquire the pathogen in one setting and, after an incubation period, develop symptoms in another; however, these schemes are highly asymmetric with regards to setting [3, 14]. While only hospital-onset cases with symptom onset within 2 or 3 days of hospital admission are considered community-acquired, all community-onset cases with symptom onset within 4 or 12 weeks of hospital discharge are classified as hospital-acquired. Empirical estimates of the median incubation period vary considerably from 18 to 33 days [36] but lie between the two extremes of these cut-offs. Therefore, it is likely that typical cut-offs for classifying hospital-onset cases as community-acquired are too short and typical cut-offs for classifying community-onset infections as hospital-acquired are too long. We confirmed this with our model by demonstrating that, to balance the sensitivity and specificity of classification, these cut-offs should be approximately 6 and 12 days, respectively. We also demonstrated that any scheme based on time since most recent hospital discharge cannot adequately distinguish hospital- and community-acquired cases. Even our balanced scheme had very poor precision for hospital-acquired cases: half or more of all cases classified as hospital-acquired were actually community-acquired. This can be understood by noting that the mean length of hospital stay (between 4 and 10 days in most high-income countries [37]) is short compared with the duration of colonisation, which may be several weeks [38]. Consequently, more than 60% of patients who were colonised at hospital discharge in our model were not exposed in the preceding hospitalisation, but rather in the community prior to admission. Therefore, even patients who developed CDI very soon after hospital discharge were more likely to be community-acquired than hospital-acquired. Adjusting the cut-off times cannot correct this flaw in the existing classification schemes. Alternative schemes, such as classification based on the total number of days spent in hospital in the weeks leading up to the onset of symptoms, should be considered. Our model demonstrates that the classification scheme recommended by IDSA and SHEA has very high sensitivity for hospital-acquired cases, and therefore may be useful if all hospital-acquired cases need to be identified or excluded. However, the asymmetry and unrealistic timescales in existing classification schemes inadvertently reinforce the *a priori* assumptions upon which they are based: that the colonisation pressure in hospitals far exceeds the colonisation pressure in the community.

Our study has several limitations. The data used to fit the model were incomplete and were gathered from many different sources, countries and settings. In particular, the published estimates of colonisation prevalence vary significantly between studies and it is unclear to what extent this reflects genuine differences between study populations or variations associated with different detection methods or small sample sizes [13]. We addressed this by considering a range of scenarios that reflected the possible range of colonisation prevalence. The relative infectiousness of infants and adults is also unknown, so we allowed the relative infectiousness of infants and adults to vary in our sensitivity analysis but were therefore unable to provide a precise estimate of the amount of transmission attributable to infants. However, our broad sensitivity analysis and wide variety of input data improve the global applicability of the model. National estimates in the USA suggest that nearly all people with CDI had received some form of healthcare soon before onset of symptoms if outpatient and primary care were included [3]. This does not imply that most CDIs are healthcare-acquired, since antibiotic exposure is a causative factor for CDI and antibiotics are prescription-only medicines in many countries, including the USA. However, we were unable to model pathogen acquisition from other sources of healthcare, because the hospital in our model consisted only of admitted patients, with the community including patients receiving all other forms of healthcare (including residents of long-term care facilities). Long-term care facilities contain subpopulations of individuals at high risk for CDI [3]. As we have not modelled long-term care facilities separately, we are likely to have underestimated the heterogeneity and thus the reproduction number in the community. The model population is well-mixed and does not capture heterogeneity in hospital admission rates or heterogeneity in the contact rates of infants, adults and the elderly, which may also affect reproduction number estimates. Finally, we did not differentiate between the many strains of *C. difficile* [39], so it is possible the hospital reproduction numbers, community reproduction numbers and animal-driven thresholds differ by strain.

Under-reporting of community-onset CDIs and the misclassification of many community-acquired infections obscure and underestimate the extent of transmission in the community. It seems likely that unreported community-onset cases will be less severe and that the classification (or misclassification) of individual cases as hospital-acquired or community-acquired will not affect the treatment or outcomes of patients. Therefore, at the level of individual cases, even large-scale under-reporting and misclassification may not be very harmful if those with severe disease receive appropriate care. However, to prevent infections, we must understand when, where and how transmission occurs at the level of the population. We have demonstrated that most infections (hospital-onset and community-onset alike) are acquired outside of hospitals, but only a small fraction are reported. Therefore, interventions that prevent acquisition outside hospitals, or prevent patients admitted with asymptomatic colonisation from developing symptoms should be considered and assessed. Merely reducing transmission between hospitalised patients will not be sufficient to prevent the spread of this important pathogen. Further investigation into the relative infectiousness of infants is required before the proportion of transmission from infants can be estimated. However, we have demonstrated that a high degree of transmission from infants is consistent with available data on spore shedding [8] and colonisation prevalence [8, 9]. Similarly, though the frequency of food and animal-to-human transmission is unknown for *C. difficile*, we have demonstrated that even very modest and plausible frequencies of exposure may imply that *C. difficile* is sustained in human populations by transmission from animals or contaminated food. If this is the case, *C. difficile* can be eradicated from the human population if and only if animal-to-human transmission is reduced.

Author ORCIDs. A. McLure 0000-0003-2890-2703, K. Glass, 0000-0001-5905-1310

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0950268819000384.

Financial support. This work was supported by an Australian National Health and Medical Council Senior Research Fellowship (#1058878 to A.C.A.C.) and an Australian Government Research Training Program Scholarship to A.M.

Conflict of interest. None.

References

1. **Jarvis WR *et al.*** (2009) National point prevalence of *Clostridium difficile* in US health care facility inpatients, 2008. *American Journal of Infection Control* **37**, 263–270.
2. **Slimings C *et al.*** (2014) Increasing incidence of *Clostridium difficile* infection, Australia, 2011–2012. *The Medical Journal of Australia* **200**, 272–276.
3. **Lessa FC *et al.*** (2015) Burden of *Clostridium difficile* infection in the United States. *The New England Journal of Medicine* **372**, 825–834.
4. **Magill SS *et al.*** (2014) Multistate point-prevalence survey of health care-associated infections. *New England Journal of Medicine* **370**, 1198–1208.
5. **Scallan E *et al.*** (2006) Factors associated with seeking medical care and submitting a stool sample in estimating the burden of foodborne illness. *Foodborne Pathogens and Disease* **3**, 432–438.
6. **Hensgens MPM *et al.*** (2014) Diarrhoea in general practice: when should a *Clostridium difficile* infection be considered? Results of a nested case-control study. *Clinical Microbiology and Infection* **20**, O1067–O1074.
7. **Adlerberth I *et al.*** (2014) Toxin-producing *Clostridium difficile* strains as long-term gut colonizers in healthy infants. *Journal of Clinical Microbiology* **52**, 173–179.
8. **Kubota H *et al.*** (2016) Longitudinal investigation of carriage rates, counts, and genotypes of toxigenic *Clostridium difficile* in early infancy. *Applied and Environmental Microbiology* **82**, 5806–5814.
9. **Rousseau C *et al.*** (2012) *Clostridium difficile* carriage in healthy infants in the community: a potential reservoir for pathogenic strains. *Clinical Infectious Diseases* **55**, 1209–1215.
10. **Holst E, Helin I and Mårdh PA** (1981) Recovery of *Clostridium difficile* from children. *Scandinavian Journal of Infectious Diseases* **13**, 41–45.
11. **Hensgens MPM *et al.*** (2012) *Clostridium difficile* infection in the community: a zoonotic disease? *Clinical Microbiology and Infection* **18**, 635–645.
12. **Knetsch CW *et al.*** (2017) Zoonotic transfer of *Clostridium difficile* harboring antimicrobial resistance between farm animals and humans. Fenwick B (ed.). *Journal of Clinical Microbiology* **56**, e01384-17.
13. **Furuya-Kanamori L *et al.*** (2015) Asymptomatic *Clostridium difficile* colonization: epidemiology and clinical implications. *BMC Infectious Diseases* **15**, 516.
14. **Cohen SH *et al.*** (2010) Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). *Infection Control and Hospital Epidemiology* **31**, 431–455.
15. **Lambert PJ *et al.*** (2009) Population-based surveillance of *Clostridium difficile* infection in Manitoba, Canada, by using interim surveillance definitions. *Infection Control and Hospital Epidemiology* **30**, 945–951.
16. **Longtin Y *et al.*** (2016) Effect of detecting and isolating *Clostridium difficile* carriers at hospital admission on the incidence of *C difficile* infections. *JAMA Internal Medicine* **176**, 796–804.
17. **McLure A *et al.*** (2018) *Clostridium difficile* classification overestimates hospital-acquired infections. *Journal of Hospital Infection* **99**, 453–460.
18. **Durham DP *et al.*** (2016) Quantifying transmission of *Clostridium difficile* within and outside healthcare settings. *Emerging Infectious Diseases* **22**, 608–616.
19. **McLure A *et al.*** (2017) Healthcare-associated *Clostridium difficile* infections are sustained by disease from the community. *Bulletin of Mathematical Biology* **79**, 2242–2257.
20. **Riggs MM *et al.*** (2007) Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic *Clostridium difficile* strains among long-term care facility residents. *Clinical Infectious Diseases* **45**, 992–998.
21. **Freeman J** (2003) Effects of cefotaxime and desacetylcefotaxime upon *Clostridium difficile* proliferation and toxin production in a triple-stage chemostat model of the human gut. *Journal of Antimicrobial Chemotherapy* **52**, 96–102.
22. **Sethi AK *et al.*** (2010) Persistence of skin contamination and environmental shedding of *Clostridium difficile* during and after treatment of C. difficile infection. *Infection Control & Hospital Epidemiology* **31**, 21–27.
23. **Viscidi R *et al.*** (1983) Serum antibody response to toxins A and B of *Clostridium difficile*. *Journal of Infectious Diseases* **148**, 93–100.
24. **Healthcare Cost and Utilization Project (HCUP)** (2014) *HCUP National Inpatient Sample (*NIS*)*. Rockville, MD: Agency for Healthcare Research and Quality.
25. **Van Cauteren D *et al.*** (2012) Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiology and Infection* **140**, 697–705.
26. **Eyre DW *et al.*** (2012) *Clostridium difficile* mixed infection and reinfection. *Journal of Clinical Microbiology* **50,** 142–144.
27. **Diekmann O, Heesterbeek JAP and Roberts MG** (2010) The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society, Interface* **7**, 873–885.
28. **Lanzas C *et al.*** (2011) Epidemiological model for *Clostridium difficile* transmission in healthcare settings. *Infection Control & Hospital Epidemiology* **32**, 553–561.
29. **Jamal W, Pauline E and Rotimi V** (2015) A prospective study of community-associated *Clostridium difficile* infection in Kuwait: epidemiology and ribotypes. *Anaerobe Elsevier* **35**, 28–32.
30. **Wilcox MH *et al.*** (2008) A case-control study of community-associated *Clostridium difficile* infection. *The Journal of Antimicrobial Chemotherapy* **62**, 388–396.
31. **Stoesser N *et al.*** (2017) Epidemiology of Clostridium difficile in infants in Oxfordshire, UK: risk factors for colonization and carriage, and genetic overlap with regional C. difficile infection strains. Deshpande A (ed.). *PLoS oNE* **12**, e0182307.
32. **Kirk M *et al.*** (2014) Foodborne illness, Australia, Circa 2000 and Circa 2010. *Emerging Infectious Diseases* **20**, 1857–1864.
33. **Rodriguez-Palacios A and Lejeune JT** (2011) Moist-heat resistance, spore aging, and superdormancy in *Clostridium difficile*. *Applied and Environmental Microbiology* **77**, 3085–3091.
34. **Rodriguez-Palacios A *et al.*** (2013) *Clostridium difficile* in foods and animals: history and measures to reduce exposure. *Animal Health Research Reviews* **14**, 11–29.
35. **Reigadas E *et al.*** (2015) Missed diagnosis of *Clostridium difficile* infection; a prospective evaluation of unselected stool samples. *The Journal of Infection* **70**, 264–272.
36. **Otete EH *et al.*** (2013) Parameters for the mathematical modelling of Clostridium difficile acquisition and transmission: a systematic review. DeRiemer K (ed.). *PLoS ONE* **8**, e84224.
37. **OECD**. *Health at a Glance 2015. Health at a Glance 2015: OECD Indicators*. Paris: OECD Publishing, 2015.
38. **Johnson S *et al.*** (1992) Treatment of asymptomatic *Clostridium difficile* carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. *Annals of Internal Medicine* **117**, 297–302.
39. **Knight DR *et al.*** (2015) Diversity and evolution in the genome of *Clostridium difficile*. *Clinical Microbiology Reviews* **28**, 721–741.

Epidemiology and Infection

Modelling diverse sources of *Clostridium difficile* in the community: importance of animals, infants and asymptomatic carriers – Supplementary Material

Angus McLure, Archie C. A. Clements, Martyn Kirk and Kathryn Glass

Further details of model structure and parameters

Demographics

The demographic structure of the model is summarised in **Supp. Figure 1**. All individuals are born into the infant class without immunity or colonisation. We assume the population is closed without immigration or emigration. The birth rate is chosen to balanced deaths such that the equilibrium population is N_{Comm} (100,000). Infants age to become (non-suppressed) adults at rate ζ (ζ^{-1} = 1 year). Non-suppressed adults and infants age slowly to become suppressed adults at rate ψ such that the median time to age is 65 years ($\psi = \frac{\ln 2}{65\times365}\text{day}^{-1}$).

There are two death rates: one rate for infants and non-supressed adults (ϕ) and a much higher rate for supressed/elderly adults (ϕ_U). The two death rates are chosen so that proportion in the suppressed class is 13.7% (the proportion of US population over 65 [1]) and proportion of deaths in the suppressed class is 72.4% (proportion of deaths that are in people over 65 in the US [1]): $\varphi_U = \psi\frac{1-0.137}{0.137}$ and $\varphi = \psi\frac{1-0.724}{0.724}$.

Immunity

The immune structure of the model is summarised in **Supp. Figure 1**. Non-supressed adults, being treated for CDI are conferred immunity upon end of treatment. Colonised, non-supressed, adults not receiving treatment develop immunity at rate δ (1/ δ = 10 days) as determined in our previous model from sero-conversion rate in response to *C. difficile* toxoid vaccine trials [2,3]. To simplify the model infants develop immunity immediately on colonisation. All immune individuals have their immunity wane at rate σ, which is estimated with our model.

Admissions and discharges

The hospital and community structure of the model is summarised in **Supp. Figure 1**. The hospital discharge rate for suppressed/elderly and non-supressed individuals is the inverse of the mean length of stay in US Hospitals for those over 65 (κ_U = 1/5.2 days) and those under 65 (κ = 1/4.15 days) respectively (2012 data) [4]. However, those receiving treatment for CDI in hospital are not discharged. Similarly admission rates are determined from US hospital admission rates which are once every 11.4 years for those under 65 (ν) and once every 3.4 years for those over 65 (ν_U) (2012 data) [4]. We assume that those who have symptoms of CDI in the community are admitted at a much faster rate (ν_{CDI}) estimated with our model.

Colonisation and Gut Flora

The model structure for gut flora disruption and *C. difficile* colonisation is summarised in **Figure 1** in the main text. Gut flora is disturbed at different rates in the hospitals and communities to reproduce the reported proportion admissions in hospitals [5] or proportion of adults in the community each

year [6] that receive antibiotics. In the community those in the suppressed/elderly received antibiotics at a higher rate such that $\alpha_{Comm} < \alpha_{U,Comm} < \alpha_{Hosp} = \alpha_{U,Hosp}$. The time take for the recovery of gut flora [7], and the duration of heightened risk of *C. difficile* following antibiotic exposure [8] depends on the antibiotic but we chose a single recovery rate $\lambda = 0.03$ day^{-1} which sits in the middle of the range. As it has been observed that 20% of hospitalised CDIs recover without specific treatment [9], the gut flora recovery rate for those with overgrowth is $\lambda_o = \frac{\rho_{Hosp}}{1/0.2 - 1}$.

The colonisation clearance rate is the same (γ) in all colonised adults and was determined in our previous model of *C. difficile* transmission in a hospital [10] based on the clearance rate in the control group in a trial for treating asymptomatic colonisation with vancomycin and metronidazole [11]. The clearance rate in infants (γ_{infant}) is estimated in our model to the colonisation profile for infants. The rate at which *C. difficile* overgrows in those with damaged gut flora (ω) is the same in all adults and was determined for our previous model from observations of *C. difficile* overgrowth in a chemostat model of gut flora in human gastrointestinal tract [12].

Transmission

Transmission is well mixed within each of the hospital and community but there is no transmission between these two locations (only movement of individuals), so there is a separate force of colonisation for the community and hospital. Person-to-person transmission comes from colonised adults with disrupted gut flora ($\beta_{Disrupt}$), colonised adults with intact gut flora (β_{Intact}) and from infants ($\beta_{infants}$; community only). The transmission parameters from adults are estimated with our model. The base assumption is that infants are half as infectious as disrupted adults ($\beta_{infants} = 0.5\ \beta_{Disrupt}$), but we consider $\beta_{infants}$ in the range 0 - $\beta_{Disrupt}$ in our sensitivity analysis. Contact precautions (for patients receiving treatment) reduces transmission from these individual by factor q, determined for our previous model from contact precaution adherence rates [13]. (We assume no contact precautions in community, i.e. $q_{Comm} = 1$). Foodborne transmission adds to the force of colonisation in the community. Infants have different susceptibility to colonisation given by the factor θ. Therefore the force of colonisation for infants is θ times the force of colonisation in the community.

CDI treatment and outcomes

The rate at which individuals with CDI seek treatment in the community was inferred from studies on treatment seeking behaviour for those with diarrhoea. Van Cauteren et al report that 33% of those who have diarrhoea seek treatment, with mean time to treatment seeking being 1.5 days [14]. Assuming a competing hazards model of treatment seeking and recovery, this means that the treatment seeking rate (ρ_{Comm}) is 0.33/1.5 = 0.22 day^{-1}. Treatment seeking rate in the hospital is much faster ($\rho_{Hosp} = 1$ day^{-1}). Treatment rate (τ) is same in hospital and community (mean time is 10 days) [15]. Treatment success proportion (p_t) is the same in hospital and community and was estimated in our previous model [10]. We assume that patients do not have a greater hazard of death with CDI as death due to CDI is sufficiently infrequent so as not to significantly affect population level outcomes.

Details of parameter estimation

We used maximum likelihood estimation to determine the value of eight parameters. The likelihood function was composed of the product of likelihood functions for the colonisation prevalence in infants, the proportion seropositive for *C. difficile* toxin antibodies by age, the incidence of CDIs in hospitals and communities and number of patients admitted to hospital with CDI as a proportion of all admissions.

Colonisation Prevalence in Infants and Proportion Seropositive by Age

Kubota et al (2016) collected stool samples from 111 Belgian neonates at approximately one, three, eight, 31, 91, 143 and 182 days after birth. Since the exact number of days since birth was not reported for each sample and infant, for the purposes of likelihood calculations we assumed that each set of samples was taken exactly at exactly one, three, eight, 31, 91 143 and 182 days after birth. Kubota et al. tested each sample for carriage of *C. difficile* and presence of genes for toxins A and B. They reported the sequence of sample results (including missing samples) for the 55 infants who had at least one *C. difficile* positive sample. For the 56 samples that had no positive tests (for which the sequence of negative and absent samples were not reported) we assumed that there were no missing samples and that all samples were negative. Since we are interested only in toxigenic *C. difficile,* we considered all *C. difficile* positive without either toxin genes, as negative samples.

Rousseau et al (2012) collected monthly stool samples from ten French infants in their first year of life starting at approximately one month. The exact number of days since birth were not given for each sample so we assumed that samples were taken (or were missed) at exactly 30-day intervals. They performed strain typing on each positive stool sample, typing either one or five colonies. They reported the sequence of sample results including typing and missing results. Since we are interested only in toxigenic *C. difficile*, we only considered a sample to be positive if at least one of the typed isolates was of a toxigenic strain.

Rousseau et al also collected a single stool sample from 85 French children aged 1.5-36.2 months from two day-care centres. They reported the number of samples positive for *C. difficile* and including the number positive for toxigenic strains of *C. difficile.* Again, we considered only the samples with a toxigenic strain to be positive and samples with non-toxigenic strains or no strain of *C. difficile* to be negative.

Holst et al. (1981) collected stool samples from 130 infants aged 1-12 months and 88 children aged 1-15 years (total 218 samples) and tested for *C. difficile* carriage. All *C. difficile* positive samples were for toxigenic strains. The proportion colonised was reported by monthly age brackets for infants under 12 months and larger brackets for older children. For likelihood calculations we assumed that the age of all subjects was exactly in the middle of their age bracket (measured in 30-day months).

Adlerberth et al. (2014) also reported longitudinal *C. difficile* colonisation data in infants, but did not provide denominator data, so their data were not used for fitting the model. However, they did demonstrate that infants can be long-term carriers toxigenic strains of *C. difficile,* finding that a third of colonised infants were still colonised by the same strain six months later.

Viscidi et al. (1983) report the age-related prevalence of antibodies to toxin A and toxin B of *C. difficile* in 98 paediatric in-patients and 242 outpatients from a US hospital. The trends for toxins A and B were similar. Since toxin A and B antibodies were measured from the same set of patients, the two trends cannot be considered independent, so we considered only the data for toxin-B prevalence.

The probability that an individual would be in each state (i.e. colonised/non-colonised or seropositive/seronegative) at a given age was calculated for a given set of parameters using the Markov chain approximation for an individual, with the force of colonisation at the equilibrium value. For the cross-sectional datasets, the model likelihood was derived from a product of binomially distributed random variables with n_i trials (the number of samples for age group *i*), x_i

successes (i.e. the number of *C. difficile* positive or seropositive samples) and success probability p_i (the model predicted probability that an individual of that age group would be *C. difficile* positive or seropositive). For the datasets reporting the longitudinal sequence of sample results, the model likelihood a given sequence of results was the model probability that Markov chain of the individual would pass through those states at the sample ages.

The older, cross-sectional study by Holst et al. [16] found much higher prevalence of toxigenic colonisation amongst infants than recent studies [17,18], lead to poor model fit when combined with the recent studies or when used as the only infant colonisation data set and was therefore excluded.

Incidence of healthcare-associated and community-associated CDI

Lessa et al. (2015) estimated the incidence of CDIs in the United States based on surveillance data from several US counties. They used a three-day cut-off for hospital onset CDIs and a single twelve-week cut-off for community onset CDIs to differentiate healthcare and community acquired CDIs. In their article they further subdivided healthcare associated cases into community-onset healthcare acquired, nursing-home onset and hospital-onset hospital-acquired cases. We simulated the application of the three-day and twelve-week cut-off to determine the model incidence of hospital-onset hospital-acquired CDIs, community-onset hospital acquired CDIs and community acquired CDIs. However, because our model did not differentiate between nursing home and the general community, we included nursing home onset cases in with community acquired cases for the construction of the likelihood function. Recurrent cases – defined as any case with a period of symptoms in the 8 weeks prior to onset of symptoms (community-onset cases) or hospital admission (hospital-onset cases) – were excluded for the purposes of incidence estimation. The likelihood functions for the incidence of each category (e.g. community-onset hospital acquired) of CDI were normal and independent with mean equal to the reported incidence and standard deviation equal to width of the reported confidence interval divided by 2×1.96.

Proportion of Admissions with CDI

HCUP provide information on hospitalisations in the Unites States stratified by diagnoses. We extracted the total number of hospital inpatient discharges and the number of hospital inpatient discharged where the principal diagnosis was *C. difficile* infection (Diagnoses--ICD-9-CM Codes (ICD9), Principal Diagnosis: 008.45 Int Inf Clstrdium Dfcile) for the year 2014 (the most recent available data at time of extraction) [19]. The likelihood function for the proportion of admissions with CDI was normally distributed with mean equal to the proportion of all discharges where CDI was the primary diagnosis and standard deviation equal to the reported confidence interval divided by 2×1.96.

Alternate fitting assumptions

We considered multiple sets of assumptions when trying to estimate the transmission rates which are best expressed in terms of the transmission parameters giving the transmission rate from each type of colonised individual by setting (hospital or community): $\beta_{Comm,Intact}$, $\beta_{Hosp,Intact}$, $\beta_{Comm,Disrupt}$, $\beta_{Hosp,Disrupt}$ and β_{infant}. Throughout we assumed that β_{infant} was some fraction 0-1 of $\beta_{Comm,Disrupt}$, however we considered multiple relations between the other parameters.

We tried assuming $\beta_{Comm,Intact,} = \beta_{Hosp,Intact,} = 0$ and $\beta_{Comm,Disrupt,} = \beta_{Hosp,Disrupt}$ with generally poor results. Model fit was reasonable when adult colonisation prevalence was low and infants as infectious as adults, however nearly all transmission came from infants in this scenario. For high colonisation prevalence and equal infant infectiousness, though the model fit was still good, this could only be

achieved with very high transmission rates in the community and very low (~10%) antibiotics disruption probability and still nearly all transmission came from infants. With lower infant infectiousness in the range 0-0.2 times adult infectiousness, the model had a very poor fit with the incidence in the community being very low (<1/3 of data), and infant colonisation rates being high, transmission rates from adults being excessive (>3) and probability that antibiotics disrupt being very low (<10%).

We tried assuming a single transmission rate $\beta_{Comm,Intact,}$= $\beta_{Hosp,Intact,}$= 0 with $\beta_{Comm,Disrupt,}$ and $\beta_{Hosp,Disrupt}$ independent. This allowed for arbitrary and independent forces of colonisation in hospital and community and had good model fit for all assumptions. However, in the resulting model nearly all transmission in the community came from infants for most assumed values of infant infectiousness and adult colonisation prevalence. When infant infectiousness was restricted to lower levels (<0.2 $\beta_{Comm,Disrupt}$), $\beta_{Comm,Disrupt}$ was up to 50 times greater than $\beta_{Hosp,Disrupt}$, which is highly implausible.

Finally, our default assumptions ($\beta_{Comm,Intact,}$= $\beta_{Hosp,Intact}$ and $\beta_{Comm,Disrupt,}$= $\beta_{Hosp,Disrupt}$) resulted in equally good model fit to the previous set of assumptions, but with much more believable transmission rates and proportion of transmission attributable to infants under most combinations of assumptions. However, when infant infectiousness was high and adult colonisation prevalence was low the estimated value of $\beta_{Comm,Intact,}$ and $\beta_{Hosp,Intact,}$ were 0, leading to poorer model fit and unbelievably high proportion of transmission from infants and so the result from these extremes were omitted from our sensitivity analyses in the main text and the supplementary figures.

Further details of model fit

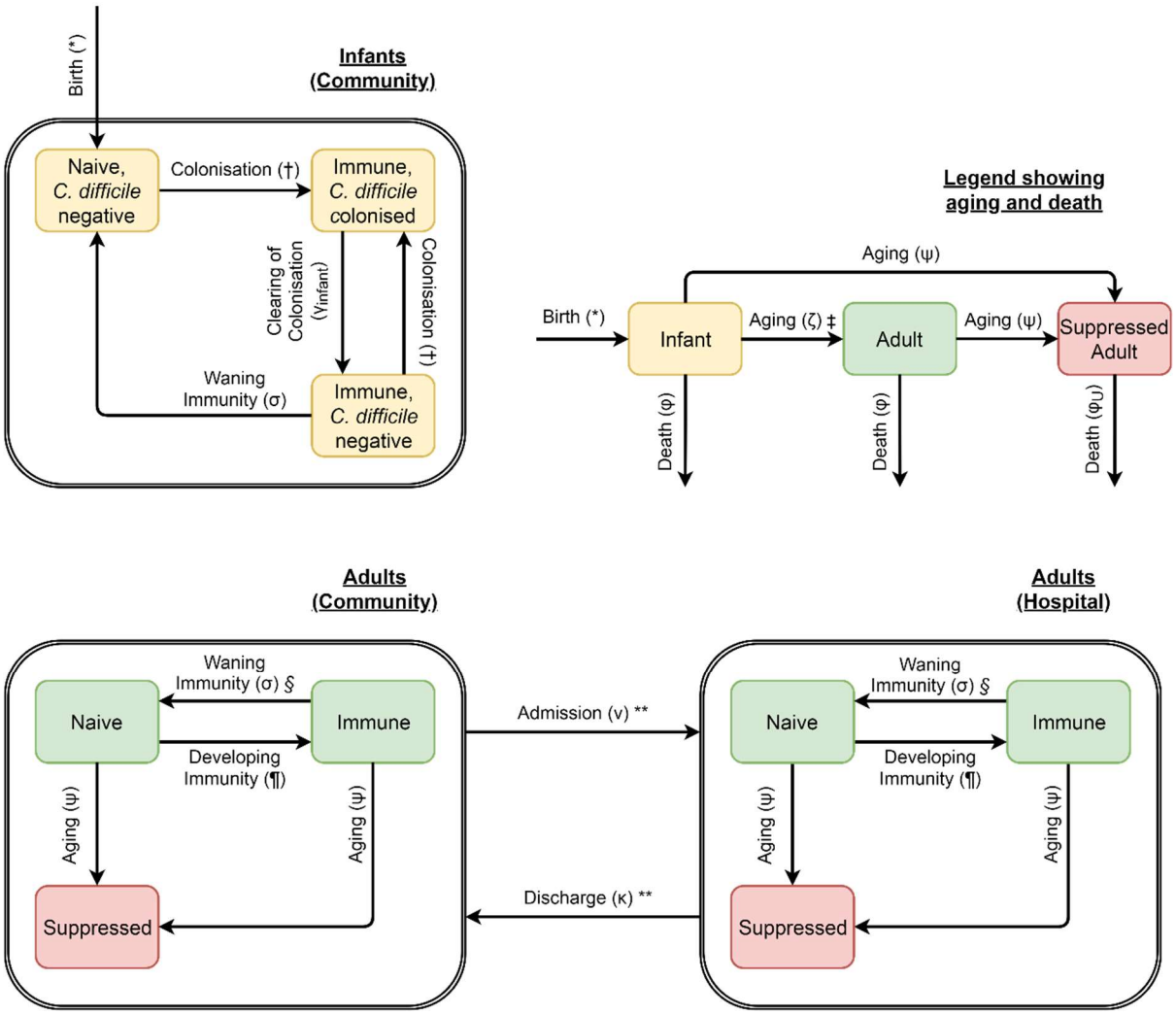
The model fit the data well reproducing the observed age profile of colonisation, immunity, reported incidence of infection, proportion of admissions for CDI (**Supp. Figure 2**). The model was also verified by outcomes not used to fit the model. The reported recurrence proportion is approximately 20% for hospital-acquired cases and approximately 10% of community-acquired cases [20]. We estimated that 18% (range 13-30%) of transmission in hospitals was from patients receiving treatment for CDI. A study using whole genome sequencing to compare isolates from CDI cases in Oxfordshire hospitals estimated that approximately one quarter of cases could be linked to ward-based transmission from another identified symptomatic carrier [21]. Using the same reporting and treatment seeking assumptions used for initial cases and defining a recurrence as a return to symptomatic colonisation within 8 weeks of resolution of symptoms, the model predicted a 19.4% recurrence proportion in hospital-acquired cases (range: 18.5-21.5%) and a 14.0% recurrence proportion in community-acquired cases (range: 13.3-15.8%). Notably, the true recurrence proportions (i.e. without simulating underreporting and misclassification) were higher for both hospital-acquired (24.6-26.8%) and community-acquired (17.8-18.5%) infections. Adlerberth et al [22] found that a third of all infants colonised with *C. difficile* were colonised by the same strain when sampled at least six months later, while the model predicted that 40% (33-42%) of colonised six-month-old infants remain colonised at 12 months of age. A previous estimate of the force of colonisation in a hospital setting [10] derived from the reported risk of colonisation and infection as a function of days of hospital stay [23] was 0.007 day^{-1} (95% CI: 0.004-0.011). In this model the estimated force of colonisation was a little lower: 0.0033 day^{-1} (range: 0.0031-0.0035 day^{-1}). However the previous estimate was drawn from a hospital during a period high incidence. In our model, a higher proportion (89.6%, range: 87.9-90.4%) of those classified as hospital-acquired CDI were elderly/immune-supressed compared to cases classified as community-acquired CDI (78.4% range: 77.6-78.6%), in agreement with the observation that community-acquired cases are younger, with fewer comorbidities [24].

Supplementary Tables and Figures

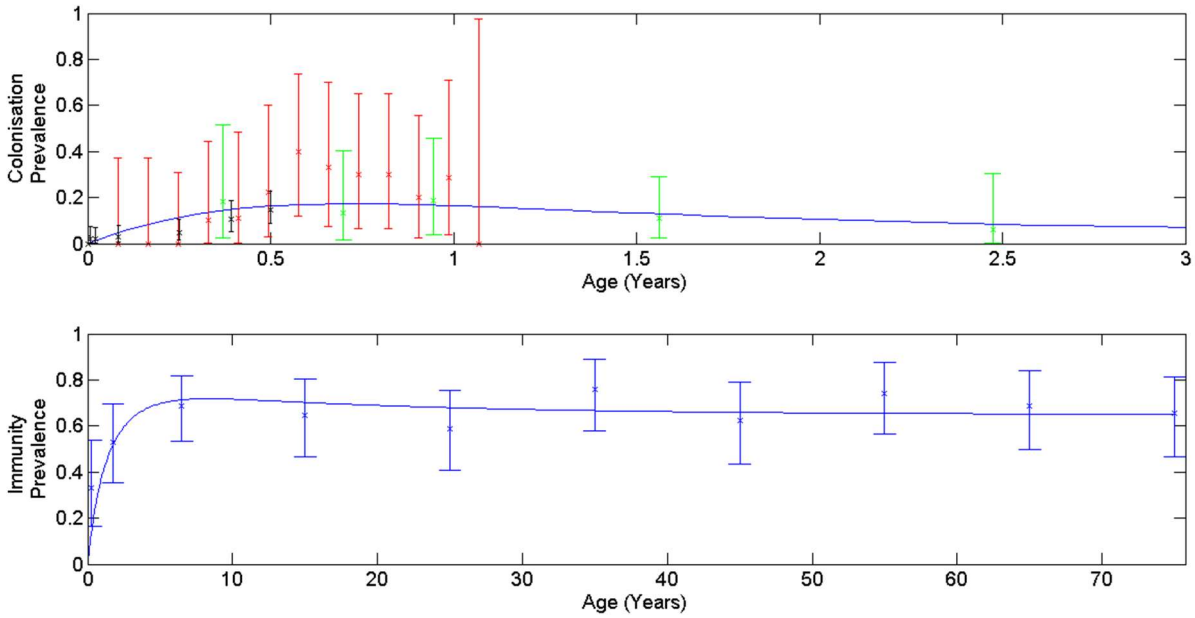
Supp. Table 1 Definitions, values and references for all parameters used in the model. All rates are in units of day^{-1}. *These parameters were fit to the model, with the range indicating values over sensitivity analysis. †Only these parameters were affected by assumptions around infant infectiousness, being estimated under the assumption that $\beta_{Infant} = k \times \beta_{Disrupt}$ for k in the range 0-1. ‡ These parameter values are the same as our previous model of hospital transmission [10].

