UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

Mathematical modelling of co-colonization and within-host abundance ratios in multi-type pathogen dynamics

Maria Rocha Peixoto Azevedo Gaivão

DISSERTAÇÃO
MESTRADO EM BIOINFORMÁTICA E BIOLOGIA COMPUTACIONAL
ESPECIALIZAÇÃO EM BIOLOGIA COMPUTACIONAL

Dissertação orientada pela Dra. Erida Gjini e pelo Prof. Dr. Francisco Dionisio

2015

Abstract

In recent years our understanding of infectious-disease epidemiology has been greatly increased through mathematical modelling. The major goal of any mathematical study in epidemiology is to develop understanding of the interplay between the variables that determine the course of infection within an individual, and the variables that control the pattern of infections within communities of people. The epidemiology of multi-type pathogen systems, such as dengue, malaria and pneumococcus are notoriously challenging. Direct and indirect interactions between multiple strains shape pathogen population processes, both at the level of a single host and at the population level. Quantifying these interactions is crucial, and the new technologies that are now available to detect multiple infections with different pathogen types are opening new avenues in this endeavour.

In this thesis, motivated by the pneumococcus system, we study the colonization dynamics by a multi-type pathogen and focus particularly on co-colonization phenomena, which reflects the simultaneous colonization/infection (terms used in this thesis interchangeably) by two antigenic types of the same pathogen. We pretend to introduce strain ratios, first quantified by Brugger *et al.* (2010), when modelling the co-colonization phenomena. Therefore, a mathematical epidemiological model is constructed using ordinary differential equations to examine the prevalence and distribution of the co-colonization in the population. Interestingly, we find one scenario where the infection can still persist despite the basic reproduction number R_0 being below 1. The phenomena of backward bifurcation is also observed. Moreover, the proportion of each double infected class, at equilibrium, is independent of the size of susceptible or single infected class.
Based on a static epidemiological point of view, we also develop an

within-host model to study the distribution of co-colonization in an average host. Both models show a clear equal abundance ratio (1:1) prevalence and this seems to be robust despite varying the parameters.

Keywords: epidemiology, co-colonization, mathematical modelling, ordinary differential equations, *streptococcus pneumoniae*

Resumo Alargado

A Epidemiologia é uma ciência que estuda quantitativamente a distribuição dos fenómenos de saúde/doença, e seus factores condicionantes e determinantes, nas populações humanas. Esta permite ainda avaliar a eficácia das intervenções realizadas no âmbito da saúde pública. O fundador da teoria epidemiológica moderna é Ronald Ross cujo estudo no ciclo de vida da malária concedeu-lhe o Nobel em 1902. Este utilizou a modelação matemática para investigar a eficácia das intervenções na prevenção desta doença. No entanto, foi só no final do século XX que a modelação matemática se tornou mais popular.

Nos últimos anos o nosso conhecimento relativo à epidemiologia das doenças infecciosas desenvolveu-se bastante devido à modelação matemática. O principal objectivo de qualquer estudo matemático em epidemiologia é melhorar o nosso entendimento relativo às relações das variáveis que determinam o curso de uma infecção quer ao nível do indivíduo como ao nível das comunidades. No entanto, devemos ter sempre em conta que os modelos são sempre abstracções/simplificações dos fenómenos em estudo e os resultados obtidos aproximações do sistema real. A modelação têm sido aplicada para o estudo de diversas doenças infecciosas tal como a sarampo, HIV ou a dengue. Estes modelos revelam-se ferramentas essenciais para compreender a dinâmica das doenças infecciosas e auxiliar no planeamento e controlo das mesmas.

Nesta tese, estou interessada em estudar as dinâmicas das doenças infecciosas, mas mais precisamente, explorar através da modelação matemática o fenómeno de co-colonização ou também designado por múltipla colonização. Esta significa a colonização simultânea do hospedeiro por vários microorganismos (da mesma espécie ou diferente).

É sabido desde há décadas que a co-colonização é um fenómeno comum na natureza e com importantes consequências para o hospedeiro e parasita. Para o hospedeiro, representa um desafio extra para o seu sistema imunitário. Para o parasita, conduz a interacções directas e indirectas entre as diversas estirpes alterando a sua dinâmica e transmissão. Geralmente este fenómeno agrava o estado de saúde do individuo em comparação com as infecções simples, ou seja, quando o individuo é unicamente colonizado por um parasita. Quantificar a interacção entre as diversas estirpes envolvidas revela-se por isso fundamental, e as novas tecnologias que estão hoje em dia disponíveis para detectar os diferentes patogénios envolvidos, estão a abrir caminho nesta área.

Recentemente, Brugger *et al.* (2010) revelou com os seus estudos na bactéria *Streptococcus pneumoniae*, também conhecida por pneumoccocus, que a co-colonização tem uma prevalência de 7.9%. Aparentemente, é também mais comum para o hospedeiro apresentar sensivelmente a mesma proporção, usualmente designada por 1:1, entre as duas estirpes da bactéria. Esta prevalência foi também observada independentemente por Valente *et al.* (2012), mas desta vez em indivíduos saudáveis. Este padrão parece ser, por isso, independente do estado de saúde do indivíduo.

O pneumococcus é uma bactéria gram-positiva que normalmente vive assimptomáticamente na nasofaringe e cuja prevalência está aumentada nos primeiros cinco anos de vida de um indivíduo. Ocasionalmente, esta pode migrar para outras regiões do corpo e potencialmente causar uma série de doenças, desde infecções respiratórias ligeiras (otites, etc.) até doenças mais invasivas (pneumonia, septicémia, meningite, etc.). O fenómeno da co-colonização parece também ser um importante factor para a evolução desta espécie, uma vez que representa uma oportunidade para a transferência horizontal de genes. Incorporar esta informação sobre os rácios nos modelos é relevante, uma vez que pode auxiliar na compreensão da sua dinâmica de

transmissão e potencialmente prever o impacto de políticas de intervenção, tal como a vacinação. Para um organismo tão diverso como o penumococcus, com mais de 90 estirpes diferentes identificadas, a compreensão da sua biologia está longe de estar completa, e formular modelos reais ainda representa um desafio.

Nesta tese foi feito um estudo detalhado acerca do padrão de co-colonização na nasofaringe por múltiplas estirpes do pneumococcus. Mais precisamente, pretendo compreender os factores que justificam a sua prevalência na população e a distribuição dos rácios de co-colonização no caso do hospedeiro apresentar duas estirpes. O principal objectivo deste estudo foi desenhar um modelo matemático que representasse adequadamente a infecção pelo pneumococcus para que o seu *output* fosse suficientemente preciso para explicar as características da distribuição das estirpes no hospedeiro.

Nesse sentido, usei duas abordagens diferentes (mas complementares) para modelar a co-colonização. Em primeiro lugar, usando equações diferenciais ordinárias, construí um modelo epidemiológico determinístico com estrutura nos tipos de co-colonização. Esta abordagem parte da dinâmica de uma população com vista a estudar a distribuição num único indivíduo. Portanto caracteriza-se como uma abordagem *top-down*. Numa segunda abordagem, criei um modelo probabilístico que a partir da dinâmica da infecção no indivíduo, permite observar a distribuição das estirpes na população. Esta abordagem caracteriza-se como *bottom-up*.

Em ambos os modelos, os resultados que obtive evidenciaram os mecanismos imunitários e estocásticos responsáveis pela distribuição dos rácios de co-colonização. Foi observada uma clara predominância dos rácios 1:1 e este resultado parece ser robusto quando se variam os parâmetros dos modelos. Foram identificados os equilibrios do sistema (trivial e endémico) e avaliada a sua estabilidade. Curiosamente, no modelo epidemiológico, encontrei um cenário em que a infecção pode persistir apesar do número básico de reprodução R_0

ser inferior a 1. Este fenómeno tem o nome de *backward bifurcation* e consiste numa alteração estrutural da estabilidade dos equilíbrios, que deve-se essencialmente ao facto do modelo desenvolvido estruturar os hospedeiros co-colonizados em classes. Estas em média apresentam um número básico de reprodução superior aos hospedeiros colonizados por uma única estirpe. Assim, contribuem em média para uma maior transmissão da infecção na população. Também a proporção de cada classe de hospedeiros duplamente infectados relativamente ao total de hospedeiros infectados, no equilíbrio, é independente da magnitude da classe dos susceptíveis ou dos infectados apenas por uma estirpe. Isto significa que quando o hospedeiro é infectado por uma segunda estirpe tem uma probabilidade fixa de apresentar um determinado rácio. Neste modelo epidemiológico foi também possível verificar, que o mecanismo responsável por desviar a distribuição em torno do rácio 1:1 baseia-se no pressuposto que cada classe de co-colonizados ter taxas de recuperação diferentes, onde umas classes recuperam mais rapidamente que outras. Este rácio traduz como os diferentes patogénios, como um "todo", estão expostos ao sistema imunitário do hospedeiro. Todas as simulações numéricas foram realizadas usando a linguagem de programação *Python* e o *software* cientifico *Mathematica*.

Construir modelos epidemiológicos que reflictam o fenómeno de co-colonização é fundamental para melhor compreender determinadas doenças, mas também apresenta muitos desafios técnicos. Nomeadamente, quanto mais factores biológicos forem tidos em conta na modelação, no sentido de os tornar mais realistas, mais parâmetros serão introduzidos e mais complexa será a sua análise. No entanto, seria interessante no futuro incorporar factores como: a identidade das estirpes, a heterogeneidade dos hospedeiros e as variações na sua resposta imunitária. Para além disso, poderíamos ter ainda em conta o fenómeno de co-transmissão, ou seja, a infecção do hospedeiro por mais de um parasita durante o mesmo evento de transmissão. Com isto poderíamos, potencialmente, contribuir para o estudo da evolução

da virulência destes patogénios. No entanto, é fundamental que hajam mais resultados experimentais para se fazer uma comparação e validação dos resultados teóricos com vista à criação de modelos biológicos mais representativos da realidade.

Palavras Chave: epidemiologia, co-colonização, modelação matemática, equações diferenciais ordinárias, *streptococcus pneumoniae*

Acknowledgements

I would like to express my sincere gratitude to my external advisor Dr. Erida Gjini for enlightening me in my first glance of research. Her enthusiasm was contagious and motivational for me. She helped me all the time and it was a pleasure to be her first master student. My sincere thanks goes to Prof. Dr. Francisco Dionísio, my internal advisor, for his dedication of time, helpful discussions and relevant comments.

Last but not the least, I would like to thank my family: my husband and best friend Zé Pedro, for his love and immense encouragement and for my two sweethearts who inspire me everyday, José Maria and António Maria. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Epidemiology is the study of the spread of disease, in space and time, with the objective to trace factors that are responsible or contribute to their occurrence O. Diekmann (2010). The founding father of the modern epidemic theory is Sir Ronald Ross. His study on the life cycle of the malaria parasite, using mathematical modelling to investigate the effectiveness of various intervention strategies, gave him the Nobel price in 1902 (Krämer A, Kretzschmar M, 2010). However, it was only towards the end of the twentieth century that mathematical modelling came into more widespread use.
The first aim of any mathematical modelling is to summarize the available knowledge and construct a formal representation of the system. The second aim is to assess the relative importance of each of the various mechanisms involved in the system dynamics. However, one should always remember that models are abstractions/simplifications of the phenomena under study and the results obtained only approximations of the real system.

Mathematical models have been applied to study almost all infectious diseases, from measles to chickenpox, HIV, dengue, etc. They have been instrumental to understand the dynamics of these diseases and design control.

In this thesis, we are interested in studying the dynamics of infectious diseases, more precisely we are going to explore through mathematical models the co-colonization phenomenon, which is the simultaneous colonization or infection

by several microorganisms (either from the same or different species) (Balmer & Tanner, 2011). It has been known for decades that multiple colonizations (or infections) are actually a common phenomena, with a major impact on the efficacy of anti-infection drugs at the population scale (Balmer & Tanner, 2011). With the recent advances in technology, is now possible to better detect and quantify all different serotypes colonizing the host simultaneously (Brugger *et al.*, 2010) .

Recently, Brugger have shown in his studies on *Streptococcus pneumoniae* (the pneumococcus) that the carriage of two different serotypes occurs in 7.9% of colonized hosts (Brugger *et al.*, 2010). Apparently, it is also more likely for the host to harbour two strains at similar proportions than presenting them at asymmetric ratios(Brugger *et al.*, 2010). These observations were also reported by (Valente *et al.*, 2012).

Co-colonization seems to be a particularly important event for pneumococcal evolution as it represents an opportunity for horizontal gene transfer, the main mechanism of evolution in this species (Shak *et al.*, 2013). Incorporating this additional information in the mathematical models seems relevant, as through such models we can gain a better understanding of the transmission dynamics of polymorphic infectious organisms, we can investigate the observed pattern of disease or epidemiology and we can predict the potential impact of alternative public health interventions, such as vaccination. For extremely diverse organisms, such as *S. pneumoniae* with more than 90 serotypes (Andrews *et al.*, 2014), for which a comprehensive understanding of the biology is far from achieved, real models are not always easy to formulate and still represent a challenge.

1.2 Objectives

In this thesis we study in detail patterns of nasopharyngeal o-colonization by multiple serotypes. More precisely, we want to understand what drives the overall prevalence in the population and the abundance ratio distribution of 2 co-colonizing strains/serotypes. We try to incorporate in the models the new biological element of serotype co-colonization ratios recently described in (Brugger *et al.*, 2010). Our main goal is to design models that are adequate representation

of the pneumococcus infection so that its outputs are sufficiently accurate and precise to explain the characteristics of the distribution of co-infection.

1.3 Contributions

To pursue our goal, we use two different but complementary approaches to model co-colonization. First, we construct a deterministic epidemiological model with structure in the types of co-colonization. It uses fixed values for parameters and generates a single "average" or expected outcome at the population level. Secondly, we create a probabilistic within-host model where the population dynamics of two co-colonizing strains is explicitly simulated. Here, we incorporate some natural variability in the infection process and generated a range of possible outcomes from explicit simulations in different individual hosts. Our results highlight immunity and stochasticity mechanisms that give rise to the co-colonization ratio distribution.

1.4 Overview

The thesis structure is as follows. Chapter 2 provides a brief overview of the process of mathematical modelling in infectious diseases using the simple Susceptible-Infected (SI) model. It also introduces a detailed description of all the parameters involved and important epidemiological concepts such as the basic reproduction number R_0. In Chapter 3 we introduce multiple strain infections, addressing super-infection and co-infection. We focus essentially in the co-infection process and briefly mention the two main topics that have concerned the scientific community so far: the evolution of virulence and the process of cross-immunity. We provide also a general description of *S.pneumoniae* bacteria. In Chapter 4 we present the Susceptible-Infected-Double infected (SID) model for co-infection together with the new data in serotype ratios distribution first mentioned by (Brugger *et al.*, 2010) and then latter by (Valente *et al.*, 2012). This is then followed by a brief description of the first proposed model (structured epidemiological model) designed to incorporate this new information in the process of pneumococcal infection. Latter in this chapter we present some numerical simulations. Chapter

5 is dedicated to the analysis of the second proposed model for co-infection, the within-host model, together with a stochastic simulation component. In both models, several numerical computations were performed using either *Python* or *Mathematica* and the programming code can be found in the Appendix. In the final chapter we summarize the conclusions and future perspectives.