Parameter	Description	Value Hospital/ Admission	Value Community/ Discharge	References
α	Rate at which patients begin new antimicrobial treatment (proxy for rate of gut flora disruption)	0.33	1.5×10^{-3}	[5,6]
α_U	Rate at which patients begin new antimicrobial treatment (proxy for rate of gut flora disruption) for elderly/suppressed	0.33	2.5×10^{-3}	[5,6]
λ	Rate at which damaged commensal gut flora recovers	0.03‡		[7,8]
λ_o	Rate at which damaged commensal gut flora recovers for those with CDI overgrowth	0.25		[9]
γ	Rate at which *C. difficile* is cleared in those with intact gut flora	0.023‡		[11]
ω	Rate of *C. difficile* overgrowth in patients with disrupted gut flora	0.2‡		[12,25]
ρ	Rate at which treatment for CDI is given (hospital) or sought (community)	1	0.22	[14]
τ	Rate at which the treatment of patients with CDI progresses towards resolution	0.1‡		[15,26]
p_t	Probability that CDI treatment will remove all *C. difficile*	0.55‡		[10]
q	Efficacy and coverage of special contact precautions (0: total prevention of transmission; 1: no reduction in transmission)	0.45‡	1	[13,27]
δ	Rate of development of immunity in adults	0.1‡		[2,3]
ζ	Rate at which infants age to develop adult-like gut-flora	N/A	1/365	[28]

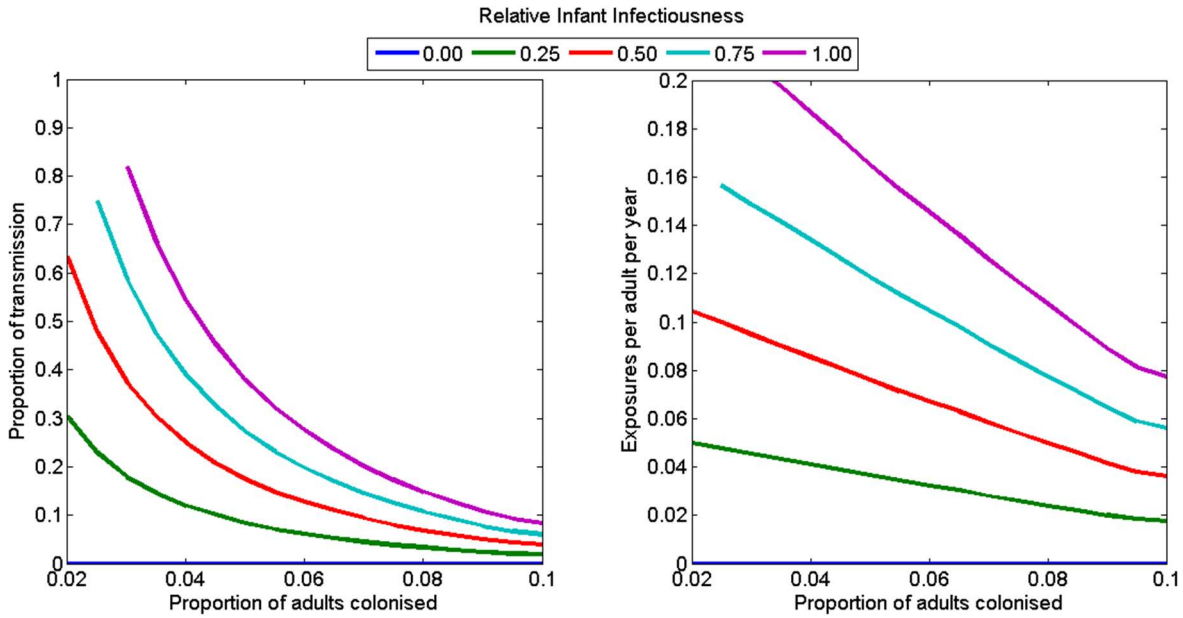
ψ	Rate at which people age to suppressed/elderly class	2.92×10^{-5} (Median time: 65 years)		
φ	Death rate	1.11×10^{-5}		[1]
φ_U	Death rate for elderly/suppressed	1.84×10^{-4}		[1]
N	Overnight Hospital Beds / Community Size	150	100,000	
σ	Rate at which immunity wanes	2.9×10^{-4} (0.9×10^{-4} – 6.9×10^{-4})		*
ϑ	Multiplicative factor for colonisation susceptibility of infants	N/A	1.4 (0.6-4.4)	*
γ_{infant}	Rate at which *C. difficile* is cleared in infants	0.0020 (0.0020-0.0025)		*
$p_{disturb}$	Proportion of antibiotics that disturb gut flora	0.22 (0.12-0.48)		*
p_{report}	Proportion of all community-treated CDIs that are reported	0.63 (0.57-0.76)		*
$\beta_{Disrupt}$	Transmission rate coefficient for colonised adults with disrupted gut flora (due to recent antibiotic exposure)	0.128 (0.071-0.174)		*†
β_{Intact}	Transmission rate coefficient for colonised adults with intact gut flora (no recent antibiotic exposure)	0.019 (0.001-0.026)		*†
β_{infant}	Transmission rate coefficient from infants	N/A	0.064 (0-0.174)	*†
ν / κ	Hospital admission/discharge rate	2.40×10^{-4}	0.241	[4]
ν_U / κ_U	Hospital admission/discharge rate for elderly	8.06×10^{-4}	0.192	[4]
ν_{CDI} / κ_t	Hospital admission rate for CDI / discharge rate for those receiving CDI treatment	0.0142 (0.0128-0.0174)	0 (Not discharged)	*



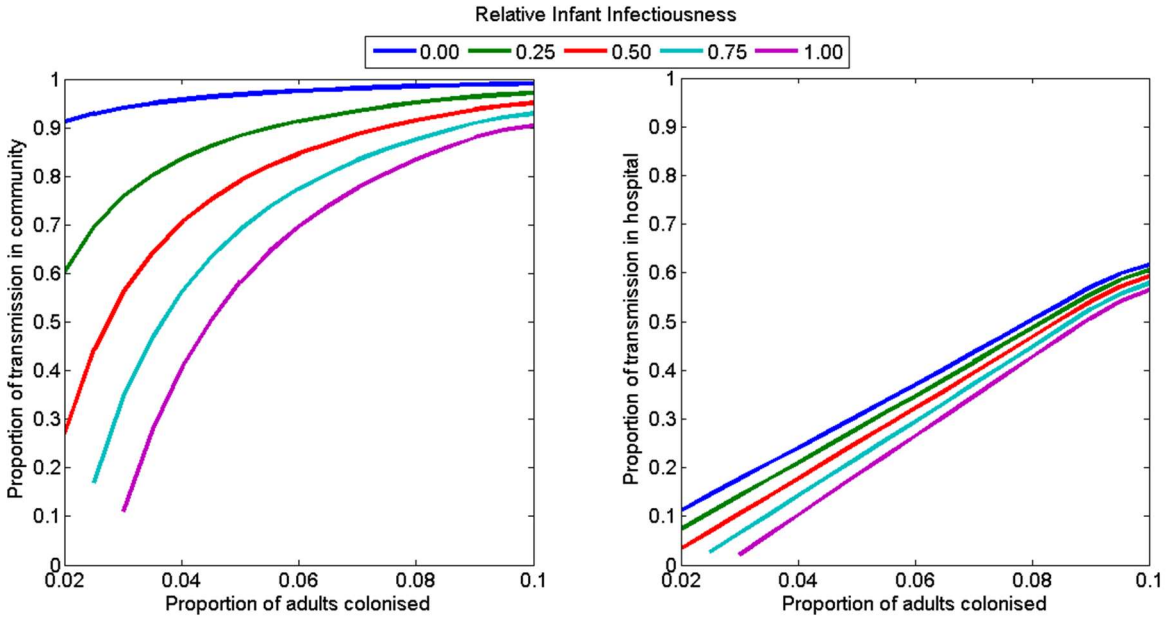
Supp. Figure 1 Model structure, showing immune states, aging, births, deaths, hospital admission and discharge and infant classes. *Birth rate matches death rate from whole population. †The force of colonisation depends on the number of infectious individuals in the population. Infants are θ times more susceptible to colonisation than adults. ‡Infants retain their immunity and colonisation status when they age to become non-suppressed adults. §Only non-colonised individuals can have their immunity wane. ¶Those with active CDI develop immunity upon recovery. Asymptomatically colonised individuals develop immunity at rate δ. Non-colonised individuals do not develop immunity. **Admission discharge rates vary by immunity and CDI status. Patients receiving treatment for CDI are not discharged and are admitted at a much higher rate.



Supp. Figure 2 Maximum likelihood model fit to infant colonisation data and immunity prevalence data, assuming 5% colonisation prevalence in adults. Blue crosses indicate *C. difficile* toxin B antibody sero-prevalence [29], green crosses indicate cross-sectional study infants and toddlers in childcare [17]. Red crosses [17] and black crosses [18] indicate longitudinal studies of *C. difficile* colonisation in infants. All error bars are 95% binomial confidence intervals.



Supp. Figure 3 *C. difficile* transmission in the community from infants under various plausible assumptions for the *C. difficile* colonisation prevalence in adults, and the relative infectiousness of infants as (A) a proportion of all transmission in the community and (B) as rate of exposure to adults in the community.



Supp. Figure 4 *C. difficile* transmission from colonised adults with intact gut flora under various assumptions for the *C. difficile* colonisation prevalence in adults, and the relative infectiousness of infants as (A) the proportion of all transmission in the community and (B) the proportion of all transmission in the hospital.

References

1. **Centers for Disease Control and Prevention - National Center for Health Statistics.** *CDC Wonder: Underlying Cause of Death, Multiple Cause of Death Files, 1999-2016*. (http://wonder.cdc.gov/ucd-icd10.html). Accessed 18 December 2017.

2. **Foglia G, *et al.*** *Clostridium difficile*: Development of a novel candidate vaccine. *Vaccine* Elsevier Ltd, 2012; **30**: 4307–4309.

3. **de Bruyn G, *et al.*** Defining the optimal formulation and schedule of a candidate toxoid vaccine against *Clostridium difficile* infection: A randomized Phase 2 clinical trial. *Vaccine* Elsevier Ltd, 2016; **34**: 2170–2178.

4. **Weiss (Truven Health Analytics) AJ, Elixhauser (AHRQ) A**. *Overview of Hospital Stays in the United States, 2012: Statistical Brief #180*. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville, MD: Agency for Healthcare Research and Quality, 2014.

5. **MacDougall C, Polk RE**. Variability in rates of use of antibacterials among 130 US hospitals and risk-adjustment models for interhospital comparison. *Infection control and hospital epidemiology* 2008; **29**: 203–11.

6. **Australian Commonwealth Department of Health - Drug utilisation sub-committee (DUSC)**. *Antibiotics : PBS / RPBS utilisation*. Canberra, 2015.

7. **Rafii F, Sutherland JB, Cerniglia CE**. Effects of treatment with antimicrobial agents on the human colonic microflora. *Therapeutics and clinical risk management* 2008; **4**: 1343–58.

8. **Dial S, *et al.*** Patterns of antibiotic use and risk of hospital admission because of *Clostridium difficile* infection. *Canadian Medical Association Journal* 2008; **179**: 767–772.

9. **Bartlett JG**. Treatment of Antibiotic-Associated Pseudomembranous Colitis. *Clinical Infectious Diseases* 1984; **6**: S235--S241.

10. **McLure A, *et al.*** Healthcare-Associated *Clostridium difficile* Infections are Sustained by Disease from the Community. *Bulletin of Mathematical Biology* 2017; **79**: 2242–2257.

11. **Johnson S, *et al.*** Treatment of asymptomatic *Clostridium difficile* carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. *Annals of Internal Medicine* 1992; **117**: 297–302.

12. **Freeman J**. Effects of cefotaxime and desacetylcefotaxime upon *Clostridium difficile* proliferation and toxin production in a triple-stage chemostat model of the human gut. *Journal of Antimicrobial Chemotherapy* 2003; **52**: 96–102.

13. **Muto CA, *et al.*** A Large Outbreak of *Clostridium difficile*–Associated Disease with an Unexpected Proportion of Deaths and Colectomies at a Teaching Hospital Following Increased Fluoroquinolone Use. *Infection Control & Hospital Epidemiology* 2005; **26**: 273–280.

14. **Van Cauteren D, *et al.*** Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiology and Infection* 2012; **140**: 697–705.

15. **Leffler DA, Lamont JT**. Treatment of *Clostridium difficile*-associated disease. *Gastroenterology* AGA Institute American Gastroenterological Association, 2009; **136**: 1899–912.

16. **Holst E, Helin I, Mårdh PA**. Recovery of *Clostridium difficile* from children. *Scandinavian Journal of Infectious Diseases* 1981; **13**: 41–5.

17. **Rousseau C, *et al.*** *Clostridium difficile* Carriage in Healthy Infants in the Community: A Potential Reservoir for Pathogenic Strains. *Clinical Infectious Diseases* 2012; **55**: 1209–1215.

18. **Kubota H, *et al.*** Longitudinal Investigation of Carriage Rates, Counts, and Genotypes of Toxigenic *Clostridium difficile* in Early Infancy. *Applied and Environmental Microbiology* 2016; **82**: 5806–14.

19. **Healthcare Cost and Utilization Project (HCUP)**. *HCUP National Inpatient Sample (NIS)*. Rockville, MD: Agency for Healthcare Research and Quality, 2014.

20. **Gerding DN, Lessa FC**. The epidemiology of *Clostridium difficile* infection inside and outside health care institutions. *Infectious disease clinics of North America* 2015; **29**: 37–50.

21. **Walker AS, *et al.*** Characterisation of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. *PLoS medicine* 2012; **9**: e1001172.

22. **Adlerberth I, *et al.*** Toxin-producing *Clostridium difficile* strains as long-term gut colonizers in healthy infants. *Journal of Clinical Microbiology* 2014; **52**: 173–9.

23. **Loo VG, *et al.*** Host and Pathogen Factors for *Clostridium difficile* Infection and Colonization. *New England Journal of Medicine* 2011; **365**: 1693–1703.

24. **Khanna S, *et al.*** The Epidemiology of Community-Acquired *Clostridium difficile* Infection: A Population-Based Study. *The American Journal of Gastroenterology* Nature Publishing Group, 2012; **107**: 89–95.

25. **Freeman J**. Comparison of the efficacy of ramoplanin and vancomycin in both in vitro and in vivo models of clindamycin-induced *Clostridium difficile* infection. *Journal of Antimicrobial Chemotherapy* 2005; **56**: 717–725.

26. **Khan FY, *et al.*** Epidemiological features of *Clostridium difficile* infection among inpatients at Hamad General Hospital in the state of Qatar, 2006-2009. *Travel medicine and infectious disease* Elsevier Ltd, 2012; **10**: 179–85.

27. **Jarvis WR, *et al.*** National point prevalence of *Clostridium difficile* in US health care facility inpatients, 2008. *American Journal of Infection Control* 2009; **37**: 263–270.

28. **Jangi S, Lamont JT**. Asymptomatic Colonization by *Clostridium difficile* in Infants: Implications for Disease in Later Life. *Journal of Pediatric Gastroenterology and Nutrition* 2010; **51**: 2–7.

29. **Viscidi R, *et al.*** Serum Antibody Response to Toxins A and B of *Clostridium difficile*. *Journal of Infectious Diseases* 1983; **148**: 93–100.

Details of Classifying Cases as Hospital or Community-Acquired

The simulation of CDI classification is a key component of the current study. The definitions used to classify CDIs rely on a knowledge of the patient's history of previous CDI (to rule out recurrent cases) hospital admission and discharge prior to onset of symptoms (to distinguish hospital and community-acquired cases). However, standard compartmental epidemiological models are memoryless and do not explicitly model (or record) the events occurring to individuals, only how an event (i.e. infection or recovery) affects the total number of individuals of any given compartment. An individual-based model could be used to simulate these details, but we employ a computationally much simpler approximation that approximates individual in the population as a simple Markov chain.

To make the individual Markov assumption, the only non-linear interaction term, the force of colonisation, is fixed as a constant in time (equal to the mean force of colonisation at equilibrium) so the population can be viewed as a large ensemble of independent Markov chains (individual people). The state space of each Markov chain is the union of S, the set of living states corresponding to the compartments of the model and Δ the death state (which is the only absorbing state). New Markov chains are initialised (births and immigration) at the jump times of a Poisson process of rate $\mu(t)$ (the birth/immigration rate) which we will assume is independent of the individuals in the model and is homogeneous (i.e. a constant birth rate). At birth these Markov chains are randomly assigned to one of the living states in S (the compartments in the model) according to the vector of probabilities $\boldsymbol{\pi}$. In our model there is no immigration (or emigration) and all births are in the same state, so $\boldsymbol{\pi}$ is the standard basis vector corresponding to the non-immune, non-colonised infant class. If there was immigration in our model, $\boldsymbol{\pi}$ would describe the probability that a new arrival was of any given compartment. These Markov chains then progress through the transient states in S, before being absorbed into the death state. For each individual Markov chain, the age a is the time since it was initialised. So if Q is the transition rate matrix between states in S, then the vector of probabilities that a given individual is in each state at age a is

$$\left[P(X(a) = s)\right]_{s\in S} = e^{Qa}\boldsymbol{\pi}. \tag{1}$$

In a whole population of these Markov chains starting with $\boldsymbol{x_0}$ people in each (living) state at time 0, the expected number of people alive in each state at time t is given by the vector

$$E[\boldsymbol{x}(t)] = e^{Qt}\boldsymbol{x_0} + \int_0^t \mu(t-a)e^{Qa}\boldsymbol{\pi}da. \tag{2}$$

If the birthrate μ is constant over time then

$$E[\boldsymbol{x}(t)] = e^{Qt}\boldsymbol{x_0} + \mu Q^{-1}\left(e^{Qt} - I\right)\boldsymbol{\pi}. \tag{3}$$

Since Q is the transition matrix for transient states (i.e. there are no immortal states), the limit for large t when the system approaches its equilibrium distribution is

$$\lim_{t\to\infty} E[\boldsymbol{x}(t)] = -Q^{-1}\boldsymbol{\pi}\mu \tag{4}$$

which is independent of t and $\boldsymbol{x_0}$. The total population is then sum over the vector, $-\mathbf{1}^T Q^{-1}\boldsymbol{\pi}\mu$. Note that $-\mathbf{1}^T Q^{-1}\boldsymbol{\pi}$ is the total of the mean dwelling time in each state (i.e. the total average time each person spends alive) and so is equal to the mean life expectancy of individuals, L.

We have an expression for the number of individuals of a given type at equilibrium but we need to know the number individuals at equilibrium which have a given history of CDI and hospitalisation. So we are interested in the number of people at equilibrium which have (or have not) been in some set of states $S_1 \subset S$, in the past T_1 units of time. For instance we may be interested in the population which haven't been in the hospital states (S_1) in the last 12 weeks (T_1). So consider the vector of probabilities that a individual Markov chain of age a is in each state s but has not been any state in S_1 in the past T_1 units of time:

$$\left[P(X(a) = s \text{ and } X(\tau) \notin S_1 \ \max\{0, a - T_1\} \leq \tau \leq a\})\right]_{s\in S}. \tag{5}$$

To derive an expression for these probabilities we consider a modified Markov chain X_a which until time $\max\{0, a-T_1\}$ behaves as the original Markov chain, but at time $\max\{0, a-T_1\}$ individuals in S_1 are moved to the absorbing 'death' state, and after $\max\{0, a-T_1\}$ all state transitions entering states in S_1 are now redirected to the absorbing 'death' state. For X_a the transition rate matrix between states in S after time $\max\{0, a-T_1\}$ is Q with the rows and columns corresponding to states in S_1 set to zero. If we assume, without loss of generality, that the states are ordered such that states in S_1 are last we can write this in block matrix form as $\begin{bmatrix} Q_1 & 0 \\ 0 & 0 \end{bmatrix}$, where is Q_1 is the sub-matrix of Q corresponding to states in $S \setminus S_1$. Therefore

$$[P(X(a)=s \text{ and } X(\tau) \notin S_1 \ \max\{0,a-T_1\} \le \tau \le a)]_{s\in S} = [P(X_a(a)=s)]_{s\in S} \tag{6}$$

$$= \begin{cases} \mathbb{P}_1^a \boldsymbol{\pi}, \ a < T_1 \\ \mathbb{P}_1^{T_1} \mathbb{P}^{a-T_1} \boldsymbol{\pi}, \ a \ge T_1, \end{cases} \tag{7}$$

where $\mathbb{P} := e^Q$ and $\mathbb{P}_1 := \begin{bmatrix} e^{Q_1} & 0 \\ 0 & 0 \end{bmatrix} = \exp\begin{bmatrix} Q_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ are matrices the same size as Q.

Therefore the expected number of people at equilibrium which have not been in some set of states $S_1 \subset S$, in the past T_1 units of time is the vector

$$\lim_{t\to\infty} \int_0^t [\mu(t-a)[P(X_a(a)=s)]_{s\in S}\, da = \lim_{t\to\infty} \left[\int_0^{T_1} \mu(t-a)\mathbb{P}_1^a \boldsymbol{\pi} da + \int_{T_1}^t \mu(t-a)\mathbb{P}_1^T \mathbb{P}^{a-T_1} \boldsymbol{\pi} da \right] \tag{8}$$

If μ is constant then this simplifies to

$$\mu \left[\int_0^{T_1} \mathbb{P}_1^a da + \int_{T_1}^{\infty} \mathbb{P}_1^{T_1} \mathbb{P}^{a-T_1} da \right] \boldsymbol{\pi} = \mu \left[\int_0^{T_1} \mathbb{P}_1^a da + \int_0^{\infty} \mathbb{P}_1^{T_1} \mathbb{P}^a da \right] \boldsymbol{\pi} \tag{9}$$

$$= \mu \left[\mathbb{Q}_1^{-1} \left(\mathbb{P}_1^{T_1} - I \right) - \mathbb{P}_1^{T_1} Q^{-1} \right] \boldsymbol{\pi} \tag{10}$$

where by an abuse of notation, $\mathbb{Q}_1^{-1} := \begin{bmatrix} Q_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}$ is a square matrix the size of Q.

The same reasoning can be extended to count the individuals at equilibrium which have not been in the sets of states $S_1, S_2, \ldots, S_n$ in the past $T_1 > T_2 > \ldots T_n$ units of time respectively. The vector of probabilities that an individual of age a is in each state s and satisfies these requirements is

$$\left[P\left(X(a)=s \text{ and } \bigcup_{i=1}^n \{X(\tau) \notin S_i \quad \max\{0,a-T_i\} \le \tau \le a\} \right) \right]_{s\in S} = \begin{cases} \mathbb{P}_n^a \boldsymbol{\pi}, & a < T_n \\ \mathbb{P}_n^{T_n} \mathbb{P}_{n-1}^{T_{n-1}-T_n} \ldots \mathbb{P}_i^{T_i - T_{i+1}} \mathbb{P}_{i-1}^{a-T_i} \boldsymbol{\pi}, & T_i \le a < T_{i-1} \\ \mathbb{P}_n^{T_n} \mathbb{P}_{n-1}^{T_{n-1}-T_n} \ldots \mathbb{P}_1^{T_1-T_2} \mathbb{P}^{a-T_1} \boldsymbol{\pi}, & T_1 \le a \end{cases} \tag{11}$$

and the expected number of each type of individual in the population at equilibrium (assuming constant birth rate μ) is

$$\mu[\mathbb{Q}_n^{-1}(\mathbb{P}_n^{T_n} - I) + \mathbb{P}_n^{T_n}\mathbb{Q}_{n-1}^{-1}(\mathbb{P}_{n-1}^{T_{n-1}} - I) + \mathbb{P}_n^{T_n}\mathbb{P}_{n-1}^{T_{n-1}-T_n}\mathbb{Q}_{n-2}^{-1}(\mathbb{P}_{n-2}^{T_{n-2}} - I) + \ldots$$
$$+ \mathbb{P}_n^{T_n}\mathbb{P}_{n-1}^{T_{n-1}-T_n} \ldots \mathbb{P}_2^{T_2-T_3}\mathbb{Q}_1^{-1}(\mathbb{P}_1^{T_1} - I) - \mathbb{P}_n^{T_n}\mathbb{P}_{n-1}^{T_{n-1}-T_n} \ldots \mathbb{P}_1^{T_1-T_2} Q^{-1}]\boldsymbol{\pi}$$

where $\mathbb{P}_i := \begin{bmatrix} e^{Q_i} & 0 \\ 0 & 0 \end{bmatrix}$ and $\mathbb{Q}_i^{-1} := \begin{bmatrix} Q_i^{-1} & 0 \\ 0 & 0 \end{bmatrix}$ are matrices the same size as Q with Q_i being the sub-matrices of Q corresponding to the states in $S \setminus \cup_{j=i}^n S_j$.

Now we have a way to calculate the incidence of (non-recurrent) CDIs which would be classified as hospital or community-acquired if standard definitions were used. We will briefly illustrate this for community onset cases for the system recommended by IDSA and SHEA.

First calculate the equilibrium number of people in each class in the community that haven't been in symptomatic states (S_1) in the past 8 weeks (T_1). At this equilibrium point calculate the rate at which the transitions corresponding to onset of CDIs occur in the community, i.e. the total rate of transitions from

asymptomatically colonised community states to symptomatic community states. This is the incidence of non-recurrent community-onset CDI.

The rate of non-recurrent CDI classified as community-acquired is the total rate of transitions corresponding to the onset of CDI for the equilibrium number of people who haven't been in the hospital states (S_1) in the past 12 weeks (T_1) or any symptomatic state (S_2) in the past 8 weeks (T_2).

The rate of non-recurrent CDI classified as indeterminate is the total rate of transitions corresponding to the onset of CDI for the equilibrium number of people who haven't been in the hospital states (S_2) in the past 4 weeks (T_2) or any symptomatic state (S_1) in the past 8 weeks (T_1) minus the rate of non-recurrent CDIs classified as community-acquired.

Finally, the rate of non-recurrent CDI classified as hospital-acquired is the rate of non-recurrent CDI minus the rate of non-recurrent CDIs classified as community-acquired or indeterminate.

The first steps of classifying of hospital-onset is similar, however the history of recurrence and previous hospitalisation is considered from the point of hospital admission. In other other words, we determine the proportion of admissions which have or have not been in a hospital in the past 12 weeks etc. Given the history of patients at the point of admission, one can simply simulate the course of hospitalisations going forward and (e.g. how many CDIs occur within the first 2 days) and classify all cases as hospital or community- acquired according to the history at admission *and* what occurs during the hospital stay.

To account for possible unreported cases of CDI in the community, once we calculate the incidence of community-onset CDIs, we then calculate (using standard Markov chain calculations for absorption probabilities) the number of community-onset cases that seek treatment in the community or in the hospital. A fraction p_{Report} of patients that seek treatment in the community and all patients that do not seek treatment in the community but are admitted to hospital are considered 'reported' and count towards the incidence calculation when fitting the model to Lessa et al. A small number patients seek treatment in the community and then are also admitted to hospital before recovery; for simplicity only p_{Report} of these are considered reported. Patients that do not seek any form of treatment before recovering are not counted towards the 'reported' incidence calculation. Note that this adjustment for 'reporting' is not made when we determine whether a case is recurrent or not; that is any CDI event in the patient's recent history (even if they didn't seek treatment) excludes the current CDI as recurrent. This close approximation greatly simplifies the calculations for excluding recurrent cases.

To see the true (not classified) location of acquisition for each CDI, we use a modified model which splits every *C. difficile*-positive compartment into separate compartments for hospital and community-acquired. The incidence of community-acquired CDI is then the equilibrium transition rate from community-acquired, asymptomatically colonised states to the corresponding community-acquired, infected (symptomatic) states.

To compare the classification system to the true incidence, we used the methods described above on the modified model. For instance, we could calculate the incidence of community-acquired, community-onset CDI that are incorrectly classified as hospital-acquired, community-onset CDI: the rate at which people in community-acquired asymptomatically colonised states (that have been in any of the hospital states in the past 4 weeks) transition into symptomatic states in the community. Note, we assume that the time and place of the onset of symptoms is always known accurately, i.e. community-onset cases are never misclassified as hospital-onset or vice versa.

5.3 Model Equations

This section provides an explicit description of the models that are used in this chapter and serve as the starting point for the model in Chapter 6. Though this section does not provide any information that cannot be inferred from the preceding sections in this chapter, the equations are written out in full for the convenience of the interested reader.

The set of compartments for the whole-population model in this chapter is larger than the set of compartments for the hospital-only model in Chapters 3 and 4 for three reasons. First, the entire set of compartments in the hospital-only population is duplicated for the community. Second, to explicitly capture different treatment seeking rates between hospital and community, the symptomatic infection compartments are split off into treated and non-treated. Finally, three additional compartments are added for infants in the community. Furthermore, since the current model captures processes at the time-scale of decades and lifetimes, rather than on the time-scale of hospital admissions, more kinds of transitions have been included (*e.g.* , birth, death, aging, waning immunity). Therefore, while the whole-population model can be seen as consisting of sub-models very similar to the hospital-only model, there the system of equations contains many more variables and terms.

Let $\mathbf{X}$ be the vector of the number of persons in each compartment ordered as $H^H, H^H_a, H^H_c, H^H_{ac}, H^H_{ao}, H^H_{aot}, U^H, U^H_a, U^H_c, U^H_{ac}, U^H_{ao}, U^H_{aot}, R^H, R^H_a, R^H_c, R^H_{ac}, R^H_{ao}, H^C, H^C_a, H^C_c, H^C_{ac}, H^C_{ao}, H^C_{aot}, U^C, U^C_a, U^C_c, U^C_{ac}, U^C_{ao}, U^C_{aot}, R^C, R^C_a, R^C_c, R^C_{ac}, R^C_{ao}, H^I, R^I_c, R^I,$ where – similar to the notation for the model in Chapter 3 – the main symbol indicates immune status (H healthy, U immune-suppressed/elderly, R resistant/immune), subscript a indicates gut flora disrupted by antibiotics, subscript c and o indicate colonisation and overgrowth by *C. difficile*, subscript t indicates treatment, superscripts H and C indicate adults in hospital and the community and superscript I indicates infants (in the community). Then the ODE formulation of the model used to calculate the endemic equilibrium and reproduction numbers for different model parameters can be stated as

$$\mathbf{X}' = A(\mathbf{X})\mathbf{X} + \mathbf{b}, \tag{5.1}$$

Here $\mathbf{b}$ is a vector for the constant birth term which is zero for all (37) entries except for the entry for healthy non-colonised infants (H^I, entry 35) which is $N^C(\phi+\psi)\phi_U/(\phi_U+\psi)$. $A(\mathbf{X})$ is a large (37×37) matrix which can be written in block matrix form, split up by

immune status and hospital/community/infant status as

$$\begin{bmatrix} A_{H^H H^H} & 0 & A_{H^H R^H} & A_{H^H H^C} & 0 & 0 & 0 \\ A_{U^H H^H} & A_{U^H U^H} & A_{U^H R^H} & 0 & A_{U^H U^C} & 0 & 0 \\ A_{R^H H^H} & 0 & A_{R^H R^H} & 0 & 0 & A_{R^H R^C} & 0 \\ A_{H^C H^H} & 0 & 0 & A_{H^C H^C} & 0 & A_{H^C R^C} & A_{H^C I} \\ 0 & A_{U^C U^H} & 0 & A_{U^C H^C} & A_{U^C U^C} & A_{U^C R^C} & A_{U^C I} \\ 0 & 0 & A_{R^C R^H} & A_{R^C H^C} & 0 & A_{R^C R^C} & A_{R^C I} \\ 0 & 0 & 0 & 0 & 0 & 0 & A_{II} \end{bmatrix} \tag{5.2}$$

The main diagonal sub-matrices – which contain information on gut flora status, colonisation, overgrowth and treatment – are much the same for adults of the same immune state in hospital and community and similar to the main diagonal sub-matrices in the hospital-only model. The main difference here is the addition of the treatment compartments for U- and H-type persons and and the novel infant classes. For compactness each sub-matrix is written as the difference of two matrices containing the diagonal and off-diagonal terms:

$$A_{H^H H^H} = \begin{bmatrix} 0 & \lambda & \gamma & 0 & 0 & 0 \\ \alpha^H & 0 & 0 & 0 & 0 & 0 \\ f^{H*} & 0 & 0 & \lambda & \lambda_o & 0 \\ 0 & f^{H*} & \alpha^H & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho^H & 0 \end{bmatrix} - \operatorname{diag} \begin{bmatrix} \alpha^H + f^H + \kappa + \psi + \phi \\ \lambda + f^H + \kappa + \psi + \phi \\ \alpha^H + \gamma + \delta + \kappa + \psi + \phi \\ \lambda + \omega + \delta + \psi + \kappa + \phi \\ \lambda_o + \rho^H + \delta + \psi + \kappa + \phi \\ \tau + \psi + \phi \end{bmatrix}, \tag{5.3}$$

$$A_{U^H U^H} = \begin{bmatrix} 0 & \lambda & \gamma & 0 & 0 & 0 \\ \alpha_U^H & 0 & 0 & 0 & 0 & -\tau p_t \\ f^H & 0 & 0 & \lambda & \lambda_o & 0 \\ 0 & f^H & \alpha_U^H & 0 & 0 & -\tau(1-p_t) \\ 0 & 0 & 0 & \omega & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho^H & 0 \end{bmatrix} - \operatorname{diag} \begin{bmatrix} \alpha_U^H + f^H + \kappa_U + \phi_U \\ \lambda + f^H + \kappa_U + \phi_U \\ \alpha_U^H + \gamma + \kappa_U + \phi_U \\ \lambda + \omega + \kappa_U + \phi_U \\ \lambda_o + \rho^H + \kappa_U + \phi_U \\ \tau + \phi_U \end{bmatrix}, \tag{5.4}$$

$$A_{R^H R^H} = \begin{bmatrix} 0 & \lambda & \gamma & 0 & 0 \\ \alpha^H & 0 & 0 & 0 & 0 \\ f^H & 0 & 0 & \lambda & \lambda_o \\ 0 & f^H & \alpha^H & 0 & 0 \\ 0 & 0 & 0 & \omega & 0 \end{bmatrix} - \operatorname{diag} \begin{bmatrix} \alpha^H + f^H + \sigma + \kappa + \psi + \phi \\ \lambda + f^H + \sigma + \kappa + \psi + \phi \\ \alpha^H + \gamma + \kappa + \psi + \phi \\ \lambda + \omega + \kappa + \psi + \phi \\ \lambda_o + \kappa + \psi + \phi \end{bmatrix}, \tag{5.5}$$

$$A_{H^C H^C} = \begin{bmatrix} 0 & \lambda & \gamma & 0 & 0 & 0 \\ \alpha^C & 0 & 0 & 0 & 0 & 0 \\ f^C & 0 & 0 & \lambda & \lambda_o & 0 \\ 0 & f^C & \alpha^C & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho^C & 0 \end{bmatrix} - \operatorname{diag} \begin{bmatrix} \alpha^C + f^C + \nu + \phi \\ \lambda + f^C + \nu + \psi + \phi \\ \alpha^C + \gamma + \delta + \nu + \psi + \phi \\ \lambda + \omega + \delta + \nu + \psi + \phi \\ \lambda_o + \rho^C + \delta + \nu_{CDI} + \psi + \phi \\ \tau^C + \nu_{CDI} + \psi + \phi \end{bmatrix}, \tag{5.6}$$

$$A_{U^C U^C} = \begin{bmatrix} 0 & \lambda & \gamma & 0 & 0 & 0 \\ \alpha_U^C & 0 & 0 & 0 & 0 & \tau p_t \\ f^C & 0 & 0 & \lambda & \lambda_o & 0 \\ 0 & f^C & \alpha_U^C & 0 & 0 & \tau(1-p_t) \\ 0 & 0 & 0 & \omega & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho^C & 0 \end{bmatrix} - \text{diag} \begin{bmatrix} \alpha_U^C + f^C + \nu_U + \phi_U \\ \lambda + f^C + \nu_U + \phi_U \\ \alpha_U^C + \gamma + \nu_U + \phi_U \\ \lambda + \omega + \nu_U + \phi_U \\ \lambda_o + \rho^C + \nu_{CDI} + \phi_U \\ \tau + \nu_{CDI} + \phi_U \end{bmatrix}, \tag{5.7}$$

$$A_{R^C R^C} = \begin{bmatrix} 0 & \lambda & \gamma & 0 & 0 \\ \alpha^C & 0 & 0 & 0 & 0 \\ f^C & 0 & 0 & \lambda & \lambda_o \\ 0 & f^C & \alpha^C & 0 & 0 \\ 0 & 0 & 0 & \omega & 0 \end{bmatrix} - \text{diag} \begin{bmatrix} \alpha^C + f^C + \sigma + \nu + \psi + \phi \\ \lambda + f^C + \sigma + \nu + \psi + \phi \\ \alpha^C + \gamma + \nu + \psi + \phi \\ \lambda + \omega + \nu + \psi + \phi \\ \lambda_o + \nu + \psi + \phi \end{bmatrix}, \tag{5.8}$$

$$A_{II} = \begin{bmatrix} 0 & 0 & \sigma \\ f^C\theta & 0 & f^C\theta \\ 0 & \gamma_{infant} & 0 \end{bmatrix} - \text{diag} \begin{bmatrix} f^C\theta + \zeta + \psi + \phi \\ \gamma_{infant} + \zeta + \psi + \phi \\ f^C\theta + \sigma + \zeta + \psi + \phi \end{bmatrix}, \tag{5.9}$$

where f^H and f^C are the forces of colonisation in the hospital and community – the only non-constant terms in the matrix – with definitions

$$\begin{aligned} f^H = \frac{\beta_{Disrupt}}{N^H}(q^H(H_{aot}^H + U_{aot}^U) + H_{ac}^H + U_{ac}^H + R_{ac}^H + H_{ao}^H + U_{ao}^H + R_{ao}^H) \\ + \frac{\beta_{Intact}}{N^H}(H_c^H + U_c^H + R_c^H) \end{aligned} \tag{5.10}$$

and

$$\begin{aligned} f^C = \frac{\beta_{Disrupt}}{N^C}(q^C(H_{aot}^C + U_{aot}^U) + H_{ac}^C + U_{ac}^C + R_{ac}^C + H_{ao}^C + U_{ao}^C + R_{ao}^C) \\ + \frac{\beta_{Intact}}{N^C}(H_c^C + U_c^C + R_c^C) + \frac{\beta_{Infant}}{N^C}(R_c^I) + r^C, \end{aligned} \tag{5.11}$$

where r^C is the force of colonisation from animal reservoirs in the community. The off-diagonal sub-matrices for hospital admission are

$$A_{H^H H^C} = A_{U^H U^C} = \begin{bmatrix} \nu & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu & 0 & 0 \\ 0 & 0 & 0 & 0 & \nu_{CDI} & 0 \\ 0 & 0 & 0 & 0 & 0 & \nu_{CDI} \end{bmatrix}, \tag{5.12}$$

$$A_{R^H R^C} = \nu I_{5\times 5}, \tag{5.13}$$

the off-diagonal sub-matrices for hospital discharge are

$$A_{H^C H^H} = \begin{bmatrix} \kappa & 0 & 0 & 0 & 0 & 0 \\ 0 & \kappa & 0 & 0 & 0 & 0 \\ 0 & 0 & \kappa & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{5.14}$$

$$A_{U^C U^H} = \begin{bmatrix} \kappa_U & 0 & 0 & 0 & 0 & 0 \\ 0 & \kappa_U & 0 & 0 & 0 & 0 \\ 0 & 0 & \kappa_U & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa_U & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa_U & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{5.15}$$

$$A_{R^C R^H} = \kappa I_{5\times 5}, \tag{5.16}$$

the off-diagonal sub-matrices for gaining of immunity are

$$A_{R^H H^H} = A_{R^C H^C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\tau p_t \\ 0 & 0 & \delta & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta & 0 & -\tau(1-p_t) \\ 0 & 0 & 0 & 0 & \delta & 0 \end{bmatrix}, \tag{5.17}$$

the off-diagonal sub-matrices for waning of immunity are

$$A_{H^H R^H} = A_{H^C R^C} = \begin{bmatrix} \sigma & 0 & 0 & 0 & 0 \\ 0 & \sigma & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{5.18}$$

the off-diagonal sub-matrices for aging of adults are

$$A_{U^H H^H} = A_{U^C H^C} = \psi I_{6\times 6}, \tag{5.19}$$

$$A_{U^H R^H} = A_{U^C R^C} = \begin{bmatrix} \psi & 0 & 0 & 0 & 0 \\ 0 & \psi & 0 & 0 & 0 \\ 0 & 0 & \psi & 0 & 0 \\ 0 & 0 & 0 & \psi & 0 \\ 0 & 0 & 0 & 0 & \psi \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{5.20}$$

and the off-diagonal sub-matrices for aging of infants are

$$A_{R^H I} = \begin{bmatrix} \zeta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{5.21}$$

$$A_{U^H I} = \begin{bmatrix} \psi & 0 & \psi \\ 0 & 0 & 0 \\ 0 & \psi & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{5.22}$$

$$A_{R^H I} = \begin{bmatrix} 0 & 0 & \zeta \\ 0 & 0 & 0 \\ 0 & \zeta & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{5.23}$$

When the individual Markov chain approximation was required for individual level outcomes, an individual was modelled as a CTMC on a state-space of the same 37 compartments. The Kolmogorov forward equation for this Markov chain was

$$\mathbf{p}' = Q\mathbf{p}, \tag{5.24}$$

where Q is A evaluated at the endemic equilibrium point $\mathbf{X}^*$, *i.e.* $Q = A(\mathbf{X}^*)$. Note this only effects the force of colonisation terms f^H and f^C in the main diagonal sub-matrices.

Chapter 6

Seasonality and community interventions in a mathematical model of *Clostridium difficile* transmission

6.1 Introduction

This chapter consists of an article in press at the *Journal of Hospital Infection* and the accompanying supplementary materials. In this paper I use the model of *C. difficile* introduced in Chapter 5 to determine the effect of reducing transmission and antibiotic prescription rates. I consider various groups within and outside the hospital and the effect of reducing antibiotic prescriptions to these groups or reducing transmission from these groups. As in Chapter 5 I use the ODE formulation of the model, assuming the population is at endemic equilibrium, and use the individual Markov chain approximation to emulate the classification of infections as hospital and community-acquired. I find that while reductions of transmission from the historical targets of infection control (hospitalised patients and those with symptoms) have a relatively small effect, small reductions in transmission from people residing in the community could completely eliminate the disease in the absence of reintroduction from outside sources. I extend the model using an ODE formulation with seasonally forced parameters to consider two possible mechanisms that might explain the observed seasonality of *C. difficile* infections: seasonal antibiotic prescriptions and seasonal transmissibility. I argue that the observed degree of seasonal prescription of antibiotics causes a seasonal pattern in infection incidence similar to the observed pattern. I argue that seasonal transmissibility could also lead to seasonal infection incidence, but unlike seasonal antibiotic prescriptions, would result in seasonal variation in community colonisation prevalence of a similar magnitude to infection incidence seasonality. There is currently only weak evidence around community colonisation prevalence seasonality but it suggests that colonisation prevalence is not seasonal, supporting antibiotic seasonality as the main mechanism driving seasonal *C. difficile* infections. I estimate the reductions in CDI incidence that could be achieved by reducing the seasonal excess of antibiotic

prescriptions. The supplementary materials provide a table of model parameters, two figures that summarise the model structure and two figures that display the full range of sensitivity analysis.

6.2 Article and Supplementary Materials

Angus McLure, Luis Furuya-Kanamori, Archie C. A. Clements, Martyn Kirk and Kathryn Glass. Seasonality and community interventions in a mathematical model of *Clostridium difficile* transmission. *Journal of Hospital Infection*, (In Press) 2019.

Journal of Hospital Infection xxx (xxxx) xxx

Available online at www.sciencedirect.com

Journal of Hospital Infection

journal homepage: www.elsevier.com/locate/jhin

Seasonality and community interventions in a mathematical model of *Clostridium difficile* transmission

A. McLure [a,*], L. Furuya-Kanamori [a,b], A.C.A. Clements [c], M. Kirk [a], K. Glass [a]

[a] Research School of Population Health, Australian National University, Canberra, Australian Capital Territory, Australia
[b] Department of Population Medicine, College of Medicine, Qatar University, Doha, Qatar
[c] Faculty of Health Sciences, Curtin University, Perth, Western Australia, Australia

ARTICLE INFO

Article history:
Received 3 December 2018
Accepted 4 March 2019
Available online xxx

Keywords:
Clostridium difficile
Mathematical model
Seasonal infections
Hospital-acquired infections
Community-acquired infections

SUMMARY

Background: *Clostridium difficile* infection (CDI) is the leading cause of antibiotic-associated diarrhoea with peak incidence in late winter or early autumn. Although CDI is commonly associated with hospitals, community transmission is important.

Aim: To explore potential drivers of CDI seasonality and the effect of community-based interventions to reduce transmission.

Methods: A mechanistic compartmental model of *C. difficile* transmission in a hospital and surrounding community was used to determine the effect of reducing transmission or antibiotic prescriptions in these settings. The model was extended to allow for seasonal antibiotic prescriptions and seasonal transmission.

Findings: Modelling antibiotic seasonality reproduced the seasonality of CDI, including approximate magnitude (13.9—15.1% above annual mean) and timing of peaks (0.7—1.0 months after peak antibiotics). Halving seasonal excess prescriptions reduced the incidence of CDI by 6—18%. Seasonal transmission produced larger seasonal peaks in the prevalence of community colonization (14.8—22.1% above mean) than seasonal antibiotic prescriptions (0.2—1.7% above mean). Reducing transmission from symptomatic or hospitalized patients had little effect on community-acquired CDI, but reducing transmission in the community by $\geq$7% or transmission from infants by $\geq$30% eliminated the pathogen. Reducing antibiotic prescription rates led to approximately proportional reductions in infections, but limited reductions in the prevalence of colonization.

Conclusion: Seasonal variation in antibiotic prescription rates can account for the observed magnitude and timing of *C. difficile* seasonality. Even complete prevention of transmission from hospitalized patients or symptomatic patients cannot eliminate the pathogen, but interventions to reduce transmission from community residents or infants could have a large impact on both hospital- and community-acquired infections.

* Corresponding author. Address: Research School of Population Health, Australian National University, 62 Mills Road, 2601, ACT, Australia.
E-mail address: angus.mclure@anu.edu.au (A. McLure).

https://doi.org/10.1016/j.jhin.2019.03.001
0195-6701/

Introduction

Clostridium difficile infection (CDI) is among the most common hospital-associated infections and antibiotic-associated infections in the world [1–7]. Infection is characterized by mild-to-severe diarrhoea, frequent recurrences and considerable mortality [8]. However, asymptomatic colonization is also common, especially in infants [9]. *C. difficile* transmission occurs both within and outside hospitals [10]. Recent whole-genome sequencing studies suggest that up to 25% of hospital-onset infections and up to 19% of all infections can be attributed to contact with another symptomatic patient in hospital [11,12].

Approximately 5% of hospitalized patients are already asymptomatically colonized at the time of admission [9,13]. Modelling studies have highlighted the importance of these colonized admissions, with some demonstrating that within-hospital transmission alone is insufficient to sustain the presence of the pathogen [14–16]. However, much less is known about transmission outside hospitals. Nearly all interventions to reduce transmission have been hospital-based, and the effect of these interventions on community-acquired and community-onset cases has only rarely been studied or modelled [17,18]. Seasonal variation in antibiotic prescriptions [19] and the reduction of fluoroquinolone prescription in the UK [20] have been shown to correlate with CDI; however, the contributions of prescriptions in the community and hospital have not been disentangled. Only one other modelling study has considered the effect of small reductions in prescriptions to transmission outside hospitals [17], but the potential impact of large reductions is unknown.

CDI is moderately seasonal, with incidence peaking in late winter or early autumn [21]. The seasonality of CDI correlates with seasonal antibiotic prescription rates [20], seasonal rainfall and temperature [22], seasonal incidence of influenza [20,23] and seasonal incidence of respiratory syncytial virus (RSV) [20]. However, like other seasonal infections such as RSV and influenza, the mechanisms driving *C. difficile* seasonality are not well understood [24]. Many mechanisms have been proposed for seasonal respiratory infections, including seasonal transmissibility (through seasonal contact rates or pathogen survival) and seasonal host susceptibility (through seasonal changes in immunity) [25]. Similar biological or behavioural factors may contribute to the seasonality of *C. difficile*, but this has not been demonstrated. Prior use of antibiotics is a key risk factor for developing CDI [13,26]. As antibiotic prescriptions rates are seasonal, they are likely to contribute to the observed seasonality of CDI [27].

This study used an existing mathematical model of *C. difficile* transmission in a population that includes a hospital and the surrounding community [28] to explore two inter-related gaps in the *C. difficile* literature. The seasonal patterns in *C. difficile* colonization and infections produced by seasonal antibiotic prescription rates and seasonal transmissibility or susceptibility of the pathogen were explored. In addition, the potential impact on the incidence of CDI and the prevalence of colonization of reducing the transmission of *C. difficile* from various subpopulations or reducing antibiotic prescriptions to various subpopulations within and outside of hospitals were evaluated.

Methods

Description of the model

The existing mathematical model is described in detail elsewhere [28] and summarized diagrammatically in Figures A.1 and A.2 (see Appendix A, online supplementary material). Briefly, the model has a compartmental structure with compartments differentiating patient setting (hospital or community), *C. difficile* status (negative, colonized or infected), gut flora status (disrupted or intact) and immune response to toxins (naïve, immune or immune suppressed/elderly). The model also includes separate compartments to represent infants in the community. All *C. difficile*-positive people in each model setting (hospital or community) can infect those in the same setting. Disruption of gut flora allows the overgrowth of *C. difficile*, and therefore increases infectiousness and leads to the development of symptoms in non-immune patients in the model. Hospitalized patients with symptoms are assumed to be treated with additional contact precautions that reduce their infectiousness. Colonized infants do not develop symptoms but are infectious.

Parameters and sensitivity analyses

The parameters for the base scenario for the intervention analyses and the seasonal average parameters in the seasonality analyses were estimated in a previous paper [28], and are summarized in Table A.1 (see Appendix A, online supplementary material). As the parameters were estimated using Western European infant colonization prevalence data [29,30], nationwide estimates for the incidence of CDI in the USA [31] and CDI hospital admissions data from the USA [32], they reflect the current epidemiology in high-income countries as closely as possible. In the sensitivity analysis (SA), estimated model parameters were refit for different assumptions of the prevalence of colonization and infant infectiousness. Estimates of the prevalence of colonization in the general community are highly variable [9]. The base assumption was that general prevalence of colonization in adults was 5%, but a range of 2–10% was considered in the SA. The relative infectiousness of infants and adults has not been well quantified, although asymptomatically colonized infants may be as infectious as adults with CDI [30]. For the base scenario in this article, it was assumed that infants were 0.5 times as infectious as symptomatic adults (SA 0.1–1 times as infectious), which is equivalent to infants being 3.3 times as infectious as asymptomatically colonized adults with intact gut flora (SA 0.5–9.8 times as infectious). Higher infant infectiousness, especially when the prevalence of colonization in adults was low, led to poor model fit and an implausibly high proportion of infections attributable to infants, and so was excluded.

Modelling the seasonality of C. difficile

Two mechanisms that may account for seasonal CDI rates were modelled: seasonal antibiotic prescription rates and seasonal transmissibility. The transmission rate parameter

combined human and pathogen factors that influence transmissibility and susceptibility, and may be affected by seasonal changes in the environment and human behaviour: spore shedding rate, contact rates, pathogen survival, and human susceptibility to colonization. Therefore, increasing the transmissibility of the pathogen from all carriers and increasing the susceptibility of the whole population were mathematically equivalent in the model. The amplitude of antibiotic prescription seasonality was extracted from seasonal prescription data for different classes of antimicrobials in the USA [27]. In the absence of data on seasonal transmissibility/susceptibility, and to make fair comparisons between the mechanisms, the same amplitude was applied to both seasonal antibiotic prescription and transmission rate parameters. Each mechanism was simulated independently. Under the assumption that current seasonal CDI is entirely due to the estimated amplitude of seasonal antibiotic prescriptions, estimates were made of the reduction in the annual incidence of CDI, peak incidence of CDI and mean prevalence of colonization that would be achieved if antibiotic prescription rates were reduced to their seasonal low levels all year round.

Modelling the reduction of transmission or antibiotic prescriptions

The remainder of the intervention analyses considered the effect of reducing the transmission and antibiotic prescription rate parameters in a non-seasonal version of the model. Five overlapping target groups/populations were considered for reducing transmission: hospitalized patients, community residents (including infants), symptomatic carriers, recipients of antibiotics, and infants. Four overlapping target groups were considered for reducing events that disrupt gut flora: hospitalized patients, community residents, the elderly or immunosuppressed population who are at high risk of developing CDI, and adults who are not in the high-risk group. For each scenario, reductions of 0–100% in the relevant rate parameter(s) were considered, and the reduction in the incidence of CDI or the prevalence of colonization was calculated as a percentage of the base incidence or prevalence of colonization.

The prevalence of colonization in the hospital and community subpopulations was calculated, and the incidence of CDI was separated into hospital-acquired and community-acquired cases in two ways: (1) the actual location of colonization simulated in the model; and (2) the apparent source of acquisition as classified by surveillance definitions similar to those recommended by the Society for Healthcare Epidemiology of America and the Infectious Diseases Society of America [34]. It has been shown previously that the recommended definitions misclassify many community-acquired cases as hospital-acquired [35], so this enabled comparison of the apparent effect on interventions with the actual effect. The model used a minor variant of recommended definitions employed by Lessa *et al.* [31], and accounted for underreporting of community-onset cases, as implemented in the authors' previous paper [28].

Results

Seasonal variation in antibiotic prescription rates and seasonal transmissibility/susceptibility produced different patterns of *C. difficile* seasonality (Figure 1). The estimated amplitude of the variation in antibiotic prescriptions was 16.2%

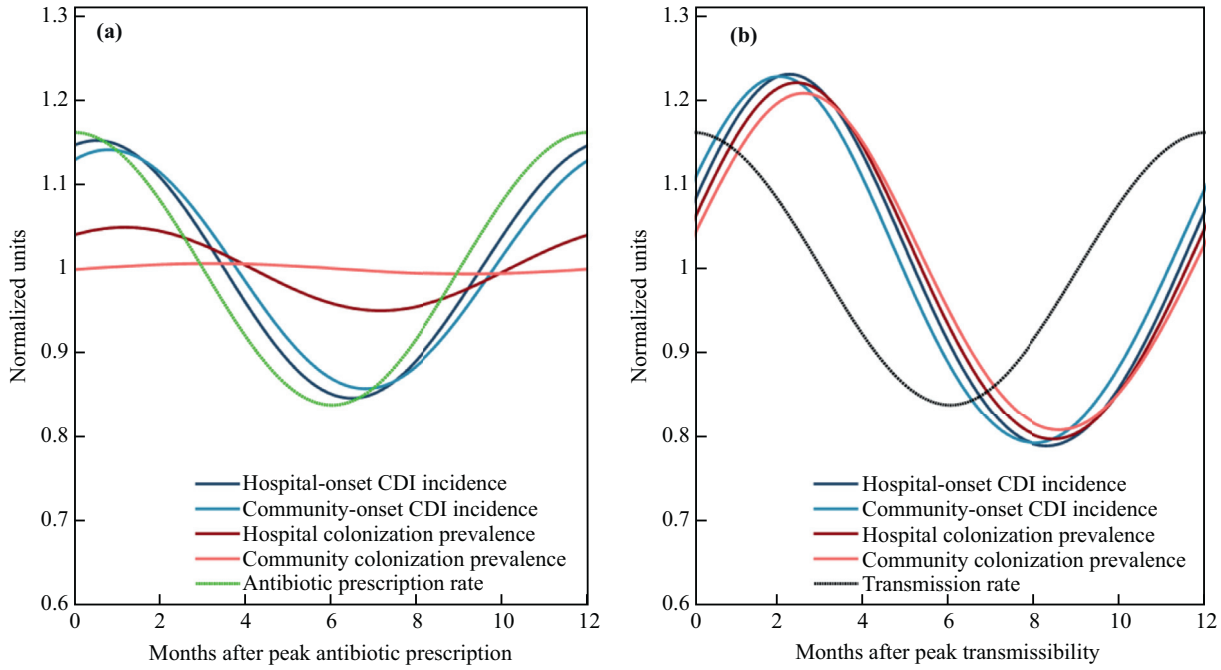


Figure 1. Comparison of seasonal patterns in the incidence of *Clostridium difficile* infection (CDI) and the prevalence of colonization produced by (a) seasonal antibiotic prescription rates and (b) seasonal transmissibility or susceptibility. Incidence, prevalence of colonization and seasonal rate parameters have been rescaled so that the annual mean is 1. The base assumptions for infant infectiousness (half as infectious as adults with CDI or 3.3 times as infectious as asymptomatically colonized adults with intact gut flora) and general prevalence of colonization in adults (5% in the community) were used for this figure.

(i.e. the seasonal high and seasonal low were 116.2% and 83.8% of the annual mean). Both seasonal antibiotics and seasonal transmissibility/susceptibility led to large, annually repeating variation in the total incidence of infection, with peak incidence 14% (SA 13.9–15.1%) and 23% (SA 18.2–25.3%) above annual mean incidence, respectively. The timing and magnitude of infection seasonality were similar for hospital-onset and community-onset infections. The annual peak in total (i.e. hospital and community combined) incidence of infection was 2.0 months (SA 1.4–2.3 months) after peak transmissibility in the seasonal transmissibility model, and 0.8 months (SA 0.7–1.0 months) after peak prescriptions in the seasonal antibiotics model.

Seasonal transmissibility/susceptibility led to large seasonal variations in the prevalence of colonization in both the hospital (peak 22.0% above mean; SA 20.3–25.1%) and the community (peak 20.8% above mean; SA 14.8–22.1%), with timing similar to the seasonality of the incidence of infection (Figure 1B). In contrast, seasonal antibiotic prescriptions led to seasonal variation in the prevalence of colonization in hospitals with similar timing (peak prevalence 1.2 months after peak prescriptions; SA 1.1–1.4 months), but less than half the amplitude (peak prevalence 4.8% above mean; SA 2.2–10.1%) of hospital-onset CDI seasonality. Seasonal antibiotic prescriptions led to very little seasonal variation in the prevalence of colonization in the community (0.6% above mean; SA 0.2–1.7%), which peaked 3.2 months (SA 2.4–3.6 months) after peak prescriptions (Figure 1A). In a sensitivity analysis for the amplitude of antibiotic prescription seasonality, the timing of peaks was independent of amplitude (results not shown).

Reducing antibiotic prescription rates to the seasonal low reduced peak incidence of CDI by 30% (SA 27–41%), annual incidence by 20% (SA 16–32%), and mean prevalence of colonization by 6% (SA 1–21%). These reductions were approximately linear, so halving seasonal excess prescriptions led to approximately half the above reductions in incidence and prevalence.

The effect of reducing transmission in the non-seasonal model is summarized in Figure 2, and further explored in Figures B.1 and B.2 (see Appendix B, online supplementary material). Modest reductions in transmission from community residents (10%, SA 7–27%) eliminated all CDI in the hospital and the community. For the base assumption, a 47.5% reduction in transmission from infants eliminated all CDI in the population. However, this finding was very sensitive to the parameter assumptions. At one extreme — with infants one tenth as symptomatic as adults and the prevalence of colonization in adults at 10% — preventing all transmission from infants only reduced the incidence of CDI by 5%. At the other extreme — with infants as infectious as symptomatic adults and the prevalence of colonization in adults at 2% — a 30% reduction in transmission from infants eliminated *C. difficile* from the population. Reducing hospital-based transmission led to approximately proportional reductions in hospital-acquired

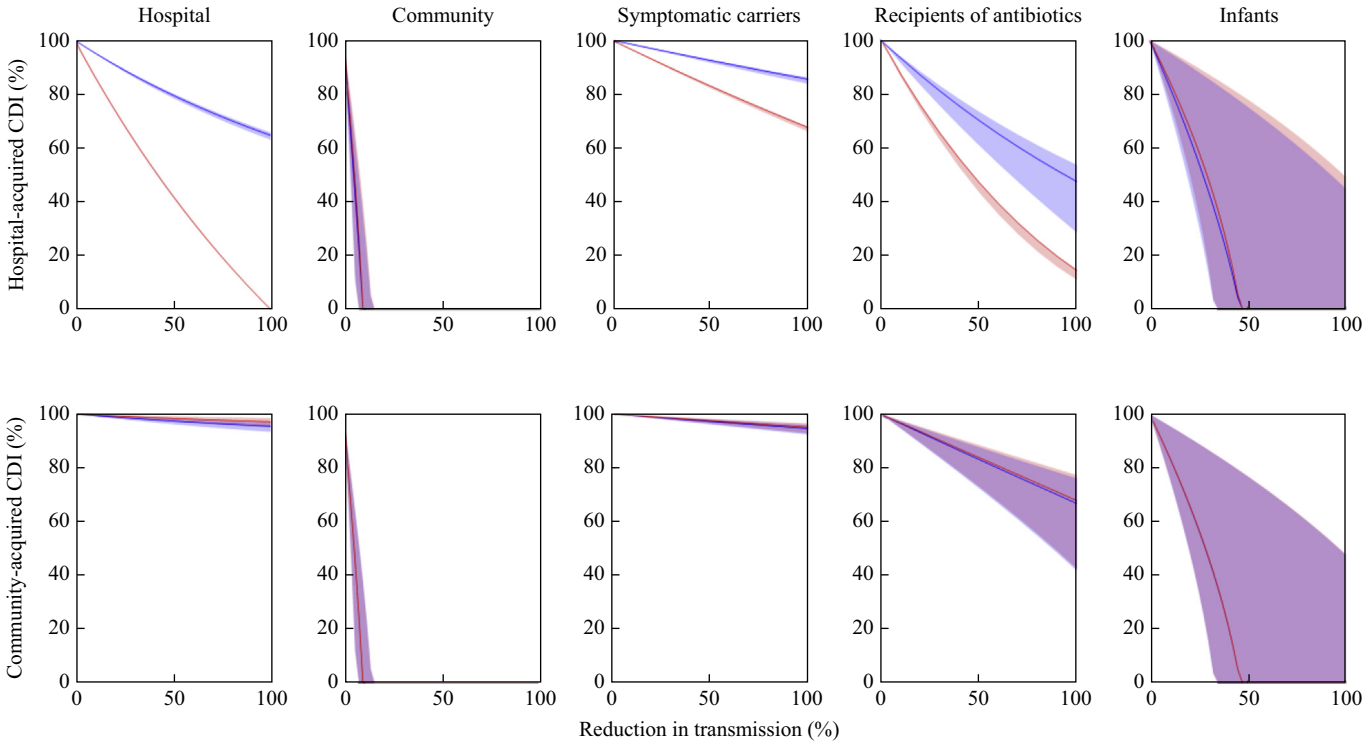


Figure 2. Comparison of the effect in the hospital (top row) and community (bottom row) of reducing *Clostridium difficile* transmission from various overlapping target populations (columns). Each figure compares the actual reductions (red) and the apparent reductions (blue), which differ due to misclassification of cases as hospital- or community-acquired. The shaded region around each line indicates the range in the sensitivity analysis for infant infectiousness. A purple shaded region indicates overlapping sensitivity analysis ranges for apparent and actual reductions. The general prevalence of colonization in adults is 5%. See Figure B.1 (Appendix B, online supplementary material) for further sensitivity analyses.

CDI, but only modest reductions in community-acquired cases. Reducing transmission from symptomatic carriers alone was much less effective than reducing transmission from all those who had taken antibiotics recently; however, neither could eliminate hospital- or community-acquired infections.

In the non-seasonal model, reducing the disruption of gut flora reduced the prevalence of colonization (due to reduced transmission from those colonized without disruption of gut flora), but this reduction was not sufficient to interrupt transmission in the population (Figure 3). Reducing the disruption of gut flora led to reductions in incidence through the combination of reduced transmission (and hence prevalence of colonization) and reduced risk of developing symptoms (Figure 3). Reducing the disruption of gut flora in either setting (hospital or community) led to reduced infections and colonization in both the hospital and the community. Comparing the reduction of prescriptions for the high-risk elderly population and the remaining low-risk population, the former led to approximately twice the reduction in incidence and a comparable reduction in the prevalence of colonization.

When reducing transmission or antibiotic prescriptions, the apparent and actual reductions in incidence were similar for community-acquired cases (Figures 2 and 3). However, when transmission was reduced in hospital residents alone, symptomatic carriers alone or recent recipients of antibiotics alone, the apparent intervention effect was much smaller than the true effect. For instance, even the complete prevention of all hospital-based transmission appeared to prevent <40% of hospital-acquired cases due to the misclassification of community-acquired cases.

Discussion

Modelling the observed seasonal variation in antibiotic prescription rates reproduced the observed seasonal pattern of CDI, including a delay of approximately one month between peak antibiotic prescription rates and peak incidence of CDI (in agreement with correlative time-series analysis studies [20,23]) and the size of the peak in the incidence of CDI [21]. According to the model, if biological and behavioural mechanisms influencing transmissibility or susceptibility are the major factors driving seasonal incidence of CDI, one should also expect to see large seasonal variation in the prevalence of colonization in both the hospital and the community. On the other hand, if seasonal antibiotic prescription rates are the primary or only mechanism, one should expect to see little to no seasonal variation in the prevalence of colonization in the community and only moderate seasonality in hospitals. Consistent with either mechanism, a study in two Australian hospitals found that the prevalence of colonization was seasonal [36]. However, the present authors are not aware of any published studies investigating seasonality of the prevalence of

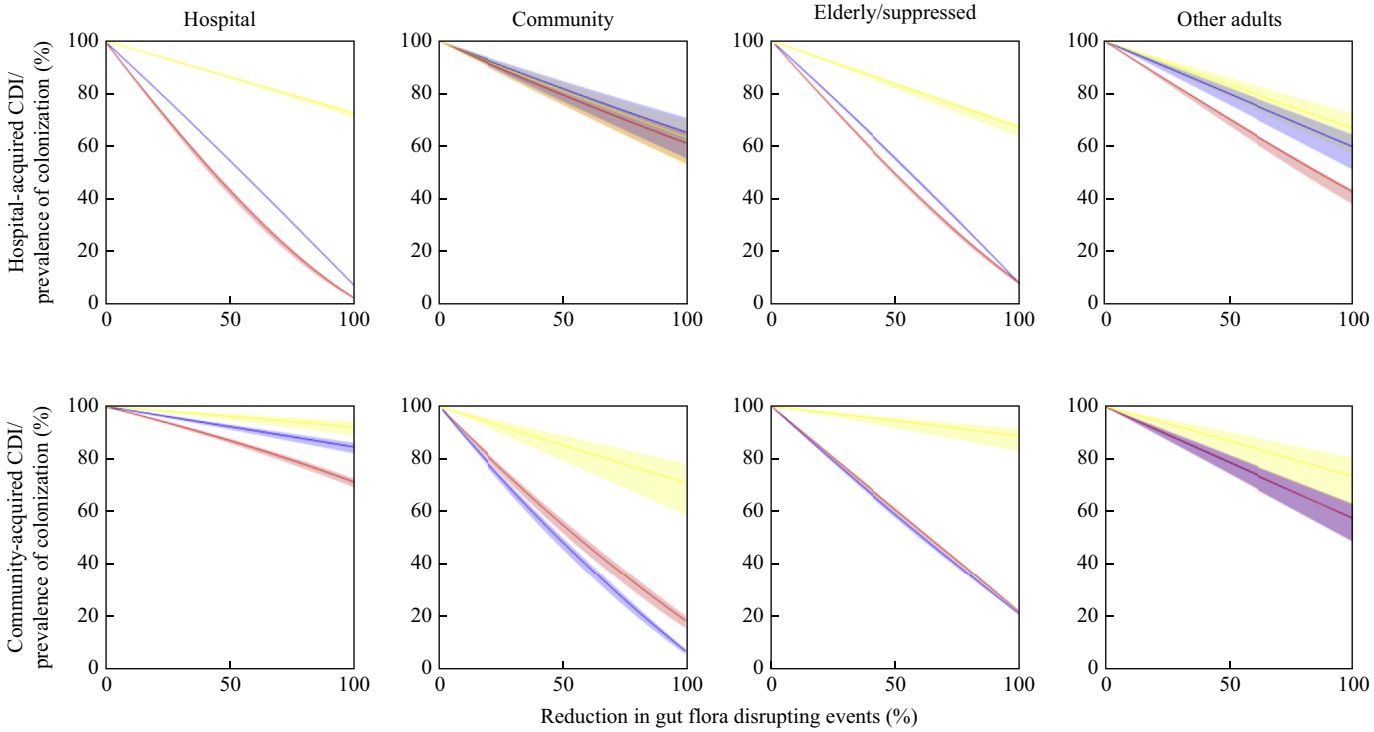


Figure 3. Comparison of the effect in the hospital (top row) and community (bottom row) of a reduction in antibiotic prescriptions to various overlapping target populations (columns). Each row displays reductions in infections acquired in that setting, and the prevalence of all colonization (yellow) in that setting (including colonizations acquired in the other setting). The actual reductions (red) and the apparent reductions (blue) in the incidence of infection differ due to misclassification of cases as hospital- or community-acquired. The shaded region around each line indicates the range in the sensitivity analysis for infant infectiousness. Purple and brown shaded regions indicate overlapping sensitivity analysis ranges. The general prevalence of colonization in adults is 5%. See Figure B.2 (Appendix B, online supplementary material) for further sensitivity analyses.

colonization in the community. Longtin *et al*. screened hospital admissions for *C. difficile* colonization as part of an intervention to reduce transmission from asymptomatically colonized patients [37]. Their unpublished results suggest that the prevalence of colonization on admission is not significantly seasonal (personal communication). Although hospital admissions may not be entirely representative of the community, this observation suggests that seasonal variation in transmissibility and susceptibility is minor, and that *C. difficile* seasonality is largely driven by antibiotics. However, seasonal mechanisms affecting transmissibility or susceptibility cannot be ruled out without further study into seasonality of the prevalence of colonization in the community.

Many antibiotic prescriptions for seasonal respiratory tract infections are clinically inappropriate [38], so some of the excess prescriptions in winter months could be avoided with improved prescribing practices. The model suggests that halving excess seasonal prescriptions would decrease the annual incidence of CDI by 8–16%, which is equivalent to approximately 36,000–72,000 infections per year in the USA alone [31].

In the model, reducing hospital-based transmission alone had a small effect on the incidence of community-acquired cases. Moreover, the systematic misclassification of community-acquired cases as hospital-acquired cases meant that only a fraction of those cases classified as hospital-acquired could be prevented by reducing hospital-based transmission. Consequently, the true proportion of hospital-acquired cases prevented by reducing a given amount of hospital-based transmission was approximately twice the apparent reduction. Importantly, even if the complete prevention of transmission within hospitals could be achieved, misclassification would maintain the appearance of continuing within-hospital transmission.