Chapter 2

Modelling the dynamics of infectious diseases

Infection is the term that defines the entrance and development of an infectious agent (either micro-organisms or macro-organisms) in a human or animal body, whether or not it develops into a disease (Barreto *et al.*, 2006).

Epidemiological modelling of infectious diseases is typically based on compartmental SI/SIR models, in which the host population is divided into a small number of compartments, each containing individuals that are identical in terms of their status with respect to the disease in question. In the SI/SIR models, there are three compartments: susceptible S, infected I , and recovered-and-immune R individuals. This approach to epidemiological processes dates back to Ronald Ross's modelling of malaria at the beginning of the 20th century (Dieckmann *et al.*, 2005). This was soon followed by the work of A. G. McKendrick and W. O. Kermack, whose paper *A Contribution to the Mathematical Theory of Epidemics* was published in 1927 with a simple deterministic (compartmental) model. The basic model is the SI model. This model assumes certain conditions as: homogeneous population (S individuals have identical susceptibility and all I individuals have identical infectiousness), homogeneous contacts between individuals and the transition from I class to S does not depend on the time since infection. These SI models are typically applied to HIV or other chronic pathogens. SIR models, where in addition there is a recovered R class, are typical for influenza or measles transmission where individuals recover with some immunity against the pathogen.

When recovery is without immunity we have an SIS model, the third classical epidemiological model. The transmission dynamics are formalized through ordinary differential equations (Brauer & Castillo-Chavez, 2012), which represent rates of change in the densities (number of individuals per unit area) of susceptible S and infective hosts I over time. For example, for the SIS setting, we have:

$$\begin{aligned} \frac{dS}{dt} &= B + \gamma I - \beta \frac{SI}{N} - dS \\ \frac{dI}{dt} &= \beta \frac{SI}{N} - (\alpha + d + \gamma) I \end{aligned} \tag{2.1}$$

where $N = S + I$ is the total population density. An average member of the population makes contact to transmit the infection with β other members of the population per unit time. The parameter β is known as the per capita transmission rate of the infection disease. The probability that a random contact by an infected member is with a susceptible member is $\frac{S}{N}$. So the number of new infections generated per infective per unit time is $\beta\frac{S}{N}$. Since there are I infective members, the rate at which new infections are generated is $\beta\frac{S}{N}I$. The force of infection is $\beta\frac{I}{N}$ and is defined as the rate at which susceptible individuals acquire an infectious disease. Among the infected, γI individuals return to the susceptible class through recovery. The parameter γ is called the clearance rate. New susceptible hosts arise at a birth rate B, possibly depending on S and I. The total population may be diminished by natural and disease-induced mortality, given by the terms dS and $(\alpha + d)I$. Here d is the natural mortality rate and α the virulence rate of the infection (Dieckmann *et al.*, 2005).

2.1 The basic reproduction number - R_0

In epidemiology, it is essential to quantify the severity of actual (or potential) outbreaks of infectious diseases. One of the most informative features of epidemiological models is the basic reproduction number denoted by R_0 that characterizes the potential of an outbreak to cause an epidemic (Heffernan *et al.*, 2005). The established definition of R_0, as phrased by Anderson and May, is "the expected

number of secondary cases that would arise from the introduction of a single primary case into a fully susceptible population during the entire infectious period" (Dieckmann *et al.*, 2005) . It is important to note that R_0 is a dimensionless number (Jones, 2007). Formally, R_0 is defined as follows. An infective individual makes β contacts per unit time, all of which are with susceptible individuals and thus produce new infections. The mean infective period is

$$\frac{1}{\alpha + d + \gamma}.$$

. So, the expected number of new infections is

$$R_0 = \frac{\beta}{d + \gamma + \alpha}. \tag{2.2}$$

This dimensionless coefficient is the basic reproduction number (Jones, 2007) for this model and we can see how it increases with β and decreases with the clearance rate γ.

In the case when there is more than one class of infectives or in any situation in which the population is divided into discrete, disjoint classes, the next generation matrix, introduced by (Diekmann *et al.*, 2010), is a general method of deriving R_0 because it accounts for different types of infected. The next generation matrix can be used for models with underlying age structure or spatial structure (Heffernan *et al.*, 2005).

R_0 defines an important threshold in the dynamics of an infectious disease. When $R_0 < 1$ each successive "infection generation" is smaller than its predecessor and the infection can not persist. Conversely, when $R_0 > 1$ successive "infection generations" are larger than their predecessors and the number of cases in the population will initially increase (Dieckmann *et al.*, 2005). This increase does not continue indefinitely. The infection process reduces the "pool of susceptibles" and hence reduces the probability that an infectious individual contacts a susceptible within its period of infectiveness. This non-linear effect can only be neglected at the beginning of an epidemic. The basic reproduction number depends on the rate of contact between individuals, the probability of transmission given contact and the time an infected remains able to transmit the infection. All these three

components are subject of disease control methods: isolating those with infection from the rest of the community, reduces their rate of contact with others, hygiene measures reduce either the contact rate or the probability of transmission given contact; and drug treatment reduces the probability of transmission and/or the length of the infectious period (Roberts & Heesterbeek, 2003). Another way of controlling the spread of an infection is by reducing the size of the available pool of susceptible, either through vaccination or prophylactic treatment (Roberts & Heesterbeek, 2003).

To see how the total population size evolves over time, we can sum the equations for $\frac{dS}{dt}$ and $\frac{dI}{dt}$ to get

$$\frac{dN}{dt} = B - dN - \alpha I\,. \tag{2.3}$$

A common assumption that simplifies substantially the mathematical analysis is to suppose that the total population remains constant over time (Dieckmann *et al.*, 2005), that is birth rate equals death rate:

$$B = dN + \alpha I. \tag{2.4}$$

In addition, one may suppose that both birth and death rates are equal to zero. This hypothesis can be justified by saying that the time scale of the disease is much faster than the time scale of births and deaths, so that demographic effects on the population may be ignored. An alternative point of view is that we are interested only in studying the dynamics of a single epidemic outbreak.

When the population size is assumed constant, we can divide the equations by the total population size N and obtain a SIS model for the proportions $s = \frac{S}{N}$ and $i = \frac{I}{N}$ of susceptible and infected, respectively,

$$\begin{aligned} \frac{ds}{dt} &= d + \alpha i + \gamma i - \beta s i - d s \\ \frac{di}{dt} &= \beta s i - (\alpha + d + \gamma) i \end{aligned} \tag{2.5}$$

In fact, since $s + i = 1$ we can reduce the above system of two equations to a

single one,

$$\frac{di}{dt} = \beta(1-i)i - (\alpha + d + \gamma)i\,. \tag{2.6}$$

This equation is simpler to analyse. Solving it, we find two equilibrium points: the **disease-free equilibrium** $i^* = 0$ and a **non-trivial equilibrium**

$$i^* = 1 - \frac{\alpha + d + \gamma}{\beta} = 1 - \frac{1}{R_0}. \tag{2.7}$$

When $R_0 > 1$ the non-trivial equilibrium exists between 0 and 1 and so it is called the **endemic equilibrium**, because the desiase persists in the population. When $R_0 < 1$ the endemic equilibrium no longer exists and the disease will go extinct in time. At $R_0 = 1$, there is a transcritical bifurcation separating the trivial clearance equilibrium from the endemic persistence equilibrium.

For simplicity, in the rest of this thesis we will maintain the initial notation for the susceptible S, infected I, although from now on these correspond directly to scaled variables with respect to total population, hence proportions. We will also focus on infections that are primarily non-virulent and therefore, will neglect from our models the death caused by the pathogen virulence, assuming $\alpha = 0$.

Chapter 3

Background on multiple strain infections

Studying the dynamics of infectious diseases often involves the interaction of multiple strain pathogens (e.g *Neisseria meningitidis*, Hepatitis, *Plasmodium falciparum*). As a general definition, strains are homogeneous groups within species (Balmer & Tanner, 2011). Strains are important because they can differ greatly in many traits, including growth rate, virulence, infectivity, antigenicity or drug resistance (Balmer & Tanner, 2011). The classification of pathogens into strains is thus of practical value. Typically the interactions between multiple strains can alter infection outcomes, such as duration and virulence.

Multi-strain models have been widely used in epidemiology however, developing and using multi-strain models is a challenging procedure due to numerous parameters such as death rate, birth rate, force of infection, and transmission rate, which are commonly assumed to be strain-specific.

According to May and Nowak (1994) we can distinguish two different instances of multiple infections: *superinfection* and *co-infection* (Dieckmann *et al.*, 2005). In the first case, a competitive hierarchy among the different parasite strains is assumed, such that a more dominant strain can infect and take over the host already infected by a less dominant strain. Multiply infected hosts transmit only the most virulent of their strains. In the opposite scenario (co-infection), which we will focus on this thesis, we have two pathogen species or strains coexisting simultaneously within the host (Dieckmann *et al.*, 2005). The occurrence of

multiple-strain infections is well documented in malaria, where it has been shown that 85% of the infected population is co-infected by either two strains of the same pathogen, or two different pathogen species (Balmer *et al.*, 2009). About 11% of all infections by human pathogens contain multiple strains (Balmer & Tanner, 2011). Multiple-strain infections are not a rare occurrence. The prevalence values vary because sample sizes are frequently small in many studies and because rates will vary in space and time and with methods used.

When we have two or more strains simultaneously colonizing the host, there are often pathogen interactions between them. In the particular case of competition, it can take three distinct forms. In direct interference competition, strains excrete substances that harm each other (eg, in case of *Escherichia coli* produce *colicins*). In resource competition, which is an indirect form of competition, one strain uses limited host resources like nutrients or space that are then no longer available to the other strain. Competition can also be mediated by the immune system which is called the apparent competition (Balmer & Tanner, 2011). By activating an immune response, strain A affects strain B if the response cross-reacts with strain B (Balmer & Tanner, 2011). Also, many pathogens have immunosuppressive properties and by suppression of the immune system, one strain reduces the effect of immunity on others (Balmer & Tanner, 2011). The relative amount of suppression and density of the strains will determine how much different strains benefit. But to distinguish each different type of competition in not always easy.

Multiple-strain infections can, however, have advantageous effects for the host compared with single infections as shown in the study of (Balmer & Caccone, 2008) on two *Trypanosoma brucei* strains, the causal agent of human African sleeping sickness. Here, hosts infected with both strains survived significantly longer than did those infected with the more virulent strain alone. Analysis of the strain dynamics reveals that this is due to the suppression of the density of the more virulent strain (Balmer *et al.*, 2009).

The primary focus of Evolutionary epidemiology has been to explore the consequences of multi-strain infections for the evolution of pathogen virulence (Alizon *et al.*, 2013) and the cross-immunity induced by different strains (Gomes *et al.*, 2002). In the following sections we discuss these two topics in more detail.

3.1 Evolution of virulence

As previously mentioned, multiple infections have major consequences for the spread of parasites in a population as well as the decrease in host fitness due to the infection (referred to here as virulence). However, even though the importance of considering the diversity of infections is often acknowledged, our understanding of how multiple infections affect the evolution of virulence is still limited. Virulence can be defined as disease-induced host death rate (Alizon *et al.*, 2009) and is a trait controlled in part by the parasite and can evolve to higher or lower levels.

The first theories on virulence suggested that pathogens would evolve to avirulent commensals since harming the host would be a poor long-term survival strategy (Alizon *et al.*, 2009). This view has changed in the mid 20th century as evolutionary biologists considered how competition among multiple different strains of a given pathogen would influence the evolution of virulence (Alizon *et al.*, 2009). Here, the superiority of one strain over another would depend on its ability to replicate within a host, the length of time that the host is infected (recovery rate), and successful transmission to a new host. These measures of pathogen fitness are easily integrated into a single term, the basic reproductive number R_0, which was modelled by Anderson and May (Minus van Baalen, 1995). R_0 gives us a measure of fitness, but not of its evolution. If each parameter in the formula would evolve independently of the others, a virus could increase its fitness by simply evolving any or all of the following: a lower host mortality rate, a lower recovery rate (longer infectious period), and a higher transmission rate. Instead, most models assume that these parameters may be coupled to each other in terms of trade-offs. A trade-off is a constraint that forces one parameter to change with another. Pathogens are assumed to evolve to an optimal balance of these factors subject to the constraints of the trade-off. The trade-off hypothesis states that virulence is an unavoidable consequence of parasite transmission (Alizon *et al.*, 2009). The presence of multiple genotypes of the same parasite within the host and their relatedness can affect the level of virulence of a pathogen (Martin A. Nowak & Robert M.May, 1994),(Frank, 1996) .

As we mentioned before, there are three main types of interaction between parasites : resource competition, cooperation for public goods and interference

competition (spite).

In the beginning of an epidemic, multiple infections are rare and pathogens adopt prudent host exploitation strategies that lead to reduced virulence. But when the pathogen becomes endemic, the frequency of multiple infections rises, and the pathogens have to share hosts more frequently. Hosts that harbour many pathogens are more likely to favour increased virulence but also single infected hosts may also favour increased virulence, because subsequent infections have to be anticipated (Minus van Baalen, 1995).

A classic result of virulence theory is that intensity of exploitation/damage to host correlates negatively with kinship among parasites infecting the host. This is because, lower relatedness leads to greater competition for resources, which selects for faster growth rates and consequently higher virulence (Gardner *et al.*, 2004). On the contrary, a positive relationship between relatedness and virulence is predicted if host exploitation is dependent on parasites cooperating with each other by producing public goods. In case of spiteful interactions (behaviours that harm both the actor and the recipient), virulence is predicted to peak at high and low levels of relatedness. Spiteful behaviours found in nature are surprisingly common, and one example is the production of bacteriocins. These anti-competitor toxins produced by bacteria are proteinaceous toxins and their lethal activity are often limited to members of the same species as the producer (Gardner *et al.*, 2004). Here, a U-shaped relationship between kinship and virulence was found (Gardner *et al.*, 2004) contrary to the previous models that predicted a linear increase or decrease in virulence as kinship (r) is increased. We refer to frequency of a particular genotype within the interaxting population as Kinship. Because bacteriocin production is expected to correlate with low bacterial growth rates, virulence will tend to be minimized at intermediate kinship and maximized when bacteria compete only with non-kin(r=0) or only with kin(r=1) (Gardner *et al.*, 2004).

This demonstrates that virulence evolution depends on biological aspects such as: if parasites are able to improve their success through prudent growth, cooperative contribution to public goods or through anti-competitor toxin production. How relatedness (which refers to similarity between actor and recipient relative to the competing population as a whole) affects virulence crucially depends on

the type of social interaction. A negative relationship is expected if cooperation is prudent resource use; a positive relationship if cooperation is the production of public goods; and an inverse unimodal if competition is mediated by spiteful behaviours. However, these relationships can be altered by other factors.