On the other hand, modest reductions in transmission in the community were found to reduce incidence dramatically, and could even interrupt transmission. This is in agreement with the only other modelling article to address this topic, which found that the incidence of CDI was more responsive to changes in community-based transmission than hospital-based transmission [17]. In practice, targeted interventions may be more achievable than community-wide improvements in hygiene. This study found that reducing transmission from symptomatic patients in hospitals and communities would only lead to small reductions in incidence. On the other hand, it was estimated that halving transmission from recent recipients of antibiotics would reduce the incidence of hospital- and community-acquired CDI by half (SA 24–86%) and one-sixth (SA 4–66%), respectively. Encouraging patients taking antibiotics (who are mostly unaware of the association between antibiotics and CDI [46]) to adopt improved hygiene has the potential to reduce the transmission of *C. difficile* and other pathogens to and from these patients, so this may underestimate the true effect. Remarkably, this study found that reducing transmission from asymptomatically colonized infants by as little as 30% could be sufficient to eliminate infections and colonization from the entire population. However, the required reduction was highly sensitive to the uncertain relative infectiousness of infants and adults. Although a high prevalence of colonization in infants is well established [29,30,40], there has been little research investigating transmission from infants [41–45], and the proportion of transmission attributable to infants is unknown. A greater understanding of the contribution of infants should be a research priority.

The study model suggested that year-round reductions in antibiotic prescriptions in the hospital or in the community would lead to approximately proportional reductions in infections in the same setting. In agreement with a recent meta-analysis, most of the improvement was attributable to the reduction of prescriptions to elderly or immunosuppressed individuals [47]. However, no plausible reduction in antibiotic consumption was enough to interrupt transmission in the hospital or community. Therefore, antimicrobial stewardship should be combined with interventions to reduce transmission in the community.

There is evidence of *C. difficile* transmission between human and livestock populations [48], and that contamination of meat products with *C. difficile* may be seasonal [49]. It has been shown previously that if the proportion of infections attributable to animals is sufficiently high (>3.5–26.0%), preventing transmission from animals could eliminate all CDI in humans [28,50]. However, the proportion of human infections that are attributable to direct or indirect transmission from livestock is unknown. Therefore, the model does not account for this source of transmission or capture its seasonality. This omission may mean that the model overestimated the impact of reducing person-to-person transmission.

Another limitation of this study is that the model does not distinguish between strains of *C. difficile* [51] or differentiate between antibiotics with different risk profiles for CDI [26]. This may influence the analysis of *C. difficile* seasonality and the effect of reducing prescription rates. A 40% reduction in fluoroquinolone prescriptions in the UK coincided with near elimination of CDI caused by fluoroquinolone-resistant strains [19]. This non-linear effect contrasts with the authors' prediction of proportional reduction, but could be due to concurrent improvements in transmission control [52], other strain-specific factors, or strain competition factors that were not captured in the model.

This analysis supports the hypothesis that seasonal prescription of antibiotics is the main driver of seasonal CDI. Further research into seasonality of the prevalence of colonization in the community and the extent of transmission from animals could clarify the role of seasonal transmissibility, susceptibility or exposure to livestock reservoirs. The authors have provided an estimate of the potential gains for *C. difficile* control that could be achieved by reducing inappropriate seasonal antibiotic prescriptions. The model supports the use of antimicrobial stewardship to reduce infections, but highlights the need to explore interventions to reduce transmission from the large population of asymptomatically colonized individuals in the community.

Conflict of interest statement
None declared.

Funding sources
This work was supported by National Health and Medical Research Council to A.C.A.C. (#1058878) and M.K. (GNT1145997), and an Australian Government Research Training Program Scholarship to A.M. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhin.2019.03.001.

References

[1] Magill SS, Edwards JR, Bamberg W, Beldavs ZG, Dumyati G, Kainer MA, et al. Multistate point-prevalence survey of health care-associated infections. N Engl J Med 2014;370:1198–208.
[2] Bartlett JG. Antibiotic-associated diarrhea. N Engl J Med 2002;346:334–9.
[3] Fernandez Canigia L, Nazar J, Arce M, Dadamio J, Smayevsky J, Bianchini H. *Clostridium difficile* diarrhea: frequency of detection in a medical center in Buenos Aires, Argentina. Rev Argent Microbiol 2001;33:101–7.
[4] Herrera P, Cotera A, Fica A, Galdo T, Alvo M. High incidence and complications of *Clostridium difficile* diarrhea among patients with renal diseases. Rev Med Chil 2003;131:397–403.
[5] Dhawan B, Chaudhry R, Sharma N. Incidence of *Clostridium difficile* infection: a prospective study in an Indian hospital. J Hosp Infect 1999;43:275–80.
[6] Sadeghifard N, Salari M, Ghassemi M, Shirazi M, Feizabadi M, Kazemi B, et al. Prevalence of *Clostridium difficile*-associated diarrhea in hospitalized patients with nosocomial diarrhea. Iran J Public Health 2005;34:67–72.
[7] Garcia C, Samalvides F, Vidal M, Gotuzzo E, Dupont HL. Epidemiology of *Clostridium difficile*-associated diarrhea in a Peruvian tertiary care hospital. Am J Trop Med Hyg 2007;77:802–5.
[8] Leffler DA, Lamont JT. Treatment of *Clostridium difficile*-associated disease. Gastroenterology 2009;136:1899–912.
[9] Furuya-Kanamori L, Marquess J, Yakob L, Riley TV, Paterson DL, Foster NF, et al. Asymptomatic *Clostridium difficile* colonization: epidemiology and clinical implications. BMC Infect Dis 2015;15:516.
[10] Gerding DN, Lessa FC. The epidemiology of *Clostridium difficile* infection inside and outside health care institutions. Infect Dis Clin N Am 2015;29:37–50.
[11] Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med 2013;369:1195–205.
[12] Walker AS, Eyre DW, Wyllie DH, Dingle KE, Harding RM, O'Connor L, et al. Characterisation of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. PLoS Med 2012;9:e1001172.
[13] Loo VG, Bourgault A-M, Poirier L, Lamothe F, Michaud S, Turgeon N, et al. Host and pathogen factors for *Clostridium difficile* infection and colonization. N Engl J Med 2011;365:1693–703.
[14] Lanzas C, Dubberke ER, Lu Z, Reske KA, Gröhn YT. Epidemiological model for *Clostridium difficile* transmission in healthcare settings. Infect Control Hosp Epidemiol 2011;32:553–61.
[15] Yakob L, Riley TV, Paterson DL, Clements ACA. *Clostridium difficile* exposure as an insidious source of infection in healthcare settings: an epidemiological model. BMC Infect Dis 2013;13:376.
[16] McLure A, Clements ACA, Kirk M, Glass K. Healthcare-associated *Clostridium difficile* infections are sustained by disease from the community. Bull Math Biol 2017;79:2242–57.
[17] Durham DP, Olsen MA, Dubberke ER, Galvani AP, Townsend JP. Quantifying transmission of *Clostridium difficile* within and outside healthcare settings. Emerg Infect Dis 2016;22:608–16.
[18] Price J, Cheek E, Lippett S, Cubbon M, Gerding DN, Sambol SP, et al. Impact of an intervention to control *Clostridium difficile* infection on hospital- and community-onset disease; an interrupted time series analysis. Clin Microbiol Infect 2010;16:1297–302.
[19] Dingle KE, Didelot X, Quan TP, Eyre DW, Stoesser N, Golubchik T, et al. Effects of control interventions on *Clostridium difficile* infection in England: an observational study. Lancet Infect Dis 2017;17:411–21.
[20] Polgreen PM, Yang M, Bohnett LC, Cavanaugh JE. A time-series analysis of *Clostridium difficile* and its seasonal association with influenza. Infect Control Hosp Epidemiol 2010;31:382–7.
[21] Furuya-Kanamori L, McKenzie SJ, Yakob L, Clark J, Paterson DL, Riley TV, et al. *Clostridium difficile* infection seasonality: patterns across hemispheres and continents – a systematic review. PLoS One 2015;10:e0120730.
[22] Furuya-Kanamori L, Robson J, Soares Magalhães RJ, Yakob L, McKenzie SJ, Paterson DL, et al. A population-based spatio-temporal analysis of *Clostridium difficile* infection in Queensland, Australia over a 10-year period. J Infect 2014;69:447–55.
[23] Gilca R, Fortin E, Frenette C, Longtin Y, Gourdeau M. Seasonal variations in *Clostridium difficile* infections are associated with influenza and respiratory syncytial virus activity independently of antibiotic prescriptions: a time series analysis in Quebec, Canada. Antimicrob Agents Chemother 2012;56:639–46.
[24] Lofgren E, Fefferman NH, Naumov YN, Gorski J, Naumova EN. Influenza seasonality: underlying causes and modeling theories. J Virol 2007;81:5429–36.
[25] Tamerius J, Nelson MI, Zhou SZ, Viboud C, Miller MA, Alonso WJ. Global influenza seasonality: reconciling patterns across temperate and tropical regions. Environ Health Perspect 2011;119:439–45.
[26] Furuya-Kanamori L, Stone JC, Clark J, McKenzie SJ, Yakob L, Paterson DL, et al. Comorbidities, exposure to medications, and the risk of community-acquired *Clostridium difficile* infection: a systematic review and meta-analysis. Infect Control Hosp Epidemiol 2015;36:132–41.
[27] Suda KJ, Hicks LA, Roberts RM, Hunkler RJ, Taylor TH. Trends and seasonal variation in outpatient antibiotic prescription rates in the United States, 2006 to 2010. Antimicrob Agents Chemother 2014;58:2763–6.
[28] McLure A, Clements ACA, Kirk M, Glass K. Modelling diverse sources of *Clostridium difficile* in the community: importance of animals, infants, and asymptomatic carriers. Epidemiol Infect 2019;147:e152.
[29] Rousseau C, Poilane I, De Pontual L, Maherault A-C, Le Monnier A, Collignon A. *Clostridium difficile* carriage in healthy infants in the community: a potential reservoir for pathogenic strains. Clin Infect Dis 2012;55:1209–15.
[30] Kubota H, Makino H, Gawad A, Kushiro A, Ishikawa E, Sakai T, et al. Longitudinal investigation of carriage rates, counts, and genotypes of toxigenic *Clostridium difficile* in early infancy. Appl Environ Microbiol 2016;82:5806–14.
[31] Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, et al. Burden of *Clostridium difficile* infection in the United States. N Engl J Med 2015;372:825–34.
[32] Healthcare cost and utilization project (HCUP). HCUP national inpatient sample (NIS). Rockville, MD: Agency for Healthcare Research and Quality; 2014.
[34] Cohen SH, Gerding DN, Johnson S, Kelly CP, Loo VG, McDonald LC, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). Infect Control Hosp Epidemiol 2010;31:431–55.
[35] McLure A, Clements ACA, Kirk M, Glass K. *Clostridium difficile* classification overestimates hospital acquired infections. J Hosp Infect 2018;99:453–60.
[36] Furuya-Kanamori L, Clements ACA, Foster NF, Huber CA, Hong S, Harris-Brown T, et al. Asymptomatic *Clostridium difficile* colonization in two Australian tertiary hospitals, 2012–2014:

prospective, repeated cross-sectional study. Clin Microbiol Infect 2017;23:48:e1—7.
[37] Longtin Y, Paquet-Bolduc B, Gilca R, Garenc C, Fortin E, Longtin J, et al. Effect of detecting and isolating *Clostridium difficile* carriers at hospital admission on the incidence of *C difficile* infections. JAMA Intern Med 2016;176:796.
[38] Fleming-Dutra KE, Hersh AL, Shapiro DJ, Bartoces M, Enns EA, File TM, et al. Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010—2011. JAMA 2016;315:1864—73.
[40] Holst E, Helin I, Mårdh PA. Recovery of *Clostridium difficile* from children. Scand J Infect Dis 1981;13:41—5.
[41] Chitnis AS, Holzbauer SM, Belflower RM, Winston LG, Bamberg WM, Lyons C, et al. Epidemiology of community-associated *Clostridium difficile* infection, 2009 through 2011. JAMA Intern Med 2013;173:1359—67.
[42] Wilcox MH, Mooney L, Bendall R, Settle CD, Fawley WN. A case—control study of community-associated *Clostridium difficile* infection. J Antimicrob Chemother 2008;62:388—96.
[43] Jamal W, Pauline E, Rotimi V. A prospective study of community-associated *Clostridium difficile* infection in Kuwait: epidemiology and ribotypes. Anaerobe 2015;35:28—32.
[44] Stoesser N, Eyre DW, Quan TP, Godwin H, Pill G, Mbuvi E, et al. Epidemiology of *Clostridium difficile* in infants in Oxfordshire, UK: risk factors for colonization and carriage, and genetic overlap with regional *C. difficile* infection strains. PLoS One 2017;12:e0182307.
[45] Na'amnih W, Adler A, Miller-Roll T, Cohen D, Carmeli Y. Incidence and risk factors for community and hospital acquisition of *Clostridium difficile* infection in the Tel Aviv Sourasky Medical Center. Infect Control Hosp Epidemiol 2017;38:912—20.
[46] Collins JPW, Roberts DMC, Evans CFM, De'Ath HDI, Galland RB. Public perception of *Clostridium difficile*. J Hosp Infect 2009;72:80—1.
[47] Feazel LM, Malhotra A, Perencevich EN, Kaboli P, Diekema DJ, Schweizer ML. Effect of antibiotic stewardship programmes on *Clostridium difficile* incidence: a systematic review and meta-analysis. J Antimicrob Chemother 2014;69:1748—54.
[48] Knetsch CW, Kumar N, Forster SC, Connor TR, Browne HP, Harmanus C, et al. Zoonotic transfer of *Clostridium difficile* harboring antimicrobial resistance between farm animals and humans. J Clin Microbiol 2018;56.
[49] Rodriguez-Palacios A, Reid-Smith RJ, Staempfli HR, Daignault D, Janecko N, Avery BP, et al. Possible seasonality of *Clostridium difficile* in retail meat, Canada. Emerg Infect Dis 2009;15:802—5.
[50] McLure A, Glass K. Some simple rules for estimating reproduction numbers in the presence of reservoir exposure or imported cases. 2018. arXiv:1809.00809.
[51] Knight DR, Elliott B, Chang BJ, Perkins TT, Riley TV. Diversity and evolution in the genome of *Clostridium difficile*. Clin Microbiol Rev 2015;28:721—41.
[52] van Kleef E, Kuijper EJ, Bonten MJM, Cooper BS. *Clostridium difficile* in England: can we stop washing our hands? Lancet Infect Dis 2017;17:478.

Seasonality and community interventions in a mathematical model of *Clostridium difficile* transmission

Appendix A: Model Description

Angus McLure, Luis Furuya-Kanamori, Archie C. A. Clements, Martyn Kirk and Kathryn Glass

This supplementary file contains an additional table (Table A.1) and two additional figures (A.1 and A.2) that summarise the model structure, parameter definitions and parameter values used in the main article. These have been adapted from a previous article which first presented the baseline, non-seasonal version of the model [1]. It also contains some additional details of how seasonality was modelled.

Details of seasonality

We modelled two mechanisms that may account for seasonal CDI rates – seasonal antibiotic prescription rates and seasonal transmissibility – by making the rate parameter associated with each mechanism oscillate sinusoidally with an annually repeating pattern of the form

$$Parameter(d) = AnnualMeanParameter \times \left(1 + Amplitude \times \cos\frac{2\pi(d-p)}{365}\right),$$

where d is the day of the year and p is the day of peak parameter value (counting from January 1st). Though the transmission rate parameters in our model depend on the age, disease status and gut flora status of the carrier, we assumed the transmission rate parameter peaked at the same time of year with the same amplitude for all groups. Similarly, though the prescription rates in our model depend on the age and location (hospital vs. community) of the recipient, we assumed prescriptions peaked at the same time of year with the same amplitude for all groups. The seasonal patterns that would be expected from each mechanism was determined by numerically solving the ordinary differential equation form of the model forward in time until the annual seasonal patterns converged to stable annual patterns.

The amplitude of antibiotic prescription seasonality was extracted from seasonal prescription data for different classes of antimicrobials in the USA [27], using the open source tool WebPlotDigitizer [33]. Combining all classes of antibiotics, we determined the amplitude of the best-fit sinusoidal curve to the prescription data, using simple least squares.

Table A.1 Definitions, values, and references for all parameters used in the model. All rates are in units of day^{-1}. *These parameters were fit to the model in our previous article [1], with the range indicating values over sensitivity analysis. †Only these transmission rate parameters were affected by assumptions around infant infectiousness, being estimated under the assumption that $\beta_{Infant} = k \times \beta_{Disrupt}$ *for* k in the range 0.1-1.0. Note that in our previous article we considered a slightly broader range for k (0.0-1.0), so the ranges of these parameters are narrower in this article. ‡ These parameters varied seasonally in the seasonal extensions of the model. The listed values are the annual mean (and the range is the range of the annual mean in the sensitivity analysis). The seasonal high and low values were 16.2% higher or lower than the annual mean.

Parameter	Description	Value Hospital/ Admission	Value Community/ Discharge	References
α	Rate at which patients begin new antimicrobial treatment (proxy for rate of gut flora disruption)	0.33‡	1.5×10^{-3}‡	[2,3]
α_U	Rate at which patients begin new antimicrobial treatment (proxy for rate of gut flora disruption) for elderly/suppressed	0.33‡	2.5×10^{-3}‡	[2,3]
λ	Rate at which damaged commensal gut flora recovers	0.03		[4,5]
λ_o	Rate at which damaged commensal gut flora recovers for those with CDI overgrowth	0.25		[6]
γ	Rate at which *C. difficile* is cleared in those with intact gut flora	0.023		[7]
ω	Rate of *C. difficile* overgrowth in patients with disrupted gut flora	0.2		[8,9]
ρ	Rate at which treatment for CDI is given (hospital) or sought (community)	1	0.22	[10]
τ	Rate at which the treatment of patients with CDI progresses towards resolution	0.1		[11,12]
p_t	Probability that CDI treatment will remove all *C. difficile*	0.55		[13]
q	Efficacy and coverage of special contact precautions (0: total prevention of transmission; 1: no reduction in transmission)	0.45	1	[14,15]
δ	Rate of development of immunity in adults	0.1		[16,17]
ζ	Rate at which infants age to develop adult-like gut-flora	N/A	1/365	[18]

ψ	Rate at which people age to suppressed/elderly class	2.92×10^{-5} (Median time: 65 years)		
φ	Death rate	1.11×10^{-5}		[19]
φ_U	Death rate for elderly/suppressed	1.84×10^{-4}		[19]
N	Overnight Hospital Beds / Community Size	150	100,000	
σ	Rate at which immunity wanes	2.9×10^{-4} (0.9×10^{-4} – 6.9×10^{-4})		*
ϑ	Multiplicative factor for colonisation susceptibility of infants	N/A	1.4 (0.6-4.4)	*
γ_{infant}	Rate at which *C. difficile* is cleared in infants	0.0020 (0.0020-0.0025)		*
$p_{disturb}$	Proportion of antibiotics that disturb gut flora	0.22 (0.12-0.48)		*
p_{report}	Proportion of all community-treated CDIs that are reported	0.63 (0.57-0.76)		*
$\beta_{Disrupt}$	Transmission rate coefficient for colonised adults with disrupted gut flora (due to recent antibiotic exposure)	0.128 (0.071-0.164) ‡		*†
β_{Intact}	Transmission rate coefficient for colonised adults with intact gut flora (no recent antibiotic exposure)	0.019 (0.001-0.026) ‡		*†
β_{infant}	Transmission rate coefficient from infants	N/A	0.064 (0.007-0.164) ‡	*†
ν / κ	Hospital admission/discharge rate	2.40×10^{-4}	0.241	[20]
ν_U / κ_U	Hospital admission/discharge rate for elderly	8.06×10^{-4}	0.192	[20]
ν_{CDI} / κ_t	Hospital admission rate for CDI / discharge rate for those receiving CDI treatment	0.0142 (0.0128-0.0174)	0 (Not discharged)	*

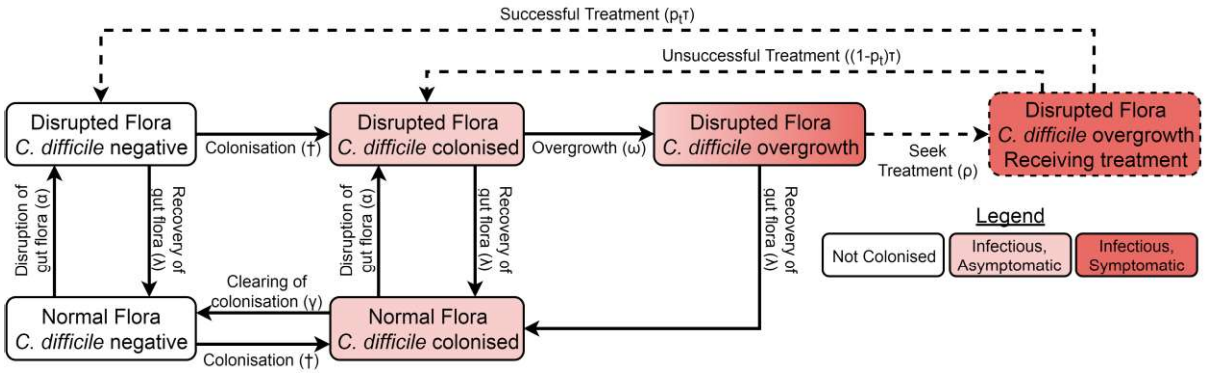


Figure A.1 The structure of the base model (introduced in [1]) showing including colonisation, gut flora status, symptoms, and treatment. The definitions and values of the

parameters associated with each transition can be found in **Table S1**. The details for infants, immunity, demographics, and hospital/community structure are summarised in **Fig S2**. Adults in the immune classes do not have symptoms and therefore not all individuals with overgrowth seek or receive treatment or contact precautions (dashed arrows and box). †The force of colonisation for an individual in one setting (hospital or community) depends on the proportions of the population in the same setting that are in each of the infectious classes, weighted by transmission rate parameters that depend on the infectious class (infants, adults with contact precautions, adults with intact gut flora, adults with disturbed gut flora).

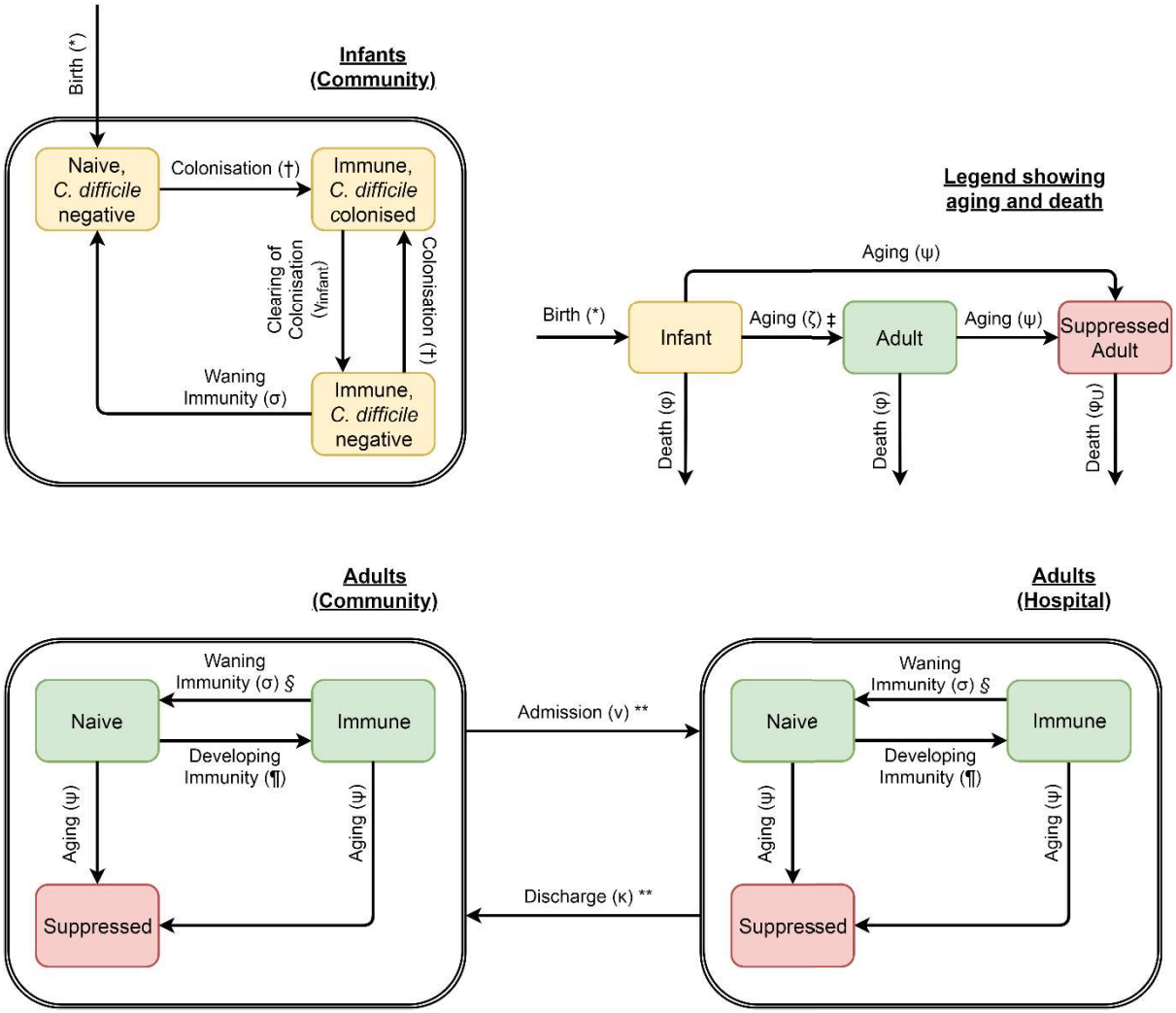


Figure A.2 Model structure, showing immune states, aging, births, deaths, hospital admission and discharge, and infant classes. The details of adult colonisation, gut flora status, symptoms, and treatment are summarised in **Figure A.1**. The definitions and values of the parameters associated with each transition can be found in **Table A.1**. *The birth rate matches the total death rate across the whole population, which were chosen to reproduce the proportion of deaths in persons younger/older than 65 years [19]. †The force of colonisation for an individual in one setting (hospital or community) depends on the proportions of the population in the same setting that are in each of the infectious classes, weighted by transmission rate parameters that depend on the infectious class (infants, adults with contact precautions, adults with intact gut flora, adults with disturbed gut flora). Each setting is assumed to be well mixed and homogenous. Additionally, infants are θ times more susceptible to colonisation (are subject to a higher force of colonisation) than adults.

‡Infants retain their immunity and colonisation status if they age to become non-suppressed adults. §Only non-colonised individuals can have their immunity wane. ¶Naïve individuals with CDI develop immunity upon recovery. Asymptomatically colonised, naïve individuals develop immunity at rate δ. Non-colonised individuals, naïve individuals, and all suppressed individuals do not develop immunity. **Admission discharge rates vary by immunity and CDI status. Patients receiving treatment for CDI are not discharged and are admitted at a much higher rate.

References for Parameter Values

1. McLure A, Clements ACA, Kirk M, Glass K. Modelling diverse sources of Clostridium difficile in the community: importance of animals, infants and asymptomatic carriers. 2018; Available: http://arxiv.org/abs/1809.00759

2. MacDougall C, Polk RE. Variability in rates of use of antibacterials among 130 US hospitals and risk-adjustment models for interhospital comparison. Infect Control Hosp Epidemiol. 2008;29: 203–11. doi:10.1086/528810

3. Australian Commonwealth Department of Health - Drug utilisation sub-committee (DUSC). Antibiotics : PBS / RPBS utilisation. Canberra; 2015.

4. Rafii F, Sutherland JB, Cerniglia CE. Effects of treatment with antimicrobial agents on the human colonic microflora. Ther Clin Risk Manag. 2008;4: 1343–58.

5. Dial S, Kezouh A, Dascal A, Barkun A, Suissa S. Patterns of antibiotic use and risk of hospital admission because of Clostridium difficile infection. CMAJ. 2008;179: 767–72. doi:10.1503/cmaj.071812

6. Bartlett JG. Treatment of Antibiotic-Associated Pseudomembranous Colitis. Clin Infect Dis. 1984;6: S235--S241. doi:10.1093/clinids/6.Supplement_1.S235

7. Johnson S, Homann SR, Bettin KM, Quick JN, Clabots CR, Peterson LR, et al. Treatment of asymptomatic Clostridium difficile carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. Ann Intern Med. 1992;117: 297–302. doi:10.7326/0003-4819-117-4-297

8. Freeman J. Effects of cefotaxime and desacetylcefotaxime upon Clostridium difficile proliferation and toxin production in a triple-stage chemostat model of the human gut. J Antimicrob Chemother. 2003;52: 96–102. doi:10.1093/jac/dkg267

9. Freeman J. Comparison of the efficacy of ramoplanin and vancomycin in both in vitro and in vivo models of clindamycin-induced Clostridium difficile infection. J Antimicrob Chemother. 2005;56: 717–725. doi:10.1093/jac/dki321

10. Van Cauteren D, De Valk H, Vaux S, Le Strat Y, Vaillant V. Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. Epidemiol Infect. 2012;140: 697–705. doi:10.1017/S0950268811000999

11. Leffler DA, Lamont JT. Treatment of Clostridium difficile-associated disease. Gastroenterology. AGA Institute American Gastroenterological Association; 2009;136: 1899–912. doi:10.1053/j.gastro.2008.12.070

12. Khan FY, Abu-Khattab M, Anand D, Baager K, Alaini A, Siddique MA, et al. Epidemiological features of Clostridium difficile infection among inpatients at Hamad General Hospital in the state of Qatar, 2006-2009. Travel Med Infect Dis. Elsevier Ltd; 2012;10: 179–85. doi:10.1016/j.tmaid.2012.06.004

13. McLure A, Clements ACA, Kirk M, Glass K. Healthcare-Associated Clostridium difficile Infections are Sustained by Disease from the Community. Bull Math Biol. 2017;79: 2242–2257. doi:10.1007/s11538-017-0328-8

14. Jarvis WR, Schlosser J, Jarvis AA, Chinn RY. National point prevalence of Clostridium difficile in US health care facility inpatients, 2008. Am J Infect Control. Elsevier Ltd; 2009;37: 263–270. doi:10.1016/j.ajic.2009.01.001

15. Muto CA, Pokrywka M, Shutt K, Mendelsohn AB, Nouri K, Posey K, et al. A Large Outbreak of Clostridium difficile–Associated Disease with an Unexpected Proportion of Deaths and

Colectomies at a Teaching Hospital Following Increased Fluoroquinolone Use. Infect Control Hosp Epidemiol. 2005;26: 273–280. doi:10.1086/502539

16. Foglia G, Shah S, Luxemburger C, Pietrobon PJF. Clostridium difficile: Development of a novel candidate vaccine. Vaccine. Elsevier Ltd; 2012;30: 4307–4309. doi:10.1016/j.vaccine.2012.01.056

17. de Bruyn G, Saleh J, Workman D, Pollak R, Elinoff V, Fraser NJ, et al. Defining the optimal formulation and schedule of a candidate toxoid vaccine against Clostridium difficile infection: A randomized Phase 2 clinical trial. Vaccine. Elsevier Ltd; 2016;34: 2170–2178. doi:10.1016/j.vaccine.2016.03.028

18. Jangi S, Lamont JT. Asymptomatic Colonization by Clostridium difficile in Infants: Implications for Disease in Later Life. J Pediatr Gastroenterol Nutr. 2010;51: 2–7. doi:10.1097/MPG.0b013e3181d29767

19. Centers for Disease Control and Prevention - National Center for Health Statistics. CDC Wonder: Underlying Cause of Death, Multiple Cause of Death Files, 1999-2016 [Internet]. [cited 18 Dec 2017]. Available: http://wonder.cdc.gov/ucd-icd10.html

20. Weiss AJ, Elixhauser A. Overview of Hospital Stays in the United States, 2012: Statistical Brief #180. Healthc Cost Util Proj Stat Briefs. 2006; 1–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/25506966

Seasonality and community interventions in a mathematical model of *Clostridium difficile* transmission

Appendix B: Additional Figures

Angus McLure, Luis Furuya-Kanamori, Archie C. A. Clements, Martyn Kirk and Kathryn Glass

This supplementary file contains Figures B.1 and B.2 that complement Figures 2 and 3 from the main text depicting the results under the assumption of higher or lower colonisation prevalence in the adult population in the community.

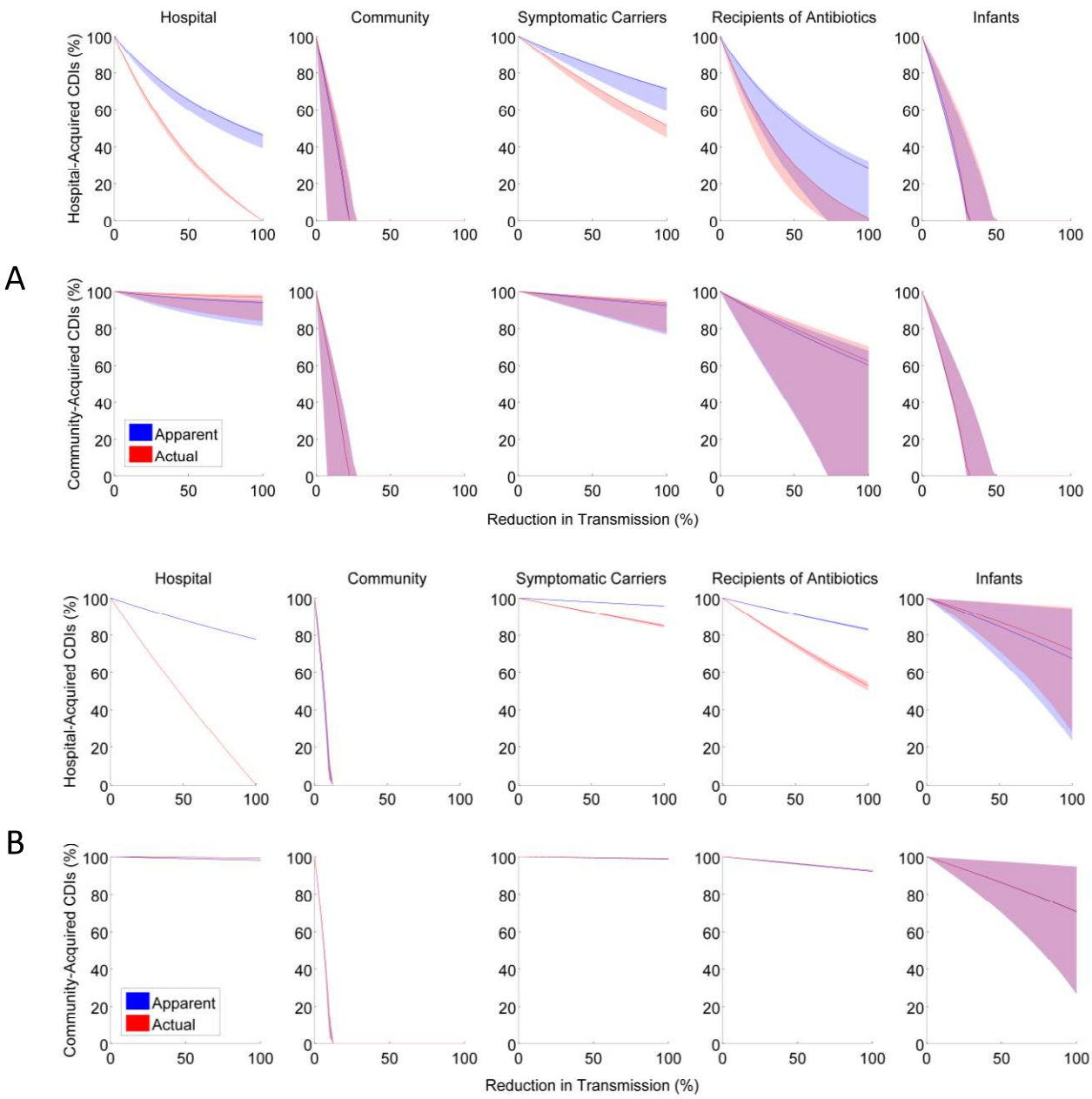


Figure B.1 Further sensitivity analysis for Figure 2 in the main text with adult colonisation prevalence set to 2% (A) and 10% (B). Comparison of the effect in the hospital and community (rows) of reducing C. difficile transmission from various overlapping target populations (columns). Each figure compares the actual reductions (red) and the apparent reductions (blue), which differ due to misclassification of cases as hospital or community-acquired. The shaded regions around each line indicates the range in the sensitivity analysis for infant infectiousness. Purple shaded region indicates overlapping sensitivity-analysis ranges for apparent and actual reductions.

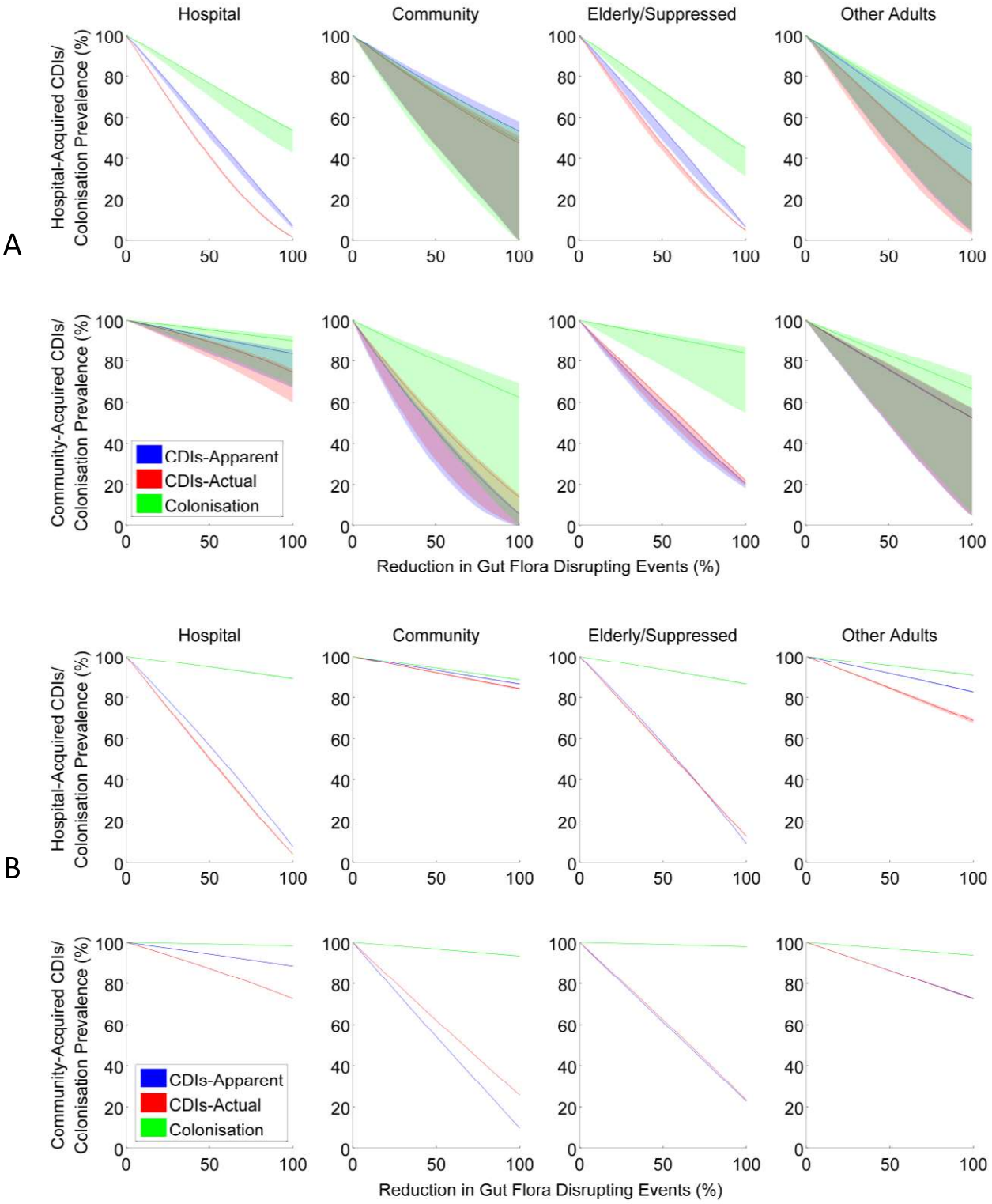


Figure B.2 Further sensitivity analysis for Figure 3 in the main text with adult colonisation prevalence set to 2% (A) and 10% (B). Comparison of the effect in the hospital and community (rows) of the reduction of antibiotic prescription to various overlapping target populations (columns). Each row displays reductions in infections acquired in that setting and prevalence of all colonisation in that setting (including colonisations acquired in the other setting). The actual reductions (red) and the apparent reductions (blue) in infection incidence differ due to misclassification of cases as hospital or community-acquired. The shaded regions around each line indicates the range in the sensitivity analysis for infant infectiousness. Purple and brown shaded regions indicate overlapping sensitivity-analysis ranges.

6.3 Model Equations

This section provides some more explicit details on the equations used in this chapter. The structure of the non-seasonal models used to assess the impact of possible reductions in transmission and antibiotic prescriptions in this chapter are identical to those presented in Section 5.3 – only the parameter values were varied. The two seasonal models for seasonal antibiotic prescriptions and seasonal transmission adapt the time-invariant ordinary differential equations to produce systems with direct time-dependence, *i.e.* a system of the form

$$\mathbf{X}(t)' = A^*(\mathbf{X}(t), t)\mathbf{X}(t) + \mathbf{b}, \tag{6.1}$$

where, as before, $\mathbf{X}$ is the vector of number of persons in each compartment and $\mathbf{b}$ is the birth term. The matrices A^* in the two seasonal models differ from the non-seasonal A matrix in Section 5.3 only in the seasonal parameters α^C, α^C_U, α^H and α^H_U in the seasonal antibiotic prescription model and $\beta_{Intact}, \beta_{Disrupt}$ and β_{Infant} in the seasonal transmission model. The seasonal parameters are sinusoidal functions of time with a period of one year as described in the methods section of the above journal article.

Chapter 7

Some simple rules for estimating reproduction numbers in the presence of reservoir exposure or imported cases

7.1 Introduction

This chapter consists of an article currently under review at *Theoretical Population Biology* that develops the notion of an animal-driven threshold introduced in Chapter 5. The article considers how a general source of new infected cases – either exposure to a reservoir or importation via travel or migration – can drive transmission in a population even if the reproduction number in that population is less than one. Considering a wide range of generic ordinary differential equation models, the article provides simple rules to determine whether a disease could be sustained by local, person-to-person transmission alone or if reservoir-exposure or importation drives transmission. The article applies these rules to two case studies in *C. difficile* to come up with estimates of the local person-to-person reproduction number or the reservoir-driven threshold. The first case study estimates the reproduction number for within-hospital transmission of *C. difficile*, corroborating the findings of Chapters 3 and 5. The second case study calculates the plausible range for the animal-driven threshold using only *C. difficile* colonisation prevalence estimates, corroborating Chapter 5.

7.2 Article

Angus McLure and Kathryn Glass. Simple rules for estimating reproduction numbers in the presence of reservoir exposure or imported cases. Submitted to *Theoretical Population Biology* (under review), 2018.

Some simple rules for estimating reproduction numbers in the presence of reservoir exposure or imported cases

Angus McLure[1]*, Kathryn Glass[1]

[1] Research School of Population Health,
Australian National University,
62 Mills Rd, Acton,
0200, ACT, Australia
* Corresponding author: angus.mclure@anu.edu.au

Abstract

The basic reproduction number (R_0) is a threshold parameter for disease extinction or survival in isolated populations. However no human population is fully isolated from other human or animal populations. We use compartmental models to derive simple rules for the basic reproduction number in populations where an endemic disease is sustained by a combination of local person-to-person transmission *and* exposure from some other source: either a reservoir exposure or imported cases. We introduce the idea of a *reservoir-driven* or *importation-driven* disease: diseases that would become extinct in the population of interest without reservoir exposure or imported cases (since $R_0 < 1$), but nevertheless may be sufficiently transmissible that many or most infections are acquired from humans in that population. We show that in the simplest case, $R_0 < 1$ if and only if the proportion of infections acquired from the external source exceeds the disease prevalence and explore how population heterogeneity and the interactions of multiple strains affect this rule. We apply these rules in two cases studies of *Clostridium difficile* infection and colonisation*: C. difficile* in the hospital setting accounting for imported cases, and *C. difficile* in the general human population accounting for exposure to animal reservoirs. We demonstrate that even the hospital-adapted, highly-transmissible NAP1/RT027 strain of *C. difficile* had a reproduction number <1 in a landmark study of hospitalised patients and therefore was sustained by colonised and infected admissions to the study hospital. We argue that *C. difficile* should be considered reservoir-driven if as little as 13.0% of transmission can be attributed to animal reservoirs.

1 Introduction

Many pathogens affecting humans circulate between humans and animals through contact, food or indirectly through common disease vectors in the environment. Other pathogens move across population boundaries due to the movement of people. In the absence of transmission from other populations or reservoirs, the basic reproduction number – the average number of secondary cases arising from each primary case in a susceptible population – determines whether a disease will die out or persist through ongoing person-to-person transmission. Effective interventions can interrupt transmission by reducing the basic reproduction number below 1 causing the disease to die out in that population. However, any reservoir exposure or imported cases will continue to replenish the infected population, and so a disease will die out in a population if and only if the basic reproduction number is <1 *and* all reservoir exposure and importation are avoided. There is a rich literature in meta-population models that capture the interactions of populations that introduce or reintroduce pathogens to one another (e.g. [1–4]). However, one often only has data or interest in a single population but needs to account for external sources of infections. It is in this context that we wish to generate some simple principles or rules for estimating the reproduction number.

Methods have been developed to estimate the human reproduction numbers of emerging zoonoses with limited person-to-person spread [5,6]. Others have developed methods to account for the often large proportion of imported cases at the beginning of new epidemics, which if excluded or treated as if locally acquired would overestimate the reproduction number [7]. Though the term 'elimination' has been defined in many different ways [8], reducing the local reproduction number below one is one measure of this progress [9], and is a necessary step towards global eradication. Methods have been developed to estimate the reproduction number that account for the potentially large proportion of imported cases in settings where progress is being made towards elimination [9]. However none of these methods account for susceptible depletion and so are restricted to diseases with very low prevalence [5,6,9] or calculate the *effective* reproduction number [7], which is not a threshold parameter for disease persistence. Starting with simple models and incorporating heterogeneity or multiple strains, we have derived simple rules for estimating the reproduction number in a population where the disease is at endemic equilibrium due to a combination of local

person-to-person transmission *and* reservoir exposure or imported cases. Many of these rules only require knowledge of disease prevalence and the proportion of infections attributable to the external source. We have applied these rules in two case studies of *C. difficile* infections.

2 The SIS Model

We begin by adapting the simplest possible compartmental model: the standard SIS model with a homogeneous, well-mixed population without demographics. We include two sources of infection: (1) person-to-person transmission which is proportional to the number of people infected (rate: βi) and (2) constant transmission from some reservoir that does not depend on the number of people infected (rate: f). Person-to-person transmission could be through direct contact, or mediated via fomites, airborne droplets, water or food provided this transmission scales with the infectious population. For our purposes a reservoir is anywhere where the pathogen persists apart from the human population, for instance a population of wild animals or livestock animals that carry the disease. The disease in the human population can be described by a system of ODEs for the proportion of the population that is susceptible (s) and infected (i):

$$\begin{aligned} \boldsymbol{s'} &= -\boldsymbol{\lambda(t)s} + \boldsymbol{\gamma i} \\ \boldsymbol{i'} &= \boldsymbol{\lambda(t)s} - \boldsymbol{\gamma i} \end{aligned} \quad , \qquad \mathbf{1}$$

where $\lambda(t) = \beta i + f$ is the force of infection and γ is the rate at which infected individuals recover.

Diseases that are acquired entirely from food or animals and diseases that are spread entirely by person-to-person transmission, are extreme cases of this model with $f = 0$ and $\beta = 0$ respectively. Many diseases lie between these two extremes. Almost all human cases of H7N9 avian influenza have been acquired from birds, but there has been some person-to-person transmission which is not enough to maintain endemic disease [10]. Meanwhile human-adapted seasonal influenza (H1N1, H3N2) are mainly transmitted to humans by other humans, though there are low frequency transmission events from animal reservoirs (e.g. [11]). Middle-eastern respiratory syndrome coronavirus sits somewhere in the middle of the spectrum with significant human-to-human and animal-to-human transmission [12].

The reproduction number for this simple model in the next-generation sense [13] is the same as for the standard SIS model (β/γ) but is a threshold parameter for extinction of the disease only when there is no transmission from the reservoir ($f = 0$), i.e. when the model reduces to the standard SIS model. Otherwise, the reservoir will continually replenish the infected population whatever the value of R_0. If there is no transmission from the reservoir we have the well-known relationship between the basic reproduction number (R_0) and the proportion of the population susceptible at equilibrium (S): $R_0 = 1/S$. The model parameters are difficult to measure directly and so we wish to estimate R_0 through observable quantities by generalising this rule. Let I, S and $\Lambda = \beta I + f$ be the non-trivial (i.e. $I, \Lambda \neq 0$) equilibrium values of i, s and λ. As equilibrium points of **1** they satisfy

$$\Lambda S = \gamma I$$

or equivalently

$$\gamma = S\frac{\Lambda}{I}.$$

Now the proportion of transmission that is from the reservoir at equilibrium, π, is

$$\pi = \frac{f}{\Lambda} = 1 - \frac{\beta I}{\Lambda},$$

which re-arranged for β gives

$$\beta = (1 - \pi)\frac{\Lambda}{I}.$$

Substituting these expressions for β and γ into the expression for the reproduction number we get

$$R_0 \equiv \frac{\beta}{\gamma} = \frac{1 - \pi}{S}.$$

2

We can also write this in terms of the proportion infected (which is usually what is reported rather than the proportion susceptible).

$$R_0 = \frac{1 - \pi}{1 - I}.$$

3

These expressions simplify to $R_0 = 0$ if the disease is only acquired from the reservoir ($\beta = 0, \pi = 1$) or to $R_0 = \frac{1}{S} = \frac{1}{1-I}$ when all transmission is person-to-person ($f = 0, \pi = 0$). The general cases of these expressions lead to a simple rule for the reproduction number: $R_0 > 1$ if and only if $I > \pi$. The disease can be maintained by person-to-person transmission in the absence of reservoir exposure if and only if the prevalence exceeds the proportion of transmission from the reservoir.

This simple rule has surprising implications. For diseases with low prevalence (e.g. 2%), if a small but larger portion (e.g. 3%) of transmission is from the reservoir, then the disease cannot be sustained in the population by person-to-person transmission alone (since $R_0 = \frac{1-0.03}{1-0.02} < 1$). Preventing the small proportion of transmission from the reservoir (reducing f and π to 0) will cause the disease to become extinct in the population. Nevertheless, names like 'food-borne' or 'zoonotic' may be misleading for such diseases because the source of transmission is another human in most (e.g. 97%) individual infections. Instead we call these diseases *reservoir-driven*. We define the *reservoir-driven threshold* as the minimum proportion of transmission which must be from the reservoir for the disease to be considered reservoir-driven (I in our simple SIS model).

The rest of this article will consider variants and extensions of the simple SIS model to demonstrate which assumptions do and do not affect the above expressions for the reproduction number and reservoir-driven threshold. We will also show that an equivalent rule of thumb and threshold exists when a disease is driven by imported cases due to travel or immigration. We will however not relax the key assumptions that the disease is at endemic equilibrium in the population, so the rules we derive are *at best* only approximately valid for diseases where the prevalence varies substantially over time. We will then consider how this rule of thumb can be applied to case studies of real diseases.

3 Simple Extensions of the SIS Model

3.1 Births and Deaths

Simple demographics does not change our rule for the reproduction number. A modified model including deaths from both classes at rate δ and births that balance deaths is described by the equations

$$\begin{aligned} s' &= -\lambda(t)s + \gamma i - \delta s + \delta \\ i' &= \lambda(t)s - \gamma i - \delta i \end{aligned},$$

where $\lambda(t) = \beta i + f$ is the force of infection. In this model $R_0 = \frac{\beta}{\gamma+\delta}$. Let I, S and $\Lambda = \beta I + f$ be the non-trivial (i.e. $I, \Lambda \neq 0$) equilibrium values of i, s and λ. Then

$$\Lambda S = (\gamma + \delta)I,$$

or equivalently

$$\gamma + \delta = S\frac{\Lambda}{I}.$$

The force of infection terms are the same as for our original model so again we have $\beta = (1-\pi)\frac{\Lambda}{I}$. Substituting this into the expression for the reproduction number we get the same result as **2** and **3**:

$$R_0 \equiv \frac{\beta}{\gamma+\delta} = \frac{1-\pi}{S} = \frac{1-\pi}{1-I}$$

and the reservoir-driven threshold is still I. We have assumed that the death rates are the same for infected and susceptible persons, but it is simple to show that a higher (or lower) death rate for infected individuals does not affect the reasoning.

3.2 Recovered Classes and Other Common Extensions

The simplest SIR model without birth and deaths or waning immunity does not have an endemic equilibrium point so our method for estimating the reproduction number is not applicable to these models. Instead, consider the SIR model with births and deaths:

$$\begin{aligned} s' &= -\lambda(t)s - \delta s + \delta \\ i' &= \lambda(t)s - \gamma i - \delta i \\ r' &= \gamma i - \delta r \end{aligned},$$

where $\lambda(t) = \beta i + f$ the force of infection. Note that adding the recovered class to the SIS model with births and deaths does not change the reproduction number, the equation governing the number of infected individuals or the force of infection and so the reasoning is identical to that in the previous section. However, since there are more than two classes, $S + I \neq 1$. Therefore **3** does not hold but instead,

$$R_0 = \frac{1-\pi}{\mathrm{S}} = \frac{1-\pi}{1-(I+R)}.$$

The reservoir-driven threshold here is $I + R$, i.e. the disease can be sustained by person-to-person transmission in the absence of reservoir exposure if and only if the proportion of transmission which is due to reservoir exposure is less than the total proportion of people infected or immune/recovered.

The same reasoning can be used for models with waning immunity, vaccination, or latent/exposed classes. Since these modifications do not affect the equations governing the number of infected individuals or the force of infection, equation **2** still holds and therefore the reservoir-driven threshold is $1 - S$ in all these cases. For diseases with comprehensive vaccination programs (or common diseases with lifelong immunity), almost all the population can be immune (e.g. 95%) and the proportion susceptible very low. If reservoir exposure accounts for nearly all cases but is still less than the reservoir-driven threshold (e.g. 90%), the disease could be sustained by person-to-person transmission alone if reservoir exposure was eliminated (since $R_0 = \frac{1-0.90}{1-0.95} > 1$) and so eliminating exposure to the reservoir would not eliminate the disease from the human population.

3.3 Imported Cases

Analogous rules can be derived for settings where some infections are acquired locally and others are imported through immigration or those returning from travel. We assume that susceptible and infected individuals emigrate/leave at the same rate δ, that immigration balances emigration and that a proportion p of those entering the population are infected. The governing equations are

$$\begin{aligned} s' &= -\lambda(t)s + \gamma i - \delta s + (1-p)\delta \\ i' &= \lambda(t)s - \gamma i - \delta i + p\delta \end{aligned}$$

where $\lambda(t) = \beta i$ is the force of infection. Again, $R_0 = \beta/(\gamma + \delta)$ but R_0 is not by itself a threshold parameter for disease extinction because the continuous immigration of new infected individuals will sustain the disease (unless $p\delta = 0$). The equilibrium proportion infected (I), proportion susceptible (S) and force of colonisation ($\Lambda = \beta I$) satisfy

$$\Lambda S + p\delta = (\gamma + \delta)I,$$

or equivalently

$$\gamma + \delta = \frac{\Lambda S + p\delta}{I}.$$

Meanwhile the proportion of new cases that are imported, q, is

$$q = \frac{p\delta}{\Lambda S + p\delta} = 1 - \frac{\Lambda S}{\Lambda S + p\delta} = 1 - \frac{\beta I S}{\Lambda S + p\delta}$$

which we can rearrange for the transmission parameter giving

$$\beta = \frac{(1-q)(\Lambda S + p\delta)}{IS}.$$

Therefore, we can write the reproduction number as

$$R_0 \equiv \frac{\beta}{\gamma + \delta} = \frac{1-q}{S} = \frac{1-q}{1-I}.$$

These expressions lead to simple rules for the reproduction number analogous to those derived for diseased reservoir exposure. $R_0 > 1$ if and only if $I > q$. That is, in this simple model, the disease can be sustained without importation by local transmission if and only if the prevalence exceeds the proportion of new cases that are imported through migration or travel. By analogy to the reservoir exposure model, we call this threshold the *importation-driven* threshold. This analogy still holds when heterogeneity or multiple strains are incorporated into these models – extensions we consider in sections 4 and 5.

4 Heterogeneity

It is known that accounting for population heterogeneity tends to increase estimates of reproduction numbers [14]. Therefore, we might expect that introducing heterogeneity into models with reservoir exposure will increase the reservoir-driven threshold. Consider a general SIS model with separable mixing in a heterogeneous population consisting of people of different x-types with susceptibility $\alpha(x)$, transmission parameter $\beta(x)$ and mean infectious period $1/\gamma(x)$, distributed according to the probability density function $g(x)$. Then

$$\begin{aligned} s_t(x,t) &= -\lambda(t)\alpha(x)s(x,t) + \gamma(x)i(x,t) \\ i_t(x,t) &= \lambda(t)\alpha(x)s(x,t) - \gamma(x)i(x,t) \end{aligned}$$

and

$$s(x,t) + i(x,t) = g(x),$$

where $\lambda(t) = f + \int \beta(x)i(x,t)dx$. For this model, $R_0 = \int \beta(x)/\gamma(x)\alpha(x)g(x)dx$ [15], but as before R_0 is only a threshold parameter for disease extinction if $f = 0$. Let I and S be the non-trivial equilibrium distributions of i, s and $\Lambda = \int \beta(x)I(x)dx + f$ the equilibrium value of λ. As equilibrium points they satisfy

$$\Lambda S(x)\alpha(x) = \gamma(x)I(x),$$

or equivalently,

$$\frac{S(x)\alpha(x)}{\gamma(x)} = \frac{I(x)}{\Lambda}.$$

4

At equilibrium, the proportion of infections acquired from the reservoir, which is the proportion of force of infection attributable to the reservoir, is

$$\pi = \frac{f}{\Lambda} = 1 - \int \frac{\beta(x)I(x)}{\Lambda}dx.$$

Substituting 4 into the above gives

$$\pi = 1 - \int \frac{\beta(x)\alpha(x)S(x)}{\gamma(x)}dx.$$

If we let $\rho = \beta\alpha/\gamma$ we can write the reproduction number and the proportion of infections from the reservoir in simpler terms

$$\begin{aligned} R_0 &= \int \rho(x)g(x)dx \\ &= \overline{\rho} \end{aligned}$$

where $\overline{\rho}$ is the mean value of ρ across the population and

$$\begin{aligned}\pi &= 1 - \int \rho(x)S(x)dx \\ &= 1 - \boldsymbol{S}\int \rho(x)\frac{S(x)}{\boldsymbol{S}}dx \\ &= 1 - \boldsymbol{S}\,\overline{\rho_S}\end{aligned}$$

where $\boldsymbol{S} := \int S(x)dx$ is the total susceptible population and $\overline{\rho_S}$ is the mean value of ρ across the susceptible population. Therefore

$$R_0 = \frac{1-\pi}{\boldsymbol{S}\,\frac{\overline{\rho_S}}{\overline{\rho}}}.$$

By similar reasoning one can show that

$$R_0 = \frac{1-\pi}{1-\boldsymbol{I}\,\frac{\int \rho(x)\frac{I(x)}{\boldsymbol{I}}dx}{\int \rho(x)g(x)dx}} = \frac{1-\pi}{1-\boldsymbol{I}\,\frac{\overline{\rho_I}}{\overline{\rho}}}$$

5

where $\boldsymbol{I} := \int I(x)dx$ is the proportion of the whole population that is infected and $\overline{\rho_I}$ is the mean value of ρ across the infected population. The quantity we want to estimate, R_0, appears as $\overline{\rho}$ in the right-hand sides of each equation and the quantities $\overline{\rho_I}$ and $\overline{\rho_S}$ are unlikely to be known, so this does not provide a practical way to estimate the reproduction number. However, these statements provide some insight into how heterogeneity can affect our estimates of the reproduction number or reservoir-driven threshold. The rule of thumb is similar to the rule for a homogenous population: $R_0 > 1$ if and only if $\boldsymbol{I}\,\overline{\rho_I}/\overline{\rho} > \pi$, i.e. the reservoir-driven threshold is $\boldsymbol{I}\,\overline{\rho_I}/\overline{\rho}$. If those who are infected have higher-than-average (or lower-than-average) ρ, then accounting for this heterogeneity increases (or decreases) the reservoir-driven threshold. We derive simple expressions for the value of $\overline{\rho_I}/\overline{\rho}$ under some specific assumptions.

4.1 Variable Susceptibility or Infectious Period

If we assume that the infectiousness (β) is fixed but the product of susceptibility and length of infectious period ($\phi := \alpha/\gamma$) is heterogeneous, then the reservoir-driven threshold is always higher than for a homogenous population. Consider the ratio $\overline{\rho_I}/\overline{\rho}$

$$\frac{\overline{\rho_I}}{\overline{\rho}} = \frac{\int \frac{\beta\alpha(x)}{\gamma(x)}\frac{I(x)}{\boldsymbol{I}}dx}{\int \frac{\beta\alpha(x)}{\gamma(x)}g(x)dx} = \frac{\int \phi(x)\frac{I(x)}{\boldsymbol{I}}dx}{\int \phi(x)g(x)dx} = \frac{\overline{\phi_I}}{\overline{\phi}},$$

where $\overline{\phi}$ and $\overline{\phi_I}$ are the mean values of ϕ across the whole population and across the infected portion of the population respectively. Now we can rearrange **4** in terms of the odds of infection of an individual of type x

$$\frac{I(x)}{S(x)} = \frac{\Lambda\alpha(x)}{\gamma(x)} = \Lambda\phi(x).$$

6

Since the odds of infection for an individual of type x is proportional to $\phi(x)$, individuals with high ϕ (i.e. more susceptible individuals or individuals with longer infectious periods) will be over-represented in the infected portion of the population at equilibrium. Therefore $\overline{\phi_I} \geq \overline{\phi}$ and so the reservoir-driven threshold is at least as high as for a homogenous population.

If the prevalence is low for people of all x-types (i.e. $S(x) \approx g(x)$) there is a simple approximation for the reservoir-driven threshold. We can rearrange **4** to get

$$I(x) = S(x)\Lambda\phi(x) \approx g(x)\Lambda\phi(x).$$

and so

$$\boldsymbol{I} = \int I(x)dx \approx \Lambda \int \phi(x)g(x)dx = \Lambda\overline{\phi}$$

and

$$\overline{\phi_I} = \int \phi(x)\frac{I(x)}{\boldsymbol{I}}dx \approx \frac{1}{\overline{\phi}}\int \phi(x)^2 g(x)dx.$$

If the population variance of ϕ is $\sigma^2 := \int (\phi(x) - \overline{\phi})^2 g(x)dx$ the ratio can be written approximately as

$$\frac{\overline{\phi_I}}{\overline{\phi}} \approx \frac{1}{\overline{\phi}^2}\int \phi(x)^2 g(x)dx = 1 + \frac{\sigma^2}{\overline{\phi}^2}$$

and the reproduction number is

$$R_0 \approx \frac{1-\pi}{1 - \boldsymbol{I}\left(1 + \frac{\sigma^2}{\overline{\phi}^2}\right)}.$$

When there is no heterogeneity in ϕ (i.e. when $\sigma^2 = 0$) this simplifies to the result for the homogenous SIS model. The larger the variance for a given mean, the greater the basic reproduction number and the higher the reservoir driven-threshold, $\boldsymbol{I}(1 + \sigma^2/\overline{\phi}^2)$. For example if $\phi(x)$ and $g(x)$ are such that the distribution of ϕ across the population is gamma with mean μ and shape parameter k (a convenient and often used assumption [14]), then the reservoir-driven threshold is approximately $\boldsymbol{I}\left(1 + \frac{1}{k}\right)$ (**Figure 1**).

If the x type of an individual corresponds to some easily determined risk class – for instance if x denotes gender or smoker status – then the proportion of people in each class, $g(x)$, and the odds of infection within each group, $I(x)/S(x)$, may be known. Since the odds of infection is proportional to ϕ, we can express $\overline{\phi_I}/\overline{\phi}$ and R_0 in terms of these observed quantities:

$$\frac{\overline{\phi_I}}{\overline{\phi}} = \frac{\int \phi(x)\frac{I(x)}{\boldsymbol{I}}dx}{\int \phi(x)g(x)dx} = \frac{\int \frac{I(x)}{S(x)}\frac{I(x)}{\boldsymbol{I}}dx}{\int \frac{I(x)}{S(x)}g(x)dx}$$

and

$$R_0 = \frac{1-\pi}{1-\boldsymbol{I}\dfrac{\overline{\phi_I}}{\overline{\phi}}} = \frac{1-\pi}{1-\boldsymbol{I}\dfrac{\int \frac{I(x)}{S(x)}\frac{I(x)}{\boldsymbol{I}}dx}{\int \frac{I(x)}{S(x)}g(x)dx}} \quad .$$

7

4.2 Variable Infectiousness

If infectiousness (β) is heterogeneous, but the product of susceptibility and length of the infectious period ($\phi := a/\gamma$) is fixed, then the reservoir-driven threshold is the same as for a homogenous population. Consider the ratio $\overline{\rho_I}/\overline{\rho}$ which can be simplified as

$$\frac{\overline{\rho_I}}{\overline{\rho}} = \frac{\int \beta(x)\phi\frac{I(x)}{\boldsymbol{I}}dx}{\int \beta(x)\phi g(x)dx} = \frac{\int \beta(x)\frac{I(x)}{\boldsymbol{I}}dx}{\int \beta(x)g(x)dx} = \frac{\overline{\beta_I}}{\overline{\beta}},$$

where $\overline{\beta}$ and $\overline{\beta_I}$ are the mean values of β across the whole population and across the infected portion of the population respectively. Now by **6**, if ϕ is constant across the population the odds of infection at equilibrium is the same for people of every x-type, i.e. independent of their infectiousness. Therefore, the mean infectiousness amongst the infected population is the same as the mean infectiousness across the whole population and $\overline{\beta_I}/\overline{\beta} = \overline{\rho_I}/\overline{\rho} = 1$. Equations **5** then simplifies to

$$R_0 = \frac{1-\pi}{1-\boldsymbol{I}}$$

which is the same as the result for a homogenous population.

Heterogeneous infectiousness will affect the reservoir-driven threshold in a population which is also heterogeneous with respect to susceptibility or infectious period. If those who are more likely to be in the infected class (high ϕ) are less infectious (low β), this will reduce the reservoir-driven threshold relative to homogeneous infectiousness but heterogeneous susceptibility and infectious period. As a simple example of this, assume that $\beta(x) = 1/\phi(x)$. Then $\rho(x) = 1$, $\overline{\rho} = \overline{\rho_I} = 1$ and the reservoir-driven threshold is simply $\boldsymbol{I}$, less than what it would be if β were constant across the population. On the other hand, if those who are more likely to be colonised (high ϕ) are also more infectious (high β), the reservoir-driven threshold will increase relative to homogeneous infectiousness but heterogeneous susceptibility and infectious period. As another simple example consider the proportional mixing assumption, i.e. $\beta \propto \phi$. In this case $\rho \propto \phi^2$ and so

$$\frac{\overline{\rho_I}}{\overline{\rho}} = \frac{\int \phi(x)^2\frac{I(x)}{\boldsymbol{I}}dx}{\int \phi(x)^2 g(x)dx} \quad .$$

When the prevalence is low for people of all x-types (i.e. $S(x) \approx g(x)$) we can use the same reasoning as in the previous section to approximate this ratio as

$$\frac{\overline{\rho_I}}{\overline{\rho}} \approx \frac{\int \phi(x)^3 g(x)dx}{\overline{\phi}\int \phi(x)^2 g(x)dx} = \frac{\nu}{\overline{\phi}(\overline{\phi}^2 + \sigma^2)}$$

and the reproduction number by

$$R_0 \approx \frac{1-\pi}{1-\boldsymbol{I}\left(\frac{\nu}{\overline{\phi}(\overline{\phi}^2+\sigma^2)}\right)}.$$

where ν is the third raw moment of ϕ across the population. If for example, $\phi(x)$ and $g(x)$ are such that ϕ is gamma distributed with shape parameter k then the reservoir-driven threshold is approximately $\boldsymbol{I}(1+\frac{2}{k})$, which is higher than if β were homogeneous. **Figure 1** summarises how the reservoir-driven threshold changes for different types of heterogeneity explored so far.

4.3 Variable Exposure to Reservoir

Heterogeneous exposure to the reservoir in an otherwise homogeneous population does not change the reservoir-driven threshold. Consider an SIS model where the population consists of people of type x distributed according to $g(x)$ each with their own level of exposure to reservoir $f(x)$. Then the differential equations governing the system are

$$\begin{aligned} s_t(x,t) &= -\lambda(x,t)s(x,t)+\gamma i(x,t) \\ i_t(x,t) &= \lambda(x,t)s(x,t)-\gamma i(x,t) \end{aligned}$$

and

$$s(x,t)+i(x,t)=g(x),$$

where $\lambda(x,t)=\beta\int i(\xi,t)d\xi+f(x)$ is the force of infection acting on individuals of type x. The basic reproduction number for this model is β/γ. Let I and S be the non-trivial equilibrium distributions of i,s (i.e. $I\neq 0$), $\boldsymbol{I}$ and $\boldsymbol{S}$ be the total number of people infected and susceptible at equilibrium and $\Lambda(x)=\beta\int I(x)dx+f(x)=\beta\boldsymbol{I}+f(x)$ be the equilibrium force of infection. As equilibrium points they satisfy

$$\Lambda(x)S(x)=\gamma I(x).$$

The proportion of transmission that is acquired from the reservoir is then

$$\begin{aligned} \pi &= \frac{\int f(x)S(x)dx}{\int \Lambda(x)S(x)dx} \\ &= 1-\frac{\int \beta\boldsymbol{I}S(x)dx}{\int \Lambda(x)S(x)dx} \\ &= 1-\frac{\int \beta\boldsymbol{I}S(x)dx}{\int \gamma I(x)dx} \\ &= 1-\frac{\beta}{\gamma}\boldsymbol{S} = 1-R_0\boldsymbol{S} \end{aligned}$$

and consequently

$$R_0=\frac{1-\pi}{\boldsymbol{S}}=\frac{1-\pi}{1-\boldsymbol{I}}$$

leaving the reservoir-driven threshold unchanged. However, interactions with additional heterogeneities will affect the reservoir-driven threshold. Consider the case where both reservoir exposure (f) and the person-to-person transmission rate (β) depend on the x-state. In this case the

equilibrium force of infection is $\Lambda(x) = \int \beta(\xi) I(\xi) d\xi + f(x)$, the reproduction number is $R_0 = \int g(\xi)\, \beta(\xi)/\gamma \, d\xi = \overline{\beta}/\gamma$ where $\overline{\beta}$ is the mean value of β in the population. The proportion of infections that are acquired from the reservoir is

$$\begin{aligned}
\pi &= \frac{\int S(x) f(x)}{\int S(x)\Lambda(x)dx} \\
&= 1 - \frac{\int S(x) \int I(\xi)\beta(\xi) d\xi dx}{\int S(x)\Lambda(x)dx} \\
&= 1 - \int I(\xi)\beta(\xi) d\xi \frac{\int S(x)dx}{\int I(x)\gamma dx} \\
&= 1 - \int \frac{I(\xi)}{\boldsymbol{I}} \beta(\xi) d\xi \frac{\boldsymbol{S}}{\gamma} \\
&= 1 - \frac{\overline{\beta_I}}{\gamma}\boldsymbol{S} = 1 - R_0 \frac{\overline{\beta_I}}{\overline{\beta}} \boldsymbol{S},
\end{aligned}$$

where $\overline{\beta_I}$ is the mean value of β in the infected population. Therefore

$$R_0 = \frac{1-\pi}{\boldsymbol{S}\dfrac{\overline{\beta_I}}{\overline{\beta}}} = \frac{1-\pi}{(1-\boldsymbol{I})\dfrac{\overline{\beta_I}}{\overline{\beta}}}.$$

Those that have greater exposure to the reservoir are more likely to be infected and so will have a disproportionally large effect on $\overline{\beta_I}$. If those with more exposure to the reservoir are also on more infectious then $\overline{\beta_I} > \overline{\beta}$ and the reservoir-driven exposure is lower, and conversely if those with more exposure to the reservoir also less infectious then $\overline{\beta_I} < \overline{\beta}$ and the reservoir-driven threshold is higher (**Figure 2**). Note that this is opposite to the relationship for heterogeneous β and heterogeneous ϕ (**Figure 1**).