Unfortunately, experimental data on the relation between transmission and host mortality (due to virulence) are often hard to obtain. Present results suggest that to understand more the evolution of virulence, it may be worthwhile to investigate in more detail what happens in case of multiple infections. Although the trade-off model has suffered many criticisms, there is a huge potential in this model to address new problems because it provides a common framework to compare experimental and theoretical results. New discoveries in the virulence evolution field will continue to use the trade-off hypothesis but will have to incorporate further aspects such as pathogen fitness, immunopathology, spatial structure, and potentially, even information on abundance ratio distribution of multiple strains in mixed infections. This, together with a proper account of host-intrinsic factors, will consequently lead to more realistic models and a better understanding of virulence. A recent avenue of research that has emerged is the consideration of the multiplicity of infection: whether a host harbours one, two or an arbitrary number of different strains (Lion, 2013). For example, in his models, Lion investigates whether multiple infections lead to higher or lower virulence by the distribution of parasite reproductive values across host classes with different multiplicity of infections.

3.2 Cross-immunity

A key concept that has been extensively studied in multi-strain systems is cross-immunity, which allows infection by one strain to induce partial/perfect protection against other strains (Ahn *et al.*, 2014). Cross-immunity is a form of apparent competition between different pathogen strains. *Andersen et al.*(1997,1999) developed a model where cross-immunity acts by reducing the susceptibility to further strains, while *Gupta et al.*(1994) modelled cross-immunity as acting by leaving susceptibility unchanged but reducing transmissibility by a factor (Gog & Swinton, 2002). Cross-immunity is included in different ways in different models,

but the general idea is the same: infection with one strain of the disease produces a lasting immune memory in the host which acts to protect against subsequent infection by other strains. This increases strain-strain competition and can make coexistence at the population level difficult.

Some of the commonly studied pathogens used to study the cross-reactive immunological responses include dengue and influenza (flu) viruses. These two viruses differ significantly in their immunological response induced by strain competition. The cross-reactive antibody response for influenza follows the more common immunological response in which a previous virus exposure yields partial protection against prospective strains, as long as virus strains are antigenically similar. Unlike flu viruses, cross-reactive antibodies following a dengue infection act to enhance (rather than to restrict) the severity of subsequent infections by other dengue strains (Reich *et al.*, 2013).

In view of the complexities that arise when multiple pathogens interact and affect a host's susceptibility to infection or transmissibility of future infections, many models have been proposed to study the dynamics of co-circulating pathogens and the immunological structures by which they interact/compete.

3.3 Co-colonization: a window into strain interactions

Colonization is the term usually used to refer to avirulent/asymptomatic carriage of an infectious organism. The term co-infection or mixed infection somehow refers to more serious health scenarios. In this thesis the two terms are used interchangeably.

Colonization with more than one strain can be fundamentally different from infections with just one strain, and have important consequences both for the host and the parasite. For the host, two or more strains represent a broader challenge that complicates defence and immune response (Balmer *et al.*, 2009). For the parasite, they lead to direct or indirect interactions between strains that can alter within-host population dynamics and transmission between hosts. This in turn can lead to novel selection pressures and thus altered evolution of parasite

life-history traits. Research on the interactions of pathogen strains has shown that factors such as contact heterogeneity, age structure, stochastic effect and the contact network structure in the population can allow for long-term coexistence of strains, which in the absence of such heterogeneities, would undergo competitive exclusion (Colijna *et al.*, 2009).

Co-colonization can provide a mechanism by which strain diversity is maintained in a population and can affect the evolution of the pathogens themselves. However, the prevalence of multiple-strain infections obtained by any method is usually an underestimate because rare strains are hard to detect and false negatives can seldom be excluded completely (Balmer & Tanner, 2011). The interest in co-infection in particular, has increased in recent years, with publications on human co-infection involving hundreds of pathogen taxa across all major pathogen groups. However, co-infection is still poorly understood, both in terms of their prevalence in nature and of their effects on specific parasites or on host–parasite interactions (Balmer *et al.*, 2009). The emerging picture is that they are found in most parasite species for which the necessary genetic tools to detect them are available and researchers actively look for them (Balmer *et al.*, 2009). They appear to be the norm rather than the exception.

Recent publications tend to show that a general negative effects of co-infection on human health are more frequent than no-effect or positive effects (Griffiths *et al.*, 2011). However, the most commonly reported co-infecting pathogens differ from those infections causing highest global mortality (Griffiths *et al.*, 2011). These results raise questions concerning the occurrence and study of co-infection in humans and their implications for effective infectious disease management. Also, the overall consequence of reported co-infections was poorer host health and enhanced pathogen abundance, compared with single infections (Griffiths *et al.*, 2011). Moreover, the tendency for positive effects on pathogen abundance corroborates the negative effects on host health because larger infections are a mechanism by which disease can be exacerbated (Griffiths *et al.*, 2011).

In this work we will focus on co-infection by different serotypes of the *Streptococcus pneumoniae*, which is usually carried asymptomatically in the human nasopharynx. We will study direct interactions without invoking cross-immunity effects. We will also focus on infection clearance patterns rather than on virulence.

Simultaneous carriage of more than one strain of *Streptococcus pneumoniae* (co-infection) has been proposed to occur in 2-20% of colonized hosts (Valente *et al.*, 2012). It promotes horizontal gene transfer events and may lead to capsule switch (Brugger *et al.*, 2010). Because multiple-strain infections are a prerequisite for recombination among strains, they facilitate rapid generation of new variants that can evade drugs, vaccines or the immune response (Balmer & Tanner, 2011). Mixture of different strains increases the probability that infected hosts harbour strains that are refractory to treatment (Balmer & Tanner, 2011). In settings with regular drug administration like hospitals, multiple strains infections favour resistant strains that would otherwise be rare. Without genotyping, which confirms the presence of distinct strains, clinical interpretation might have been conversion from susceptibility to resistance of one strain (Balmer & Tanner, 2011).

Important questions to answer when studying co-colonization are: first, which strains are there and how related they are? Secondly, how they interact together? How do they impact the host immune-system or host heath?And finally, why are they more or less abundant?

3.4 A primer on *Streptococcus pneumoniae*

Among the multi-type pathogens, an especially diverse pathogen are *Streptococcus pneumoniae.*, also known as pneumococcus. This is a gram-positive bacterium naturally carried in the nasopharynx and is mostly a commensal (asymptomatic) (Shak *et al.*, 2013). The prevalence of pneumococcal carriage increases in the first few years of life, peaking at approximately 50% up to 70% in hosts 2-3 years of age (Shak *et al.*, 2013). There are more than 92 serotypes identified (Andrews *et al.*, 2014) and since 2001 there has been vaccination against some of the circulating serotypes. Serotypes are defined as distinct variations, of the cell surface antigens, within a species of bacteria or viruses. In pneumococcus each serotype refers to a distinct polysaccharide capsule. Current formulas of the pneumococcal conjugate vaccine protect against 7, 10 or 13 capsular vaccine types, but have no direct effect on the more than 79 non-vaccine types. Almost every child is colonized by

pneumococcus sometime in life and each serotype can colonize for several weeks being then replaced by another serotype or reacquired (Valente *et al.*, 2012).

Occasionally, the pneumococcus can migrate from the nasopharynx and potentially cause a range of diseases, from mild respiratory tract infection (e.g. sinusitis, otitis media) to more serious invasive or non-invasive conditions (e.g. pneumonia, septicaemia, meningitis)(Shak *et al.*, 2013). The median duration of carriage is estimated to be 31 days in adults and 60.5 days in children, but this is dependent on capsular serotype and previous immunologic exposure, host's age and immunocompetence (Shak *et al.*, 2013).

Each strain produces one of the 92 capsular polysaccharides, which are distinguished by using a set of antisera that recognise the chemical differences in the capsules. The capsule is important for virulence, but is immunogenic, and the large number of different capsular serotypes is believed to have been selected as a mechanism to evade the human immune response (Bentley *et al.*, 2006). Antibodies against capsular polysaccharide can protect against pneumococcal invasive disease (like pneumonia, otitis media, meningitis), and a highly effective protein-conjugated polysaccharide vaccine has been developed to protected children against the serotypes most commonly associated with serious diseases.Since 2009 advanced molecular tools allow the quantification and the characterization of co-colonization in pneumococcus (Brugger *et al.*, 2009). We will not examine vaccination effects in this thesis, but natural carriage dynamics of *S. pneumoniae* in the absence of interventions.

Chapter 4

Epidemiological dynamics of multi-type infections

4.1 A basic model for co-infection

In the basic model for co-infection the total population is distributed in three classes: S susceptible, I infected by a single strain and D doubly-infected by two different strains. The equations of the SID model governing the dynamics of these classes are

$$\begin{aligned}
\frac{dS}{dt} &= \mu N - \beta\frac{S(I+D)}{N} - \mu S + \gamma(I+D),\\
\frac{dI}{dt} &= \beta\frac{S(I+D)}{N} - \mu I - \gamma I - \sigma\beta\frac{I(I+D)}{N},\\
\frac{dD}{dt} &= \sigma\beta\frac{I(I+D)}{N} - \mu D - \gamma D.
\end{aligned} \tag{4.1}$$

From now on, we will use the word strain and serotype interchangeably. To derive the equations of the basic model we have assumed that:

1. New susceptible hosts enter the population at constant rate μ which is equal to the *per capita* departure (natural death) rate, leaving constant the total population size N;

2. Susceptible individuals first move to the single infected class and upon further exposure may acquire a second strain, thus moving to the double in-

fected class;

3. Individuals move from the susceptible class into the infected class according to the rate at which infectious are generated, $\beta\frac{S(I+D)}{N}$ where β is the transmission rate. I and D contribute equally to transmission.

4. An infected or double infected member of the population makes $\beta\frac{I}{N}$ contacts with single infected members to transmit the second infection. Only a portion $\sigma \in (0,1)$ of those contacts will produce a second infection. This parameter can also be interpreted as the vulnerability of an already infected individual to acquire a second infection. This means that there are competitive interactions between the resident and the new-coming strain. Therefore, the rate at which new double infected individuals are generated is $\sigma\beta\frac{I(I+D)}{N}$. We do not track explicitly the identity of the co-circulating strains. However, we assume the number of strains is sufficiently large so that typically double infection is with two different strains.

5. In this first formulation, we assume an underlying form of non-specific immunity. Infected and double infected individuals move to the susceptible class through the same recovery rate γ, also called the clearance rate. This will also simplify the analysis.

See figure 4.1 for a schematic representation of the basic co-infection model. Since the total population $N = I + S + D$ remains constant over time we can divide the equations by N and obtain a system of equations for the proportions of susceptible, infected and double infected, respectively. We will use the same letters S, I and D to denote the proportions. Since $S = 1 - I - D$, we can reduce the system (4.1) by one equation and obtain,

$$\begin{aligned} \frac{dI}{dt} &= \beta(1-I-D)(I+D) - \mu I - \gamma I - \sigma\beta I(I+D) \\ \frac{dD}{dt} &= \sigma\beta I(I+D) - \mu D - \gamma D \end{aligned} \tag{4.2}$$

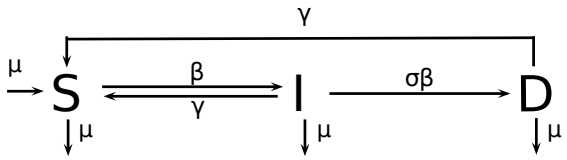


Figure 4.1: Schematic representation of the basic co-infection model.

4.1.1 Disease-free equilibrium

It is easy to see that the system has a disease-free equilibrium $(I^*, D^*) = (0, 0)$, as it is a solution of the equations

$$\frac{dI}{dt} = 0 \quad \text{and} \quad \frac{dD}{dt} = 0.$$

To determine its stability we calculate the Jacobian matrix of the system,

$$J = \begin{pmatrix} -I\beta\sigma - \beta(\sigma+2)(D+I) + \beta - \gamma - \mu & -I\beta\sigma - 2\beta(D+I) + \beta \\ I\beta\sigma + \beta\sigma(D+I) & I\beta\sigma - \gamma - \mu \end{pmatrix}.$$

When evaluated at the disease-free equilibrium we get,

$$J = \begin{pmatrix} \beta - \mu - \gamma & \beta \\ 0 & -\gamma - \mu \end{pmatrix}.$$

Since the Jacobian matrix is triangular, we can read its eigenvalues from the diagonal. All eigenvalues are negative except for $\beta - \mu - \gamma$ that can take both signs. Therefore, the sign of $\beta - \mu - \gamma$ determines the stability of the trivial equilibrium. Let

$$R_0 = \frac{\beta}{\mu + \gamma} \tag{4.3}$$

be the basic reproduction number in the basic co-colonizing model. Then, if $R_0 < 1$ means that the disease-free equilibrium is stable. Otherwise, if $R_0 > 1$, then it is unstable. This indicates the presence of another equilibrium, the so-called endemic equilibrium.

4.1.2 Endemic equilibrium

To determine the non-trivial equilibrium we have to find additional solutions of the system of equations

$$\frac{dI}{dt} = 0 \quad \text{and} \quad \frac{dD}{dt} = 0.$$

Besides having a disease-free equilibrium, the system of equations has another solution,

$$I^* = -\frac{(\mu+\gamma)(\mu-\beta+\gamma)}{\beta(\mu+\gamma-(\mu-\beta+\gamma)\sigma)} = \frac{R_0-1}{R_0(1+(R_0-1)\sigma)},$$
$$D^* = \frac{(\mu-\beta+\gamma)^2\sigma}{\beta(\mu+\gamma-(\mu-\beta+\gamma)\sigma)} = \frac{(R_0-1)^2\sigma}{R_0(1+(R_0-1)\sigma)}.$$

Since $S = 1 - I - D$ we also know the proportions of the susceptible class at the non-trivial equilibrium,

$$S^* = \frac{\mu+\gamma}{\beta} = \frac{1}{R_0}. \tag{4.4}$$

The condition $R_0 > 1$ guarantees that (S^*, I^*, D^*) is an endemic equilibrium and vice-versa. In fact, $0 < S^* < 1$ if and only if $R_0 > 1$. Moreover, $R_0 > 1$ implies that

$$0 < I^* < \frac{R_0-1}{R_0} < 1 \quad \text{and} \quad 0 < D^* < \frac{(R_0-1)^2}{R_0(R_0-1)} < 1.$$

Checking the eigenvalues of the Jacobian matrix J evaluated at the non-trivial equilibrium we can determine its stability. The eigenvalues of the Jacobian are

$$\lambda^+ = -\beta+\gamma+\mu \quad \text{and} \quad \lambda^- = -\beta\sigma+\gamma\sigma+\mu\sigma-\mu-\gamma. \tag{4.5}$$

We can rewrite both eigenvalues as follows

$$\lambda^+ = \frac{\beta}{R_0}(1-R_0) \quad \text{and} \quad \lambda^- = \frac{\beta}{R_0}(\sigma(1-R_0)-1). \tag{4.6}$$

Now, it is clear that under the condition $R_0 > 1$, both eigenvalues $\lambda^\pm$ are negative, resulting in a asymptotically stable endemic equilibrium.