5 Multiple Strains

There is frequently more than one strain of a pathogen co-circulating within human populations and the dynamics of multi-strain interactions have been modelled extensively (e.g. [16–20]). In the few simple multi-strain models we consider, accounting for host competition increases the reservoir driven threshold for each strain compared to the single strain model. Consider a simple competitive multi-strain extension of our basic SIS model with reservoir exposure. Each strain has its own transmission parameter (β_k), recovery rate (γ_k) and reservoir exposure rate (f_k). We assume that infection with one strain prevents infection from all other strains for the duration of the infection. With n strains the $n+1$ equations governing this system are

$$\begin{aligned}
s' &= -\sum_{k=1}^{n} \lambda_k(t) s + \sum_{k=1}^{n} \gamma_k i_k \\
i_k' &= \lambda_k(t) s - \gamma_k i_k, \qquad k = 1, \ldots, n
\end{aligned}$$

where $\lambda_k(t) = \beta_k i_k(t) + f_k$ is the force of infection for each strain. Each strain has its own basic reproduction number in a fully susceptible population: $R_0^k = \beta_k/\gamma_k$. Here, R_0^k are not threshold parameters for strain extinction because reservoir exposure will cause the disease to persist *and* strain competition for hosts may cause a strain without reservoir exposure to die out even if that strain's reproduction number exceeds one. Let S, be the equilibrium number of susceptible people at the non-

trivial equilibrium where the number of people infected with each strain (I_k) and the force of colonisation for each strain (Λ_k) are non-zero. For each strain we have the following relation

$$\Lambda_k S = \gamma_k I_k,$$

or equivalently

$$\gamma_k = \frac{\Lambda_k S}{I_k}.$$

The proportion of transmission of strain k that is from the reservoir is

$$\pi_k = \frac{f_k}{\Lambda_k} = 1 - \frac{\beta_k I_k}{\Lambda_k}.$$

Rearranging for β_k:

$$\beta_k = (1 - \pi_k)\frac{\Lambda_k}{I_k}.$$

We can re-write the basic reproduction number for strain k as

$$R_0^k \equiv \frac{\beta_k}{\gamma_k} = \frac{1 - \pi_k}{S} = \frac{1 - \pi_k}{1 - \sum_{j=1}^{n} I_j}.$$

Consequently $R_0^k < 1$ if $\pi_k > \sum_{j=1}^{n} I_j$. It follows that a given strain cannot persist without reservoir exposure if the proportion of transmission of that strain due to reservoir-exposure is more than the total prevalence of *all* strains.

We also want to account for strain competition which can lead to the extinction of strains that would otherwise persist in a population. Therefore, we consider the invasion reproduction number for each strain, i.e. not the reproduction number in a fully susceptible population, but in a population at endemic equilibrium with all the *other* strains. Consider the equilibrium point without any infections of strain k that exists if there is no reservoir exposure for strain k (i.e. $f_k = 0$). Let $S^k, I_1^k, \dots, I_n^k$, be the equilibrium values of s, $i_1, \dots, i_n$ when $f_k = 0$, such that $I_k^k = 0$ and $I_j^k > 0$ if $j \neq k$. The invasion reproduction number for strain k is then $R_{Invasion}^k = R_0^k S^k$. It is possible to calculate S^k in terms of $\pi_1, \dots, \pi_n$ and $I_1, \dots, I_n$ but the exact form is cumbersome (even for the $n = 2$ case) so instead we consider a simple bound. Consider that the equilibrium proportion of each strain other than k will certainly not decrease in the absence of the competition with strain k, i.e. $I_j^k \geq I_j$ for $j \neq k$. Consequently $S^k \leq S + I_k$ since

$$S^k = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} I_j^k \leq 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} I_j = 1 + I_k - \sum_{j=1}^{n} I_j = S + I_k.$$

We can bound the invasion reproduction number for strain k by

$$\begin{aligned} R^k_{Invasion} &= R^k_0 S^k \\ &\leq \frac{1-\pi_k}{S}\,(S+I_k) \\ &= \frac{1-\pi_k}{1-\frac{I_k}{S+I_k}}. \end{aligned}$$

Consequently $\frac{I_k}{S+I_k}$ is an upper bound for the reservoir-driven threshold in the presence of other strains since $R^k_{Invasion} < 1$ whenever $\pi_k > \frac{I_k}{S+I_k}$.

Our simple competitive model assumes complete exclusion, but in reality, strains are unlikely to completely exclude one another. If one allows for the possibility of coinfections or superinfection, assuming that persons infected with strains other than strain k ($I_{\neg k}$) are a_k times as susceptible to infection with strain k as those not infected with any strain (S) and that coinfecting/superinfecting strains do not affect infectiousness or infectious period for the infecting strains, then at endemic equilibrium

$$\Lambda_k(S + a_k I_{\neg k}) = \gamma_k I_k$$

where I_k is the proportion of people infected with strain k (who may also be infected with other strains) and $\Lambda_k = \beta_k I_k + f_k$. One can use the same reasoning as above to show that

$$R^k_0 \equiv \frac{\beta_k}{\gamma_k} = \frac{1-\pi_k}{S + a_k I_{\neg k}} = \frac{1-\pi_k}{1 - \mathrm{I_k} + (1-a_k) I_{\neg k}}$$

and

$$R^k_{Invasion} \leq \frac{1-\pi_k}{1-\frac{I_k}{S + a_k I_{\neg k} + I_k}} = \frac{1-\pi_k}{1-\frac{I_k}{1-(1-a_k)I_{\neg k}}}.$$

and so $\frac{I_k}{S+a_k I_{\neg k}+ I_k} = \frac{I_k}{1-(1-a_k)I_{\neg k}}$ is an upper bound for the reservoir-driven threshold. If $a_k = 0$, this reduces to the case of complete exclusion we considered above. If $a_k = 1$, that is if infection with another strain neither prevents nor predisposes a patient to infection with strain k, then the reservoir driven threshold and reproduction number are the same for as for a model with only a single strain. In general, the greater the exclusion against strain k (i.e. as $a_k \to 0$), the higher the reservoir-driven threshold and reproduction number. Consequently the case of complete exclusion is an upper bound for these quantities in these simple models.

6 Case Study: *Clostridium difficile*

Clostridium difficile is a bacterium that colonises the intestines of many mammals including humans and livestock [21]. Most human hosts do not have symptoms despite being colonised. Colonisation is typically transient, lasting approximately one month in adults [22], due to competition and interactions with other intestinal flora [23]. Disruption of the gut flora, often caused by consumption of antibiotics or proton-pump-inhibitors, allows *C. difficile* to proliferate in large numbers [23]. Toxigenic strains of *C. difficile* then produce a number of toxins that can cause diarrhoea which is often severe and sometimes life-threatening [24]. A robust immune response to these toxins is able to neutralise their effect [25] and most of the population have anti-toxin antibodies starting at a young age [26]. Immune responses protect against symptoms but not protect against colonisation [27]. Asymptomatically colonised carriers are also infectious [28] while animal models have shown that

disruption of gut flora, even in the absence of symptoms, increases spore shedding and infectiousness [29].

Since immunity does not prevent colonisation or infectiousness, a simple SIS model is an appropriate starting point for *C. difficile*, provided we identify the I-class with all *C. difficile* positive individuals (not just those with symptoms). We will use variations on the SIS model below to determine whether *C. difficile* is importation-driven in a hospital setting, and calculate the reservoir-driven threshold for *C. difficile* for the human population as a whole (where animals are the reservoir).

6.1 *C. difficile* in Hospitals

Historically, *C. difficile* has been of most concern and thus most studied in hospitalised patients where it complicates the care of many initially hospitalised for other conditions [30]. However, there is growing recognition of community-acquired cases that manifest during hospital stay. Since *C. difficile* is consistently present in many hospitals, it has been assumed that *C. difficile* is endemic in these settings and is responsible for many cases in the community.

If we begin by modelling *C. difficile* in hospitals as a (homogeneous) SIS model with very high rates of migration (hospital admission and discharge) then we can estimate the reproduction number using the method outlined in Section 3.3. In words, we will estimate the within-hospital reproduction number as

$$R_0^{\text{Hospital}} = \frac{1 - \text{Proportion of colonisations and infections acquired prior to admission}}{1 - \text{Prevalence of colonisation and infection}}.$$

One study of colonisation and infections in hospitalised patients found 184 patients colonised at admission, and another 240 patients that acquired colonisation or infection after admission [27]. They identified an additional 60 or so patients that developed a CDI within 72 hours of admission who were therefore deemed to have been exposed prior to admission. Thus, the proportion of *C. difficile* positive patients that acquired the pathogen prior to admission was approximately 50%.

In the same study 528/5422 patients were colonised or developed an infection for part of their hospital stay. Some patients were excluded from their analysis (mostly for missing data) leaving 424/4143 patients that were colonised or developed an infection for part of the hospital stay. While these do not provide an estimate of prevalence (since many of the colonised or infected patients were only colonised for part of the hospital stay) these figures provide upper bounds to the prevalence of colonisation and infection in the study hospital: 9.7% amongst all study patients and 10.2% after exclusions. Putting this into the above formula gives an upper bound for the within-hospital reproduction number of approximately 0.55.

Unlike the study cited above, most studies focus on symptomatic patients and do not test asymptomatic patients at admission. However, the proportion of patients diagnosed with a *C. difficile* infection that were admitted for a *C. difficile* infection (principal diagnosis) is routinely reported. As patients admitted with asymptomatic colonisation who subsequently develop symptoms will not have *C. difficile* infection as their principal diagnosis, this proportion is a lower bound for the total proportion of infections that are due to exposure prior to admission and thus let us estimate an upper bound for the reproduction number. In the USA in the years 1993-2014, 20-34% of admissions who had a *C. difficile* infection had it as their primary diagnosis [31]. This is in excess of typical prevalence of colonisation and infection amongst hospitalised patients: a review of colonisation prevalence reported a range of 4-29% [32]. Therefore, our upper bound for the reproduction number lies in the range 0.69-1.1.

So far, we have assumed hospitalised patients are homogeneous, but this is not the case. Patients who have recently been administered antibiotics are not more susceptible to colonisation but are more likely to develop symptoms and be more infectious [27]. Thus, an SIS model with heterogeneous infectiousness is perhaps more appropriate. However, heterogeneity in infectiousness alone does not affect the estimate of the reproduction number (Section 4.2). Factors affecting susceptibility to colonisation exist and adjusting for these will increase our estimate of the reproduction number (Section 4.1), but this is unfortunately beyond the scope of this case study. However, our simple estimates of R_0 are in agreement with more sophisticated models of *C. difficile* transmission in hospitals that have found that the reproduction number is likely to be less than one in many or most hospital settings [33,34].

There are many strains and types of *C. difficile* and it has been suggested that certain strains or types, such as NAP1/RT027, are particularly hospital-adapted [35,36]. It is possible that these strains have significantly higher reproduction numbers in the hospital than we have estimated above and thus may be self-sustaining in hospitals. Unfortunately, we do not have strain-level or type-level data for all strains or types. However, the article used to calculate our first estimate of the reproduction number report the proportion of infections and colonisations typed as NAP1/RT027 [27]. As the authors did not type all isolates, we assume that un-typed isolates were equally likely to be NAP1/RT027 as the isolates from similar patients that were typed, and that the proportion of NAP1/RT027 infections in those with onset <72h after admission (not reported) was similar to patients with colonisation at admission (13%). Under these assumptions, approximately 32 out of 150 (21%) colonisations or infections with NAP1/RT027 were present at admission. Of the approximately 10% of cases that were colonised or infected for some part of their hospital stay, approximately 3% were with NAP1/RT027 and the remaining 7% were with other types. Though colonisation with non-toxigenic strains appears to be protective against infection with toxigenic strains [37], we do not have good information about the interaction of *C. difficile* types. Nevertheless, we can use the argument we presented in section 5 to bound the invasion reproduction number. This becomes

$$\begin{aligned} R_{\text{Invasion}}^{\text{NAP1}} &\leq \frac{1 - \text{Proportion of NAP1 colonisations and infections acquired prior to admission}}{1 - \dfrac{\text{Prevalence of NAP1 colonisation and infection}}{\text{Proportion } C.\,difficile \text{ negative} + \text{Prevalence of NAP1 colonisation and infection}}} \\ &\approx \frac{1 - 0.21}{1 - \dfrac{0.03}{0.9 + 0.03}} \\ &\approx 0.8. \end{aligned}$$

The basic reproduction number (i.e. in a completely susceptible population without competition with other types) is slightly greater:

$$\begin{aligned} R_0^{\text{NAP1}} &\leq \frac{1 - \text{Proportion of NAP1 colonisations and infections acquired prior to admission}}{1 - \text{Prevalence of any } C.\,difficile \text{ colonisation and infection}} \\ &\approx \frac{1 - 0.21}{1 - 0.1} \\ &\approx 0.9. \end{aligned}$$

This suggests that even if other strains were eliminated and NAP1/RT027 did not compete for hosts, the continual importation of colonised and infected individuals would be required to sustain endemic disease in the study hospital. If we perform the same analysis for the pooled non-NAP1/RT027 strains in the study (approximately 212 of 334 colonisations and infections were present on admission) the equivalent upper bounds for the invasion reproduction number and basic reproduction number are both approximately 0.4. Therefore it appears that NAP1/RT027, though importation-driven, was better adapted for transmission in the study hospital than other strains.

6.2 *C. difficile* and Animal Reservoirs

Carriage of *C. difficile* in the general adult population is less common than in hospitals or aged-care facilities, with reported prevalence in the range 0-15%, though $\lesssim 5\%$ is perhaps most typical [32]. *C. difficile* is also commonly found colonising pets and livestock, while *C. difficile* spores are frequently isolated on meat, fresh produce and in water [21]. Crucially, there is significant overlap in strains observed in human and non-human sources [35]. However the proportion of human cases that are acquired from a non-human reservoir is unknown. Consequently, we cannot use our methods to estimate the reproduction number, but we can calculate the reservoir-driven threshold. If it is reasonable to suspect that reservoir exposure accounts for a proportion equal to or exceeding the threshold, then *C. difficile* may be sustained in the human population by exposure to animal reservoirs.

If we begin with a homogeneous SIS model with reservoir exposure, then our estimate of the reservoir-driven threshold is simply the prevalence in the community which is typically $\lesssim 5\%$ for adults (Section 2). Given the ubiquity of non-human exposure it is plausible that reservoir exposure exceeds this very low threshold. Some individuals will have higher exposure to these reservoirs (depending on diet and lifestyle factors), but this alone will not affect the reservoir-driven threshold unless those with greater exposure are also are more (or less) infectious (Section 4.3). Heterogenous infectiousness due to potential differences between patients with and without symptoms, or differences between patients with and without recent antimicrobial exposure does not affect the food driven exposure in isolation (Section 4.2). However, communities are not homogeneous with regards to *C. difficile* colonisation risk, as demonstrated by the higher rates of colonisation and infection in hospitals, aged-care facilities and the very high colonisation rates amongst infants. Accounting for this heterogeneity will increase our estimate of the reservoir-driven threshold (Section 4.1).

If we split our population into four risk categories – (A) hospitalised patients, (B) aged-care residents, (C) infants under 12 months and (D) the rest of the population – we can begin to account for some of this heterogeneity. If we assume separable mixing with heterogeneous susceptibility and infectious period, we need only the prevalence in each group and the proportion of the population that is in each group to estimate the reservoir-driven threshold (equation **7**). The reported range of colonisation prevalence in each of these groups is (A) 0-29%, (B) 0-51%, (C) 18-90% and (D) 0-15% respectively [32], while the total proportion of the population in each of these groups in a developed country like Australia is (A) <0.5% [38], (B) <1% [39], (C) <1.5% [40] and (D) >97% respectively.

If we use the upper end of the prevalence range for each risk group, though only 16.6% of the population is colonised, the reservoir-driven threshold is 48.0%. Assuming a lower colonisation prevalence in the majority population (D) decreases overall prevalence but increases heterogeneity and can increase the reservoir-driven threshold. If only 1% of the healthy adult population is colonised, then overall prevalence is 3.0% but the reservoir-driven threshold is much higher at 81.1%. These extreme values taken from across the literature are not typical and are unlikely to coincide in a single population. If we consider more typical values of colonisation prevalence, the picture is quite different. With prevalence half of the maximum reported values (i.e. (A) 14.5%, (B) 25.5%, (C) 45% and (D) 7.5%), which is still probably much higher than typical for infants in particular [41], the reservoir-driven threshold is only 13.0%. The reservoir-driven threshold is lower still if prevalence is lower in any of the high-risk minority groups (A-C). Figure 3 explores the effect of different prevalence assumptions on the reservoir-driven threshold.

This model and estimate of the reservoir-driven threshold is of course very rough. Transmission is not well mixed between or within the four risk-categories. Furthermore, the pathogen's interactions with medications, gut-flora and host immunity leads to greater complexity than can be captured with a simple SIS model. The risk-categories of individuals change over time as patients age or move in and out of hospitals and so a multi-patch with age structure would provide better estimates. Nevertheless,

this very simple calculation serves as a back-of-the-envelope estimate for the plausible range of the reservoir-driven threshold, demonstrating that under a range of reasonable assumptions a relatively small amount of transmission from animals could be sustaining endemic disease in human populations. Our simple calculations with figures from the middle of the reported prevalence range agree with a detailed, model of hospitals and communities that found the reservoir-driven threshold was between 3.5% and 26.0% for a wide range of plausible assumptions.

There are many strains or types of *C. difficile* that circulate in human populations and the arguments set out in section 5 could be used to determine whether individual types are reservoir-driven. It could be the case that some strains are sustained by exposure to animals, while other strains – though also present in animal populations – are sufficiently transmissible between humans to persist without transmission from animals. *C. difficile* PCR ribotype 078 (RT078) is a particularly good candidate to consider as a reservoir-driven strain. Though it is not known what proportion of human RT078 cases can be attributed to transmission from an animal source, whole-genome sequencing of isolates of this strain from livestock and humans strongly suggest frequent transmission between these groups [42]. On the other hand NAP1/RT027 which is found in livestock but appears to be more transmissible between people than other strains, might have some human cases attributable to animal sources but is less likely to be animal-driven [43]. Finally RT001, which accounts for many human infections in European settings, appears to be uncommon in livestock [43].

7 Conclusion

We have outlined the theory and application of very simple rules to estimate reproduction numbers in the presence of reservoir-exposure or imported cases. The rules require minimal information about the population and the pathogen of interest and could be a useful starting point or alternative to more complex models tailored to a population or pathogen. Churcher et al. have developed a statistical test using branching process theory to infer whether $R_0 < 1$ in a population nearing disease elimination but with many imported cases [9]. Cauchemez et. al use a similar approach that accounts for incomplete case detection and the overrepresentation of larger outbreaks to estimate the reproduction number for emerging zoonoses [5]. However, their models assume almost all the population is susceptible and so are not suitable for situations where the prevalence of infection or immunity is far from zero. Moreover, the latter method assumes that the reproduction is less than one so is not appropriate in settings where is there is genuine uncertainty as to whether the reproduction number is above or below one [5]. Our model accounts for susceptible depletion and works for infections where the reproduction number is above or below one, but relies on estimates of prevalence to do so. This can pose a potential difficulty as incidence rather than prevalence is usually reported. Reliable estimates of prevalence either requires near-perfect case acquisition or surveys with large sample sizes especially when prevalence is low. Indeed a good deal of the variability in colonisation prevalence reported for *C. difficile* outside hospitals might be attributed to the relatively small sample sizes involved [32].

Some caution is required when using the reservoir-driven and importation-driven thresholds. It does not follow that if a disease is reservoir-driven or importation driven, then interventions targeting the external source and transmission from the external source will be most effective or 'best'. The 'best' control strategy will depend on the relative effort required to prevent each kind of exposure, the impact of these interventions and metric used to compare these. If it is equally feasible and desirable to eliminate all (or most) exposure from either source, eliminating transmission from the reservoir or importation is clearly the better choice for a reservoir-driven or importation disease as this will prevent all local human cases, while preventing all person-to-person transmission will prevent only the proportion of human cases spread locally by humans. However, if only modest reductions are feasible, then targeting local human transmission may be more effective. One can calculate the

normalised derivatives of equilibrium prevalence to estimate the reduction in prevalence achieved by a small reduction in person-to-person transmission or exposure to the external source. For example, in the homogenous SIS model with reservoir-exposure, a greater reduction in prevalence is achieved by reducing person-to-person transmission whenever less than half of cases are acquired from the reservoir [1]. This is true whether or not the disease is reservoir-driven. A similar rule can be derived for the SIS model with imported cases.

The major limitation of our method is the assumption that the disease and population are at equilibrium. Many diseases, including our case study disease *C. difficile,* exhibit seasonal variation [44]. It is possible that an infection is sufficiently transmissible to be locally sustained in high-transmission seasons, but reservoir-driven or importation-driven in low-transmission seasons [9]. Similarly, it possible that exposure to the reservoir is seasonal [45]. It is possible that an epidemic in one setting is driven by exposure to a population or reservoir where an epidemic is ongoing. Our model does not account for these kinds of temporal variability when estimating reproduction numbers and reservoir-driven thresholds.

The simplicity, minimal data requirements, generality and extensibility of the method we have presented here make it useful starting point for understanding the impact and interaction of transmission sources both internal and external to a population.

8 Acknowledgements

AM is supported by an Australian Government Research Training Program Scholarship. We thank Laith Yakob for suggested reading that proved valuable for framing our analysis of models with imported cases and providing helpful comments on the manuscript.

9 References

1. Hufnagel L, Brockmann D, Geisel T. 2004 Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci.* **101**, 15124–15129. (doi:10.1073/pnas.0308344101)
2. Becker NG, Dietz K. 1995 The effect of household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* **127**, 207–19. (doi:10.1016/0025-5564(94)00055-5)
3. Balcan D, Colizza V, Goncalves B, Hu H, Ramasco JJ, Vespignani A. 2009 Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**, 21484–21489. (doi:10.1073/pnas.0906910106)
4. Ball F, Neal P. 2002 A general model for stochastic SIR epidemics with two levels of mixing. *Math. Biosci.* **180**, 73–102. (doi:10.1016/S0025-5564(02)00125-6)
5. Cauchemez S, Epperson S, Biggerstaff M, Swerdlow D, Finelli L, Ferguson NM. 2013 Using routine surveillance data to estimate the epidemic potential of emerging zoonoses: application to the emergence of US swine origin influenza A H3N2v virus. *PLoS Med.* **10**, e1001399. (doi:10.1371/journal.pmed.1001399)
6. Kucharski A, Mills H, Pinsent A, Fraser C, Van Kerkhove M, Donnelly CA, Riley S. 2014 Distinguishing Between Reservoir Exposure and Human-to-Human Transmission for Emerging

[1] For this simple model the normalised derivatives w.r.t the person-to-person transmission rate and reservoir exposure rate can be written $\frac{\beta}{I}\frac{\partial I}{\partial \beta} = \frac{1-\pi}{\frac{I}{1-I}+\pi}$ and $\frac{f}{I}\frac{\partial I}{\partial f} = \frac{\pi}{\frac{I}{1-I}+\pi}$. Hence $\frac{\beta}{I}\frac{\partial I}{\partial \beta} > \frac{f}{I}\frac{\partial I}{\partial f}$ whenever $\pi < 1/2$.

Pathogens Using Case Onset Data. *PLoS Curr.* , 1–32. (doi:10.1371/currents.outbreaks.e1473d9bfc99d080ca242139a06c455f)
7. Mercer GN, Glass K, Becker NG. 2011 Effective reproduction numbers are commonly overestimated early in a disease outbreak. *Stat. Med.* **30**, 984–94. (doi:10.1002/sim.4174)
8. Cohen JM, Moonen B, Snow RW, Smith DL. 2010 How absolute is zero? An evaluation of historical and current definitions of malaria elimination. *Malar. J.* **9**, 213. (doi:10.1186/1475-2875-9-213)
9. Churcher TS, Cohen JM, Novotny J, Ntshalintshali N, Kunene S, Cauchemez S. 2014 Measuring the path toward malaria elimination. *Science (80-.).* **344**, 1230–1232. (doi:10.1126/science.1251449)
10. Zhou L *et al.* 2018 Clusters of Human Infection and Human-to-Human Transmission of Avian Influenza A(H7N9) Virus, 2013-2017. *Emerg. Infect. Dis.* **24**, 397–400. (doi:10.3201/eid2402.171565)
11. Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team *et al.* 2009 Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N. Engl. J. Med.* **360**, 2605–15. (doi:10.1056/NEJMoa0903810)
12. Zumla A, Hui DS, Perlman S. 2015 Middle East respiratory syndrome. *Lancet* **386**, 995–1007. (doi:10.1016/S0140-6736(15)60454-8)
13. Diekmann O, Heesterbeek JAP, Roberts MG. 2010 The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface* **7**, 873–885. (doi:10.1098/rsif.2009.0386)
14. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
15. Diekmann O, Heesterbeek JA, Metz JA. 1990 On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–82. (doi:10.1007/BF00178324)
16. Andreasen V, Lin J, Levin SA. 1997 The dynamics of cocirculating influenza strains conferring partial cross-immunity. *J. Math. Biol.* **35**, 825–42. (doi:10.1007/s002850050079)
17. Kamo M, Sasaki A. 2002 The effect of cross-immunity and seasonal forcing in a multi-strain epidemic model. *Phys. D Nonlinear Phenom.* **165**, 228–241. (doi:10.1016/S0167-2789(02)00389-5)
18. Minayev P, Ferguson N. 2009 Incorporating demographic stochasticity into multi-strain epidemic models: application to influenza A. *J. R. Soc. Interface* **6**, 989–996. (doi:10.1098/rsif.2008.0467)
19. Ferguson N, Anderson R, Gupta S. 1999 The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 790–4. (doi:10.1073/pnas.96.2.790)
20. Wikramaratna PS, Kucharski A, Gupta S, Andreasen V, McLean AR, Gog JR. 2015 Five challenges in modelling interacting strain dynamics. *Epidemics* **10**, 31–34. (doi:10.1016/j.epidem.2014.07.005)
21. Hensgens MPM, Keessen EC, Squire MM, Riley TV, Koene MGJ, de Boer E, Lipman LJA, Kuijper EJ. 2012 Clostridium difficile infection in the community: a zoonotic disease? *Clin. Microbiol. Infect.* **18**, 635–645. (doi:10.1111/j.1469-0691.2012.03853.x)
22. Johnson S, Homann SR, Bettin KM, Quick JN, Clabots CR, Peterson LR, Gerding DN. 1992 Treatment of asymptomatic Clostridium difficile carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. *Ann. Intern. Med.* **117**, 297–302.
23. Britton RA, Young VB. 2012 Interaction between the intestinal microbiota and host in Clostridium difficile colonization resistance. *Trends Microbiol.* **20**, 313–319. (doi:10.1016/j.tim.2012.04.001)
24. Voth DE, Ballard JD. 2005 Clostridium difficile Toxins: Mechanism of Action and Role in Disease. *Clin. Microbiol. Rev.* **18**, 247–263. (doi:10.1128/CMR.18.2.247-263.2005)

25. Kelly CP, Kyne L. 2011 The host immune response to Clostridium difficile. *J. Med. Microbiol.* **60**, 1070–1079. (doi:10.1099/jmm.0.030015-0)
26. Viscidi R, Laughon BE, Yolken R, Bo-Linn P, Moench T, Ryder RW, Bartlett JG. 1983 Serum Antibody Response to Toxins A and B of Clostridium difficile. *J. Infect. Dis.* **148**, 93–100. (doi:10.1093/infdis/148.1.93)
27. Loo VG *et al.* 2011 Host and Pathogen Factors for Clostridium difficile Infection and Colonization. *N. Engl. J. Med.* **365**, 1693–1703. (doi:10.1056/NEJMoa1012413)
28. Riggs MM, Sethi AK, Zabarsky TF, Eckstein EC, Jump RLP, Donskey CJ. 2007 Asymptomatic Carriers Are a Potential Source for Transmission of Epidemic and Nonepidemic Clostridium difficile Strains among Long-Term Care Facility Residents. *Clin. Infect. Dis.* **45**, 992–998. (doi:10.1086/521854)
29. Lawley TD *et al.* 2009 Antibiotic Treatment of Clostridium difficile Carrier Mice Triggers a Supershedder State, Spore-Mediated Transmission, and Severe Disease in Immunocompromised Hosts. *Infect. Immun.* **77**, 3661–3669. (doi:10.1128/IAI.00558-09)
30. Gerding DN, Lessa FC. 2015 The epidemiology of Clostridium difficile infection inside and outside health care institutions. *Infect. Dis. Clin. North Am.* **29**, 37–50. (doi:10.1016/j.idc.2014.11.004)
31. Healthcare Cost and Utilization Project (HCUP). 2014 HCUP National Inpatient Sample (NIS).
32. Furuya-Kanamori L, Marquess J, Yakob L, Riley T V, Paterson DL, Foster NF, Huber CA, Clements ACA. 2015 Asymptomatic Clostridium difficile colonization: epidemiology and clinical implications. *BMC Infect. Dis.* **15**, 516. (doi:10.1186/s12879-015-1258-4)
33. McLure A, Clements ACAA, Kirk M, Glass K. 2017 Healthcare-Associated Clostridium difficile Infections are Sustained by Disease from the Community. *Bull. Math. Biol.* **79**, 2242–2257. (doi:10.1007/s11538-017-0328-8)
34. Lanzas C, Dubberke ER, Lu Z, Reske KA, Gröhn YT. 2011 Epidemiological Model for Clostridium difficile Transmission in Healthcare Settings. *Infect. Control Hosp. Epidemiol.* **32**, 553–561. (doi:10.1086/660013)
35. Knight DR, Elliott B, Chang BJ, Perkins TT, Riley T V. 2015 Diversity and Evolution in the Genome of Clostridium difficile. *Clin. Microbiol. Rev.* **28**, 721–41. (doi:10.1128/CMR.00127-14)
36. Fawley WN, Davies KA, Morris T, Parnell P, Howe R, Wilcox MH, Clostridium difficile Ribotyping Network (CDRN) Working Group. 2016 Enhanced surveillance of Clostridium difficile infection occurring outside hospital, England, 2011 to 2013. *Euro Surveill.* **21**. (doi:10.2807/1560-7917.ES.2016.21.29.30295)
37. Gerding DN *et al.* 2015 Administration of spores of nontoxigenic Clostridium difficile strain M3 for prevention of recurrent C. difficile infection: a randomized clinical trial. *JAMA* **313**, 1719–27. (doi:10.1001/jama.2015.3725)
38. OECD. 2015 Healthcare Activities: Hospital Beds. In *Health at a Glance 2015*, pp. 104–105. Paris: OECD Publishing.
39. Australian Institute of Health and Welfare. 2010 Residential aged care in Australia 2010–11: A statistical overview. *Aged care Stat. Ser. no. 36. Cat. no. AGE 68*.
40. Australian Bureau of Statistics. 2017 Australian Demographic Statistics, 'TABLE 59. Estimated Resident Population By Single Year Of Age, Australia ', time series spreadsheet, cat. no. 3101.0.
41. Kubota H *et al.* 2016 Longitudinal Investigation of Carriage Rates, Counts, and Genotypes of Toxigenic Clostridium difficile in Early Infancy. *Appl. Environ. Microbiol.* **82**, 5806–14. (doi:10.1128/AEM.01540-16)
42. Knetsch CW *et al.* 2018 Zoonotic Transfer of Clostridium difficile Harboring Antimicrobial Resistance between Farm Animals and Humans. *J. Clin. Microbiol.* **56**, JCM.01384-17. (doi:10.1128/JCM.01384-17)
43. Rodriguez-Palacios A, Borgmann S, Kline TR, LeJeune JT. 2013 Clostridium difficile in foods

and animals: history and measures to reduce exposure. *Anim. Heal. Res. Rev.* **14**, 11–29. (doi:10.1017/S1466252312000229)
44. Furuya-Kanamori L, McKenzie SJ, Yakob L, Clark J, Paterson DL, Riley T V., Clements AC. 2015 Clostridium difficile Infection Seasonality: Patterns across Hemispheres and Continents – A Systematic Review. *PLoS One* **10**, e0120730. (doi:10.1371/journal.pone.0120730)
45. Rodriguez-Palacios A *et al.* 2009 Possible Seasonality of Clostridium difficile in Retail Meat, Canada. *Emerg. Infect. Dis.* **15**, 802–805. (doi:10.3201/eid1505.081084)

Figures

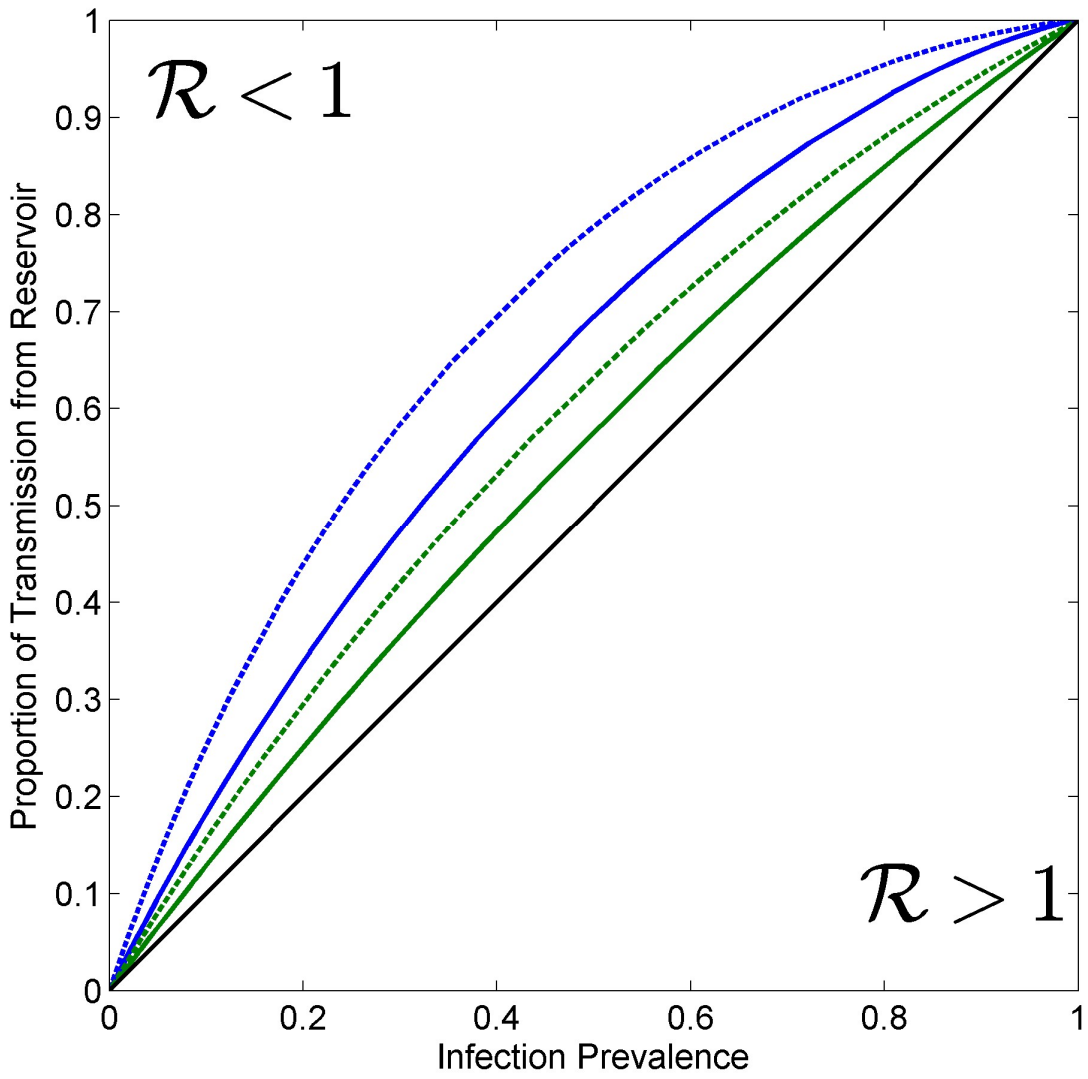


Figure 1 The reservoir-driven threshold (RDT) – the minimum proportion of transmission attributable to the reservoir above which the basic reproduction number is <1 – as a function of disease prevalence. Each curve indicates the RDT for different population heterogeneity assumptions for infectiousness ($\boldsymbol{\beta}$) and the product of susceptibility and infectious period ($\boldsymbol{\phi} \coloneqq \boldsymbol{a}/\boldsymbol{\gamma}$). The RDT for a homogenous population is equal to the disease prevalence (black line). Heterogeneous $\boldsymbol{\beta}$ alone does not change the RDT (black line). The RDT is higher if $\boldsymbol{\phi}$ heterogeneous and $\boldsymbol{\beta}$ homogenous (solid curves). The size of the effect increases with increasing heterogeneity (green curves: $\boldsymbol{\phi} \sim \boldsymbol{\Gamma}(\mathbf{3}, \boldsymbol{\mu})$, blue curves: $\boldsymbol{\phi} \sim \boldsymbol{\Gamma}(\mathbf{1}, \boldsymbol{\mu})$). Heterogeneity in $\boldsymbol{\beta}$ interacts with heterogeneity in $\boldsymbol{\phi}$, further increasing the RDT if $\boldsymbol{\beta} \propto \boldsymbol{\phi}$ (dashed curves) but decreasing the RDT if $\boldsymbol{\beta} \propto \mathbf{1}/\boldsymbol{\phi}$ (black line).