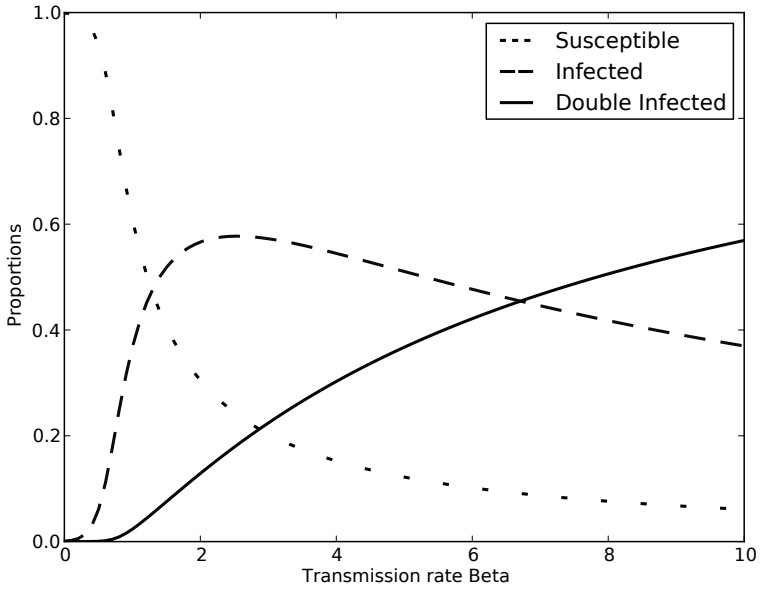
As we can see in figure 4.2a an endemic equilibrium starts to exist (be positive and stable) only if β is sufficiently high, in fact $\beta > \mu + \gamma$. As β increases, co-infection becomes dominant in the population. In figure 4.2b, as expected, only for values of σ above 0 we begin to have a double infected class and furthermore, for $\sigma > \frac{1}{R_0 - 1}$ the co-infection becomes more prevalent than single infection.

In the figure 4.3 there is a geometric representation of the trajectories of the system in the phase plane. Each curve or trajectory is obtained using different initial conditions which are represented by dots. As the figure shows, there is numerical evidence that when $R_0 > 1$, the endemic equilibrium is a global attractor.

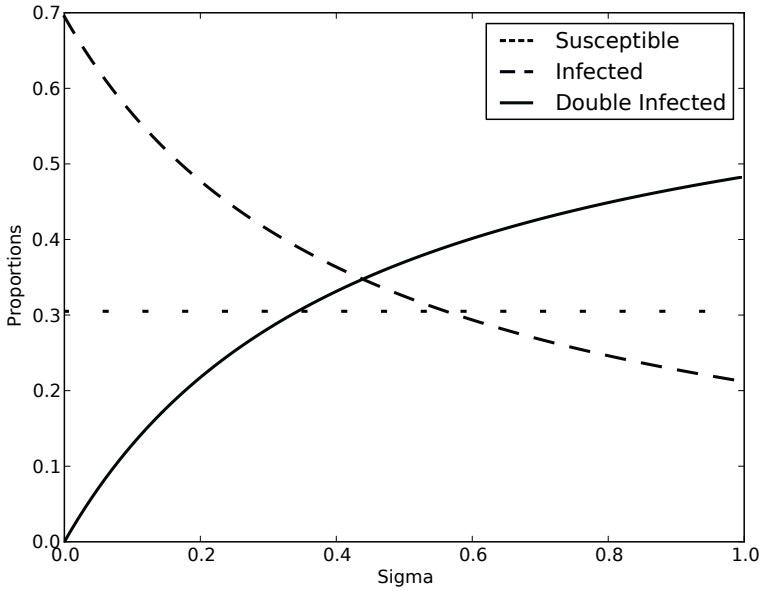
4.2 Co-colonizing serotype ratio

As mentioned before, sometimes more than one pneumococcal strain colonizes the nasopharynx (process known as co-colonization) and this is probably required for horizontal gene transfer between different pneumococcal strains, leading to capsular switching and acquisition of multidrug resistance. Such genetic exchange has been shown to occur for the capsule gene locus and the main mechanism of evolution in this species. There are few data on rates of multiple colonization, but existing estimates range from 1.3 up to 20% (Brugger *et al.*, 2010). Not only geographical variations in pneumococcal epidemiology but also the different techniques used to detect pneumococcus may explain the differences in reported prevalence estimates for co-colonization (Brugger *et al.*, 2010).

With the introduction of *multivalent* vaccines, there has been a renewed interest in the study of co-colonization since it is important to understand serotype changes among carriers following vaccination and whether this will shift evolutionary patterns of *S.pneumoniae* (Valente *et al.*, 2012). Conventional culture-based techniques are biased to detect the most-abundant serotype and are prone to miss co-colonization with a less-abundant type, an effect referred to as unmasking (Brugger *et al.*, 2009) This probable underestimation of multiple colonization can lead to a false interpretation of strain distribution, especially under selective pressure due to antibiotics or vaccines. It has been argued that observed changes



(a) Endemic equilibrium varying the transmission rate β. Parameters: $\gamma = 0.7$, $\sigma = 0.5$ and $\mu = 0.02$.



(b) Endemic equilibrium varying the competition parameter σ. Parameters: $\gamma = 0.7$, $\beta = 3$ and $\mu = 0.02$.

Figure 4.2: Endemic equilibrium in the basic SID model with single and co-colonization. See equations (4.1)

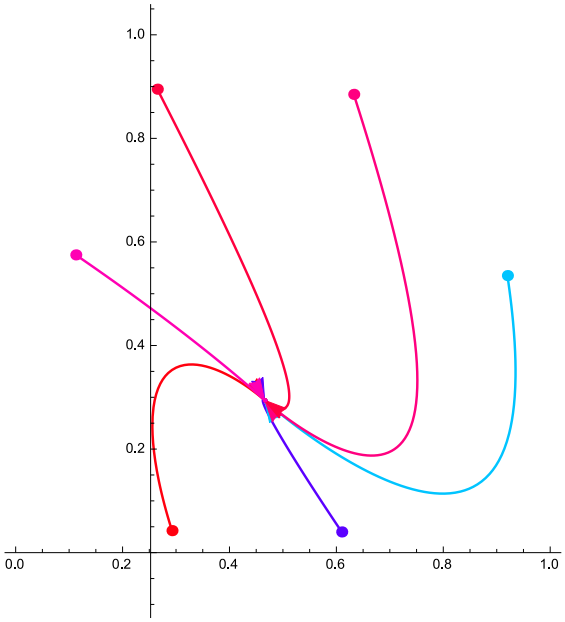


Figure 4.3: Phase portrait of the system. The x-axis represents the class D and the y-axis represents the class I. Parameters: $\gamma = 0.7$, $\beta = 3$, $\sigma = 0.5$ and $\mu = 0.02$ ($R_0 > 1$). See equation (4.1)

in serotype distribution may reflect unmasking of multiple colonization rather than true redistribution.

There has been much attention to serotype distribution but less attention to serotype interactions and co-occurrence. (Brugger *et al.*, 2010) have developed a novel DNA-based method for the detection of colonization with multiple *S. pneumoniae* strains directly in nasopharyngeal swabs. The advantages of this technique are that it can be applied directly to nasopharyngeal swabs, that it gives the number and relative amounts of cocolonizing strains in a given sample, and that it is relatively easy and economical to perform (Brugger *et al.*, 2010). As a consequence, it has been observed recently with this method of detection, the abundance ratio of each strains. His work on clinical samples suggests that is more likely to find a ratio of 1:1 between co-colonizing strains (see figure 4.5). These ratios refer to the number of PCR cycles needed to amplify the DNA from each strain.

The study of (Valente *et al.*, 2012), this time performed on healthy individuals, reveals similar patterns of within-host 2 serotype co-occurrence. See figure 4.4.

The balanced coexistence in a ratio less than 3 is seen in 50% of the subjects. So the ratio of co-colonizing strains (whether close to 1:1 or when one strain clearly prevails is abundance) does not seem to vary with disease/health status of an individual and seems to be also independent of the identity of the serotypes. Although both studies describe just a snapshot of the within-host dynamics that may be going on on co-colonized individuals, and report observations in a small sample, such results hint a potential at intrinsic regulatory mechanisms in strain-strain interaction, that may persist over time and thus be characteristic of the pneumococcal infections in general. Also the distribution of such ratios does not differ much across the two studies. These recent technologies for strain differentiation and typing have made possible to detect genetic diversity of pathogens, and there is a need for new epidemiological models to take into account and explain this new level of detail. Therefore, this thesis will explore 2 main models to try to incorporate these new biological observations.

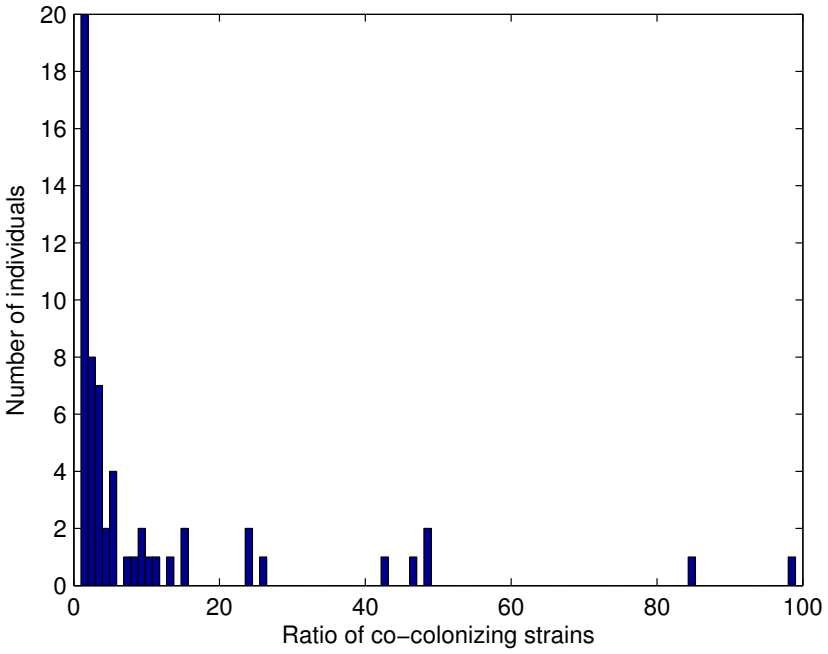


Figure 4.4: Strain ratios were obtained by comparative quantification in real-time PCR, which focuses on fold differences of expression $= 2^{\delta Ct}$ (Valente *et al.*, 2012).

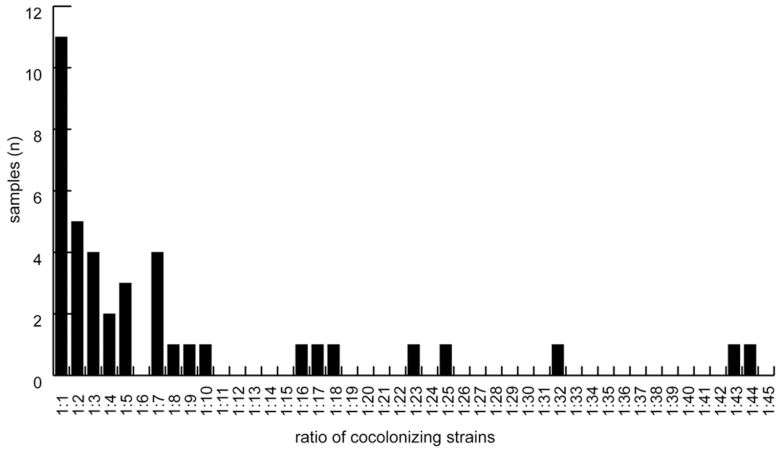


Figure 4.5: Strain ratios were determined from the peak heights in the chromatograms obtained by terminal-restriction fragment length polymorphism analysis (T-RFLP). Two different strains were present in 38 of the 41 samples and three samples contained three strains. The relative ratio of strains present in the same sample ranged from 1:1 to a maximum of 1:45, with a median ratio of 1:3.8 (Brugger *et al.*, 2010).

4.3 Structured epidemiological model

This model is similar to the previous one in the sense that we have again three different compartments: susceptible S, infected by a single strain I and a double infected class D. However, here we structure the double infected class by assuming that the individuals can harbour two arbitrary strains at different proportions each. We denote by $x \in [0, 1]$ the abundance ratio of the two strains in the co-infected host. We assume that this ratio is immediately determined upon acquisition of a second serotype and it does not change through the infection/colonization period of the host. This means that, the within-host dynamics reach an equilibrium on a fast time-scale relative to epidemiological dynamics. We denote by D_x the density of co-infected individuals harbouring two strains whose ratio between the strain less dominant and the strain more dominant is x,

$$x = \frac{\text{Abundance of less dominant strain}}{\text{Abundance of more dominant strain}}$$

The equations for the structure model are

$$\begin{aligned}
\frac{dS}{dt} &= \mu N - \beta\frac{S(I+D)}{N} - \mu S + \gamma I + \sum \gamma_x D_x, \\
\frac{dI}{dt} &= \beta\frac{S(I+D)}{N} - \mu I - \sigma\beta\frac{I(I+D)}{N} - \gamma I, \\
\frac{dD_x}{dt} &= q_x\sigma\beta\frac{I(I+D)}{N} - \gamma_x D_x - \mu D_x, \quad x \in X
\end{aligned} \tag{4.7}$$

In addition to the assumptions 1-4 presented in section 4.1, we have derived the equations of the structure model assuming that

1. The co-infection abundance ratio x varies in the set

$$X = \left\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\};$$

2. There are $n+1$ co-infected classes D_x and D is the sum of all co-infected densities of hosts,

$$D = \sum_{x \in X} D_x;$$

3. The rate at which new double infected individuals are generated is

$$\sigma\beta\frac{I(I+D)}{N}$$

and only a q_x portion of these will enter the co-infected class D_x;

4. Upon co-colonization, hosts are uniformly distributed among the D_x classes with probability $q_x = \frac{1}{n}$. Sub-sequentially depending on x, some double infections may be cleared faster/slower than others;

5. Singly infected individuals move to the susceptible class through the the recovery rate γ and double infected individuals of class D_x clear the infection at rate γ_x;

6. $\gamma_x \leq \gamma$ for every $x \in X$. This means that it is easier to clear one infection than two. Also, clearance γ_x decreases with x. This appears to be a more

Table 4.1: Model parameters and interpretation.

Parameter	Interpretation
γ	Clearance rate of a single infection
γ_x	Clearance rate of a double infection where strains display ratio x
β	Per capita transmission rate
σ	Competition parameter [0-1]. Factor reducing the relative acquisition rate of a second strain once colonized with one.
μ	Birth (death) rate, related to age group

likely assumption as co-infections seem to generally worsen the host health in comparison with single infections (Griffiths *et al.*, 2011).

A summary of all the parameters are shown in Table 4.1.

As before, we divide the equations of the model by the total population size N, and obtain a system of equations for the proportions of susceptible, infected and double infected D_x of co-infected ratio x, respectively. Again, we will use the same letters S, I and D_x to denote these proportions. Since $S = 1 - I - D$ we can reduce the system by one equation and obtain,

$$\begin{aligned} \frac{dI}{dt} &= \beta(1-I-D)(I+D) - \mu I - \sigma\beta I(I+D) - \gamma I \\ \frac{dD_x}{dt} &= q_x\sigma\beta I(I+D) - \gamma_x D_x - \mu D_x, \quad x \in X \end{aligned} \tag{4.8}$$

4.3.1 Disease-free equilibrium

The system has a disease-free equilibrium $(I^*, D_x^*) = (0, 0)$ (also known as trivial equilibrium) which is (locally) asymptotically stable provided $R_0 < 1$ and unstable provided $R_0 > 1$. Indeed, it is easy to check that $(I^*, D_x^*) = (0, 0)$ is an equilibrium of the system. To determine its stability we look at the Jacobian matrix evaluated at the equilibrium,

$$J = \begin{pmatrix} \beta-\mu-\gamma & \beta & \beta & \cdots & \beta \\ 0 & -\mu-\gamma_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\mu-\gamma_1 \end{pmatrix}.$$

Since the Jacobian matrix is triangular, we can read its eigenvalues from the diagonal. All eigenvalues are negative except from $\beta - \mu - \gamma$ that can take both signs. Therefore, the sign of $\beta - \mu - \gamma$ (and hence R_0) determines the stability of the trivial equilibrium.