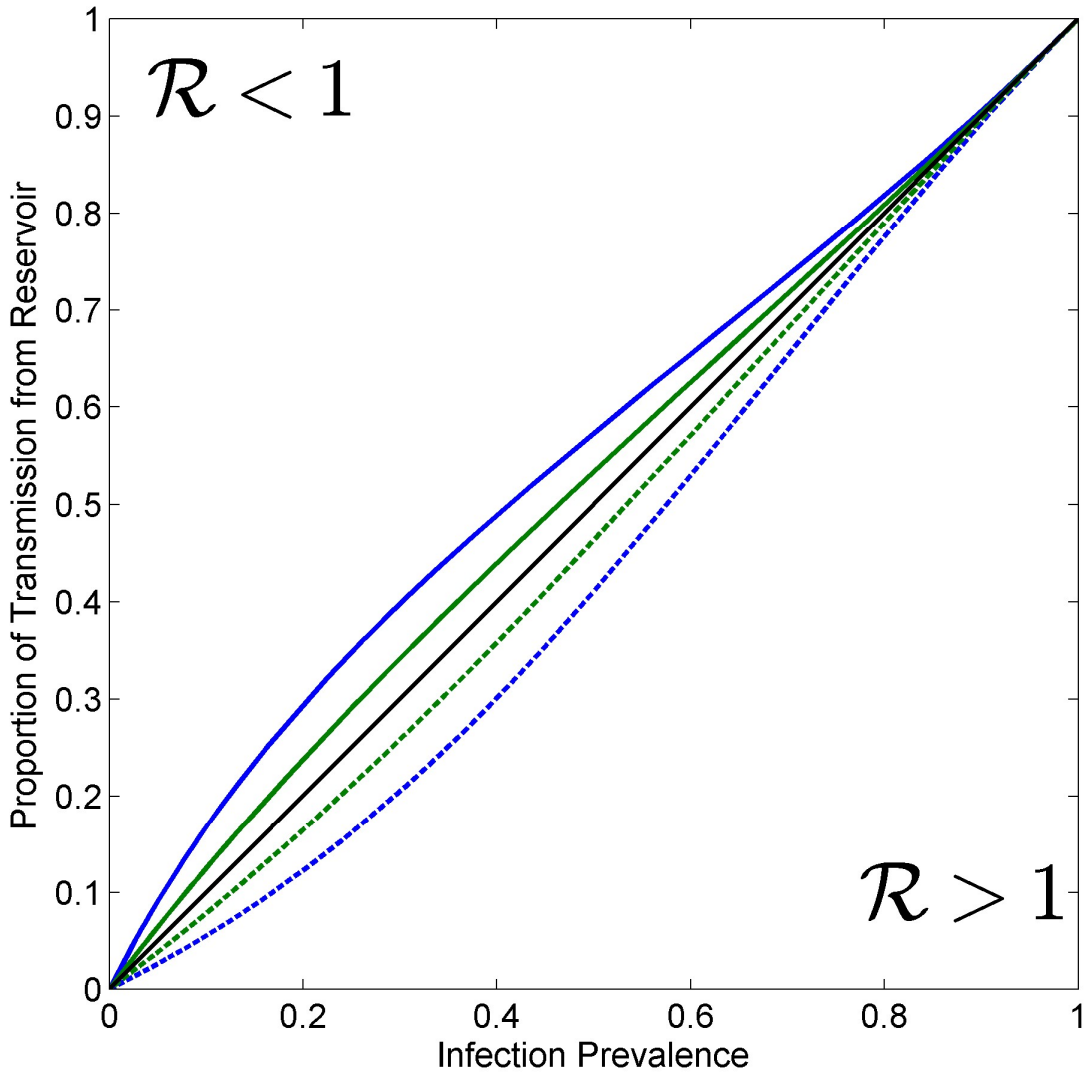


Figure 2 The reservoir-driven threshold (RDT) for different assumptions for heterogeneity of reservoir exposure ($\boldsymbol{f}$) and person-to-person transmission ($\boldsymbol{\beta}$) across the population. The RDT for a homogenous population is equal to the disease prevalence (black line). The RDT does not change if only $\boldsymbol{f}$ or only $\boldsymbol{\beta}$ is heterogeneous (black line). The RDT is lower if both are heterogeneous and $\boldsymbol{\beta} \propto \mathbf{f}$ (dashed curves). The RDT is higher if $\boldsymbol{\beta}$ decreases with increasing $\boldsymbol{f}$ (solid curves: $\boldsymbol{\beta} \propto \boldsymbol{e}^{-f}$). The size of the effect increases with increasing heterogeneity (green curves: $\boldsymbol{f} \sim \boldsymbol{\Gamma}(\mathbf{3}, \boldsymbol{\mu})$, blue curves: $\boldsymbol{f} \sim \boldsymbol{\Gamma}(\mathbf{1}, \boldsymbol{\mu})$).

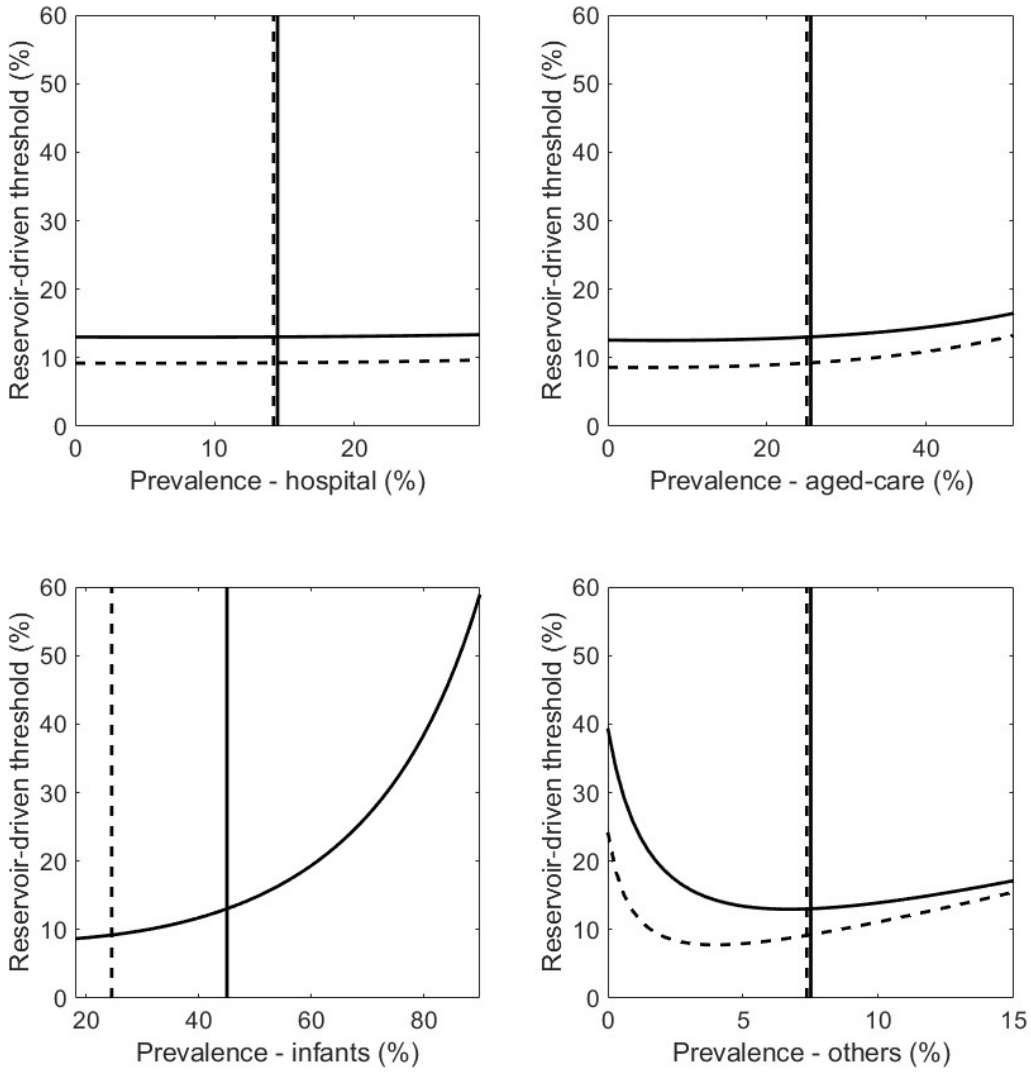


Figure 3 Estimates of the reservoir-driven threshold for C. difficile in human populations and its dependence on the prevalence of each of four risk groups. In each subfigure, the prevalence in one risk group is varied across the reported range [32] (x-axes) while the other three prevalences are fixed at the values indicated by the vertical lines in the other subfigures. We consider two scenarios; one where each of the fixed prevalences is assumed to be in the middle of the reported range (solid lines and curves); the other the same except the prevalence in infants is only 25% (dotted lines and curves). We assume that 0.5%, 1%, 1.5% and 97% of the population are in the hospital, aged-care, infant and 'other' risk groups respectively.

Chapter 8

Discussion

The aim of this research was to use mathematical models to further the understanding of the transmission and epidemiology of *C. difficile* in hospitals and communities. I began by developing a compartmental model of *C. difficile* carriage and infection in a hospitalised population (Chapter 3). This model was used to assess common definitions for the classification of hospital-onset infections as hospital or previously-acquired (Chapter 4). The modelling framework was extended to a two-patch model of hospital and community and adapted to include infants, demographics, and the potential for transmission from animals (Chapter 5). This extended model became the basis of an investigation into the seasonality of *C. difficile* and hospital and community-based interventions that could reduce disease burden (Chapter 6). Finally, I developed an independent method to estimate the local reproduction number in settings with both local transmission and infections acquired from an external source or imported into the population. This method corroborated the findings of Chapters 3 to 6 using simple arguments to demonstrate that transmission within hospitals is insufficient to sustain endemic disease and that transmission from animal reservoirs may sustain disease in humans if it exceeds a low threshold.

This discussion chapter summarises the contributions of this thesis to the *C. difficile* and mathematical modelling literature; considers some of the key strengths, weaknesses, and implications of my work; and highlights areas which require further research. The discussion begins with three sections that survey the key research themes that extend throughout the thesis.

8.1 Theme 1 – Improving mathematical models of *C. difficile*

Chapters 3 to 6 of this thesis present bottom-up, multi-dimensional models of *C. difficile* transmission. The models in this thesis are multi-dimensional in that they capture the possible combinations and interactions of many factors: *C. difficile*, gut flora status, immunity, age (infant *vs.* adult *vs.* elderly), and setting (hospital *vs.* community). This contrasts with many existing models, which omit key factors (*e.g.* immunity [96, 97, 99], host gut flora [97] or community [93–106]) or consider only limited combinations of these factors (*e.g.* many models assume that only those with disrupted gut flora can be colonised [98, 104, 106]). The models are bottom-up in that the important epidemiological states and phenomena (*e.g.* infection, asymptomatic colonisation, and recurrent infection)

are emergent from the underlying physiological factors (immunity, gut flora, and pathogen) or demographic attributes (age and setting) encoded into the model. This contrasts with many of the existing models of *C. difficile* where the population is divided into compartments to correspond directly with the epidemiological states or phenomena (*e.g.* infection, recurrence, and asymptomatic carriage). I will demonstrate some of the advantages of my approach by way of comparison to other published models.

In contrast to to my bottom-up approach, Durham *et al.* modelled the recurrence of CDI by including a series of six classes differentiating those patients currently having an initial CDI, a first recurrence or an additional recurrence of CDI and those patients who have just recovered from an initial CDI, a first recurrence of a CDI or an additional recurrence of CDI [92]. Using these additional compartments the authors were able to accurately reproduce the observation that the probability of recurrence is high and increases with subsequent recurrences [113]. However, the base model presented in Chapter 3 captured the same phenomena by accounting for the underlying physiological factors: incomplete clearance of *C. difficile* following treatment and the development of immunity to toxins following exposure. Moreover, when I extended the model to the community, the model reproduced the observed difference in recurrence probability between hospital-onset and community-onset CDIs without additional fitting to this outcome (Chapter 5).

Another example of the bottom-up approach can be seen in the different outcomes that can follow asymptomatic carriage. Because the models in Chapters 3 to 6 are multi-dimensional, there are many combinations of factors that manifest as asymptomatic carriage. For instance, an individual could be asymptomatic in these models because they have intact gut flora or are immune to the toxins. These individuals will eventually clear their colonisation and will not develop symptoms unless the relevant protective factors are removed. In the same models, patients that have recently been exposed but lack protection from either immunity or intact gut flora will develop symptoms after an asymptomatic (latent) period. In contrast, in many existing models, asymptomatic carriage is only ever a precursor to symptomatic disease (*e.g.* [95, 96, 99, 100, 102]). Some models do not model the protective effect of intact gut flora in asymptomatically colonised patients because they assume that only those with disrupted gut-flora can be colonised [95, 98, 103, 104, 106, 107]. Many models that do capture the protective effect of immunity do not have compartments for immune, *C. difficile* negative individuals so assume immune and asymptomatically colonised individuals do not clear their colonisation [98, 103, 104, 106, 107].

One contribution of this thesis to the *C. difficile* modelling literature is the inclusion of dynamic and persistent immunity. Chapter 3 introduced immunity that can develop following exposure to *C. difficile*, in agreement with the much reduced likelihood of initial and recurrent infection in those with robust immune response to toxins A and B [36, 45]. This addition improved the biological realism of the model and reproduced the observed epidemiology of recurrent infections. Moreover, accounting for the large proportion of individuals immune to infection but not *C. difficile* colonisation, reduced the population at

risk of infection and helped reconcile the relatively high prevalence of colonisation and low prevalence of infection. Other authors have accounted for this by assuming that patients colonised before hospital admission develop symptomatic disease much more slowly than those exposed in hospitals [96, 102] or that CDI develops from colonisation at different rates depending on age, comorbidities or setting (hospital *vs.* community) [92]. Others reduced the population at risk by assuming all patients under 65 years of age are completely immune to colonisation and infection [97]. Chapter 5 introduces two processes that can remove immunity: gradual waning of immunity in healthy individuals and immunosenesence associated with ageing. Using a bottom-up approach, these dynamic processes reproduced the age profile of seroprevalence which increases with age amongst children but levels off below 100% for adults. The dynamic nature of immunity has been omitted from hospital-based models because the time-scale of hospital stays (days or weeks) is short compared to demographic processes and the probable duration of immunity (decades) [98]. The few existing hospital models that have incorporated some aspects of immunity only track immunity amongst colonised or vaccinated individuals and therefore cannot be easily translated to a whole-population model.

The model framework presented in Chapter 3 was extended to include adults in hospitals and the community. The required changes were minimal because the multi-dimensional, bottom-up approach used in the base model captured the main factors of *C. difficile* epidemiology relevant to both settings. Consistent with this approach, the differences between hospital and community emerged by emulating hospital admission and discharge rates for elderly and non-elderly adults [114] (which led to an over-representation of elderly or immune suppressed individuals in hospitals) and the antibiotic prescription rates in each setting [115, 116] (which led to a high proportion of patients with gut flora disruption in hospitals). Though it was assumed that advanced contact precautions for patients with CDI only occurred in the hospital, the per-person transmission rates were otherwise assumed to be the same in the hospital and the community. Nevertheless, the above emergent differences between settings led the model to reproduce the incidence of hospital and community-acquired CDIs once differences in treatment seeking, reporting and misclassification of cases were taken into account. This approach contrasts with the only other explicit model of hospital and community, where the transmission rates and the rate of progression from asymptomatic to symptomatic colonisation had to be much higher in hospital than in the community to account for the differences between the two settings [92]. Though the incidence of hospital and community-acquired infections was used to fit model parameters, other epidemiological phenomena arose naturally from the model. The recurrence proportion, colonisation prevalence and proportion of infections amongst elderly patients were all higher for hospitalised individuals, in agreement with empirical studies [13, 32, 117]. The extension of the model to the community allowed an assessment of hospital and community-based interventions on hospital and community-acquired infections. The whole-population model demonstrated that *C. difficile* infections in hospitals are more strongly dependent on transmission in the community than previ-

ously thought, a theme which is discussed at further in the next section.

The models in this thesis are the first to capture the role of infants. This is a major omission of the modelling literature as infants have colonisation prevalence that far exceeds that of adults whether in the hospital, long-term-care facility or the broader community [13, 33, 35]. Modelling the role of infants has been hampered by the paucity of studies assessing infants as a source of transmission. Though the spore shedding rate of infants has been shown to be high [35], there are very few studies that consider exposure to asymptomatically colonised infants as a risk factor for colonisation or infection amongst adults or older children [118–121]. Using conservative assumptions and broad sensitivity analysis for the infectiousness of asymptomatically colonised infants, I demonstrated that infants could reasonably account for a large portion – perhaps even the majority – of transmission in the community (Chapter 5). Consequently, under default assumptions, reducing transmission from infants by 50% or more would be sufficient to interrupt person-to-person transmission in the community in the absence of external sources (Chapter 6).

To my knowledge, the models in this thesis are the first mechanistic models of *C. difficile* transmission to incorporate seasonality. *C. difficile* infections are moderately seasonal, peaking in late winter in temperate northern-hemisphere countries [52]. Though CDI incidence has been shown to correlate with seasonal antibiotic prescription rates, rainfall, and the incidence of influenza, respiratory syncytial virus, and norovirus infections, the mechanisms that cause *C. difficile* seasonality are uncertain [54–58]. The model in Chapter 6 suggests that seasonal variation in antibiotic prescriptions would cause a different pattern of *C. difficile* seasonality than seasonal variation in pathogen transmissibility or host susceptibility. Both mechanisms could explain the seasonal variation in infection incidence, but seasonal antibiotic prescriptions would create much less variation in colonisation prevalence, especially in the community. Therefore studying colonisation prevalence seasonality in the community could help determine whether transmissibility or susceptibility are seasonal and help identify potential environmental, social or biological mechanisms.

8.2 Theme 2 – Populations that sustain *C. difficile*

Hospitals cannot sustain *C. difficile* on their own, but require importation from the community to sustain disease. This fact was demonstrated using different models and arguments in Chapters 3, 5 and 7. In Chapter 3 a broad sensitivity analysis of a model of a hospital population with admissions and discharges showed that the reproduction number was significantly less than one for nearly all reasonable scenarios. Only hospitals or wards with exceptionally long mean length of stay and/or very high transmission rates could have a reproduction number in excess of one. In Chapter 5 an integrated model of hospital and community again found that the reproduction number in the hospital was less than one (in fact < 0.5) for a broad sensitivity analysis in which the colonisation prevalence in the community and the relative infectiousness of adults and infants were varied. The general rules

developed in Chapter 7 to estimate reproduction numbers in populations with both locally acquired and imported cases was applied to the case of *C. difficile* in hospitals. Though the simple rules did not account for the complex interactions of immunity, gut flora, and pathogen that were incorporated into the models in Chapters 3 and 5, the simple rules came to the same conclusion. An extension of these rules derived to account for multi-strain competition were applied to the 'epidemic' *C. difficile* strain NAP1/RT027. This demonstrated that the within-hospital reproduction number was higher for NAP1/RT027 than for other strains, but still less than one. It must be noted that the latter calculations did not account for heterogeneity in colonisation and infection risk and therefore may have underestimated the reproduction numbers.

Chapter 6 argues that reducing transmission in the community by as little as 7-27% could be sufficient to eliminate the disease in the entire population including the hospital. This again demonstrates that hospitals are unable to sustain *C. difficile* transmission without the admission of patients colonised in the community – within-hospital transmission and the readmission of patients previously colonised in the hospital are not enough to sustain transmission in hospitals. On the other hand, communities do not require transmission in hospitals to sustain *C. difficile*. When reservoirs external to the human population were assumed not to contribute to human colonisation and infection, the community-only reproduction number was greater than one for nearly all scenarios considered in a broad sensitivity analysis (Chapter 5). In the interventions analysis of Chapter 6, even complete elimination of transmission in the hospital reduced the incidence of infections in the community by less than 20%.

It is important to note that I argue that hospitals are probably not disease sustaining populations, but maintain that hospitals nevertheless amplify the burden of *C. difficile* in the general population. The models in Chapters 3 and 5 found that the high antibiotic prescription rate and transmission rate in hospitals create an environment with a high force of colonisation, such that the prevalence of colonisation is higher at discharge than at admission, and higher amongst hospitalised patients than adults in the general population. In other words, the models presented in this thesis do not contradict the empirical evidence that those who are currently or have recently been hospitalised are at higher risk of colonisation and infection [13, 42]. Instead this thesis challenges the prevailing paradigm of *C. difficile* as a *predominantly* or *essentially* hospital-acquired pathogen.

There is good reason to believe that zoonotic transmission plays an important role in the epidemiology of *C. difficile*. Others have shown that there is significant overlap in the types of *C. difficile* colonising humans and animals and that these types of *C. difficile* are also present on retail meats and vegetables [19, 122]. Moreover whole genome sequencing of RT078 *C. difficile* from animals and humans suggest frequent transmission between the two populations including the transfer of antibiotic resistance traits [66]. The lack of any estimate of the proportion of human cases attributable to animal reservoirs or even confirmed cases of *C. difficile* acquired from animals has hampered an accurate modelling assessment of zoonotic transmission in this thesis. However, it has been possible to es-

timate threshold proportions of transmission that, if exceeded, imply that *C. difficile* in the human population is driven by exposure to an animal reservoir (*i.e.* requires continual transmission from animals to be sustained). Chapter 7 uses arguments analogous to those used to estimate the within-hospital reproduction number to estimate the animal-driven threshold to be approximately 13% of human infections. Though the model used to derive this estimate is simple and general, it begins to account for some of the heterogeneity in colonisation prevalence observed in infants, long-term care facilities, and hospitals. The model in Chapter 5 is a much more sophisticated model of *C. difficile* transmission but arrived at a similar estimate of the animal-driven threshold that ranged from 3.5% to 26% across scenarios. This was equivalent to, on average, one exposure to *C. difficile* from an animal source leading to colonisation per person every 9.4-175.5 years. In comparison, Australians have on average an episode of food-borne gastroenteritis (*i.e.* not counting asymptomatic exposure) once every five years [123]. Given the relatively high prevalence of *C. difficile* in meat and produce [124], it is plausible that these thresholds are exceeded. More research is required to determine with confidence whether particular strains of *C. difficile* (*e.g.* RT078) are animal-driven. If so, eliminating these strains in animals or preventing transmission from animals to humans would eradicate these animal-driven strains in the human population.

These findings from Chapters 3, 5, 6 and 7 show that we must look beyond the hospital if we are to understand and control the spread of *C. difficile*. They demonstrate that interventions to reduce within-hospital transmission will never be enough to prevent all hospital-onset CDIs let alone community-onset CDI. Since colonised admissions are essential for the continued presence of CDI in hospitals, these finding support interventions that reduce transmission from those colonised prior to hospital admission and the use of novel interventions to reduce transmission in the community. Longtin and colleagues screened all hospital admissions to identify asymptomatically colonised patients, who were then treated with additional contact precautions [27]. The incidence of hospital-onset infections decreased dramatically following this intervention. It is difficult to determine whether this reduction was due to reduced transmission from patients admitted with colonisation or because identifying asymptomatically colonised patients reduced the probability of progressing to symptomatic infection. Either way, this intervention demonstrated the efficacy of interventions that address the large proportion of patients colonised prior to admission. The UK saw a dramatic 80% decrease in hospital and community-onset CDI incidence following the introduction of national interventions [51]. It has been argued that the reduction was due primarily to falling fluoroquinolone prescriptions in the hospital and the community rather than reduced within-hospital transmission [51], again supporting interventions which go beyond reducing hospital-based transmission.

The methods used and developed in this thesis have broad applications for determining disease-sustaining populations for infections other than CDI. Chiefly, the simple rules developed in Chapter 7 could be used to estimate the within-hospital reproduction numbers for other hospital-acquired infections with significant importation of cases such as

methicillin-resistant *Staphylococcus aureus* [125]. Similarly, in the context of disease elimination programs, a significant proportion of infections in a country or region are often imported from other countries or regions, and this proportion is often monitored and reported. If coupled with estimates of infection prevalence, these could be used to estimate the local reproduction number and establish whether local transmission would persist in the absence of imported infections.

8.3 Theme 3 – Classifying *C. difficile* infections

Classification of *C. difficile* infections as hospital or community-acquired typically uses a variant of surveillance definitions endorsed by the Infectious Diseases Society of America (IDSA) and the Society for Hospital Epidemiology of America (SHEA) [23], which are based on interim definitions proposed in 2007 [25]. The most up-to-date recommendations by IDSA and SHEA have not updated the classification schemes [49]. The authors recommending the surveillance definitions acknowledged that they were not based on evidence but hoped standardised definitions would make for consistent comparisons over time and between hospitals [23]. To my knowledge, the articles contained in this thesis are the first to assess this system of classification.

In Chapters 4 and 5, I assessed two of the three components of the classification system. I did not assess the third component, the recommended definition of a recurrent infection (a positive assay in the eight weeks prior to infection). The remaining components of the recommended definitions (and its variants) classify non-recurrent infections as hospital or community-acquired using a cut-off for time since hospital admission and a much longer cut-off for time since hospital discharge (summarised in Figure 2, Chapter 5). I have demonstrated that the commonly used cut-offs for recent hospital admission lead to gross overestimation of the proportion of cases that are hospital acquired (Chapters 4 and 5). Many patients that acquire *C. difficile* prior to hospital admission only begin to have symptoms after the cut-off (usually two or three days after admission [6, 8, 23]). Though a small proportion of these patients will have acquired the pathogen in a different healthcare setting, the majority have not and are thus incorrectly classified as hospital-acquired. On the other hand, the same cut-off misclassifies very few community-acquired cases as hospital-acquired. A longer cut-off of approximately five or six days better balances the sensitivity, specificity, and precision, reducing both the number of cases misclassified and the overestimation of the proportion of cases that are hospital-acquired.

The cut-off for recent hospital discharge is unable to adequately discriminate between hospital and community-acquired cases for any choice of cut-off. The problem lies in the fact that *any* hospital exposure in the cut-off period prior to onset of symptoms, whatever the duration, results in a classification as hospital-acquired. However, in my model that captured movement between hospital and community (Chapter 5), the majority of patients colonised at discharge had acquired the colonisation prior to hospital admission and had remained colonised for the duration of their hospital stay. This is possible as the

proportion of patients colonised at admission is high and the mean hospital length of stay is approximately five days in many high-income countries [126] – much shorter than the duration of colonisation of approximately one month [30]. Consequently, even infections that develop immediately after admission are more likely to be community-acquired than hospital-acquired.

The overall effect of the recommended classification system and its variants is to systematically overestimate the proportion of cases that are hospital-acquired and overemphasise the importance of transmission within hospitals (Chapters 4 and 5). When assessing interventions to reduce hospital-based transmission, the classification system may lead researchers to underestimate the proportion of hospital-acquired infections averted, since many of the cases thought to be preventable with better hospital hygiene (*i.e.* cases classified as hospital-acquired) may in fact have been community-acquired (Chapter 6).

Adjusting the classification cut-off for recent hospital admission to five or six days will improve classification of hospital-onset cases. Though the specificity of a given cut-off to identify hospital-acquired infections is largely independent of the extent of within-hospital transmission, a given cut-off will have worse precision in hospitals with less within-hospital transmission and thus further over-estimate the proportion of cases that are hospital-acquired when compared with hospitals with more within-hospital transmission (Chapter 4). This is particularly problematic if classification systems of this type are used to compare hospitals, as the misclassification error will be most acute for hospitals with lower within-hospital transmission (Chapter 4).

No simple adjustment of the cut-off for recent hospital discharge adequately fixes the deficiencies of the classification system as long as brief healthcare exposures within the cut-off period results in a classification as hospital-acquired (Chapter 5). Though this thesis has not evaluated such a system, a classification system that accounts for the total duration of hospital or healthcare exposures in the weeks leading up to onset of symptoms may be able to better distinguish hospital and community-acquired cases. Determining a classification scheme that is simple, easy to implement, robust to differences in setting, and adequately distinguishes hospital and community-acquired cases remains an open challenge requiring further research from epidemiologists and mathematical modellers.

The models in this thesis accounted for classification error by emulating the classification system during the model-fitting process (Chapters 3 and 5). Some authors have not accounted for misclassification and therefore have used inflated estimates of hospital-acquired CDIs to fit or validate their models (*e.g.* [93, 99]). This may have lead to potentially significant distortions in model outcomes. Other authors have circumvented this issue by fitting their models to the incidence of hospital-onset infections (*e.g.* [102]) or to the incidence of infections stratified by location of onset (*e.g.* [92]) rather than the putative location of acquisition. Since symptom onset is more easily observed than pathogen acquisition, these data are likely to be more accurate. However, the reporting of location of acquisition may also be compromised. IDSA and SHEA recommend that CDI cases with onset of symptoms up to 48 hours after hospital admission who have had a hospital

discharge in the previous 4 weeks should be classified as 'community-onset, healthcare-facility associated disease' [23]. That is, it is recommended that some cases who were in hospital at the time of symptom onset should be classified as community-onset infections.

Many common hospital-associated infections such as methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococci [125, 127, 128] are classified as hospital or community-acquired with classification schemes similar to those commonly used for *C. difficile*. Like *C. difficile*, carriage of these pathogens is often asymptomatic and occurs both in hospitals and in the community. It is plausible that the cut-offs used for these infections also lead to systematic misclassification of the place of acquisition. Mathematical models may be appropriate tools for informing classification schemes for these pathogens.

8.4 Limitations

The main difficulties and limitations of the research in this thesis stemmed from a lack of data. The complex, mechanistic models in Chapters 3 to 6 required information on many individual-level, hospital-level or community-level variables and outcomes, including asymptomatic colonisation, symptomatic infections, recurrences, hospital length of stay, antibiotic prescription rates, and immunity to toxins. I did not have access to a dataset with all these variables for a single population let alone for individuals within a single population – such a dataset may not exist for any community. Instead my approach, common in mathematical modelling, has been to draw on the best available research to identify the value of parameters that are likely to be similar across all populations (such as mean time to *C. difficile* overgrowth or to development of immunity), infer the typical value of other parameters from population-level data (such as hospital length of stay and antibiotic prescription rates), and fit remaining parameters (*e.g.* transmission parameters) indirectly to other observations such as incidence and prevalence estimates. When parameter values or assumptions were particularly uncertain, influential or variable between settings, extensive sensitivity analysis was used to explore the effect of these parameters and assumptions. Therefore, while the models in Chapters 3 to 6 are unlikely to be accurate representations of any single population, they are expected to reflect general trends and the breadth of epidemiology. The much simpler modelling framework in Chapter 7 avoided these problems by omitting much of the complexity of CDI, limiting its usefulness the estimation of the reproduction number and animal-driven thresholds.

It was difficult to obtain accurate estimates of the incidence or prevalence of hospital and community-acquired infections and colonisations, which were essential to estimate the contributions of hospital-based and community-based transmission. The available estimates of hospital-acquired and community-acquired infection incidence use indirect means to classify infections as hospital or community-acquired (*e.g.* [6, 8]). Moreover, it is likely that community-onset *C. difficile* infections – like other gastrointestinal diseases – go unreported and so the incidence of community-onset infections is underestimated [129–132].

Accounting for the bias introduced by indirect classification of the location of acquisition is a major theme of this thesis (Theme 3). However, the ability of the models to account for this bias is dependent on the accuracy with which the model captures movement between healthcare facilities and community and the timing of infections relative to these events. While extensive sensitivity analysis was used to assess the robustness of these findings (Chapter 4), inaccuracies in these parts of the models may have introduced their own biases. In particular, the model of hospital admissions and discharges was highly simplified. Though the hospital admission rates differed by patient type (dependent on CDI status, age and immune state), the model did not account for transfers between hospitals, the higher rate of hospital admission amongst recently discharged patients and other heterogeneities [133–135] which would tend to recirculate some patients through the hospital system much more frequently than others. This may have affected the assessment of the classification system and underestimated the ability of *C. difficile* to persist in hospitals.