4.3.2 Endemic equilibrium

In order to find the endemic equilibrium we set the system equations (4.8) equal to 0. From the equations of D_x we get

$$D_x = \sigma I(I + D)q_x R_x \tag{4.9}$$

where

$$R_x = \frac{\beta}{\mu + \gamma_x}, \tag{4.10}$$

can be interpreted as the basic reproduction number of the class D_x. We define the **averaged reproduction number of** D (all doubly-infected hosts) as,

$$\bar{R} = \sum q_x R_x. \tag{4.11}$$

Note that $\bar{R} > R_0$ since $\gamma \geq \gamma_x$ and γ_x is strictly decreasing with x, by assumption. Then summing equation (4.9) over all x we conclude that the endemic equilibrium is a solution of the following system of equations

$$\begin{cases} R_0(1 - I - D)(I + D) - I - \sigma R_0 I(I + D) = 0 \\ \sigma \bar{R} I(I + D) - D = 0. \end{cases} \tag{4.12}$$

Using the build-in function `Solve` from *Mathematica 10*, we have computed the non-zero solutions of the previous system,

$$I_{\pm} = \frac{1}{2}\left(\frac{\sigma((R_0 - 2)\bar{R} + R_0) + R_0}{\bar{R}\sigma^2(R_0 - \bar{R})} \pm \sqrt{\frac{R_0\left(R_0(\bar{R} - 1)^2\sigma^2 + 2\sigma((R_0 - 2)\bar{R} + R_0) + R_0\right)}{\bar{R}^2\sigma^4(R_0 - \bar{R})^2}}\right) \tag{4.13}$$

$$D_{\pm} = \frac{{I_{\pm}}^2 \bar{R}\sigma}{1 - I_{\pm}\bar{R}\sigma} \tag{4.14}$$

We are interested in the **admissible solutions** (I^*, D^*), that is when both I^* and D^* belong to the interval $[0, 1]$ and $0 < I^* + D^* < 1$. Using the build-in function `Reduce` we have determined the conditions on the parameters of the model which yield the endemic equilibrium. When $R_0 > 1$ there is a single admissible solution which corresponds to (I_+, D_+), see figure 4.6a. When $R_0 < 1$ both solutions $(I_\pm, D_\pm)$ are admissible if and only if

$$\bar{R} > \frac{2\sqrt{1 - R_0 + R_0\sigma - R_0^2\sigma} + R_0\sigma - R_0 + 2}{R_0\sigma}. \tag{4.15}$$

This scenario is interesting because although the classical basic reproduction number R_0 is less than one, the infection persists when the average reproduction number $\bar{R}$ of the co-infected class is very large, see figure 4.7. The fact that the infected population is stratified into several co-infection classes, the persistence of the infection depends on the balance of reproduction numbers as shows formula (4.15). As $\bar{R}$ increases (or equivalently γ_x and μ decreases) the equilibrium (I_-, D_-) gets closer to $(0, 0)$ whereas the equilibrium (I_+, D_+) gets closer to $(0, 1)$, as figure 4.7 suggests.

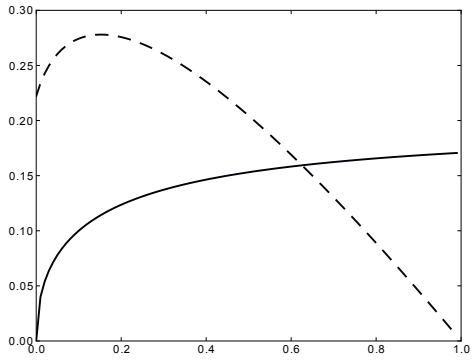
To compute the endemic equilibrium of each double infected class D_x we can use equation (4.9).

4.3.3 Distribution of co-infected hosts at equilibrium

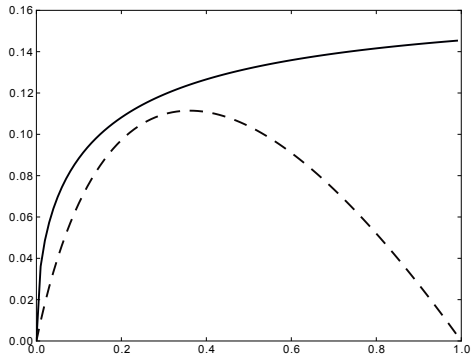
The proportion of each double infected class D_x at equilibrium only depends on $q_x R_x$ (see equation (4.9)). Moreover $D^* = \sigma\bar{R}I^*(I^* + D^*)$. So, dividing equation (4.9) by the equation for D^* we conclude that the proportion of co-infected individuals at equilibrium in the class D_x with respect to D is

$$\frac{D_x^*}{D^*} = q_x \frac{R_x}{\bar{R}}. \tag{4.16}$$

Note that this proportion is independent of the size of the susceptible or infected class. Moreover, it does not depend on the transmission rate β nor depends on the competition parameter σ. The number (4.16) can be interpreted as the **probability of an individual displaying a certain abundance ratio of co-colonizing strains, x, given that it is coinfected**. So, it is a conditional

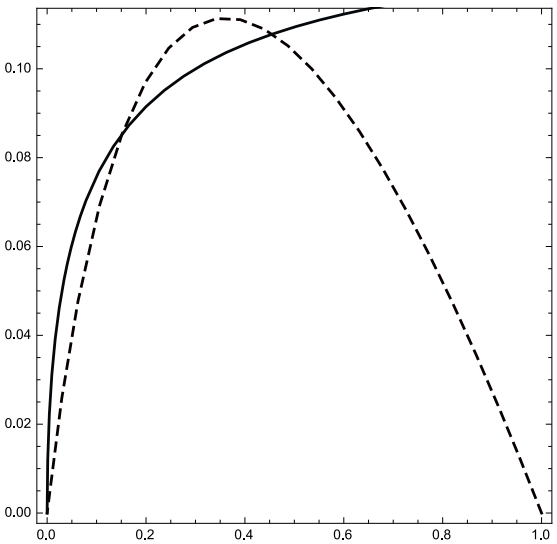


(a) $\sigma = 0.5$, $R_0 = 1.5$ and $\bar{R} = 10$.

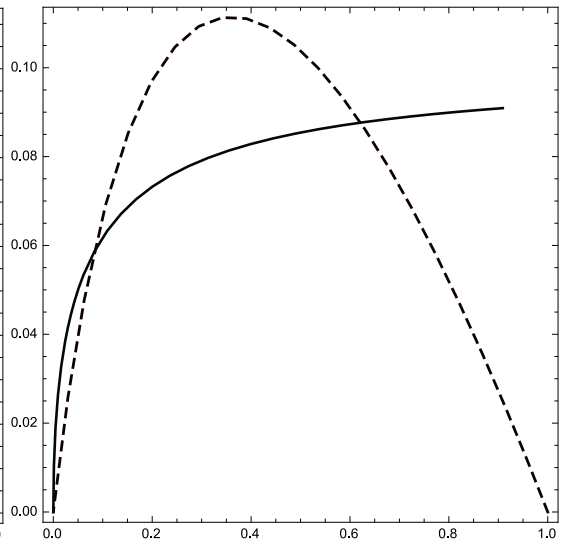


(b) $\sigma = 0.5$, $R_0 = 0.5$ and $\bar{R} = 12$.

Figure 4.6: Plots of the curves defined by the system (4.12). The variable D is on the horizontal axis and I on the vertical. The dashed curve corresponds to the first equation of the system and the solid curve corresponds to the second equation. In figure (a) we can observe the trivial equilibrium $(I = 0, D = 0)$ and the single endemic equilibrium (I_+, D_+). When $R_0 < 1$ and $\bar{R}$ is not big enough (not satisfy the equation 4.15) figure (b), there exists only the trivial equilibrium.



(a) $\sigma = 0.5$, $R_0 = 0.5$ and $\bar{R} = 15$.



(b) $\sigma = 0.5$, $R_0 = 0.5$ and $\bar{R} = 20$.

Figure 4.7: Plots of the curves defined by the system (4.12). The variable D is on the horizontal axis and I on the vertical. The dashed curve corresponds to the first equation of the system and the solid curve corresponds to the second equation. The leftmost (resp. rightmost) intersection point of the dashed and solid curve corresponds to (I_-, D_-) (resp. (I_+, D_+)).

probability of being in a double infected class D_x and is independent of the distribution of the population at equilibrium (it just depends on the parameters that define R_x and $\bar{R}$).

4.4 Numerical simulations

We now report several simulations on the structured co-infected model to understand the effect of each parameter in the distribution of D_x at equilibrium. We have used *Python 3.3* to perform our numerical simulations.

4.4.1 Clearance profiles

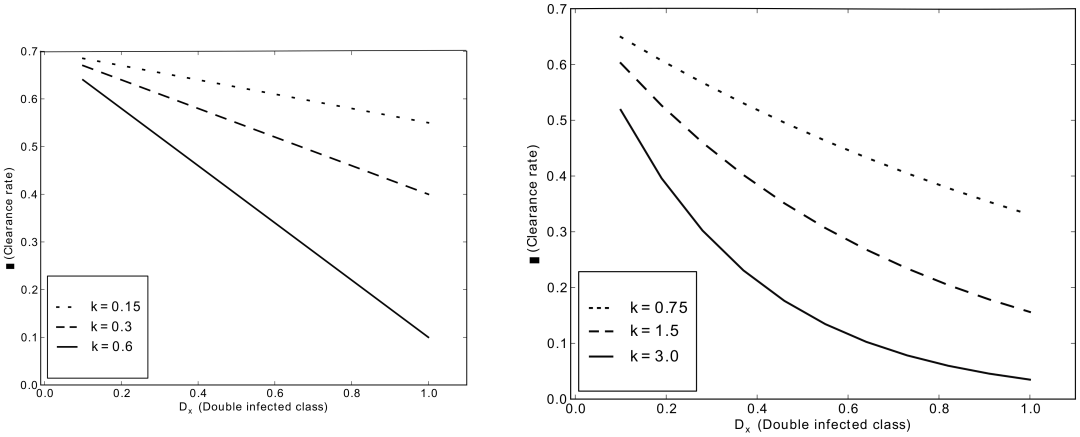
We have considered a **uniform distribution** of x, i.e. $q_x = \frac{1}{n}$, which means the same probability of entering any double infected class D_x upon co-infection. Also, we have considered the following scenarios: linear clearance and exponential clearance. An assumption of the model is that the **clearance rate γ_x decreases with** x. This means that it is easier for the individual to clear a co-infection where the ratios of the two serotypes are more asymmetric (x small) in comparison with a more balanced ratio (x close to 1). As we will see from the following simulations, at equilibrium the ratios around $1:1$ are more frequently observed in the population.

To perform the simulations we have consider the following clearance profiles:

- **Linear clearance**: clearance rate is a linear function of the abundance ratio of two strains. $\gamma_x = \gamma - kx$, where $0 < k \leq \gamma$,

- **Exponential clearance**: clearance rate is a exponential function of the abundance ratio of two strains. $\gamma_x = \gamma e^{-kx^s}$, where $k > 0$ and $s > 0$.

See figure 4.8 for the profile of the negative linear and the negative exponential clearance.



(a) Linear clearance rates.

(b) Exponential clearance rates($s = 2$).

Figure 4.8: Illustration of the clearance function γ_x varying for each doubly-infected host. $k = 0$ corresponds to the model without structure that we considered initially. Parameters: $n = 10$ and $\gamma = 0.7$.

4.4.2 Dynamics and endemic equilibria

We have computed and proved the existence of two endemic equilibria $(I_\pm, D_\pm)$ which co-exist for a certain range of the parameters. Determining its stability analytically is a difficult problem since the structured system (4.8) does not reduce to an (I, D) system where the stability is easy to determine analytically as we did with the system (4.2). Therefore, we have performed several numerical simulations to study the stability of the endemic equilibria. In the figures below, we plot several trajectories of the system given certain initial conditions. This is called the phase portrait of the system. In figure 4.9 we have considered $R_0 > 1$ and 4 different initial conditions (dots). The progression, in time, of the system is represented by the trajectories of the solid curves. It is easy to observe that, in all 4 scenarios, the system progresses towards the unique endemic equilibrium (I_+, D_+) (since $R_0 > 1$) which is represented here by the point of interception of the solid and dashed curves. Therefore, we assert that (I_+, D_+) is asymptotically stable.

In figure 4.10, we have considered $R_0 < 1$ and $\bar{R}$ sufficiently large so that condition (4.15) is fulfilled. In this scenario, we have 2 endemic equilibria and a trivial equilibrium. As we can observe from figure 4.10, with 4 different initial

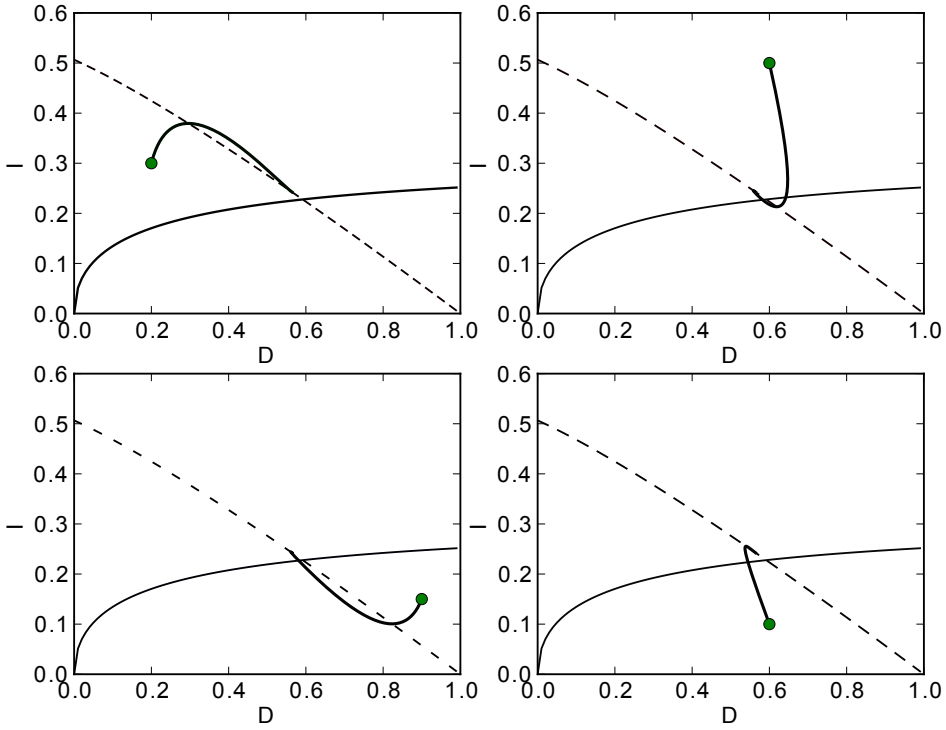


Figure 4.9: The phase portrait of the model when $R_0 \approx 4.17$ ($R_0 > 1$) and $\bar{R} \approx 6.34$. The dashed curve corresponds to the first equation of the system (4.12) and the solid curve corresponds to the second equation. We have considerer the following initial conditions (dots) around the (I_+, D_+) equilibrium. Parameters: $n = 10$, $\beta = 3$, $\mu = 0.02$, $\sigma = 0.5$ and $\gamma_0 = 0.7$. Exponential negative clearance with $k = 1$ and $s = 2$.