A lack of data that could be used to infer the role of infants in the transmission of *C. difficile* was another limitation of the thesis. In the absence of firm estimates of infant infectiousness, Chapters 5 and 6 used broad sensitivity analysis for the relative infectiousness of infants and adults. Though these chapters demonstrated that infants are likely to be an important source of transmission in the community, the lack of data around infant infectiousness introduced a great deal of uncertainty into a number of model outcomes including the proportion of transmission in the community that is from infants or asymptomatic adults (Chapter 5), the effect of reducing transmission from infants (Chapter 6), and the value of the food-driven threshold (Chapter 5). Infants and transmission from infants were only modelled in the community. This is a reasonable simplification, as hospitalised infants often receive treatment in dedicated wards and so probably do not constitute a substantial transmission risk for hospitalised adults. Moreover, *C. difficile* rarely causes disease in infants, with recommended surveillance definitions specifically excluding infants [23, 49]. Therefore the omission of infants from the hospital sub-model is unlikely to have interfered with the comparison of the simulated and reported incidence of hospital and community-acquired infections.

C. difficile is common in livestock and pets and has been isolated in produce and water [19]. However, because the minimum infectious dose (or minimum colonising dose) is unknown [10], it is difficult to determine how often these sources lead to infection (or colonisation). This in turn has made it difficult to develop models that account for transmission from both humans and animals. For this reason, many of the results in this thesis – such as the impact of reducing person-to-person transmission rates (Chapter 6) or the estimates of reproduction number Chapter 5 – had to be calculated assuming *no* transmission from animals. Consequently, the estimates of the reproduction numbers in Chapter 5 are really estimates of upper bounds for person-to-person reproduction numbers. Arguments introduced in Chapter 5 and further developed in Chapter 7 were used to estimate the minimum frequency of transmission from animal sources that would imply that transmission from animal reservoirs drives human disease. Though it is not *implausible* that

transmission from animals exceeds this threshold, this thesis does not provide a way to determine whether it is *probable*, nor does it estimate the effect of preventing transmission from animals.

The initial model of *C. difficile* in this thesis captured only events inside hospitals (Chapters 3 and 4). Subsequent models incorporated transmission in the community but did not explicitly model transmission in long-term care facilities or the potential for exposure through outpatient care (Chapters 5 and 6). The two settings in the latter models (hospital and community) were assumed to be homogenous and well-mixed. However, since the elderly are at higher risk of infection and infants have very high colonisation rates, age-dependent mixing is likely to have a significant impact in the community [136]. Other authors have modelled multiple wards within hospitals or the contact networks of patients and healthcare workers [95–97]. In general, accounting for population heterogeneity and non-random mixing increases estimates of the reproduction number. Consequently, the estimates of the reproduction number and the effort required to interrupt transmission may have been somewhat underestimated in this thesis.

C. difficile has numerous strains with different toxin profiles, antibiotic susceptibilities, and epidemiology [137]. However, most of the models in this thesis (and most models in the literature) are single-strain models. There is some evidence that strains differ in their relative frequency of isolation between adults and infants [18], hospitals and communities [138], and humans and animals [122]. Whole genome sequencing of European *C. difficile* has identified two distinct transmission patterns, with genetically related isolates of some (predominantly fluoroquinolone-resistant) strains clustering locally or regionally, but isolates from other (predominantly fluoroquinolone-susceptible) strains sharing close genetic relationships across long distances [139]. By pooling all toxigenic strains, the models in Chapters 3 to 6 have overlooked this variability. Thus some of the key findings in this thesis – such as estimates of hospital and community reproduction numbers, animal-driven thresholds and the efficacy of hospital and community-based interventions – may have quantitatively and qualitatively different true values for individual strains. This is demonstrated in Chapter 7 where the simple framework was extended to a strain-by-strain analysis. Using this extension I concluded that in the study hospital ([15]) ribotype 027 had a higher reproduction number than other types.

This thesis relies on the threshold property of the basic reproduction number to argue that transmission in hospitals is insufficient to maintain endemicity in hospitals and calculate the animal-driven threshold. However, the threshold property of the basic reproduction number can be blurred by a backward bifurcation if the depletion of susceptibles is balanced by a mechanism that increases population susceptibility or pathogen transmissibility with increasing prevalence [86, 87]. When a mechanism of this kind exists, certain parameter combinations allow both the disease-free equilibrium and the endemic equilibrium to be locally stable if the reproduction number is less than but close to one. In such a scenario, the introduction of a small number invectives is unlikely to lead to endemic disease, but reducing the reproduction number to less than one is not a sufficient

criterion for eliminating endemicity in the absence of importation [86, 87]. *C. difficile* has a potential mechanism that may lead to a backward bifurcation; namely, the antibiotic treatment of symptomatic patients increases the number of people in the population who have disrupted gut flora and who are therefore at higher risk of subsequent infection and long-lasting colonisation. However, formal bifurcation analyses were not performed for any of the models in this thesis, so the basic reproduction numbers and animal-driven thresholds *may* not be true thresholds for disease endemicity. However, I believe this is unlikely for two reasons. First, even in populations with very high incidence of *C. difficile* infections, the antibiotic treatment of these infections only accounts for a small minority of antibiotic prescriptions. Second, if a backward bifurcation was operating for the parameter values used, one would expect to observe an abrupt (discontinuous) decrease in prevalence and incidence once the bifurcation parameter was reduced to below a critical threshold. This was not observed for either the transmission parameters or the antibiotic prescription rate parameters in the analyses in Chapter 6.

8.5 Conclusion

Clostridium difficile infections are a major concern and their burden may continue to grow as the global population ages. In this thesis I argue that we need to adopt a holistic, integrated view of *C. difficile* transmission that considers people of all ages in healthcare facilities and the community and the role of animals. Though many shortcomings and limitations remain, this thesis advances the mathematical modelling of *C. difficile* transmission. The findings from these models have applications for the design of improved standards for the classification of *C. difficile* infections as hospital or community-acquired, the design of interventions to reduce *C. difficile* infection and colonisation, and the identification of settings where *C. difficile* is sustained by transmission within the local population. The identified differences between antibiotic-driven and transmission-driven seasonality provide means to discern which mechanism is responsible for CDI seasonality. The methods I have developed and advanced have applications not only for *C. difficile* but many hospital-acquired, zoonotic, and travel-associated infections.

Bibliography

[1] Shelley S. Magill, Jonathan R. Edwards, Wendy Bamberg, et al. Multistate Point-Prevalence Survey of Health Care–Associated Infections. *New England Journal of Medicine*, 370(13):1198–1208, 2014.

[2] L Fernandez Canigia, J Nazar, M Arce, et al. Clostridium difficile diarrhea: frequency of detection in a medical center in Buenos Aires, Argentina. *Revista Argentina de microbiologia*, 33(2):101–7, 2001.

[3] Coralith Garcia, Frine Samalvides, Margot Vidal, Eduardo Gotuzzo, and Herbert L Dupont. Epidemiology of Clostridium difficile-associated diarrhea in a Peruvian tertiary care hospital. *The American journal of tropical medicine and hygiene*, 77(5):802–5, 2007.

[4] B. Dhawan, R. Chaudhry, and N. Sharma. Incidence of Clostridium difficile infection: a prospective study in an Indian hospital. *Journal of Hospital Infection*, 43(4):275–280, 1999.

[5] N Sadeghifard, MH Salari, MR Ghassemi, et al. Prevalence of Clostridium Difficile-Associated Diarrhea in Hospitalized Patients with Nosocomial Diarrhea. *Iranian Journal of Public Health*, 34(4):67–72, 2005.

[6] Claudia Slimings, Paul Armstrong, Wendy D Beckingham, et al. Increasing incidence of Clostridium difficile infection, Australia, 2011–2012. *The Medical Journal of Australia*, 200(5):272–276, 2014.

[7] Kristin E. Burke and J. Thomas Lamont. Clostridium difficile Infection: A Worldwide Disease. *Gut and Liver*, 8(1):1–6, 2014.

[8] Fernanda C Lessa, Yi Mu, Wendy M Bamberg, et al. Burden of Clostridium difficile infection in the United States. *The New England Journal of Medicine*, 372(9):825–34, 2015.

[9] Erik R. Dubberke and Margaret A. Olsen. Burden of Clostridium difficile on the healthcare system. *Clinical Infectious Diseases*, 55(Suppl 2):S88–92, 2012.

[10] Lauren E. Bloomfield and Thomas V. Riley. Epidemiology and Risk Factors for Community-Associated Clostridium difficile Infection: A Narrative Review. *Infectious diseases and therapy*, 5(3):231–51, 2016.

[11] Joyanna M. Wendt, Jessica A. Cohen, Yi Mu, et al. Clostridium difficile infection among children across diverse US geographic locations. *Pediatrics*, 133(4):651–8, 2014.

[12] Daniel A. Leffler and J. Thomas Lamont. Treatment of Clostridium difficile-associated disease. *Gastroenterology*, 136(6):1899–912, 2009.

[13] Luis Furuya-Kanamori, John Marquess, Laith Yakob, et al. Asymptomatic Clostridium difficile colonization: epidemiology and clinical implications. *BMC Infectious Diseases*, 15(1):516, 2015.

[14] Robert A Britton and Vincent B Young. Interaction between the intestinal microbiota and host in Clostridium difficile colonization resistance. *Trends in microbiology*, 20(7):313–9, 2012.

[15] Vivian G. Loo, Anne-Marie Bourgault, Louise Poirier, et al. Host and Pathogen Factors for Clostridium difficile Infection and Colonization. *New England Journal of Medicine*, 365(18):1693–1703, 2011.

[16] Luis Furuya-Kanamori, Jennifer C. Stone, Justin Clark, et al. Comorbidities, Exposure to Medications, and the Risk of Community-Acquired Clostridium difficile Infection: A Systematic Review and Meta-analysis. *Infection Control & Hospital Epidemiology*, 36(02):132–141, 2015.

[17] Michelle M Riggs, Ajay K Sethi, Trina F Zabarsky, et al. Asymptomatic Carriers Are a Potential Source for Transmission of Epidemic and Nonepidemic Clostridium difficile Strains among Long-Term Care Facility Residents. *Clinical Infectious Diseases*, 45(8):992–998, 2007.

[18] Nicole Stoesser, David W. Eyre, T. Phuong Quan, et al. Epidemiology of Clostridium difficile in infants in Oxfordshire, UK: Risk factors for colonization and carriage, and genetic overlap with regional C. difficile infection strains. *PloS One*, 12(8):e0182307, 2017.

[19] M.P.M. Hensgens, E.C. Keessen, M.M. Squire, et al. Clostridium difficile infection in the community: a zoonotic disease? *Clinical Microbiology and Infection*, 18(7):635–645, 2012.

[20] Geoffrey P. Garnett, Simon Cousens, Timothy B. Hallett, Richard Steketee, and Neff Walker. Mathematical models in the evaluation of health programmes. *The Lancet*, 378(9790):515–525, 2011.

[21] Reuben M. Granich, Charles F. Gilks, Christopher Dye, Kevin M. De Cock, and Brian G. Williams. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *The Lancet*, 373(9657):48–57, 2009.

[22] Allan Schapira, Thomas Smith, Nakul Chitnis, and Richard Steketee. Comparing the Effectiveness of Malaria Vector-Control Interventions Through a Mathematical Model. *The American Journal of Tropical Medicine and Hygiene*, 83(2):230–240, 2010.

[23] Stuart H Cohen, Dale N Gerding, Stuart Johnson, et al. Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). *Infection Control and Hospital Epidemiology*, 31(5):431–55, 2010.

[24] Eroboghene H Otete, Anand S Ahankari, Helen Jones, et al. Parameters for the mathematical modelling of Clostridium difficile acquisition and transmission: a systematic review. *PloS One*, 8(12):e84224, 2013.

[25] L Clifford McDonald, Bruno Coignard, Erik Dubberke, et al. Recommendations for surveillance of Clostridium difficile-associated disease. *Infection control and hospital epidemiology*, 28(2):140–5, 2007.

[26] European Centre for Disease Prevention and Control. European Surveillance of Clostridium difficile infections. Surveillance protocol version 2.3., 2017.

[27] Yves Longtin, Bianka Paquet-Bolduc, Rodica Gilca, et al. Effect of Detecting and Isolating Clostridium difficile Carriers at Hospital Admission on the Incidence of C difficile Infections. *JAMA Internal Medicine*, 176(6):796, 2016.

[28] Ivan C. Hall and Elizabeth O'Toole. Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, Bacillus difficilis. *American Journal of Diseases of Children*, 49(2):390–402, 1935.

[29] Daniel E. Voth and Jimmy D. Ballard. Clostridium difficile Toxins: Mechanism of Action and Role in Disease. *Clinical Microbiology Reviews*, 18(2):247–263, 2005.

[30] S Johnson, S R Homann, K M Bettin, et al. Treatment of asymptomatic Clostridium difficile carriers (fecal excretors) with vancomycin or metronidazole. A randomized, placebo-controlled trial. *Annals of Internal Medicine*, 117(4):297–302, 1992.

[31] Mukil Natarajan, Seth T. Walk, Vincent B. Young, and David M. Aronoff. A clinical and epidemiological review of non-toxigenic Clostridium difficile. *Anaerobe*, 22(734):1–5, 2013.

[32] Dale N. Gerding and Fernanda C. Lessa. The epidemiology of Clostridium difficile infection inside and outside health care institutions. *Infectious Disease Clinics of North America*, 29(1):37–50, 2015.

[33] Clotilde Rousseau, Isabelle Poilane, Loic De Pontual, et al. Clostridium difficile Carriage in Healthy Infants in the Community: A Potential Reservoir for Pathogenic Strains. *Clinical Infectious Diseases*, 55(9):1209–1215, 2012.

[34] E. Holst, I. Helin, and P A Mårdh. Recovery of Clostridium difficile from children. *Scandinavian Journal of Infectious Diseases*, 13(1):41–5, 1981.

[35] Hiroyuki Kubota, Hiroshi Makino, Agata Gawad, et al. Longitudinal Investigation of Carriage Rates, Counts, and Genotypes of Toxigenic Clostridium difficile in Early Infancy. *Applied and Environmental Microbiology*, 82(19):5806–14, 2016.

[36] Ciarán P Kelly and Lorraine Kyne. The host immune response to Clostridium difficile. *Journal of medical microbiology*, 60(Pt 8):1070–9, 2011.

[37] R Viscidi, B E Laughon, R Yolken, et al. Serum Antibody Response to Toxins A and B of Clostridium difficile. *Journal of Infectious Diseases*, 148(1):93–100, 1983.

[38] Mary C Rea, Alleson Dobson, Orla O'Sullivan, et al. Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences*, 108(S1):4639–44, 2011.

[39] Fatemeh Rafii, John B Sutherland, and Carl E Cerniglia. Effects of treatment with antimicrobial agents on the human colonic microflora. *Therapeutics and Clinical Risk Management*, 4(6):1343–58, 2008.

[40] Patrizia Spigaglia. Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection. *Therapeutic advances in infectious disease*, 3(1):23–42, 2016.

[41] Claudia Slimings and Thomas V. Riley. Antibiotics and hospital-acquired Clostridium difficile infection: update of systematic review and meta-analysis. *The Journal of Antimicrobial Chemotherapy*, 69(4):881–91, 2014.

[42] Sandra Dial, Abbas Kezouh, Andre Dascal, Alan Barkun, and Samy Suissa. Patterns of antibiotic use and risk of hospital admission because of Clostridium difficile infection. *Canadian Medical Association Journal*, 179(8):767–772, 2008.

[43] Joseph F. Rodemann, Erik R. Dubberke, Kimberly A. Reske, Da Hea Seo, and Christian D. Stone. Incidence of Clostridium difficile Infection in Inflammatory Bowel Disease. *Clinical Gastroenterology and Hepatology*, 5(3):339–344, 2007.

[44] Abhishek Deshpande, Chaitanya Pant, Vinay Pasupuleti, et al. Association Between Proton Pump Inhibitor Therapy and Clostridium difficile Infection in a Meta-Analysis. *Clinical Gastroenterology and Hepatology*, 10(3):225–233, 2012.

[45] Lorraine Kyne, Michel Warny, Amir Qamar, and Ciarán P. Kelly. Association between antibody response to toxin A and protection against recurrent Clostridium difficile diarrhoea. *The Lancet*, 357(9251):189–193, 2001.

[46] Brett A Leav, Barbra Blair, Mark Leney, et al. Serum anti-toxin B antibody correlates with protection from recurrent Clostridium difficile infection (CDI). *Vaccine*, 28(4):965–9, 2010.

[47] Niloufar Roshan, Katherine A. Hammer, and Thomas V. Riley. Non-conventional antimicrobial and alternative therapies for the treatment of Clostridium difficile infection. *Anaerobe*, 49:103–111, 2018.

[48] Paul Moayyedi, Yuhong Yuan, Harith Baharith, and Alexander C Ford. Faecal microbiota transplantation for Clostridium difficile-associated diarrhoea: a systematic review of randomised controlled trials. *The Medical journal of Australia*, 207(4):166–172, 2017.

[49] L Clifford McDonald, Dale N Gerding, Stuart Johnson, et al. Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases*, 66(7):e1–e48, 2018.

[50] Dale N Gerding, Thomas Meyer, Christine Lee, et al. Administration of Spores of Nontoxigenic Clostridium difficile Strain M3 for Prevention of Recurrent C difficile Infection. *JAMA*, 313(17):1719, 2015.

[51] Kate E. Dingle, Xavier Didelot, T. Phuong Quan, et al. Effects of control interventions on Clostridium difficile infection in England: an observational study. *The Lancet Infectious Diseases*, 17(4):411–421, 2017.

[52] Luis Furuya-Kanamori, Samantha J. McKenzie, Laith Yakob, et al. Clostridium difficile infection seasonality: patterns across hemispheres and continents - a systematic review. *PloS one*, 10(3):e0120730, 2015.

[53] L. Furuya-Kanamori, A.C.A. Clements, N.F. Foster, et al. Asymptomatic Clostridium difficile colonization in two Australian tertiary hospitals, 2012–2014: prospective, repeated cross-sectional study. *Clinical Microbiology and Infection*, 23(1):48.e1–48.e7, 2017.

[54] Philip M. Polgreen, Ming Yang, Lucas C. Bohnett, and Joseph E. Cavanaugh. A Time-Series Analysis of Clostridium difficile and Its Seasonal Association with Influenza. *Infection Control & Hospital Epidemiology*, 31(04):382–387, 2010.

[55] Rodica Gilca, Elise Fortin, Charles Frenette, Yves Longtin, and Marie Gourdeau. Seasonal variations in Clostridium difficile infections are associated with influenza and respiratory syncytial virus activity independently of antibiotic prescriptions: a time series analysis in Quebec, Canada. *Antimicrobial agents and chemotherapy*, 56(2):639–46, 2012.

[56] S.P. Barrett, A.H. Holmes, W.A. Newsholme, and M. Richards. Increased detection of Clostridium difficile during a norovirus outbreak. *Journal of Hospital Infection*, 66(4):394–395, 2007.

[57] M. Wilcox and W. Fawley. Viral gastroenteritis increases the reports of Clostridium difficile infection. *Journal of Hospital Infection*, 66(4):395–396, 2007.

[58] Luis Furuya-Kanamori, Jenny Robson, Ricardo J Soares Magalhães, et al. A population-based spatio-temporal analysis of Clostridium difficile infection in Queensland, Australia over a 10-year period. *The Journal of infection*, 69(5):447–55, 2014.

[59] Mark H. Wilcox and Warren N. Fawley. Hospital disinfectants and spore formation by Clostridium difficile. *Lancet (London, England)*, 356(9238):1324, 2000.

[60] Axel Kramer, Ingeborg Schwebke, and Günter Kampf. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infectious Diseases*, 6:1–8, 2006.

[61] Alexander Rodriguez-Palacios and Jeffrey T Lejeune. Moist-heat resistance, spore aging, and superdormancy in Clostridium difficile. *Applied and Environmental Microbiology*, 77(9):3085–91, 2011.

[62] Umair Jabbar, Julie Leischner, Douglas Kasper, et al. Effectiveness of Alcohol-Based Hand Rubs for Removal of Clostridium difficile Spores from Hands. *Infection Control & Hospital Epidemiology*, 31(06):565–570, 2010.

[63] Scott R. Curry, Carlene A. Muto, Jessica L. Schlackman, et al. Use of multilocus variable number of tandem repeats analysis genotyping to determine the role of asymptomatic carriers in Clostridium difficile transmission. *Clinical Infectious Diseases*, 57(8):1094–1102, 2013.

[64] A Sarah Walker, David W Eyre, David H Wyllie, et al. Characterisation of Clostridium difficile hospital ward-based transmission using extensive epidemiological data and molecular typing. *PLoS Medicine*, 9(2):e1001172, 2012.

[65] Thomas V. Riley, J E Adams, G L O'Neill, and R a Bowman. Gastrointestinal carriage of Clostridium difficile in cats and dogs attending veterinary clinics. *Epidemiology and infection*, 107(3):659–65, 1991.

[66] C W Knetsch, N Kumar, S C Forster, et al. Zoonotic Transfer of Clostridium difficile Harboring Antimicrobial Resistance between Farm Animals and Humans. *Journal of Clinical Microbiology*, 56(3), 2017.

[67] Alexander Rodriguez-Palacios, Henry R. Staempfli, Todd Duffield, and J. Scott Weese. Clostridium difficile in retail ground meat, Canada. *Emerging infectious diseases*, 13(3):485–7, 2007.

[68] D.S. Metcalf, M.C. Costa, W.M.V. Dew, and J.S. Weese. Clostridium difficile in vegetables, Canada. *Letters in Applied Microbiology*, 51(5):600–602, 2010.

[69] A J McMichael. Prisoners of the Proximate: Loosening the Constraints on Epidemiology in an Age of Change. *American Journal of Epidemiology*, 149(10):887–897, 1999.

[70] Roy M. Anderson and Robert M. May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

[71] Fred Brauer, Pauline van den Driessche, and Jianhong Wu, editors. *Mathematical Epidemiology*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[72] M J Keeling and P Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, New Jersey, 2011.

[73] Odo Diekmann, Hans Heesterbeek, and Tom Britton. *Mathematical tools for understanding infectious disease dynamics*. Princeton University Press, Princeton, New Jersey, 2012.

[74] Linda J. S. Allen. *An Introduction to Stochastic Processes with Applications to Biology*. Chapman and Hall/CRC, New York, 2 edition, 2010.

[75] Lawrence F. Shampine. *Numerical solution of ordinary differential equations*. CRC Press, 1994.

[76] Stewart Ethier and Thomas Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley and Sons, Hoboken, 1986.

[77] David F. Anderson. An Efficient Finite Difference Method for Parameter Sensitivities of Continuous Time Markov Chains. *SIAM Journal on Numerical Analysis*, 50(5):2237–2258, 2012.

[78] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124(4):44104, 2006.

[79] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[80] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[81] Dorothy Anderson and Ray Watson. On the Spread of a Disease with Gamma Distributed Latent and Infectious Periods. *Biometrika*, 67(1):191, 1980.

[82] Lander Willem, Frederik Verelst, Joke Bilcke, Niel Hens, and Philippe Beutels. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). *BMC Infectious Diseases*, 17(1):612, 2017.

[83] Gerardo Rubino and Bruno Sericola. Sojourn times in finite Markov processes. *Journal of Applied Probability*, 26(04):744–756, 1989.

[84] O Diekmann, J A P Heesterbeek, and M G Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society, Interface*, 7(47):873–85, 2010.

[85] Christian L. Althaus. Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. *PLoS currents*, 6, 2014.

[86] Fred Brauer. Backward bifurcations in simple vaccination models. *Journal of Mathematical Analysis and Applications*, 298(2):418–431, 2004.

[87] Jonathan Dushoff, Wenzhang Huang, and Carlos Castillo-Chavez. Backwards bifurcations and catastrophe in simple models of fatal diseases. *Journal of Mathematical Biology*, 36(3):227–248, 1998.

[88] J. A.P. Heesterbeek and M. G. Roberts. The type-reproduction number T in models for infectious disease control. *Mathematical Biosciences*, 206(1):3–10, 2007.

[89] M. G. Roberts and J. A.P. Heesterbeek. A new method for estimating the effort required to control an infectious disease. *Proceedings of the Royal Society B: Biological Sciences*, 270(1522):1359–1364, 2003.

[90] John M. Starr, Tom R. Rogers, and Mario Impallomeni. Hospital-acquired Clostridium difficile diarrhoea and herd immunity. *The Lancet*, 349(9049):426–428, 1997.

[91] J.M. Starr and A Campbell. Mathematical modeling of Clostridium difficile infection. *Clinical Microbiology and Infection*, 7(8):432–437, 2001.

[92] David P Durham, Margaret A Olsen, Erik R Dubberke, Alison P Galvani, and Jeffrey P Townsend. Quantifying Transmission of Clostridium difficile within and outside Healthcare Settings. *Emerging Infectious Diseases*, 22(4):608–16, 2016.

[93] Esther van Kleef, Sarah R. Deeny, Mark Jit, et al. The projected effectiveness of Clostridium difficile vaccination as part of an integrated infection control strategy. *Vaccine*, 34(46):5562–5570, 2016.

[94] Laith Yakob, Thomas V Riley, David L Paterson, et al. Mechanisms of hypervirulent Clostridium difficile ribotype 027 displacement of endemic strains: an epidemiological model. *Scientific Reports*, 5(1):12666, 2015.

[95] J.M. Starr, A. Campbell, E. Renshaw, I.R. Poxton, and G.J. Gibson. Spatio-temporal stochastic modelling of Clostridium difficile. *Journal of Hospital Infection*, 71(1):49–56, 2009.

[96] Michael A Rubin, Makoto Jones, Molly Leecaster, et al. A simulation-based assessment of strategies to control Clostridium difficile transmission and infection. *PloS one*, 8(11):e80671, 2013.

[97] James Codella, Nasia Safdar, Rick Heffernan, and Oguzhan Alagoz. An Agent-based Simulation Model for Clostridium difficile Infection Control. *Medical Decision Making*, 35(2):211–229, 2015.

[98] C. Lanzas, E. R. Dubberke, Z. Lu, K. A. Reske, and Y. T. Gröhn. Epidemiological Model for Clostridium difficile Transmission in Healthcare Settings. *Infection Control and Hospital Epidemiology*, 32(06):553–561, 2011.

[99] Eric T Lofgren, Rebekah W Moehring, Deverick J Anderson, David J Weber, and Nina H Fefferman. A mathematical model to evaluate the routine use of fecal microbiota transplantation to prevent incident and recurrent Clostridium difficile infection. *Infection control and hospital epidemiology*, 35(1):18–27, 2014.

[100] Laith Yakob, Thomas V. Riley, David L. Paterson, and Archie C A Clements. Clostridium difficile exposure as an insidious source of infection in healthcare settings: an epidemiological model. *BMC infectious diseases*, 13(1):376, 2013.

[101] Laith Yakob, Thomas V. Riley, David L. Paterson, John Marquess, and Archie CA Clements. Assessing control bundles for Clostridium difficile: a review and mathematical model. *Emerging Microbes & Infections*, 3(6):e43–e43, 2014.

[102] Christos A. Grigoras, Fainareti N. Zervou, Ioannis M. Zacharioudakis, Constantinos I. Siettos, and Eleftherios Mylonakis. Isolation of C. difficile Carriers Alone and as Part of a Bundle Approach for the Prevention of Clostridium difficile Infection (CDI): A Mathematical Model Based on Clinical Study Data. *PLOS ONE*, 11(6):e0156577, 2016.

[103] Cristina Lanzas and Erik R Dubberke. Effectiveness of screening hospital admissions to detect asymptomatic carriers of Clostridium difficile: a modeling evaluation. *Infection control and hospital epidemiology*, 35(8):1043–50, 2014.

[104] Jason Bintz, Suzanne Lenhart, and Cristina Lanzas. Antimicrobial Stewardship and Environmental Decontamination for the Control of Clostridium difficile Transmission in Healthcare Settings. *Bulletin of Mathematical Biology*, 79(1):36–62, 2017.

[105] Anna K. Barker, Oguzhan Alagoz, and Nasia Safdar. Interventions to Reduce the Incidence of Hospital-Onset Clostridium difficile Infection: An Agent-Based Modeling Approach to Evaluate Clinical Effectiveness in Adult Acute Care Hospitals. *Clinical Infectious Diseases*, 66(8):1192–1203, 2018.

[106] Brittany Stephenson, Cristina Lanzas, Suzanne Lenhart, and Judy Day. Optimal control of vaccination rate in an epidemiological model of Clostridium difficile transmission. *Journal of Mathematical Biology*, 75(6-7):1693–1713, 2017.

[107] Jennie H. Kwon, Cristina Lanzas, Kimberly A. Reske, et al. An Evaluation of Food as a Potential Source for Clostridium difficile Acquisition in Hospitalized Patients. *Infection control and hospital epidemiology*, 37(12):1401–1407, 2016.

[108] Lorraine Kyne, Michel Warny, Amir Qamar, and Ciarán P Kelly. Asymptomatic Carriage of Clostridium difficile and Serum Levels of IgG Antibody against Toxin A. *New England Journal of Medicine*, 342(6):390–397, 2000.

[109] Fernanda C Lessa, Carolyn V Gould, and L Clifford McDonald. Current Status of Clostridium difficile Infection Epidemiology. *Clinical Infectious Diseases*, 55(S2):S65–S70, 2012.

[110] Alexandra B Hogan. *Mathematical models for respiratory syncytial virus (RSV) transmission*. Phd, Australian National University, 2017.

[111] Sonia Altizer, Andrew Dobson, Parviez Hosseini, et al. Seasonality and the dynamics of infectious diseases. *Ecology Letters*, 9(4):467–484, 2006.

[112] Thomas Liggett. *Continuous Time Markov Processes*, volume 113 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2010.

[113] Iris Figueroa, Stuart Johnson, Susan P. Sambol, et al. Relapse Versus Reinfection: Recurrent Clostridium difficile Infection Following Treatment With Fidaxomicin or Vancomycin. *Clinical Infectious Diseases*, 55(suppl 2):S104–S109, 2012.

[114] Audrey J Weiss (Truven Health Analytics) and Anne Elixhauser (AHRQ). *Overview of Hospital Stays in the United States, 2012: Statistical Brief #180*. Agency for Healthcare Research and Quality, Rockville, MD, 2014.

[115] Conan MacDougall and Ronald E. Polk. Variability in rates of use of antibacterials among 130 US hospitals and risk-adjustment models for interhospital comparison. *Infection Control and Hospital Epidemiology*, 29(3):203–11, 2008.

[116] Australian Commonwealth Department of Health - Drug utilisation sub-committee (DUSC). Antibiotics : PBS / RPBS utilisation. Technical report, Australian Commonwealth Department of Health, Canberra, 2015.

[117] Sahil Khanna, Darrell S Pardi, Scott L Aronson, et al. The Epidemiology of Community-Acquired Clostridium difficile Infection: A Population-Based Study. *The American Journal of Gastroenterology*, 107(1):89–95, 2012.

[118] M. H. Wilcox, L. Mooney, R. Bendall, C. D. Settle, and W. N. Fawley. A case-control study of community-associated Clostridium difficile infection. *The Journal of Antimicrobial Chemotherapy*, 62(2):388–96, 2008.

[119] Amit S Chitnis, Stacy M Holzbauer, Ruth M Belflower, et al. Epidemiology of community-associated Clostridium difficile infection, 2009 through 2011. *JAMA internal medicine*, 173(14):1359–67, 2013.

[120] Wafaa Jamal, Eunice Pauline, and Vincent Rotimi. A prospective study of community-associated Clostridium difficile infection in Kuwait: Epidemiology and ribotypes. *Anaerobe*, 35(PB):28–32, 2015.

[121] Wasef Na'amnih, Amos Adler, Tamar Miller-Roll, Dani Cohen, and Yehuda Carmeli. Incidence and Risk Factors for Community and Hospital Acquisition of Clostridium difficile Infection in the Tel Aviv Sourasky Medical Center. *Infection control and hospital epidemiology*, 38(8):912–920, 2017.

[122] Alex Rodriguez-Palacios, Stefan Borgmann, Terence R. Kline, and Jeffrey T. LeJeune. Clostridium difficile in foods and animals: history and measures to reduce exposure. *Animal Health Research Reviews*, 14(1):11–29, 2013.

[123] Martyn Kirk, Laura Ford, Kathryn Glass, and Gillian Hall. Foodborne Illness, Australia, Circa 2000 and Circa 2010. *Emerging Infectious Diseases*, 20(11):1857–1864, 2014.

[124] K. Warriner, C. Xu, M. Habash, S. Sultan, and S. J. Weese. Dissemination of Clostridium difficile in food and the environment: Significant sources of C. difficile community-acquired infection? *Journal of Applied Microbiology*, 122(3):542–553, 2017.

[125] Michael Z. David, Daniel Glikman, Susan E. Crawford, et al. What Is Community-Associated Methicillin-Resistant Staphylococcus aureus? *The Journal of Infectious Diseases*, 197(9):1235–1243, 2008.

[126] OECD. *Health at a Glance 2015*. Health at a Glance. OECD Publishing, Paris, 2015.

[127] N Deborah Friedman, Keith S Kaye, Jason E Stout, et al. Healthcare-associated bloodstream infections in adults: a reason to change the accepted definition of community-acquired infections. *Annals of internal medicine*, 137(10):791–7, 2002.

[128] Adam L. Cohen, David Calfee, Scott K. Fridkin, et al. Recommendations For Metrics For Multidrug-Resistant Organisms In Healthcare Settings: SHEA/HICPAC Position Paper. *Infection Control & Hospital Epidemiology*, 29(10):901–913, 2008.

[129] M.P.M. Hensgens, O.M. Dekkers, A. Demeulemeester, et al. Diarrhoea in general practice: when should a Clostridium difficile infection be considered? Results of a

nested case-control study. *Clinical Microbiology and Infection*, 20(12):O1067–O1074, 2014.

[130] D Van Cauteren, H De Valk, S Vaux, Y Le Strat, and V Vaillant. Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiology and Infection*, 140(4):697–705, 2012.

[131] Elaine Scallan, Timothy F Jones, Alicia Cronquist, et al. Factors associated with seeking medical care and submitting a stool sample in estimating the burden of foodborne illness. *Foodborne Pathogens and Disease*, 3(4):432–8, 2006.

[132] E. Reigadas, L. Alcalá, M. Marín, et al. Missed diagnosis of Clostridium difficile infection; a prospective evaluation of unselected stool samples. *The Journal of Infection*, 70(3):264–72, 2015.

[133] Susan S Huang, Taliser R Avery, Yeohan Song, et al. Quantifying Interhospital Patient Sharing as a Mechanism for Infectious Disease Spread. *Infection Control & Hospital Epidemiology*, 31(11):1160–1169, 2010.

[134] Bruce Y. Lee, Sarah M. McGlone, Yeohan Song, et al. Social Network Analysis of Patient Sharing Among Hospitals in Orange County, California. *American Journal of Public Health*, 101(4):707–713, 2011.

[135] Rachel B. Slayton, Damon Toth, Bruce Y. Lee, et al. Vital Signs: Estimated Effects of a Coordinated Approach for Action to Reduce Antibiotic-Resistant Infections in Health Care Facilities - United States. *American Journal of Transplantation*, 15(11):3002–3007, 2015.

[136] Joël Mossong, Niel Hens, Mark Jit, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine*, 5(3):e74, 2008.

[137] Daniel R. Knight, Briony Elliott, Barbara J. Chang, Timothy T. Perkins, and Thomas V. Riley. Diversity and Evolution in the Genome of Clostridium difficile. *Clinical Microbiology Reviews*, 28(3):721–41, 2015.

[138] Warren N. Fawley, Kerrie A. Davies, Trefor Morris, et al. Enhanced surveillance of Clostridium difficile infection occurring outside hospital, England, 2011 to 2013. *Eurosurveillance*, 21(29):30295, 2016.

[139] David W Eyre, Kerrie A Davies, Georgina Davis, et al. Two Distinct Patterns of Clostridium difficile Diversity Across Europe Indicating Contrasting Routes of Spread. *Clinical Infectious Diseases*, 2018.