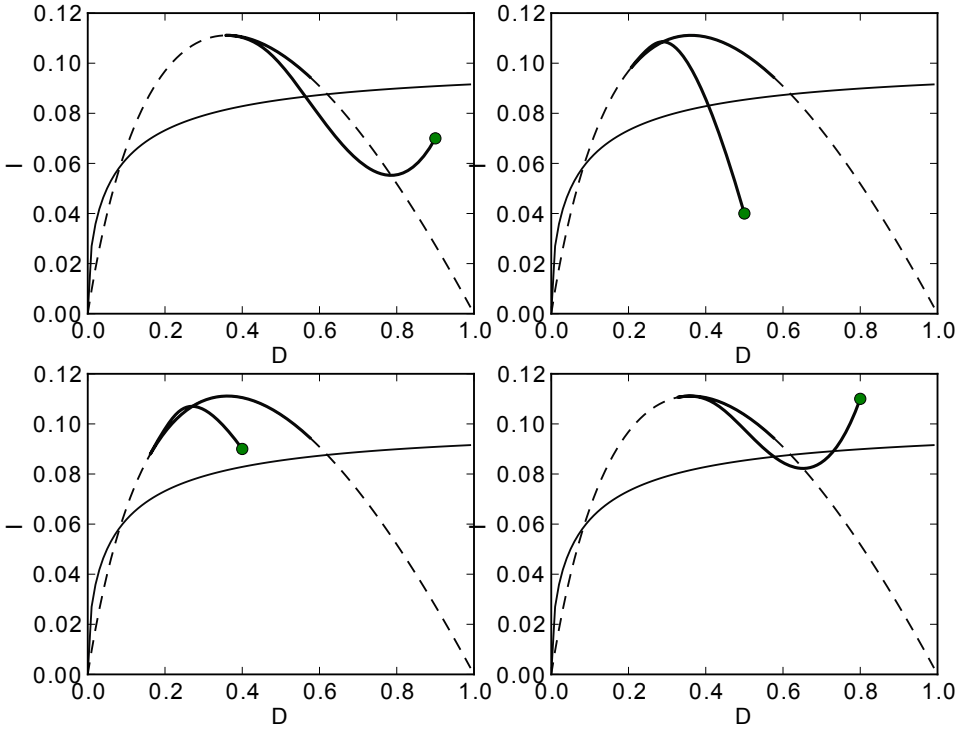


Figure 4.10: The phase portrait of the model when $R_0 \approx 0.498$ ($R_0 < 1$) and $\bar{R} \approx 20$. We have considerer 4 initial conditions (dots) around the (I_+, D_+) equilibrium. Parameters: $n = 10$, $\beta = 3$, $\mu = 0.02$, $\sigma = 0.5$ and $\gamma_0 = 6$. Exponential negative clearance with $k = 6.37$ and $s = 2$.

conditions (dots), the system progresses towards the (I_+, D_+) which is the rightmost intersection point of the dashed and solid curves. The numerics suggest that this equilibrium is asymptotically stable. This also implies that the trivial equilibrium, which we know to be asymptotically stable, is not a global attractor. In other words, the infection persists under certain initial conditions. We also have numerical evidence that the endemic equilibrium (I_-, D_-) is unstable.

When $R_0 < 1$ and $\bar{R}$ is sufficiently high there are 2 positive endemic equilibria and a **backward bifurcation** at $R_0 \approx 0.5$. The behaviour at a bifurcation may be described graphically by the bifurcation curve, which is the graph of equilibrium total infective population size $(I^* + D^*)$ as a function of the basic reproductive number R_0 (Brauer, 2004). The bifurcation curve has the form shown

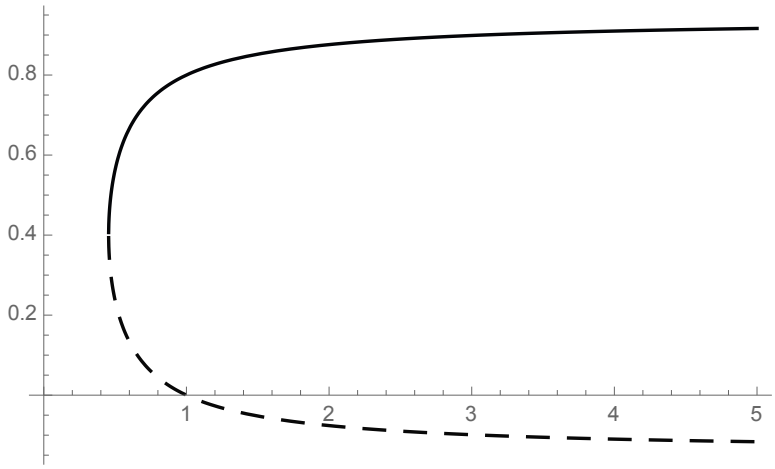


Figure 4.11: Backward bifurcation. The **y-axis** represents the total proportion of hosts infected ($I^* + D^*$) at equilibrium and the **x-axis** represents the R_0. One can see that even when $R_0 < 1$ two endemic equilibria arise: a stable (solid curve) and an unstable one(dashed curve). This is due to the high value of $\bar{R}$

in figure 4.11 with a dashed curve denoting an unstable endemic equilibrium and a solid curve denoting the stable endemic equilibrium. In figure 4.12 we can see explicitly above which value for $\bar{R}$ we begin to have 2 endemic equilibria.

4.4.3 Co-infection class at equilibrium

We will now analyse how the distribution of the co-infection class depends on the parameters. Recall that

$$\frac{D_x^*}{D^*} = q_x \frac{R_x}{\bar{R}} = \frac{\frac{q_x}{\mu+\gamma_x}}{\sum_y \frac{q_y}{\mu+\gamma_y}}. \tag{4.17}$$

The variation of $\frac{D_x^*}{D^*}$ depending on the transmission rate k can be observed from figure 4.13. In both cases of linear and exponential clearance, as we increase the parameter k, the distribution becomes more skewed favouring more balanced D_x ratios (e.g. 1:1). If we vary the parameter μ which is the birth (death) rate, it has the opposite effect since as μ increases the distribution becomes less skewed (see figure 4.14). These facts can also be deduced directly using the relation (4.17).

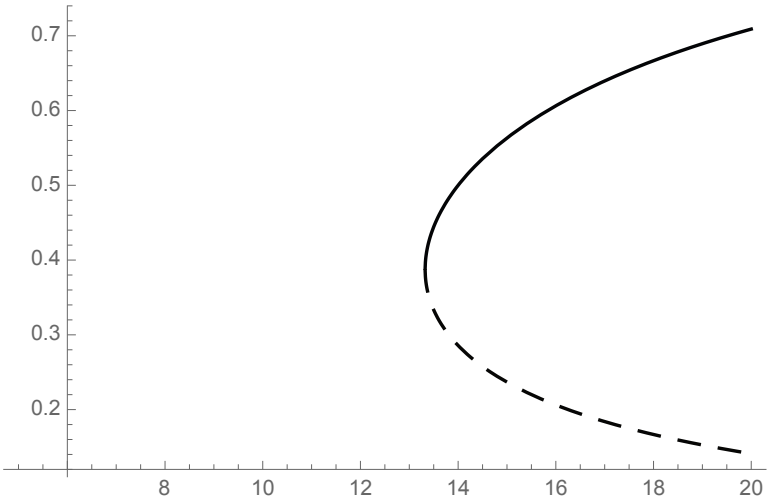
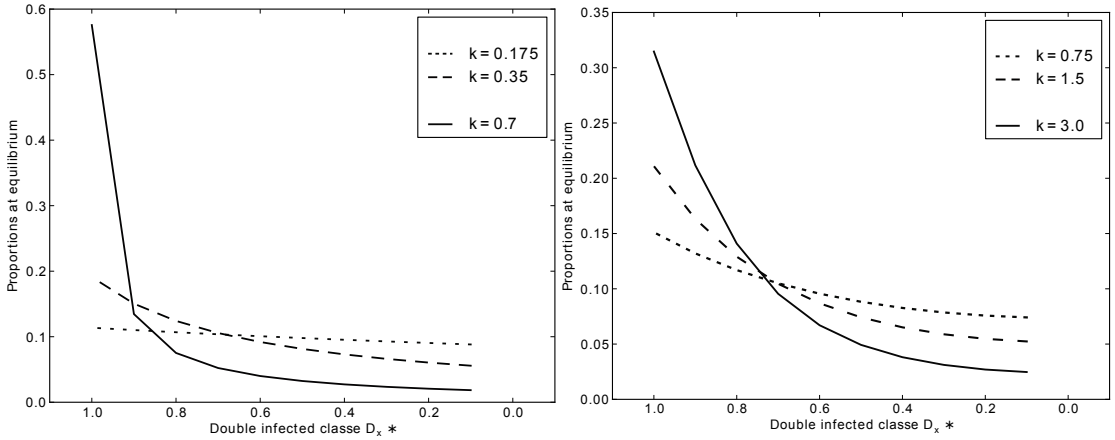


Figure 4.12: Backward bifurcation. The y**-axis** represents the total proportion of hosts infected $(I^* + D^*)$ at equilibrium and the x**-axis** represents the $\bar{R}$. Above a sufficiently high value for $\bar{R}$ an unstable equilibrium arises (dashed line) together with a stable one (solid line).

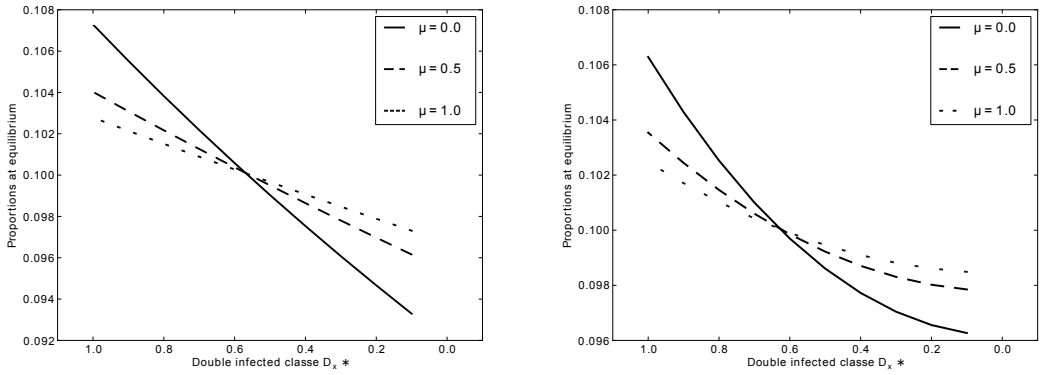


(a) Linear clearance model.

(b) Exponential clearance rates.

Figure 4.13: Impact of the parameter k on the equilibrium distribution of $\frac{D_x^*}{D}$. Graphs obtained from simulations.

We now compare the total prevalence, at equilibrium, of the D^* class in both co-infection models, the basic and the structure model. Firstly, we have to assume that $R_0 > 1$ since only under this condition we have an endemic equilibrium in



(a) Linear clearance model.

(b) Exponential clearance model.

Figure 4.14: Impact of the birth (death) rate μ on the distribution of $\frac{D_x^*}{D}$. Graphs obtained from simulations

both models. Let us denote by D_b^* and D_s^* the proportion of co-infected hosts at equilibrium in the basic and structure model, respectively. Recalling from section 4 we know that

$$D_b^* = \frac{(R_0 - 1)^2 \sigma}{R_0(1 + (R_0 - 1)\sigma)}. \tag{4.18}$$

On the other hand, we know from section 4.3

$$D_s^* = \frac{I_+^2 \bar{R} \sigma}{1 - I_+ \bar{R} \sigma} \tag{4.19}$$

where I_+ is given by formula (4.13) and depends on σ, R_0 and $\bar{R}$. In figure 4.15 we compare the total prevalence of D^* in both models. As the figure shows the proportion at equilibrium of the co-infection class of the structure model is always bigger than the basic model. This is not surprising since in the structured (4.7) model the clearance rate of double infections is smaller when compared to the basic model (4.1).

4.4.4 Conclusions

In the following we enumerate the conclusions of our study:

- From the epidemiological model, we can see that a sufficient mechanism

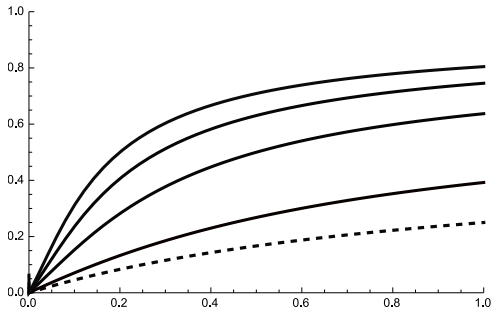


Figure 4.15: The dotted curve is the plot of D_b^* depending on σ when $R_0 = 2$. The solid curves are plots of D_s^* depending on σ when $R_0 = 2$ and from bottom to top: $\bar{R} = 3, 6, 9, 12$.

to obtain the skewed distribution of co-colonizing strain ratios towards 1:1, observed in pneumococcus carriage (Brugger *et al.*, 2010), (Valente *et al.*, 2012), is the assumption of some co-colonizations being cleared at faster or lower rates than others. The ratio of co-colonizing strains, x, can impact how exposed the "pathogen mixture" is to host immunity. A biologically plausible hypothesis is that the more asymmetric their within-host abundances, thus the more dominant one strain is, the more immunity might be stimulated, and this in turn can lead to faster clearance of both strains. This assumption was explored through the linear clearance rate and exponential clearance rate for doubly-infected hosts, as decreasing functions of x. We saw that the distributions of host in these co-colonization classes at equilibrium depend strongly on the assumptions in these clearance functions. For example, the faster γ_x decreases with x (in exponential case), the stronger the effect on the equilibrium distribution.

- When $R_0 > 1$ the infection persists and the endemic equilibrium is a global attractor. Of course, in this case the disease-free equilibrium is unstable. When $R_0 < 1$ the disease-free equilibrium becomes stable and the endemic equilibrium no longer exists. However, if $\bar{R}$ is sufficiently large then two new endemic equilibria $(I_\pm, D_\pm)$ emerge through a saddle-node bifurcation. From the numerics we conclude that (I_-, D_-) is unstable and (I_+, D_+) is asymptotically stable. Therefore, through this *backward bifurcation* phe-

nomena, the pathogen can persist via doubly-infected hosts even when it could not persist via single infection alone. This is explained by doubly-infected hosts having higher average reproductive number $\bar{R}$, than singly-infected hosts and thus contributing more to transmission on average.

- Not only have we explored the shape of this distribution, but also the total magnitude of co-colonization prevalence. The proportions, at equilibrium, of the co-infection class is always higher in the structured co-infection model compared to the basic model. If there is structure in D class, with generally lower clearance rates than single infection, the individuals in this class leave each double infecting class x at much slower rate. Consequently, the total prevalence of co-colonization goes up in the system, and as a consequence also total prevalence of carriage $(I + D)$. Also, the proportion of each double infected class D_x^* relative to D^* at equilibrium is independent of the size of the susceptible or infected class. This means that, whenever an individual acquires a second infection, he will have a fixed probability of being in a particular class x. Moreover, this probability only depends on the parameters of the model, namely: the transmission rate β, birth (death) rate μ and clearance rates γ and γ_x.

Chapter 5

Within-host co-infection dynamics model

5.1 Model description

In this section we will analyse co-infection at the host level and so, from a static epidemiological perspective, we will now describe a more detailed micro-perspective of the within-host infection processes. We will simulate h hosts that will be infected sequentially by two different strains. We will suppose that the dynamics of the density of the parasites in the host follows a logistic equation

$$\frac{dp}{dt} = rp\left(1 - \frac{p}{K}\right)$$

where r is the growth rate and K is the carrying capacity. The general solution of this equation is

$$p(t) = Kp_0\frac{e^{rt}}{K + p_0(e^{rt} - 1)}. \tag{5.1}$$

To incorporate in the model the effect of clearance by immunity we assume that r decreases monotonically with time since infection, according to

$$r(t) = r_0 e^{-\gamma t}.$$

where γ is the clearance rate. The clearance rate in this model is different from

the clearance rate of the epidemiological model. Here γ is per unit time per unit parasite contrary to previous model where is per unite time per host. Clearance of the first strain does not interfere with clearance of the second strain. The remaining parameters r_0 and p_0 are the initial growth rate of the infection and the initial parasite population size, respectively.

The host will acquire the first infection at a determined time point (e.g $t_0 = 0$). A second strain will infect the host at time $T > 0$. The identity of the strains is not followed and there is no fitness difference between strains. The dynamics of the second strain are completely independent of the dynamics of the first strain. This model reflects more a strain-specific immunity, in contrast to the epidemiological model presented in section 4, where the two strains were cleared simultaneously (γ_x was applying equally to both strains). We will assume that T is a random variable following an exponential distribution with parameter λ. This parameter λ represents the force of infection experienced by infected individual which is the number of new strain acquisitions per unit of time that an already-infected individual experiences during an endemic scenario. Mean waiting time to acquire a second strain when already colonized is given by $\frac{1}{\lambda}$. We assume λ to be constant over the population reflecting an epidemiological equilibrium.

To mimic the epidemiological sampling process we will consider an uniform random variable S taking values in $[0, \tau]$ that will represent time of survey of the co-colonized host,that will take place at time S. It is reasonable to suppose that τ is approximately equal to the duration of a single infection. To define precisely the value of τ we have considered the moment in the future when the density of parasites attains the value of p_0 with a relative error of 1%. More precisely, τ is the positive solution of the equation $p(t) = 1.01p_0$. See figure 5.1 for the behaviour of τ as γ varies. Clearly, higher values of γ (clearance of infection) gives smaller values τ (duration of infection).

To simplify the model even further, we also assume that T and S are independent.

Our goal is to determine through simulations the probability distribution of the co-infection ratio in a double infected host. For this, we will define the random variable $X(T, S)$ that represents the ratio of the two strains found in the host

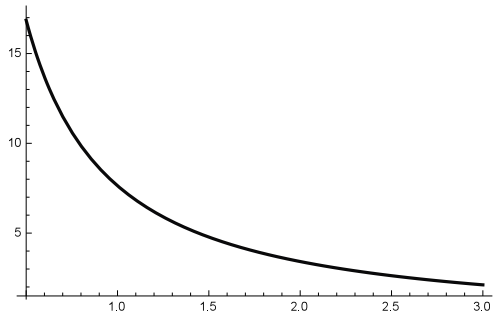


Figure 5.1: Plot of τ (on the vertical axis) depending on γ (horizontal axis). Parameters: $K = 1$, $r_0 = 3$ and $p_0 = 0.1$.

(with the most dominant one in the denominator),

$$X(T,S) = \min\left\{\frac{p(S)}{p(S-T)}, \frac{p(S-T)}{p(S)}\right\}.$$

The question that naturally arises is given the distribution of S and T what will be the distribution of X? And how the later will depend on the parameters of the infection dynamics.

5.2 Simulation results

The graphs below show the dynamics of the infection process over time in a single individual. We have varied the parameter γ:

We are more interested in scenarios where the *Streptococcus pneumoniae* carriage is present for a long period of time, thus γ small. We simulated different hosts (to replicate different co-colonization episodes). The timing of the new strain acquisition varied between different individuals, but the sampling times were the same for all. The following graph illustrates how the sampling process is performed. We can observe two curves of infection, with same shape but the second curve (yellow curve) has been shifted by T, the time at which the host have acquired a second infection. The vertical line, placed at S is a sample of a point where the proportion of serotypes are calculated.

In figure 5.4 we present the sampling distributions of the random variable

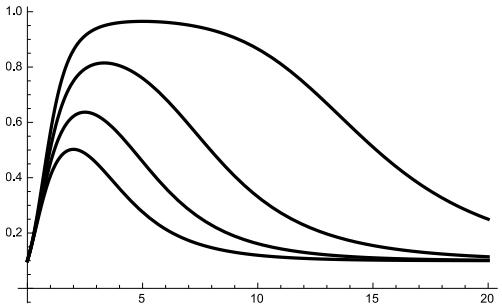


Figure 5.2: Within-host dynamics of a single strain infection. The x-axis represents time and the y-axis represents bacterial load. As γ increases from 0.2 to 0.5 the duration of the period of maximal infection decreases. Parameters: $K = 1$, $r_0 = 3$ and $p_0 = 0.1$.

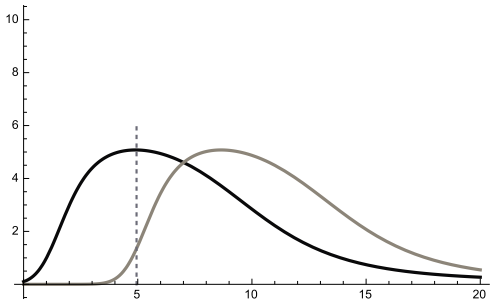


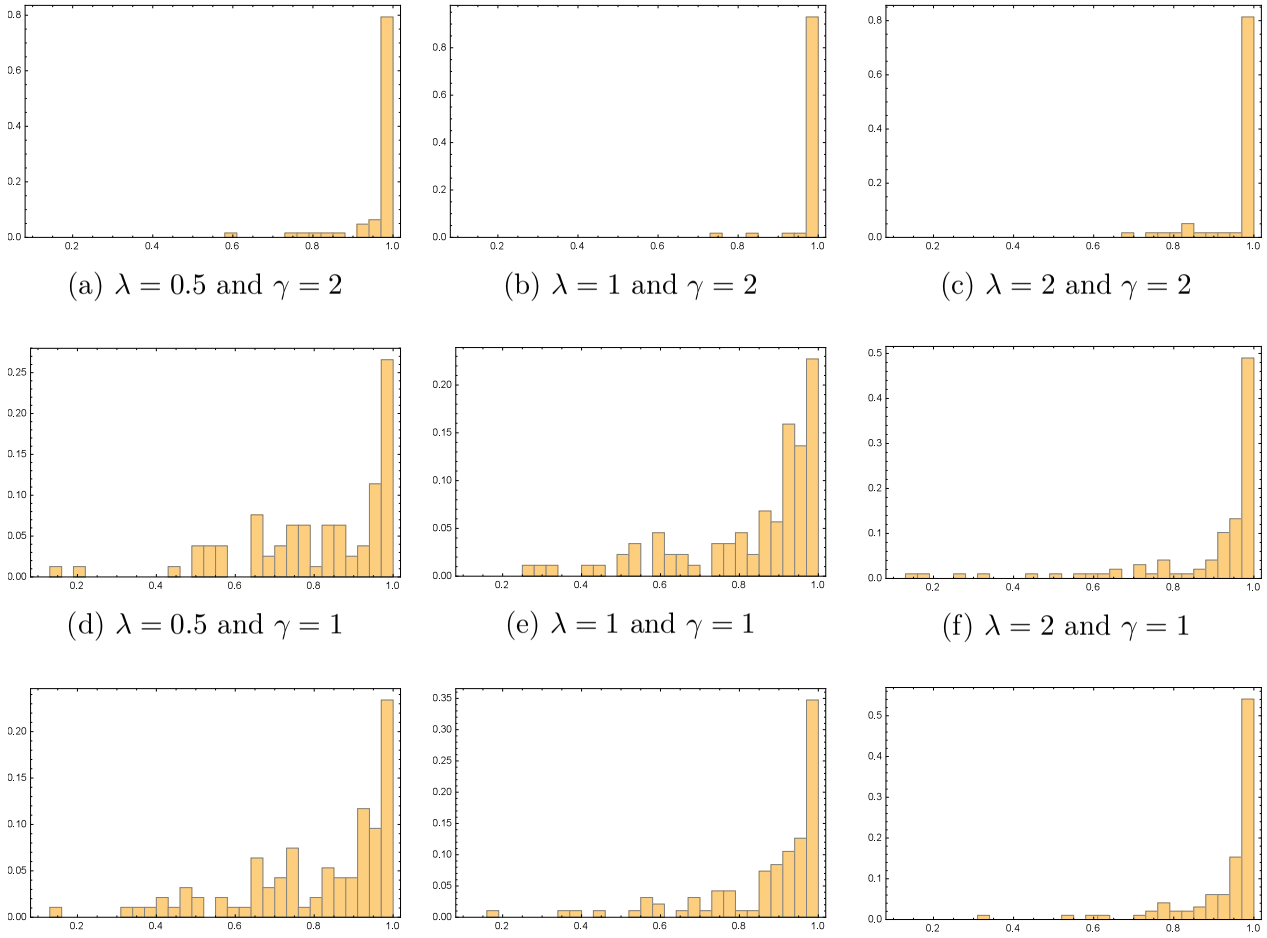
Figure 5.3: Illustration of the sampling process: the sampled co-infected hosts can display a different ratio of the two strains, depending on the time since co-infection occurred. x-axis represent the time and y-axis represent bacterial load.

X. We have varied the parameter γ and λ to see how the distribution varies its shape.

Clearly, in all scenarios the distributions are skewed towards equal abundance ratios.

We can also observe that as λ or γ increase, the distribution favours more equal abundance ratios. This can be explained as follows. Since λ is the parameter of the exponential distribution, the expected time for the start of the second infection is $1/\lambda$. This means that for high values of λ both curves are rather close, thus increasing the number of ratios closer to 1. On the other hand, for high values of γ, as figure 5.4 suggests, the plateau of the density of parasites becomes larger, thus increasing as well the number of ratios close to 1. A similar reasoning explains that for small values of λ or γ the distribution favours asymmetric abundance ratios.

Moreover, the shape of the distribution is robust with respect to small changes of the parameters.



(a) $\lambda = 0.5$ and $\gamma = 2$ (b) $\lambda = 1$ and $\gamma = 2$ (c) $\lambda = 2$ and $\gamma = 2$

(d) $\lambda = 0.5$ and $\gamma = 1$ (e) $\lambda = 1$ and $\gamma = 1$ (f) $\lambda = 2$ and $\gamma = 1$

(g) $\lambda = 0.5$ and $\gamma = 0.5$ (h) $\lambda = 1$ and $\gamma = 0.5$ (i) $\lambda = 2$ and $\gamma = 0.5$

Figure 5.4: Histogram of 100 samples of the random variable X. Parameters: $K = 1$, $r_0 = 3$ and $p_0 = 0.1$.

Chapter 6

Discussion and future perspectives

6.1 The two models of co-colonization

The first model presented in this thesis is a deterministic compartmental model. The population was subdivided into a number of compartments, and the flows between these compartments were described by a set of ordinary differential equations. We have assumed the ratio x between two strains is determined upon acquisition, which is the case of competition happening very quickly initially and it did not change over the infectious period of the co-colonized host. The recovery capacity of the host γ_x is the driving force for the structure emerging among co-colonized hosts at equilibrium. Those co-infections that are cleared faster will be less prevalent and those co-infections that are cleared slower will be more prevalent. for γ_x decreasing with x, we find that the distribution of doubly-infected hosts is skewed towards $1:1$.

As we have seen, mathematical modelling can provide many significant insights concerning the epidemiology of infectious diseases. The most notable of these include threshold conditions (like the basic reproductive number R_0) that describe when invasion and persistence of an infection is possible. Interestingly, we have shown in our first model that even if the basic reproduction number is below 1, the infection can still persist if the average reproduction number of the co-infected class is sufficiently big. Also, we have obtained the conditions for the persistence of the infection (endemic equilibrium) and detected the phenomena of backward bifurcation. Moreover, we have observed that the distribution of

the co-infected class, at equilibrium, is independent of the size of the susceptible or the single infected class. And so, when the host displays a double infection, he will present a particular strain ratio with fixed probability and this is only dependent on the transmission rate, birth (death) rate and the clearance rates. This applies to both stable and unstable endemic equilibria.

Brugger's work was pioneer in quantifying each strain involved in the co-colonization process, and understanding this phenomena serves as a motivation for modelling this work. His strain ratios were obtained by comparative quantification in real-time PCR. However, when we analyse more carefully his graphical representation of the data, we observe that it was used a non-uniform scale (i.e. with samples included in bins of different sizes) giving the reader the impression of a clear co-colonizing ratio distribution more skewed towards 1:1 (equal abundance ratios). The interpretation of his graph can be misleading if this fact is not addressed properly. Notice that even if the actual distribution of the co-colonizing hosts was uniform in a normal scale $[0, 1]$ with equally spaced bins, like the one we use to represent our results, displaying this distribution on a "fold" scale would naturally give us a skewness result towards 1:1. More precisely,

$$q_{\frac{1}{N}} = \int_{\frac{1}{N+1}}^{\frac{1}{N}} x\,dx = \frac{1+2N}{2N^2(N+1)^2}.$$

Thus, $q_{\frac{1}{N}}$ increases with N and gives higher weights to values of the ratio close to $1:1$, i.e. skewed towards $1:1$.

The second model presented in this thesis was a bottom-up approach to the co-colonization process. This was a semi-probabilistic within host model,which assumed that the ratio is dynamic over the infection period, and that the distribution emerges as a transient feature of colonization dynamics. From the different time-points over the course of co-infection when a host may be sampled. Thus the ratio was not static throughout infection, as assumed in the previous model. The infection dynamics and the acquisition of the second strain were explicitly simulated. The force of infection λ was constant, reflecting an epidemiological equilibrium. Here the two strains/serotypes were cleared independently and sequentially. In the simulations performed we have observed that when the entire range of values were divided into a series of small intervals (equal size bins) and

a histogram was obtained, the distribution showed the tendency towards equal abundance ratios for intermediate and high values of λ and γ parameters. Moreover, the simulations have shown that the shape of the distribution is robust for small perturbations these two parameters.

6.2 Challenges in co-colonization modelling

Multiple infections are becoming more studied because of their consequences to the health of the host, and also because they can potentially change the selective pressures acting on parasites. Selection shapes the evolution of parasite virulence and may have direct clinical implications. The case of influenza, norovirus, malaria and dengue are just a few of the infectious agents for which the interacting dynamics of different strains form a crucial part of their biology. The classical theory predicts that competition between strains in most cases will select the parasites to evolve towards higher virulence (Minus van Baalen, 1995) (Frank, 1996).

Given the commonness of multiple strain infections,building epidemiological models that reflect these phenomena is fundamental to help us understand many important diseases, but may also bring many technical challenges. One of the challenges is how to translate into the models the immunodynamics of the individual host to a population level? In our work, the host clearance rate can be seen as a very basic immune system action. The rate of pathogen clearance or the duration of the infectiousness is affected by partial immunity and this will also shape the pathogen population that is available for transmission. Does the overall pathogen load matters? If it does, it will probably be important to introduce in to the models the ratios of each pathogen involved in the colonization/infection process.

Another factor that could be taken into account in the future is the host heterogeneity. When we try to incorporate population structure into the transmission models, this potentially changes the dynamics of the infection. Our model assume population without immigration, except via births or deaths, and nobody enters or leaves the pool of susceptible hosts. It is important to understand how population structure and movement influences strain dynamics over a long

time-scale. Another kind of host heterogeneity is the variation in the immune response driven by host factors. We have not addressed in to these models the important effect of cross-immunity and considered strains as independent without explicit reference to their relatedness. Parasite diversity is thus shaped not only by resource competition between co-infecting parasites but also by host-driven immune mediated competition. Such interactions with the host immune system can amplify or reverse inherent differences in competitive ability of co-infecting parasites.

Most epidemiological models typically assume that only one of the co-infecting strains can be transmitted at a time. In the future, we could consider co-transmission, defined as the infection of the host by more than one parasite strains or species during the same transmission event (Alizon, 2013) . This process remains largely absent from epidemiological models and this is expected to have important consequences for virulence evolution. Recent work predicts that if co-infections are caused by different strains from the same species, increased probabilities of co-transmission favor less-virulent strains (Alizon, 2013). It would be useful in the future, to know the identity and relatedness of each strain involved in co-colonization process to predict virulence evolution.

With work we have tried to highlight the phenomena of co-colonization, particularly introducing the information on strain ratios into the epidemiological model, hoping that this provides useful insight into the multiple strain infections, specially in pneumococcus. However, more integration with biological and epidemiological data is urgently needed to motivate and support the extension towards more realistic models in the future.

Appendix A

Python code

```
from scipy.integrate import odeint
from math import gamma
from math import factorial


def CM(y,t,param):
    """ Function CM that implements an ODE system for
    the basic co-infection model."""
    beta=param[0]
    sigma=param[1]
    gamma=param[2]
    mu=param[3]
    I=[beta*y[0]*sum(y[1:])-mu*y[1]-gamma*y[1]-
    sigma*beta*y[1]*sum(y[1:])]
    D=[sigma*beta*y[1]*sum(y[1:])-mu*y[2]-gamma*y[2]]
    S=[mu-beta*y[0]*sum(y[1:])-mu*y[0]+gamma*sum(y[1:])]
    return S+I+D

def cm_pro(y0,t,beta,sigma,gamma,mu):
    """Function cm_pro which returns the solution for an
    ODE system given the initial proportions
```

```
    of infected and double infected classes (y0), time interval
    for analysis (t), beta, sigma,gamma and mu as parameters"""
    param=(beta,sigma,gamma,mu)
    return odeint(CM,y0,t,args=(param,))

def f(y,t,param):
    """Function f which implements an ODE system for
    the structured co-infection model"""
    beta=param[0]
    sigma=param[1]
    g0=param[2]
    gamma=param[3]
    q=param[4]
    d=param[5]
    I=[beta*(1-(y[0]+sum(y[1:])))*(y[0]+sum(y[1:]))-d*y[0]-
    sigma*beta*y[0]*(y[0]+sum(y[1:]))-g0*y[0]]
    doubleI=[q[i]*y[0]*(y[0]+sum(y[1:]))*sigma*beta-
    gamma[i]*y[i+1]-d*y[i+1] for i in range(0,len(q))]
    return I+doubleI

def sid_pro(y0,t,beta,sigma,g0,gamma,q,d):
    """Function sid_pro which returns the solution of an ODE
    system given the initial proportions (y0) of the infected
    I and double infected D_x classes, time interval for analysis (t),
    beta, sigma,gamma and mu as parameters"""
    param=(beta,sigma,g0,gamma,q,d)
    return odeint(f,y0,t,args=(param,))


def pdfBetaBinomial(x,n,a,b):
    """Function that returns a beta binomial mass distribution
    with n elements. a and b are the shape parameters"""
    return (gamma(x+a)*gamma(n-x+b)/gamma(n+a+b))
   *factorial(n)/(factorial(n-x)*factorial(x))/(gamma(a)
```

```
   *gamma(b)/gamma(a+b))
```

```
import sid_proportions
import matplotlib.pyplot as plt
import math
import numpy as np
import random
from pylab import *
```

```
============================    PARAMETERS    ============================
```

```
n=10
# Number of co-infected classes
tmax=50
beta=3
sigma=0.5
deltat=0.01
t=np.arange(0,tmax,deltat)
X=np.linspace(0,1,n+1)
g0=0.7
# Gamma of the Infected class
k=0.5
d=0.02
y0=[0.4]+[0 for x in X]
# Initial condition (proportions)
```

```
  1) Distribution of q_x
```

```
q=[sid_proportions.pdfBetaBinomial(x,n,0.7,2) for x in range(0,n+1)]
q=q[::-1]
# Non-uniform

q=[1/(n+1) for x in X]
# Uniform

  2) Distribution of Gamma_x

gamma=[g0*math.exp(-k*x**2) for x in X]
# Negative exponential
gamma=[g0 for x in X]
# Uniform

sol=sid_proportions.sid_pro(y0,t,beta,sigma,g0,gamma,q,d)
sol_inverse=sol[-1][1:]
sol_inverse=sol_inverse[::-1]


==========================       PLOTS      ================================

plt.plot(t,1-sol[:,0]-[sum(sol[i,1:]) for i in range(0,len(sol[:,0]))])
# Evolution of the proportion of Susceptible over time
plt.plot(t,sol[:,0])
# Evolution of the proportion of Infected over time
plt.plot(t,sol[:,1:])
# Evolution of the proportion of the co-infected over time
plt.bar(X-1/n*0.4,sol[-1][1:],width=1/n*0.8)

plt.xlim(1.1,-0.1)
plt.legend(loc="lower right")
plt.title(r"Proportions of $D_x$ at equilibrium")
plt.xlabel("Double infected classes $D_x$")
plt.ylabel("Proportions")
```

```
P=5
X_inverse=X[::-1]
Z=list()
for k in np.linspace(0,3,P):
    gamma=[g0*math.exp(-k*x**2) for x in X]
    sol=sid_proportions.sid_pro(y0,t,beta,sigma,g0,gamma,q,d)
    sol_inverse=sol[-1][1:]
    sol_inverse=sol_inverse[::-1]
    plt.plot(X_inverse,sol_inverse,'-o',label="k="+str(k))
plt.legend(loc="lower right")
plt.xlim(1.1,-0.1)
plt.title(r"$D_x*$ at equilibrium versus k")
plt.xlabel("Double infected classes $D_x*$")
plt.ylabel("Proportions")
# Proportion of D_x at the Equilibrium varying K

for beta in np.linspace(1.5,3,P):
    sol=sid_proportions.sid_pro(y0,t,beta,sigma,g0,gamma,q,d)
    sol_D=sol[-1][1:]
    plt.plot(X,sol_D/sum(sol[-1][1:]),'-o',label=r"$\beta$="+str(beta))
plt.legend(loc="lower right")
plt.title(r"$D_x$ at equilibrium versus $\beta$")
plt.xlim(1.1,-0.1)
plt.xlabel(r"Double infected classes $D_x$")
plt.ylabel(r"Proportions")
# Proportion of D_x at Equilibrium varying Beta

plt.show()
```

Appendix B

Mathematica code

```
(* Parameters *)
r0=3;P0=0.1;K=10;T=2;\[Lambda]=0.1;\[Gamma]=0.5;n=10000;

(* Strain infection curve *)
r[t_]:=r0 Exp[-\[Gamma] t];
P[t_]:=K P0 Exp[r[t]t]/(K+P0(Exp[r[t]t]-1));

(* Coinfection ratio *)
X[t_,s_]:=Min[P[s]/P[s-t],P[s-t]/P[s]]

(* Sampling the Coinfection ratio X *)
L=Table[X[RandomVariate[ExponentialDistribution[\[Lambda]]],
RandomVariate[UniformDistribution[{0,T}]]],{i,0,n}];

(* Plot of histogram *)
Histogram[L,{0.1,1,0.03},"Probability"]
```

References

Ahn, K.W., Kosoy, M. & Chan, K.S. (2014). An approach for modeling cross-immunity of two strains, with application to variants of Bartonella in terms of genetic similarity. *Epidemics*, **7**, 7–12. 15

Alizon, S. (2013). Parasite Co-Transmission and the Evolutionary Epidemiology of Virulence. *Evolution*, **67**, 921–933. 54

Alizon, S., Hurford, a., Mideo, N. & Van Baalen, M. (2009). Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of evolutionary biology*, **22**, 245–59. 13

Alizon, S., de Roode, J.C. & Michalakis, Y. (2013). Multiple infections and the evolution of virulence. *Ecology letters*, **16**, 556–67. 12

Andrews, N.J., Waight, P.a., Burbidge, P., Pearce, E., Roalfe, L., Zancolli, M., Slack, M., Ladhani, S.N., Miller, E. & Goldblatt, D. (2014). Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: a postlicensure indirect cohort study. *The Lancet. Infectious diseases*, **14**, 839–46. 2, 18

Balmer, O. & Caccone, A. (2008). Multiple-strain infections of Trypanosoma brucei across Africa. *Acta Trop.*, **107**, 275–279. 12

Balmer, O. & Tanner, M. (2011). Prevalence and implications of multiple-strain infections. *The Lancet. Infectious diseases*, **11**, 868–78. 2, 11, 12, 17, 18

REFERENCES

Balmer, O., Stearns, S.C., Schotzau, A. & Brun, R. (2009). Intraspecific competition between co-infecting parasite strains enhances host survival in African trypanosomes ¨. *Ecology*, **90**, 3367–3378. 12, 16, 17

Barreto, M.L., Teixeira, M.G. & Carmo, E.H. (2006). Infectious diseases epidemiology. *Journal of epidemiology and community health*, **60**, 192–5. 5

Bentley, S.D., Aanensen, D.M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M.a., Samuel, G., Skovsted, I.C., Kaltoft, M.S., Barrell, B., Reeves, P.R., Parkhill, J. & Spratt, B.G. (2006). Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS genetics*, **2**, e31. 19

Brauer, F. (2004). Backward bifurcations in simple vaccination models. *Journal of Mathematical Analysis and Applications*, **298**, 418–431. 39

Brauer, F. & Castillo-Chavez, C. (2012). *Mathematical Models in Population Biology and Epidemiology*. Springer. 6

Brugger, S.D., Hathaway, L.J. & Mühlemann, K. (2009). Detection of Streptococcus pneumoniae strain cocolonization in the nasopharynx. *Journal of clinical microbiology*, **47**, 1750–1756. 19, 25

Brugger, S.D., Frey, P., Aebi, S., Hinds, J. & Mühlemann, K. (2010). Multiple colonization with S. pneumoniae before and after introduction of the seven-valent conjugated pneumococcal polysaccharide vaccine. *PloS one*, **5**, e11638. v, viii, xix, 2, 3, 18, 25, 28, 30, 43

Colijna, C., Cohenc, T. & Murray, M. (2009). Latent Coinfection and the Maintenance of Strain Diversity. *Bull Math Biol.*, **71**, 247–263. 17

Dieckmann, U., Metz, J.A., Sabelis, M. & Sigmund, K. (2005). *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*. CAMBRIDGE UNIVERSITY PRESS. 5, 6, 7, 8, 11

Diekmann, O., Heesterbeek, J.A.P. & Roberts, M.G. (2010). The construction of next-generation matrices for compartmental epidemic models. *J. R. Soc. Interface*, **7**, 873–885. 7

Frank, S.A. (1996). Models of Parasite Virulence. *The Quarterly Review of Biology*, **71**, 37–78. 13, 53

Gardner, A., West, S.a. & Buckling, A. (2004). Bacteriocins, spite and virulence. *Proceedings. Biological sciences / The Royal Society*, **271**, 1529–35. 14

Gog, R. & Swinton, J. (2002). A status-based approach to multiple strain dynamics. *Mathematical Biology*, **184**, 169–184. 15

Gomes, M.G.M., Medley, G.F. & Nokes, D.J. (2002). On the determinants of population structure in antigenically diverse pathogens. *Proceedings. Biological sciences / The Royal Society*, **269**, 227–33. 12

Griffiths, E.C., Pedersen, A.B., Fenton, A. & Petchey, O.L. (2011). The nature and consequences of coinfection in humans. *The Journal of infection*, **63**, 200–6. 17, 32

Heffernan, J.M., Smith, R.J. & Wahl, L.M. (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society, Interface / the Royal Society*, **2**, 281–93. 6, 7

Jones, J.H. (2007). Notes On R 0. Tech. rep., Department of Anthropological Sciences Stanford University. 7

Krämer A, Kretzschmar M, K.K. (2010). *Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health*. Springer New York. 1

Lion, S. (2013). Multiple infections, kin selection and the evolutionary epidemiology of parasite traits. *Journal of evolutionary biology*, **26**, 2107–22. 15

Martin A. Nowak & Robert M.May (1994). Superinfection and the evolution of parasite virulence. *The Royal Society*, 81–89. 13

REFERENCES

Minus van Baalen, M.W. (1995). The Dynamics of Multiple Infection and the Evolution of Virulence. *The American Naturalist*, **146**. 13, 14, 53

O. Diekmann, J.A.P.H. (2010). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Son, Ltd. 1

Reich, N.G., Shrestha, S., King, A.a., Rohani, P., Lessler, J., Kalayanarooj, S., Yoon, I.K., Gibbons, R.V., Burke, D.S. & Cummings, D.a.T. (2013). Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *Journal of the Royal Society, Interface / the Royal Society*, **10**, 20130414. 16

Roberts, M.G. & Heesterbeek, J.A.P. (2003). Mathematical Models in Epidemiology. In M.G. Roberts & J.A.P. Heesterbeek, eds., *Encyclopedia of Life support Systems (EOLSS)*, vol. III, chap. Mathematic, EOLSS. 8

Shak, J.R., Vidal, J.E. & Klugman, K.P. (2013). Influence of bacterial interactions on pneumococcal colonization of the nasopharynx. *Trends in microbiology*, **21**, 129–35. 2, 18, 19

Valente, C., Hinds, J., Pinto, F., Brugger, S.D., Gould, K., Mühlemann, K., de Lencastre, H. & Sá-Leão, R. (2012). Decrease in pneumococcal co-colonization following vaccination with the seven-valent pneumococcal conjugate vaccine. *PloS one*, **7**, e30235. viii, xix, 2, 3, 18, 19, 25, 28, 29, 43